EKONOMI OCH SAMHÄLLE
ECONOMICS AND SOCIETY

HANKEN

Essays on Volatility Modeling

ALEXANDER BACK

Ekonomi och samhälle
Economics and Society

Skrifter utgivna vid Svenska handelshögskolan
Publications of the Hanken School of Economics

Nr 389

Alexander Back

Essays on Volatility Modeling

Helsinki 2024

Essays on volatility modeling

Keywords: Volatility, asymptotic theory, GARCH, stochastic volatility, maximum likelihood, misspecification testing

Supervised by
Niklas Ahlgren
Associate Professor
Hanken School of Economics, Finland

Timo Teräsvirta
Professor Emeritus
Aarhus BSS, Aarhus University, Denmark
C.A.S.E., Humboldt-Universität zu Berlin, Germany

Opponent
Yongmiao Hong
Professor Emeritus
Cornell University, United States
Distinguished Professor
University of Chinese Academy of Sciences, China

Alexander Back
Hanken School of Economics
Department of Finance and Economics
Subject: Finance
P.O.Box 479, 00101 Helsinki, Finland

Hanken School of Economics

ISBN 978-952-232-526-6 (Printed)
ISBN 978-952-232-527-3 (PDF)
ISSN-L 0424-7256
ISSN 0424-7256 (Printed)
ISSN 2242-699X (PDF)

Grano Oy, Helsinki 2024

To the interested reader

ACKNOWLEDGEMENTS

One of the main things I have learned as a doctoral student is that the process of doing research can appear highly nonlinear. Eight hours of input does not seem to equal eight hours of output. A workday may even yield negative output in the sense that you realize that you are further away from achieving your goal than you thought you were when you had your morning coffee. A more accurate description of reality might thus be that *perceived output* is nonlinear, while actual output is linear. You might not have much to show for every sentence you have deleted, every note you have thrown in the garbage and every ostensibly relevant article you have read, but it is all part of the process. The wheels of progress keep turning as long as you keep at it, albeit squeakily at times.

People tend to like things that are predictable and linear. Perception is often conflated with reality, so the lines between perceived and actual output can become blurred, at which point it is hard to see that you are making progress. Therein lies the struggle and therefore the education. Luckily, I have had people to help me along the way. I am thankful to my thesis supervisors Niklas Ahlgren and Timo Teräsvirta, who believed in me enough to include me in their research project on volatility modeling. The project turned out to be more challenging than any of us likely imagined at the outset. We worked through a pandemic and a dense literature on asymptotic theory and have now completed three articles within the scope of the original proposal. Your supervision has aided me in becoming a better econometrician and researcher than I otherwise could have hoped for.

I thank my pre-examiners, Professor Esther Ruiz and Professor Yongmiao Hong, for taking the time to read the manuscript and evaluate it. Your helpful comments have undoubtedly improved the thesis.

I gratefully acknowledge financial support from OP Ryhmän tutkimussäätiö, as well as travel grants from the Hanken foundation and the Graduate School of Finance.

I have had the benefit of having excellent academics as colleagues at Hanken. I am grateful to have worked alongside Niklas Ahlgren, Gunnar Rosenqvist, Jan Antell, Karl Felixson, Eva Liljeblom, Agnieszka Jach, Anders Löflund, Henrik Palmén, Benjamin Maury, Jonas Gustafsson, Markus Belfrage, Theogene Habimana, Sin Thoi Mai, Anh Nguyen, Ville Savolainen, Niclas Meyer, Naval Verma, Ali Malik, Lidan Zhang, Kevin McIntire, Alex Günsberg, Khalela Francis and many others throughout the years. I am especially indebted to my degree supervisor Gonul Colak for always supporting me and teaching me a great deal about research. Your unrelenting work ethic is an inspiration. I also want to thank the Finance department amanuensis Annika Bärlund, librarian Mona Weckström and the other people in the Hanken administration for kindly assisting me at various stages of this process.

I have greatly enjoyed being a student at the Graduate School of Finance. The standard at the GSF courses and events is truly excellent. A special thanks to Mikko Leppämäki and Reetta Ali-Alha for putting so much thought and effort into the organization, to Matthijs Lof and Henri Nyberg for discussing my papers at the GSF workshops, and to

Peter Nyberg for the especially inspiring lectures on theoretical and empirical asset pricing.

Hanken has not only made me a better researcher but also a better teacher. I want to thank Gunnar Rosenqvist who first recruited me to be a teaching assistant in his course on statistics. Teaching exercise sessions in the course is what first made me seriously consider pursuing a postgraduate degree. I am also indebted to Niklas Ahlgren, Jan Antell and Karl Felixson for the valuable experience I have gained by being part of their courses.

I would like to thank my mom, Lotta, my dad, Tom, and my brother, Felix, for their unwavering support throughout my life and studies. Good parents teach their kids to stay in school. At age 31, I realize that I might have taken the wisdom to heart a bit too literally. Still, I am happy with my decision, and I credit my family with instilling enough confidence in me to be able to make it. I am also grateful for my supportive family in Österbotten, Ösmo, Öland and Åland.

I thank my partner, Jenny, for everything. I tried to think of a better way of phrasing this but could not find anything sufficiently encompassing. I love getting to spend my time off work with you and Hobbe, who throughout the years of remote meetings has learned more about econometrics than any other Springer Spaniel on the planet. I am also thankful for my entire extended family. You have always accepted and supported me and thereby made Borgå and Lovisa feel like home.

Finally, I thank all my friends in Österbotten, Nyland and elsewhere.

Borgå, July 2024
Alexander Back

CONTENTS

PART A: SUMMARY

PART B: ESSAYS

PART A: SUMMARY

1 INTRODUCTION

There is a notable disparity between how economists approach the modeling of the first two moments of financial returns. Returns typically exhibit little to no autocorrelation, a fact that is often attributed as being due to a weak form of the efficient market hypothesis (cf. Fama, 1970). Consequently, models of the conditional mean of returns that only rely on past returns are rarely given serious consideration. Instead, mean returns are often modeled by some factor model that involves theoretically motivated inputs.

Squared returns, however, are typically highly autocorrelated. This has led to a voluminous and fruitful literature on volatility modeling. Broadly speaking, there are two main classes of volatility models: Autoregressive conditional heteroskedasticity (ARCH) models and stochastic volatility models. The autoregressive stochastic volatility (ARSV) model is due to Taylor (1986). The model features a stochastic error in the equation describing the evolution of the variance, which makes the likelihood of the model a latent quantity. This presents challenges in terms of estimation and inference. Still, the theoretical underpinnings of the model make it empirically interesting, and several authors have made compelling arguments in its favour (see e.g. Carnero et al., 2004). The original ARCH model is due to Engle (1982). It can be interpreted as an autoregressive (AR) model in squared returns. Its generalization, Generalized ARCH (GARCH), was proposed by Bollerslev (1986) and can be interpreted as an autoregressive moving average (ARMA) model in squared returns. Importantly, in the GARCH class, volatility is deterministic conditional on current information. This specification leads to a standard likelihood framework for estimation and inference, which has been a factor contributing to its popularity.

In the first three papers of this thesis, we apply the standard likelihood framework to an empirically motivated extension of the GARCH model. We propose augmenting the GARCH conditional variance by a time-varying intercept. This specification entails mainly two theoretical challenges. Firstly, the time-variation in the intercept makes the model nonstationary. This complicates the proofs of the two results that typically function as cornerstones of inference in time series models: consistency and asymptotic normality of the (quasi-) maximum likelihood estimator (QMLE). In the stationary case, these results are known to hold under mild conditions. In the first paper, we use the theory of locally stationary processes to prove that the results continue to hold for our model under slightly stronger but empirically reasonable assumptions. The second challenge is that the parametric function that we use to model the time-variation in the intercept is unidentified if the intercept is constant. This makes it necessary to test for a type of additive misspecification before attempting to fit our proposed model. However, the identification problem complicates this type of inference. The second and third papers propose solutions.

In the last paper, which is single authored, I consider a stochastic volatility model that is motivated by financial theory. GARCH and stochastic volatility models typically exploit autocorrelation in squared returns to predict that volatility tomorrow is some function of what it has been in the past. The underlying causal reasons for increases and decreases in volatility are left unexplored, leading to the aforementioned disparity between models of the conditional mean and of the conditional variance. A recent vein in the volatility modeling literature has, however, taken some steps to remedy this by using a common and theoretically supported hypothesis: financial leverage makes equity more risky. I use this observation to propose a new model. As the model is of the stochastic volatility type, inference is more complicated than in the GARCH case. The model posits latent and stochastic volatility, which means that the likelihood function cannot be written down in closed form. However, several methods have been proposed to deal with this. By a close study of the literature on estimating stochastic volatility, I find that the methods of quasi maximum likelihood (QML, Nelson, 1988, Harvey et al., 1994) and Monte Carlo maximum likelihood (MCL, Durbin and Koopman, 1997, Sandmann and Koopman, 1998) work well in my case. I use these methods to develop a framework for estimation and misspecification testing. I provide two empirical examples that highlight the utility and relevance of the model.

The literature on GARCH and stochastic volatility is too vast to be surveyed exhaustively. In what follows, I shall review some theory and concepts that have been especially important in our work. In the final section, I provide a summary of each paper included in the thesis.

2 THEORY AND LITERATURE REVIEW

In this section, I review the theory that underlies the papers in this thesis. Concomitantly, I provide a literature review. I start by covering GARCH models and how they can be estimated by QML. Along with a brief introduction to the theory of locally stationary processes, this provides a backdrop for the first paper. I continue by discussing the problem of testing models that feature unidentified nuisance parameters under the null hypothesis. This is relevant for the second and third papers. Lastly, to introduce the reader to the material in the fourth paper, I discuss stochastic volatility.

2.1 GARCH

In the (G)ARCH(p,q) ("Generalized Autoregressive Conditional Heteroskedastictity") class of models, the basic equation that descibes how returns evolve over time is given by:

$$X_t = \sigma_t \varepsilon_t, \tag{1}$$

where ε_t is typically assumed to be independent and identically distributed (IID) with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t^2) = 1$. The term σ_t describes the time t conditional volatility. By modifying its definition, we obtain different types of models in the ARCH class. The standard ARCH(p) model by Engle (1982) is given by (1) with

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \tag{2}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0, i = 1, 2, \ldots, p$. The model was *generalized* to GARCH by Bollerslev (1986):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3}$$

where $\alpha_0 > 0, \alpha_i \geq 0, i = 1, 2, \ldots, p, \beta_j \geq 0, j = 1, 2, \ldots, q$. The GARCH$(p, q)$ model is weakly stationary if

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1.$$

Since the conditional variance σ_t^2 in (3) is an unobserved quantity, the theory of estimation requires a scheme for generating the process. More specifically, σ_t^2 is dependent on $\sigma_{t-1}^2, \sigma_{t-2}^2, \ldots, \sigma_{t-q}^2$, as well as $X_{t-1}^2, X_{t-2}^2, \ldots, X_{t-p}^2$. Therefore, assuming our process starts at $t = 1$ and the first observed value occurs at $t = 0$, we need q initial values for $\left\{\sigma_t^2\right\}$ and $p - 1$ initial values for $\left\{X_t^2\right\}$ to generate the recursion.

GARCH models have gained popularity throughout the years. New research has been concerned mainly with extending the model to account for a wider class of data generating processes and features in empirically observed time series, but also with deriving theoretical results for exisiting models. An example of a popular extension of the standard GARCH is the GJR-GARCH(1,1) by Glosten et al. (1993), which is given by (1) with

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^{2+} + \alpha_2 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $X^{2+} = X^2 I_{X:x \geq 0}$. The model accounts for the asymmetric effect of negative returns on volatility. Typically, negative returns have a disproportionately large effect on volatility. Black (1976) hypothesized that this is due to financial leverage, so the phenomenon is commonly referred to as the *leverage effect*.

Theoretical developments have consisted of deriving the statistical properties of the model, as well as considering inference. An active area of research has been concerned with the properties of the (quasi-) maximum likelihood estimator of the parameters of the model and its variations. The research is typically focused on finding as weak conditions as possible for obtaining *consistent* and *asymptotically normal* parameter estimates. Early papers in this literature often focused on the GARCH$(1, 1)$-case (see Lee and Hansen, 1994 and Lumsdaine, 1996). For the standard GARCH(p, q) model, the current state of the art can be found in Berkes et al. (2003) and Francq and Zakoïan (2004).

2.2 Estimation by maximum likelihood

To see how maximum likelihood can be used to estimate a GARCH model, consider first a parameter vector $\boldsymbol{\theta} \in \Theta$, where Θ is a parameter space, and a sequence $\{X_t\}$ of data. The method of maximum likelihood is based on making a distributional assumption on the component of a model that is introducing randomness. The distributional assumption gives us a functional form for a likelihood function $f(X_1, X_2, \ldots, X_T; \boldsymbol{\theta})$, where $f(\cdot)$ is a probability density function. We choose the vector $\boldsymbol{\theta}$ so that it maximizes the value of the likelihood function, i.e. the *likelihood* of the model. The log-likelihood is typically used as it is easier to work with: the logarithm is an increasing function, so maximizing the logarithm of a quantity is the same as maximizing the quantity itself. The resulting estimator is denoted $\widehat{\boldsymbol{\theta}}$. If we have T time series observations, we can use the notation $\widehat{\boldsymbol{\theta}}_T$ to emphasize that the estimator depends on the amount of time series observations that we have. In the asymptotic theory, we are considering a sequence $\{\widehat{\boldsymbol{\theta}}_T\}$ of estimators . We use $\boldsymbol{\theta}_0$ to denote the "true" value of the parameter vector. We say that the estimator is strongly *consistent* if

$$\widehat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0$$

almost surely as $T \to \infty$. If the convergence occurs in probability rather than almost surely, the estimator is said to be weakly consistent.

Given IID data, if we choose our objective function $L_T(X_1, X_2, \ldots, X_T; \boldsymbol{\theta})$ (the function that we maximize) as the average log-likelihood, i.e.

$$L_T(X_1, X_2, \ldots, X_T; \boldsymbol{\theta}) = \frac{1}{T}\sum_{n=1}^{T} \log f(X_T|\boldsymbol{\theta}),$$

then the QMLE is an example of an *M-estimator*. There exist theorems stating general conditions under which M-estimators are consistent and asymptotically normal. Proving these properties is therefore often a matter of verifying that the conditions hold. For a GARCH model, and most other time series models, the data are not IID. We will therefore be dealing with the *conditional log-likelihood*. For notational simplicity, consider a GARCH(1,1) given by (1) and

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$. The likelihood can be factorized as follows:

$$\begin{aligned} f(X_1, \ldots, X_T; \boldsymbol{\theta}) &= f(X_1, X_2, \ldots, X_T|X_0; \boldsymbol{\theta}) \\ &= f(X_1|X_0) f(X_2, X_3, \ldots, X_T|X_1, X_0; \boldsymbol{\theta}) \\ &= \prod_{t=1}^{T} f(X_t|X_{t-1}, \ldots, X_0; \boldsymbol{\theta}) \\ &= f(X_0; \boldsymbol{\theta}) \prod_{t=1}^{T} f(X_t|X_{t-1}; \boldsymbol{\theta}). \end{aligned}$$

Now, in order to obtain an estimator, we need to make a distributional assumption on $\{\varepsilon_t\}$. It is common to assume a standard normal distribution. If the assumption is valid, then the estimator is called a maximum likelihood estimator. If it is false, it may still be possible to obtain a consistent and asymptotically normal estimator. This estimator is called a quasi maximum likelihood estimator. Recall that the $N(\mu, \sigma^2)$ density is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Assuming a GARCH(1,1) model and a normal distribution implies that, conditional on time $t-1$ information, $X_t = \varepsilon_t \sigma_t \sim N(0, \sigma_t^2)$, which gives

$$f(X_t | X_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{X_t^2}{2\sigma_t^2}\right).$$

Plugging this into the conditional likelihood yields

$$f(X_0, \ldots, X_T; \boldsymbol{\theta}) = f(X_0; \boldsymbol{\theta}) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{X_t^2}{2\sigma_t^2}\right).$$

Now we drop $f(X_0; \boldsymbol{\theta})$. This will not affect the asymptotic results. Taking the logarithm yields our log-likelihood function

$$L(X_t, \boldsymbol{\theta}) = \frac{1}{2T} \sum_{t=1}^{T} \left(-\log(2\pi) - \log(\sigma_t^2) - \frac{X_t^2}{\sigma_t^2} \right).$$

Our parameter vector is $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1)$. We go on to choose $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} L(X_t, \boldsymbol{\theta})$. This is done numerically using some optimization algorithm. It is common to exclude constants that do not affect the optimization problem. The time t term in the sum in the objective function is then

$$l_t(\boldsymbol{\theta}) := -\left(\log(\sigma_t^2) + \frac{X_t^2}{\sigma_t^2} \right). \tag{4}$$

Francq and Zakoïan (2004) give general and weak conditions under which the QMLE of GARCH processes is consistent and asymptotically normal. They show that the results hold when the GARCH recursion is initialized e.g. using the first observation of the process. They also obtain similar results for ARMA-GARCH processes. Under the following conditions, Francq and Zakoïan (2004, Theorem 2.1) showed that for a GARCH(p, q) model,

$$\widehat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0 \text{ almost surely as } T \to \infty.$$

(I) $\boldsymbol{\theta}_0 \in \Theta$ and Θ is compact.

(II) The top Lyapunov exponent (cf. Francq and Zakoïan, 2004) $\gamma < 0$ and for all $\boldsymbol{\theta} \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

(III) ε_t^2 has a non-degenerate distribution with $\mathbb{E}(\varepsilon_t^2) = 1$.

(IV) If $q > 0$, the polynomials $\mathcal{A}_{\boldsymbol{\theta}_0}(z) = \sum_{i=1}^{p} \alpha_{0i} z^i$ and $\mathcal{B}_{\boldsymbol{\theta}_0}(z) = 1 - \sum_{j=1}^{q} \beta_{0j} z^j$ have no common root, $\mathcal{A}_{\boldsymbol{\theta}_0}(1) \neq 0$ and $\alpha_{0p} + \beta_{0q} \neq 0$, where subscript 0 is used to denote the parameter evaluated at the "true" value.

Under two additional assumptions, Francq and Zakoïan (2004, Theorem 2.2) show that the estimator is asymptotically normal, i.e.,

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \to N\left(\mathbf{0}, (1-\kappa_\varepsilon)\,\boldsymbol{V}^{-1}\right) \text{ as } T \to \infty$$

where κ_ε is the kurtosis of the error term ($\mathbb{E}\varepsilon_0^4$), and

$$\boldsymbol{V} := \mathbb{E}\left(\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\right).$$

The additional assumptions are:

(V) $\boldsymbol{\theta}_0$ is in the interior $\mathrm{int}(\Theta)$ of Θ.

(VI) $\mathbb{E}(\varepsilon_t^4) < \infty$.

Consistency is often proved by verifying high-level conditions that are stated for a more general M-estimator. The conditions can vary slightly depending on the intended application. The following theorem, which can be found in Amemiya (1985, Theorem 4.1.1), gives weak consistency under a commonly used version of the conditions. It can be used to illustrate roughly from where the assumptions in Francq and Zakoïan (2004) stem. Let k be the number of parameters in a model.

Theorem 1 (Amemiya, 1985, Theorem 4.1.1)**.** *Suppose that*

(C1) The parameter space Θ is a compact subset of $\mathbb{R}^k$.

(C2) The objective function $Q_T(\boldsymbol{\theta})$ is a measurable function of the data for all $\boldsymbol{\theta} \in \Theta$, and $Q_T(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} \in \Theta$.

(C3) $Q_T(\boldsymbol{\theta})$ converges uniformly in probability to a non-stochastic function $Q(\boldsymbol{\theta})$, and $Q(\boldsymbol{\theta})$ attains a unique global maximum at $\boldsymbol{\theta}_0$.

Then

$$\widehat{\boldsymbol{\theta}}_T \xrightarrow{P} \boldsymbol{\theta}_0.$$

Note that C1 is assumed in (I). C2 is a mild regularity condition and is often seen to hold easily for the Gaussian log-likelihood. C3 requires the most effort. Its verification can often be split into three parts:

1. Use a law of large numbers to prove that the log-likelihood converges pointwise to its expectation.

2. Strengthen the condition to hold uniformly by using a criterion known as *stochastic equicontinuity.*

3. Show that the log-likelihood attains a unique global maximum at $\boldsymbol{\theta}_0$.

Condition (II) in Francq and Zakoïan (2004) ensures that the GARCH(p, q) process admits a *strictly stationary* solution, which is the main tool that enables an application of the ergodic theorem to the log-likelihood. This is used to verify the condition in the first step above. Francq and Zakoïan (2004) do not appeal to stochastic equicontinuity but use a different, more direct argument to verify that the convergence is uniform. For an equicontinuity argument in a similar context, see Berkes et al. (2003, Lemma 5.4). Assumptions (III) and (IV) in Francq and Zakoïan (2004) are identification conditions and are used to prove that the third step of the verification of C3 holds.

For asymptotic normality, note that the mean value theorem gives, for some $\bar{\boldsymbol{\theta}}$ between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$,

$$\frac{\partial l_T(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial l_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \frac{\partial^2 l_T(\bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\intercal}.$$

At the optimal solution, the left hand side is zero. Multiplying by $\sqrt{T}$ and some algebra yields

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) = -\left[\frac{1}{T}\frac{\partial^2 l_T(\bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\intercal}\right]^{-1} \frac{1}{\sqrt{T}}\frac{\partial l_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}.$$

Now we see that by Slutsky's theorem, if the first term converges in probability to some non-singular matrix, and the second term converges to a multivariate random normal variable, then the resulting distribution will be multivariate normal. These convergence results require the strengthened moment condition (VI) and the increased regularity imposed by (V). In the GARCH(1,1) case, Lumsdaine (1996) established similar asymptotic results under a stronger moment assumption based on the existence of the first 32 moments of the innovations.

2.3 Local stationarity

Verification of the high-level assumptions in the previous section requires mainly two tools from probability theory: a law of large numbers (LLN) and a central limit theorem (CLT). As the standard GARCH process analyzed by Berkes et al. (2003) and Francq and Zakoïan (2004) is strictly stationary, applicable versions of LLNs and CLTs are readily available. In this thesis, however, the GARCH process that we analyze is nonstationary. To overcome this, we will use the theory of *locally stationary processes*. A locally stationary process is a stochastic process with moments that change sufficiently slowly. Formally, the process $\{X_{t,T}\}$ is said to be locally stationary if (see e.g. Dahlhaus and Subba Rao, 2006)

$$X_{t,T} = \widetilde{X}_t(u) + O_P\left(|t/T - u| + \frac{1}{T}\right), \tag{5}$$

where $u \in [0, 1]$ and $\widetilde{X}_t(u)$ is a stationary approximation of $X_{t,T}$ at u. The double subscript t/T is referred to as *rescaled time* and is the device that facilitates the asymptotic theory. To gain some intuition on how the concept works, imagine taking a long time series and cramming it into the unit interval. As the length of the time series increases, the points in the interval become horizontally closer to each other. If the time series is nonstationary but changing slowly, as points are added it will look increasingly stationary in "zoomed-in" sub-intervals containing some fixed number of points.

The idea of approximating a nonstationary process locally by a stationary process dates back to Priestley (1965). The idea has since spawned a substantial literature. Early papers largely focused on linear models (see e.g. Dahlhaus, 1997, 2000). See Dahlhaus (2012) for a review of the early literature. More recently, several authors have used the theory of locally stationary processes to analyze non-linear processes such as GARCH. For example, Dahlhaus and Subba Rao (2006) study locally stationary ARCH processes and discuss how to estimate their parameters. Subba Rao (2006) conducts a theoretical investigation of the general class of non-linear locally stationary processes to which GARCH belongs. Chen and Hong (2016) expand on the theory in the aforementioned papers to propose methods for fitting and evaluating the need for time-varying parameter GARCH models (tvGARCH). An important and relatively recent paper in this literature is Dahlhaus et al. (2019). In the paper, the authors provide conditions under which general non-linear locally stationary processes satisfy a global LLN and CLT. The main requirement is a smoothness condition imposed on the function that generates the process. The smoothness condition is formulated in terms of *Hölder continuity*: A function f is said to be Hölder continuous if for any two points a and b in its domain it holds

$$|f(a) - f(b)| \leq C\, |a - b|^{\rho}\,, \tag{6}$$

where $\rho > 0$ and $C \geq 0$. If $\rho = 1$, the function is *Lipschitz continuous.* Note that by the triangle inequality, we can perform the decomposition

$$\left| X_{t,T} - \widetilde{X}_t(u) \right| \leq \left| X_{t,T} - \widetilde{X}_t(t/T) \right| + \left| \widetilde{X}_t(t/T) - \widetilde{X}_t(u) \right|.$$

A smoothness condition of the type (6) is then used to show that

$$\mathbb{E} \left| X_{t,T} - \widetilde{X}_t(u) \right| \leq C_1 \left(\frac{1}{T} \right) + C_2\, |t/T - u|\,, \tag{7}$$

for constants C_1 and C_2. The inequality (7) can now be compared with the definition (5) to provide intuition. For a more detailed argument, see Subba Rao (2006) and Dahlhaus et al. (2019).

2.4 Testing when nuisance parameters are unidentified under the null

Consider the non-linear regression model

$$Y_t = \boldsymbol{\beta}^{\intercal} \boldsymbol{X}_t + \lambda G(\boldsymbol{Z}_t, \boldsymbol{\pi}) + \varepsilon_t, \tag{8}$$

where ε_t is an IID error term, $\boldsymbol{X}_t$ is a vector of regressors and $G(\cdot)$ is some nonlinear function of another vector of regressors $\boldsymbol{Z}_t$. The parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\intercal}, \lambda, \boldsymbol{\pi}^{\intercal})^{\intercal}$. Under many parameterizations of $G(\cdot)$, the parameters in $\boldsymbol{\pi}$ will be unidentified if $\lambda = 0$. Intuitively, if $\lambda = 0$, then it does not matter what parameter vector $\boldsymbol{\pi}$ we choose for the function $G(\boldsymbol{Z}_t, \boldsymbol{\pi})$: it will have no impact the regression relationship. However, it is often of interest to test the null hypothesis $\lambda = 0$, i.e to test a parsimonious linear model against a non-linear alternative. Due to the unidentified parameters, standard methods cannot be applied to construct misspecification tests. The problem of constructing tests when nuisance parameters are unidentified under the null was first considered by Davies (1977) in the normally distributed test statistic case, and in the χ^2-case by Davies (1987).

A common solution in the literature has been to Taylor expand the function $G(\cdot)$ around the null hypothesis. The approximating Taylor polynomial has linear coefficients that are identified and equal to zero under the null. The asymptotic distribution of the resulting test statistic is thus standard. This type of test was suggested by Luukkonen et al. (1988). In particular, an LM-type test based on a third order Taylor polynomial will lead to a statistic with an asymptotic $\chi^2(3)$-distribution. Due to the computational simplicity of this procedure, it remains a popular way of devising misspecification tests in non-linear models. In the second paper of the thesis, we use this method to construct tests against the model that we propose in the first paper.

Another common solution to the identification problem is to maximize a suitable test statistic over fixed values of the unidentified nuisance parameters. A test constructed in this way is called a *sup test*. Typically, for each fixed $\boldsymbol{\pi}$, the resulting test statistic will have a standard asymptotic distribution. However, as we are taking the maximum over a sequence of test statistics, the resulting statistic will have a non-standard asymptotic distribution. Hansen (1996) examines this problem in depth and proves several useful results for conducting sup tests in regression models. The author shows that under regularity conditions, for common test statistics such as Wald and LM tests, which in standard cases are asymptotically distributed as $\chi^2(p)$, where p denotes the degrees of freedom of the distribution, the asymptotic test statistic will follow a distribution that is a *functional* of a χ^2-process. The test is carried out many times with varying values of the unidentified parameters, so the resulting sequence of test statistics can be thought of as a dependent stochastic process in $\boldsymbol{\pi}$. The asymptotic distribution is usually not available in closed form, so it is approximated by simulation. Hansen (1996) suggests a procedure for simulating the asymptotic distribution and proves that it will yield asymptotically correct p-values.

All things considered, the procedure for conducting a sup test is computationally demanding. First, the test statistic is found by maximization over a possibly multidimensional subspace of the parameter space. Second, its asymptotic null distribution is approximated by repeating the maximization problem many times for simulated draws of null statistics. It is therefore a great advantage if the estimation required to calculate the test statistic can be carried out by ordinary least squares (OLS): numerical maximization is slower and subject to potential errors due to e.g. local optima in some DGPs. Note that for $\boldsymbol{\pi}$ fixed, the model (8) can be fitted by OLS. This observation will be important in the third paper of the thesis, where we construct a sup test for testing a standard GARCH model against a model of the type that we propose in the first paper.

2.5 Stochastic volatility

The basic autoregressive stochastic volatility model of order one (ARSV(1), Taylor, 1986) is given by

$$X_t = \bar{\sigma} e^{\frac{1}{2} h_t} \varepsilon_t, \tag{9}$$

$$h_t = \beta_1 h_{t-1} + \sigma_\eta \eta_t, \tag{10}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are IID standard normal sequences of random variables and $\bar{\sigma}, \beta_1$ and σ_η are parameters. It is assumed that $\sigma_\eta > 0$. The model is weakly stationary under the assumption $|\beta_1| < 1$.

Like GARCH, the ARSV model is used to model heteroskedastic time series such as stock returns. There are mainly three differences between GARCH and ARSV models. First, notice that (10) is entirely latent: it does not contain any terms involving the observed returns. Second, it features a stochastic innovation η_t. Third, the term h_t enters as an exponent in (9), meaning that the latent component can be interpreted as modeling the logarithm of the conditional variance.

Estimation is complicated by the fact that the log-likelihood of the model given by (9) and (10) contains latent terms. It is therefore not possible to simply write down an analytic expression for the log-likelihood and maximize it. To overcome this, several methods of estimation have been proposed in the literature. A common starting point in many of these methods is to log-linearize the model. Squaring and taking logarithms of (9), we obtain

$$\log(X_t^2) = \log(\bar{\sigma}^2) + h_t + \log(\varepsilon_t^2). \tag{11}$$

Nelson (1988) and Harvey et al. (1994) show that by treating $\log(\varepsilon_t^2)$ in (11) as approximately normal, the ARSV model admits a linear Gaussian state space representation. A QML estimator can then be constructed via the Kalman filter.

The linear Gaussian state space formulation is given by:

$$y_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \tilde{\varepsilon}_t, \tilde{\varepsilon}_t \sim N(0, H_t) \tag{12}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t) \tag{13}$$

with $y_t = \log(X_t^2)$ and

$$\boldsymbol{\alpha}_t = \begin{bmatrix} h_t \\ \log(\bar{\sigma}^2) \end{bmatrix}, \boldsymbol{T}_t = \begin{bmatrix} \beta & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{Q}_t = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\boldsymbol{Z}_t = \begin{bmatrix} 1 & 1 \end{bmatrix}, H_t = \pi^2/2.$$

The Kalman filter is given by

$$\begin{aligned} v_t &= y_t - \boldsymbol{Z}_t \boldsymbol{\alpha}_t, & F_t &= \boldsymbol{Z}_t \boldsymbol{P}_t \boldsymbol{Z}_t^\intercal + H_t \\ \boldsymbol{K}_t &= (\boldsymbol{T}_{t+1} \boldsymbol{P}_t \boldsymbol{Z}_t^\intercal) F_t^{-1}, & \boldsymbol{L}_t &= \boldsymbol{T}_{t+1} - \boldsymbol{K}_t \boldsymbol{Z}_t, \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}_{t+1} \boldsymbol{\alpha}_t + \boldsymbol{K}_t v_t, & \boldsymbol{P}_{t+1} &= \boldsymbol{T}_{t+1} \boldsymbol{P}_t \boldsymbol{L}_t^\intercal + \boldsymbol{Q}_{t+1}. \end{aligned}$$

The log-likelihood emerges as a byproduct of the Kalman filter (see e.g. Durbin and Koopman, 2012, p. 35):

$$\log L(Y|\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} (\log(2\pi) + \log |F_t| + v_t^\intercal F_t^{-1} v_t).$$

The resulting estimator is consistent and asymptotically normal under regularity conditions, but not efficient (Ruiz, 1994). The inefficiency is caused by approximating $\log(\varepsilon_t^2)$ as normal. This was remedied by Sandmann and Koopman (1998), who propose using the method of Monte Carlo maximum likelihood (Durbin and Koopman, 1997) to estimate stochastic volatility models. The method works by using *importance sampling* to integrate the unobserved state vector out of the likelihood in a numerically efficient way. The integration problem is:

$$L(Y|\boldsymbol{\theta}) = \int p(Y, \alpha|\boldsymbol{\theta}) \mathrm{d}\alpha, \tag{14}$$

where $p(\cdot)$ is a presumably non-Gaussian density function. Now, pick an approximating Gaussian model and denote the measure of the probability taken with respect to it by $g(\cdot)$. Multiply and divide (14) by $g(\alpha|Y, \boldsymbol{\theta})$ and do some algebraic manipulation:

$$\begin{aligned} L(Y|\boldsymbol{\theta}) &= \int \frac{p(Y, \alpha|\boldsymbol{\theta})}{g(\alpha|Y, \boldsymbol{\theta})} g(\alpha|Y, \boldsymbol{\theta}) \mathrm{d}\alpha \\ &= g(Y|\boldsymbol{\theta}) \int \frac{p(Y, \alpha|\boldsymbol{\theta})}{g(Y, \alpha|\boldsymbol{\theta})} g(\alpha|\boldsymbol{\theta}) \mathrm{d}\alpha \\ &= L_g(Y|\boldsymbol{\theta}) \mathbb{E}_g w(\alpha, Y, \boldsymbol{\theta}). \end{aligned} \tag{15}$$

This shows that the likelihood can be multiplicatively decomposed into two factors. The first factor in (15) is simply the likelihood of the approximating Gaussian model and can thus be easily obtained by the Kalman filter. The second factor is the expectation of a fraction that is referred to as the *importance weight*. Importantly, the expectation is taken with respect to the measure induced by the approximating Gaussian model. This part of the likelihood is evaluated by simulation. Using an algorithm called a "simulation smoother" (see De Jong and Shephard, 1995), it is possible to obtain IID draws of the importance weights $w(\alpha, Y, \boldsymbol{\theta})$. Under regularity conditions, the sample average of these draws will converge to the expectation in (15). Sandmann and Koopman (1998) point out that this decomposition is computationally advantageous as it does not require that the entire likelihood is evaluated by simulation. The importance weights can be interpreted as measuring departures from Gaussianity. Thus, only the part of the likelihood that is not captured by the Gaussian approximating model requires evaluation by simulation.

In practice, the log-likelihood is maximized. This introduces some bias that can be approximately corrected by means of a Taylor expansion (Durbin and Koopman, 1997). The function to be maximized is thus

$$\ln \widehat{L}(Y|\boldsymbol{\theta}) = \ln_g(Y|\boldsymbol{\theta}) + \ln \bar{w} + \frac{s_w^2}{2M}, \tag{16}$$

where $\bar{w}$ is the sample mean of the importance weights, s_w^2 stems from the bias correction and is the sample variance of the importance weights and M is the number of importance samples used when evaluating the density. Sandmann and Koopman (1998) argue that a relatively small number of samples (such as $M = 5$) will suffice to accurately evaluate the log-likelihood. Due to the computational efficiency, e.g. $M = 20$ is reasonable.

3 SUMMARY OF PAPERS

3.1 Paper 1

We propose a GARCH model augmented by a parametric time-varying intercept. We call the model the additive time-varying (ATV-)GARCH model. The intercept is parameterized by a generalized logistic transition function. The model can be intepreted in two ways: as asymptotically equivalent to a multiplicative decomposition of volatility, and as a time-varying (tv)GARCH model. Due to time variation in the parameters, tvGARCH processes are nonstationary. However, under appropriate assumptions (see Subba Rao, 2006), they are *locally stationary.*

Local stationarity enables a meaningful theory of estimation and inference by (quasi-) maximum likelihood (QML). We show that the QML estimator (QMLE) is consistent and asymptotically normally distributed. The standard GARCH QMLE theory is relatively new, with many important contributions during roughly the last 20 years (see Berkes et al., 2003 and Francq and Zakoïan, 2004). We adapt the standard theory to our model. The main problem is that in order to prove consistency and asymptotic normality of the QMLE, one needs a law of large numbers and a central limit theorem. The traditional theory uses the ergodic theorem and a CLT for martingale difference sequences. Due to the nonstationarity of our model, these are not applicable. A first step towards a general theory of non-linear locally stationary processes was taken by Dahlhaus et al. (2019). We use this theory to arrive at our results. To examine the small sample properties of the estimator, we conduct a simulation study. The simulations show that the large sample approximations involved are cogent in moderate sample sizes.

We demonstrate the usefulness of the model by applying it to the returns of the American software company Oracle. We find strong evidence in favour of time-variation in the intercept of the model. As compared to a GARCH(1,1), the fitted ATV-GARCH(1,1) exhibits lower persistence. This is line with a hypothesis put forward by Diebold (1986), who argues that omitted time-variation in the GARCH intercept can overstate estimates of persistence.

3.2 Paper 2

The ATV-GARCH model is unidentified under constancy of the intercept. It is therefore imperative to test for time-variation in the intercept before attempting to fit the model. However, the identification problem complicates inference. In this paper, we solve this problem by deriving LM-type tests against time-variation in the GARCH intercept. The tests are based on the commonly utilized Luukkonen et al. (1988) methodology of approximating the unidentified non-linear function by a Taylor series around the null hypothesis. We consider the method in Wooldridge (1990) to correct for potential distributional misspecification.

We conduct an extensive simulation study to examine the size- and power properties of the test. The time-varying intercept is parameterized by a combination of logistic transition functions. We examine the performance of the test when it is used sequentially to determine the appropriate number of transition functions. We consider three different cases: testing zero against one transition function, zero against two transition functions, and one against two transitions. We find that the test has high statistical power against all specifications except when testing one against two transition functions when both transitions are monotone and in the same direction. The size varies with the persistence of the underlying series but is close to the nominal level for long time series.

We apply the test to returns on the implied volatility index VIX. We find strong statistical evidence in favour of smooth structural change in the form of an increase in the intercept of the VIX conditional variance, leading to an increase in the unconditional volatility.

3.3 Paper 3

We detail a two-step procedure that can be used to conduct misspecification tests against an omitted additive non-linear function in the conditional variance equation of GARCH-type models. First, we specify a null hypothesis and fit a standard GARCH model that agrees with the null. Second, we extend the model by a non-linear function of an additive component and rewrite the formulation in terms of an ARMA model in the squared data. Conditional on the estimate of the conditional variance from the first stage, the resulting model can be fitted by ordinary least squares. We use a Wald-type test to test the null hypothesis of a standard GARCH against the extended formulation. To correct for conditional heteroskedasticity, we consider a robust test. We examine a scenario where the non-linear function of the additive component is unidentified under the null-hypothesis. This means that tests based on standard likelihood inference cannot be easily calculated and feature non-standard asymptotic distributions. A common solution of the identification problem has been to maximize the considered test statistic over the unidentified components. A test statistic obtained in this way is called a supremum or *sup test*.

Sup tests are uncommon in the GARCH literature. One reason for their infrequent use might be that GARCH-type models are usually fitted by numerical optimization methods. In addition, quantities such as the score and the Hessian of the log-likelihood function are needed for inference, and these are also approximated numerically. The maximization of the test statistic over the unidentified parameters can be computationally intensive. Further, the asymptotic distribution of the test statistic usually needs to be approximated by simulation. Thus, the computational burden and amount of approximation might be deemed too great to fit into the sup test methodology. Perhaps consequently, a large part of the theoretical literature on sup tests is based on linear models. By writing our model in its linear ARMA representation, these issues are largely resolved.

We prove that conditional on observing the conditional variance, the test statistic is asymptotically distributed as a functional of a χ^2-stochastic process. In practice, the test relies on an initial estimate of the conditional variance. This is in essence a *generated regressor* problem. In congruence with a large part of the literature on generated regressors, we find that it is necessary to correct inference for sampling variation induced by the first-stage estimation procedure. We propose a simple way of doing so. The resulting test statistic is relatively computationally convenient and works well in practice. We compare the new test to the method proposed in the second paper of this thesis and find that it performs favourably. We use both tests in two empirical applications and conclude that they can be used in conjunction to better inform modeling choices.

3.4 Paper 4

In the fourth paper, which is single authored, I reconsider a parameterization of a relationship between equity volatility, asset volatility and leverage in Engle and Siriwardane (2018), henceforth ES. ES use the Merton (1974) observation that equity can be considered a call option on a firm's assets to derive an approximate relationship expressing equity volatility as a multiplicative decomposition of two factors: asset volatility and a so-called "leverage multiplier". Under economically sensible assumptions, the leverage multiplier is increasing and concave in leverage. The relationship thus captures a notion that is standard in financial economics: leverage makes equity more risky. Contrary to standard volatility models, models in this class are explicitly motivated by financial theory. They connect volatility to the exogenous input debt, over which the managers of a firm exercise discretion. They can thus be used in applications to provide insights beyond the capabilities of standard models.

ES model the asset volatility component as a GARCH process. I argue that the ARSV model might be a better model of asset volatility. In particular, as the market value of assets is a latent process, it makes sense to model its volatility as not depending on any inputs. Further, as the ARSV model is a model of the log-volatility, the multiplicative relationship between asset volatility and leverage becomes log-linearized. The log-linearization makes it possible to write the structural model as a stochastic volatility model with a covariate in the variance equation, which is a specification that has been studied in the literature. Consequently, tools from the standard stochastic volatility framework become available for use.

To improve the empirical utility of the model, I use the log-linearization to devise a way of conducting misspecification tests against volatility asymmetry. I propose using a connection derived by Asai (1998) between stochastic volatility and a logarithmic GARCH model. The author shows that the ARSV model can be interpreted as a log-GARCH model with a fat-tailed innovation distribution. The log-likelihood of the log-GARCH model is observed and easy to deal with. To construct a misspecification test, I fit the null ARSV model, transform it into its log-GARCH representation and derive a regression-based LM test for the transformed model. Asai (2000) uses a similar approach to test for serial correlation in a stochastic volatility framework, but otherwise this method seems unexplored. Standard regression-based LM tests require the assumption that the errors of the model are normally distributed, which is violated due to the fact that the null is stochastic volatility. To correct for this, I use a procedure proposed by Wooldridge (1990) to construct a test against volatility asymmetry in the spirit of Engle and Ng (1993). A simulation study shows that the test does reasonably well in small samples when the signal-to-noise ratio of the data is high or moderate.

The fact that the log-likelihood is unobserved complicates the statistical treatment of the new model. In the case of estimation, several methods have been proposed in the literature to deal with the latent objective function. I find that the method of Monte Carlo maximum likelihood by Sandmann and Koopman (1998) works well in my case. As the method can be used to approximate an ML estimator arbitrarily closely, it is efficient under distributional assumptions. Further, it enables testing nested models against each other by way of a likelihood ratio test. I also consider the inefficient but fast and flexible QML method by Nelson (1988) and Harvey et al. (1994).

I utilize the new model in two empirical applications. First, I use it to test if the leverage effect is due to leverage. In line with previous research, I find that financial leverage does explain a part of the leverage effect, albeit a small one. Second, I consider a case study where I show how a financial manager can use a multivariate structural stochastic volatility model to analyze counterfactual leverage scenarios. In conclusion, the paper provides highly usable tools for practitioners and academics who are interested in modeling volatility while taking leverage into account.

REFERENCES

Amemiya, T. (1985). *Advanced econometrics*. Harvard University Press.

Asai, M. (1998). A new method to estimate stochastic volatility models: A log-GARCH approach. *Journal of the Japan Statistical Society*, *28*(1), 101–114.

Asai, M. (2000). Testing for serial correlation in the presence of stochastic volatility. *Asia Pacific Financial Markets*, *7*, 321–337.

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

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business & Economic Statistics Section, 1976*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Carnero, M. A., Peña, D., & Ruiz, E. (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, *2*(2), 319–342.

Chen, B., & Hong, Y. (2016). Detecting for smooth structural changes in garch models. *Econometric Theory*, *32*(3), 740–791.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, *25*(1), 1–37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, *28*(6), 1762–1794.

Dahlhaus, R. (2012). Locally stationary processes. In T. Subba Rao, S. Subba Rao, & C. Rao (Eds.), *Handbook of statistics. Time series analysis: Methods and applications* (pp. 351–413). Elsevier.

Dahlhaus, R., Richter, S., Wu, W. B., et al. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, *25*(2), 1013–1044.

Dahlhaus, R., & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, *34*(3), 1075–1114.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*(2), 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *74*(1), 33–43.

De Jong, P., & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, *82*(2), 339–350.

Diebold, F. X. (1986). Modeling persistence in conditional variances: A comment. *Econometric Reviews*, *5*(1), 51–56.

Durbin, J., & Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, *84*(3), 669–684.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). OUP Oxford.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1777.

Engle, R. F., & Siriwardane, E. N. (2018). Structural GARCH: The volatility-leverage connection. *Review of Financial Studies*, *31*(2), 449–492.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

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

Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, *64*(2), 413–430.

Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, *61*(2), 247–264.

Lee, S.-W., & Hansen, B. E. (1994). Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, *10*(1), 29–52.

Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, *64*(3), 575–596.

Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, *75*(3), 491–499.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, *29*(2), 449–470.

Nelson, D. B. (1988). *The time series behavior of stock market volatility and returns* (Doctoral dissertation). Massachusetts Institute of Technology.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *27*(2), 204–229.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, *63*(1), 289–306.

Sandmann, G., & Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, *87*(2), 271–301.

Subba Rao, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Advances in Applied Probability*, *38*(4), 1155–1172.

Taylor, S. J. (1986). *Modelling financial time series*. Wiley.

Wooldridge, J. M. (1990). A unified approach to robust, regression-based specification tests. *Econometric Theory*, *6*(1), 17–43.

ALEXANDER BACK

Essays on Volatility Modeling

Time series of squared returns are typically highly autocorrelated. This has led to a voluminous literature on volatility modeling. Broadly speaking, there are two main classes of volatility models: Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models and stochastic volatility models.

In the first three papers of this thesis, my co-authors and I propose and examine an empirically motivated extension of the GARCH model. We augment the GARCH conditional variance by a time-varying intercept. This specification entails mainly two theoretical challenges. Firstly, the time-variation in the intercept makes the model nonstationary. This complicates the proofs of the two results that typically function as cornerstones of inference in time series models: consistency and asymptotic normality of the (quasi-) maximum likelihood estimator. In the stationary case, these results are known to hold under mild conditions. In the first paper, we use the theory of locally stationary processes to prove that the results continue to hold for our model under slightly stronger but empirically reasonable assumptions. The second challenge is that the parametric function that we use to model the time-variation in the intercept is unidentified if the intercept is constant. This makes it necessary to test for a type of additive misspecification before attempting to fit our proposed model. However, the identification problem complicates this type of inference. The second and third papers propose solutions.

In the last paper, I consider a stochastic volatility model that is motivated by financial theory. A commonly accepted and well motivated prediction of financial theory is that financial leverage makes equity more risky. The theory is based around a structural model of the firm in which equity is modeled as a call option on the firm's assets. A recent contribution in the GARCH literature has exploited this to propose a so-called structural GARCH model. I use the same reasoning to propose a parameterization in the stochastic volatility class: a structural stochastic volatility model. As the model is of the stochastic volatility type, inference is more complicated than in the GARCH case. By a close study of the literature on estimating stochastic volatility, I find that the methods of quasi maximum likelihood and Monte Carlo maximum likelihood work well for the proposed model. I use the methods to develop a framework for estimation and misspecification testing. I provide two empirical examples that highlight the utility and relevance of the model.

HANKEN SCHOOL OF ECONOMICS

HELSINKI
ARKADIANKATU 22, P.O. BOX 479,
00101 HELSINKI, FINLAND
PHONE: +358 (0)29 431 331

VAASA
KIRJASTONKATU 16, P.O. BOX 287,
65101 VAASA, FINLAND
PHONE: +358 (0)6 3533 700

BIBLIOTEKET@HANKEN.FI
HANKEN.FI/DHANKEN

EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION OF MBAS ACCREDITED

9 789522 325273

ISBN 978-952-232-526-6 (PRINTED)
ISBN 978-952-232-527-3 (PDF)
ISSN-L 0424-7256
ISSN 0424-7256 (PRINTED)
ISSN 2242-699X (PDF)

GRANO OY