Analysis of Stock Market Volatility by Continuous-time GARCH Models

Gernot Müller[1], Robert B. Durand[2], Ross Maller[3], Claudia Klüppelberg[4]

Abstract

The discrete time ARCH/GARCH model of Engle and Bollarslev has been enormously influential and successful in the modelling of financial data. Recently, Klüppelberg, Lindner, and Maller (2004) introduced the so-called "COGARCH" model as a continuous-time analogue to the GARCH model. Many aspects of the COGARCH have been investigated, including various of its theoretical properties, its relations to other continuous-time models, and the estimation of the parameters in it. We review some of these results in the present paper, and go on to apply the COGARCH to 5-minute data on the S&P500 index, in order to illustrate its ability to analyse stochastic volatility in very high-frequency, irregularly spaced, financial data.

Keywords: COGARCH; Continuous-time Models; Estimation; High-Frequency Data; Maximum Likelihood; Stochastic Volatility;

1 Introduction

1.1 Modelling Market Volatility and the Risk-Return Tradeoff

Understanding the volatility of a market is critical to our understanding of Finance. The *returns* of an equity market as a whole – where the market's returns may be proxied, for example, by the returns on an index such as the S&P500 index – are frequently modelled as a function of investors' expectations of the market's *volatility* (see, for example, Merton 1973, 1980; French, Schwert, and Stambaugh, 1987; Abel, 1988; Barsky, 1989). Ang, Hodrick, Xing, and Zhang (2006) present evidence that the volatility of the market is a candidate for inclusion as an additional factor augmenting standard multi-factor models of the cross-section of stock returns (Fama and French, 1993; Carhart, 1997). In arguing that total risk is priced, Ang et al. (2006) present a considerable challenge to paradigms which argue that only diversifiable,

[1]Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, 85747 Garching, Germany, email: mueller@ma.tum.de

[2]The University of Western Australia, 25 Stirling Highway, Crawley WA 6009, Australia, email: rdurand@biz.uwa.edu.au

[3]School of Finance & Applied Statistics and Center for Mathematics & its Applications, The Australian National University, ACT 0200, Australia, email: ross.maller@anu.edu.au

[4]Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, 85747 Garching, Germany, email: cklu@ma.tum.de

or systematic, risk is required to capture the cross-section of expected equity returns (Sharpe, 1964).

Nevertheless, while the theorized relationship of the risk of a market to its return has attracted considerable attention, empirical support for the relationship has been mixed and disappointing. To quantify such a relationship, the market's volatility must be estimated, in some way, from the market's returns. Analyses such as those of Ang et al. and Durand, Lim and Zumwalt (2007) take an indirect route by using a proxy for expected volatility – the Chicago Board of Options Exchange Volatility Index (the VIX) – which is exogenous to the market. Although the use of a proxy is one way of dealing with the issue, still the question of estimating risk directly from returns is appealing, and conditional volatility models represent a natural choice with which to model it.[5]

Hsieh (1991) represents an early attempt to model the S&P500 using high-frequency data (Hsieh uses 15 minute rather than 5 minute observation intervals), and finds support for an EGARCH (4,4) model. Later, Anderson, Benzoni and Lund (2002), and Eraker, Johannes and Polson (2003), used models admitting stochastic volatility together with jumps both in returns and in volatility, for the S&P500. Lundblad (2007) estimates daily and monthly volatility of returns on the US market using data stretching from the early 1800s to the 1990s (but the S&P500 index was not available over such a long period). While he finds some evidence of a relationship between volatility and return, he is agnostic as to which of the models he uses – GARCH (1,1), TARCH (1,1), QGARCH (1,1) or EGARCH (1,1) – best fits the data.

In the present paper we will also model the volatility of the US market using the S&P500 index to represent the returns of the market. The S&P500 index is perhaps the most widely followed measure of the US equities market, being made up of 500 of the largest stocks in the market (representing around three quarters of all US equities).[6] We are not concerned, however, with running a "horse race" to find a best model. Rather, we demonstrate the applicability of a certain continuous-time GARCH model (the "COGARCH" model of Klüppelberg, Lindner and Maller (2004)) for modelling the time-varying volatility of the S&P500. Maller, Müller and Szimayer (2008) have recently shown how to apply this kind of methodology to describe the volatility of the Australian stock market, using it to analyse ten years of daily data, mostly equally spaced in time, for the ASX200 index. This analysis does not, however, demonstrate the full potential applicability of the COGARCH. Rather than using daily data, the analysis in the present paper will use very high-frequency observations – observations taken at five-minute intervals – to better approximate the underlying continuous-time framework the methodology is designed to capture. The use of COGARCH enables the analysis of irregularly spaced data without recourse to approximations involving missing-values estimation, and, in high-frequency data such as that studied in this chapter, irregular spacing – due to weekends, holidays and breaks between market closing and opening times – is a dominant feature of the data.

[5] We note that such models have recently attracted some criticism (Ghysels, Santa-Clara and Valkanov, 2005; Durham, 2007).

[6] See www.indices.standardandpoors.com for further information on the index.

1.2 ARCH/GARCH and COGARCH Modelling

The ARCH/GARCH model paradigm introduced by Engle (1995) [15] and Bollerslev et al. (1995) [7] has been enormously influential and successful in capturing some of the most important empirical features of financial data, and is, therefore, widely used in finance research and applications. Empirical studies commonly show that volatility changes randomly in time, has distributions with heavy or semi-heavy tails, and clusters on high levels. These stylized features are well modelled by the GARCH family as has been shown in many studies. For a recent discussion concerning the GARCH(1,1) process, see Mikosch and Stărică (2000) [28].

Up till quite recently, stochastic volatility models have been investigated mostly in discrete time, but with modern easy access to voluminous quantities of very high-frequency data, a demand for continuous-time models which allow, for instance, for a more natural analysis of possibly irregularly spaced data, has arisen. A first attempt to create a continuous-time GARCH model dates back to Nelson (1990) [31], who, by taking a limit of a discrete time GARCH process, derived a bivariate diffusion driven by two independent Brownian motions. By contrast, the discrete time GARCH model incorporates only one source of uncertainty. Consequently, Nelson's continuous time limit process does not possess the feedback mechanism whereby a large innovation in the mean process produces a burst of higher volatility, which is a distinctive feature of the discrete time GARCH process. Moreover, the diffusion limit no longer has the heavy tailed distribution of returns needed for realistic modelling of returns in high-frequency financial data.

To overcome these problems, Klüppelberg, Lindner and Maller (2004) [25] suggested an extension of the GARCH concept to continuous-time processes. Their COGARCH (continuous-time GARCH) model is based on a single background driving (continuous-time) Lévy process, which preserves the essential features of discrete time GARCH processes, and is amenable to further analysis, possessing useful Markovian and stationarity properties.

The aim of this paper is to illustrate the advantages of using the continuous time COGARCH model for the analysis of very high frequency, unequally spaced, financial data. We proceed by summarising, in the following section, the currently available literature on the COGARCH, taking a detailed look at the model definition, and also outlining briefly the pseudo-maximum-likelihood method we will use to fit the model to data. Section 3 gives an illustrative data analysis, applying the COGARCH model to high-frequency data from the S&P500 stock market index, over the years 1998 to 2007. The model proves to be remarkably stable and informative. We discuss some extensions of the model, and other issues, in Section 4, and conclude with Section 5.

2 COGARCH: A Summary of the Current Literature

Since its introduction in 2004, many aspects of the COGARCH model have been studied. One field of research covered Markovian and stationarity properties, as well as extremal behaviour of the model, and its relations to other models. Model extensions, such as a COGARCH(p,q), and a multivariate COGARCH model, have been developed. Regarding data analysis, so far a couple of different approaches have been suggested for estimation of parameters.

In this section we summarize the literature on these topics, and briefly sketch the ideas. Throughout, by "COGARCH" we will usually mean the COGARCH(1,1) model, as introduced in Section 2.1, because this is the most widely applied version. Since there exists, analogously to the GARCH(1,1) and GARCH(p,q) models, a generalization to COGARCH(p,q), we will always point this out by adding the complexity tupel (p,q), when referring to this kind of model extension.

2.1 The COGARCH Model and its Theoretical Properties

We first recall the definition of the COGARCH process as introduced in Klüppelberg, Lindner, and Maller (2004) [25]. On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ satisfying the "usual hypothesis" (see Protter (2005) [32], p. 3), one is given a *background driving Lévy process* $L = (L(t))_{t\geq 0}$. See Applebaum (2004) [4], Bertoin (1996) [6], and Sato (1999) [33] for detailed results concerning Lévy processes. Throughout it is assumed that $\mathbb{E}L(1) = 0$ and $\mathbb{E}L^2(1) = 1$.

Given parameters (β, η, φ), with $\beta > 0, \eta > 0, \varphi \geq 0$, and a square integrable random variable (rv) $\sigma(0)$, independent of L, the *COGARCH variance process* $\sigma^2 = (\sigma^2(t))_{t\geq 0}$ is defined as the almost surely unique solution of the stochastic differential equation

$$\mathrm{d}\sigma^2(t) = \beta \mathrm{d}t - \eta\sigma^2(t-)\mathrm{d}t + \varphi\sigma^2(t-)\mathrm{d}[L, L](t),\ t > 0, \tag{2.1}$$

where $[L, L]$ is the bracket process (quadratic variation) of L (Protter (2005) [32], p. 66). Then one defines the *integrated COGARCH process* $G = (G(t))_{t\geq 0}$ in terms of L and σ as

$$G(t) = \int_0^t \sigma(s-)\,\mathrm{d}L(s), \quad t \geq 0. \tag{2.2}$$

As has been shown in Klüppelberg et. al. (2004) [25], Corollary 3.1, the bivariate process $(\sigma(t), G(t))_{t\geq 0}$ is Markovian. Moreover, under a certain integrability condition, and for the right choice of $\sigma^2(0)$, the process $\sigma^2(t)$ is strictly stationary (Klüppelberg et. al. 2004 [25], Thm. 3.2). As a consequence, G has stationary increments. Furthermore, [25] also provides explicit expressions for the moments and autocorrelation functions of the variance process σ^2 and for the increments $G^{(r)}(t) := G(t + r) - G(t)$ of G. These can be used to estimate the COGARCH parameters by the method of moments, cf. Section 2.3.

Fasen, Klüppelberg, and Lindner (2005) [17] show that the COGARCH model, in general, exhibits regularly varying (heavy) tails, volatility jumps upwards, and clusters on high levels. More precisely, it can be shown that both the tail of the distribution of the stationary volatility and the tail of the distribution of $G(t)$ are Pareto-like under weak assumptions (cf. Klüppelberg, Lindner, and Maller (2006) [26]). For more details on the theoretical properties of G and σ^2, we refer to [25], [17] and [26].

2.2 The Relation between GARCH and COGARCH

The discrete time GARCH(1,1) model is specified by the mean and variance equations

$$Y_i = \sigma_{i-1}\varepsilon_i$$

and

$$\sigma_i^2 = a + b\sigma_{i-1}^2\varepsilon_{i-1}^2 + c\sigma_{i-1}^2,$$

for $i = 1, 2, \ldots, n$, with σ_0^2 given, and a, b, c, as parameters. The ε_i are i.i.d. random variables with mean 0 and unit variance. By writing a discretised version of (2.2) in the form

$$G_i - G_{i-1} = \sigma_{i-1}\varepsilon_i,$$

where we replace the increment $dL(t)$ with one of an i.i.d. sequence $\varepsilon_1, \ldots, \varepsilon_n$, and, similarly, a discretised version of (2.1) in the form

$$\sigma_i^2 = \beta\Delta t_i + \varphi\sigma_{i-1}^2\varepsilon_i^2 + (1-\eta)\sigma_{i-1}^2\Delta t_i,$$

we can see a direct analogy between the discrete and continuous time models. In fact, taking $\Delta t_i = 1$ in the continuous time equations (corresponding to equally spaced data), we see that the models differ only by a non-essential alteration in parameterisation.

More generally, in the continuous time model, we do not need to assume equally spaced data. We can proceed as follows. Starting with a finite interval $[0, T]$, $T > 0$, take deterministic sequences $(N_n)_{n\geq 1}$ with $\lim_{n\to\infty} N_n = \infty$ and $0 = t_0(n) < t_1(n) < \ldots < t_{N_n}(n) = T$, and, for each $n = 1, 2, \ldots$, divide $[0, T]$ into N_n subintervals of length $\Delta t_i(n) := t_i(n) - t_{i-1}(n)$, for $i = 1, 2, \ldots, N_n$. Define, for each $n = 1, 2, \ldots$, a discrete time process $(G_{i,n})_{i=1,\ldots,N_n}$ satisfying

$$G_{i,n} = G_{i-1,n} + \sigma_{i-1,n}\sqrt{\Delta t_i(n)}\varepsilon_{i,n},\ i = 1, 2, \ldots, N_n, \tag{2.3}$$

where $G_{0,n} = G(0) = 0$, and the variance $\sigma_{i,n}^2$ follows the recursion

$$\sigma_{i,n}^2 = \beta\Delta t_i(n) + \left(1 + \varphi\Delta t_i(n)\varepsilon_{i,n}^2\right) e^{-\eta\Delta t_i(n)}\sigma_{i-1,n}^2,\ i = 1, 2, \ldots, N_n. \tag{2.4}$$

Here the innovations $(\varepsilon_{i,n})_{i=1,\ldots,N_n}$, $n = 1, 2, \ldots$, are constructed using a "first jump" approximation to the Lévy process developed by Szimayer and Maller (2007) [37], which divides a compact interval into an increasing number of subintervals and for each subinterval takes the first jump exceeding a certain threshold. Finally, embed the discrete time processes $G_{\cdot,n}$ and $\sigma_{\cdot,n}^2$ into continuous-time versions G_n and σ_n^2 defined by

$$G_n(t) := G_{i,n} \quad \text{and} \quad \sigma_n^2(t) := \sigma_{i,n}^2, \quad \text{when } t \in [t_{i-1}(n), t_i(n)),\ 0 \leq t \leq T, \tag{2.5}$$

with $G_n(0) = 0$. The processes G_n and σ_n are in $\mathbb{D}[0, T]$, the space of càdlàg real-valued stochastic processes on $[0, T]$.

Assume $\Delta t(n) := \max_{i=1,\ldots,N_n} \Delta t_i(n) \to 0$ as $n \to \infty$. As one main result of their paper, Maller et. al. (2008) [29] showed then that the discretised, piecewise constant processes $(G_n, \sigma_n^2)_{n\geq 1}$ defined by (2.5) converge in distribution as $n \to \infty$ to the continuous time processes (G, σ^2) defined by (2.1) and (2.2). Further, this result was used by Maller et. al. (2008) [29] to develop a pseudo maximum likelihood estimation procedure for the parameters in the COGARCH. We sketch this method in Subsection 2.3.

2.3 Estimation Procedures

Haug et. al. (2007) [22] suggested moment estimators for the parameters of the COGARCH process based on equally spaced observations. Using the fact that the increments of the COGARCH process are strongly mixing with exponential rate, they showed that the resulting estimators are consistent and asymptotically normal. The paper by Müller (2007) [30] shows that it is also possible to use Bayesian methods to estimate the COGARCH model. At the time of writing, this method is, however, restricted to the case where COGARCH is driven by a compound Poisson process. More generally, Maller et. al. (2008) [29] describe a straightforward and intuitive pseudo maximum likelihood (PML) method based on the GARCH approximation to COGARCH. This estimation procedure is well adapted for the analysis of unequally spaced data, as we will illustrate in the next section. By applying it, we can easily account also for a transformation to a business time rather than calendar time scale. The general strategy is as follows.

Suppose we are given observations $G(t_i)$, $0 = t_0 < t_1 < \ldots < t_N = T$, on the integrated COGARCH as defined and parameterised in (2.1) and (2.2), assumed to be in its stationary regime. The $\{t_i\}$ are assumed fixed (non-random) time points; set $\Delta t_i := t_i - t_{i-1}$. Let $Y_i = G(t_i) - G(t_{i-1})$, $i = 1, \ldots, N$, denote the observed returns. We wish to estimate the parameters (β, η, φ). From (2.2) we can write

$$Y_i = \int_{t_{i-1}}^{t_i} \sigma(s-)\mathrm{d}L(s),$$

and, because σ is Markovian (Klüppelberg et. al. (2004) [25], Thm. 3.2), Y_i is conditionally independent of $Y_{i-1}, Y_{i-2}, \ldots$, given the natural filtration of the Lévy process L, with conditional expectation 0, and a conditional variance given by Eq. (3.2) of [29]. To ensure stationarity, we take $\mathbb{E}\sigma^2(0) = \beta/(\eta - \varphi)$, with $\eta > \varphi$.

We can then apply the PML method, as in [29], assuming at first that the Y_i are conditionally $N(0, \rho_i^2)$, and using recursive conditioning and a GARCH-type recursion for the variance process to write a pseudo-log-likelihood function for $Y_1, Y_2, \ldots, Y_N$. Taking as starting value for $\sigma^2(0)$ the stationary value $\beta/(\eta - \varphi)$, one can maximise the function $\mathcal{L}_N$ in Eq. (3.3) of [29] to get PMLEs of (β, η, φ).

3 Data Analysis

In this section we illustrate how to apply the COGARCH model to some market data. After a brief description of the raw data on the S&P500 index, we discuss data cleaning and preprocessing. Then we report on results of fitting the COGARCH model, using the PML method described in Section 2.3, after rescaling calendar time to a business time scale.

3.1 Description of Raw Data

Intra-day data on the value of the S&P500 index was obtained from the TAQTIC database maintained by the Securities Industry Research Centre of Asia-Pacific (SIRCA). Based on this tick-by-tick index data, we compute and analyse 5-minutes log-returns from the index, separately for the years 1998 to 2007. Table 3.1 reports, for each year, the number of trading days and the total number of observations, i.e. the number of log-returns. On average, we have around 19169 log-returns per year. However, even for years with the same number of trading days, the values can differ quite significantly. The first reason is that on a couple of trading days each year the NYSE opens later or closes earlier for annual memorials and holidays. For example, in Table 3.1 Christmas Eve, as long as it does not coincide with Saturday or Sunday, is counted as a full trading day, although the NYSE closes at 1 pm on December 24. The second reason for irregularities are special events. Here we just note, as an example, one of many which occurred within our time frame since 1998: on September 11, 2002, the NYSE did not open until 11 am due to the memorial events commemorating the one-year anniversary of the attack on the World Trade Center. However, in Table 3.1 this day is again counted as a full trading day. For a complete list of trading hours exceptions at NYSE since 1885 see www.nyse.com/pdfs/closings.pdf. The third reason is that each year a few observations are missing, usually since some data is obviously erroneous, and has to be removed from the data set. However, taking advantage of our continuous-time approach, we will not use interpolation to fill in the missing values, but instead take the time difference to the previous available observation into account.

3.2 Data Cleaning: Local Trends and Local Volatility Weights

Before applying the COGARCH model to the data, one has to think carefully about which features of the data are to be captured by the COGARCH, and which are not. The COGARCH is mainly designed to describe the behaviour of the volatility in the data. Moreover, since we assume the COGARCH parameters β, η, and φ to be constant over time, we must check first whether the data indeed shows a stationary volatility pattern over the whole time frame. In our data, this is definitely not the case. See, for instance, the 2002 returns (c.f. Figure 1, first and second row). Therefore we first pre-process the data by estimating local trends and

year	trading days	observations	year	trading days	observations
1998	245	18934	2003	251	19260
1999	250	19359	2004	249	19373
2000	251	19311	2005	252	19617
2001	245	18468	2006	251	19460
2002	251	19028	2007	251	19458

Table 1: *Number of trading days at NYSE and number of log-returns based on 5-minute index data from the S&P500, for years 1998 to 2007.*

local volatility weights. This procedure is necessary since we have around 20000 observations per year, so that we cannot expect the COGARCH parameters to be constant over the whole time frame. On the other hand, we must be careful not to destroy the volatility structure which we want to describe by the COGARCH model. Therefore we aim at a standardization procedure which uses trends and volatility weights over longer periods such as one month, which corresponds to around 1600 observations in our setup.

In the following we denote the observed log-returns by y_i, $i = 1, \ldots, T$, whereas D denotes the number of trading days in the year under consideration (e.g., in 2002 we have $T = 19028$ and $D = 251$). Next we have to introduce a few functions, to be able to cover all aspects of the data within our subsequent formulas. First, let $d : \{1, \ldots, T\} \to \{1, \ldots, D\}$, $i \mapsto d(i)$, denote a function which returns the trading day for observation i, $A : \{1, \ldots, D\} \to \mathbb{N}$, $d \mapsto A(d)$, denote a function which returns the number of available observations on trading day d, and $N : \{1, \ldots, D\} \to \mathbb{N}$, $d \mapsto N(d)$, a function which returns the number of missing observations on trading day d. On a regular trading day we usually have $A(d) = 78$ and

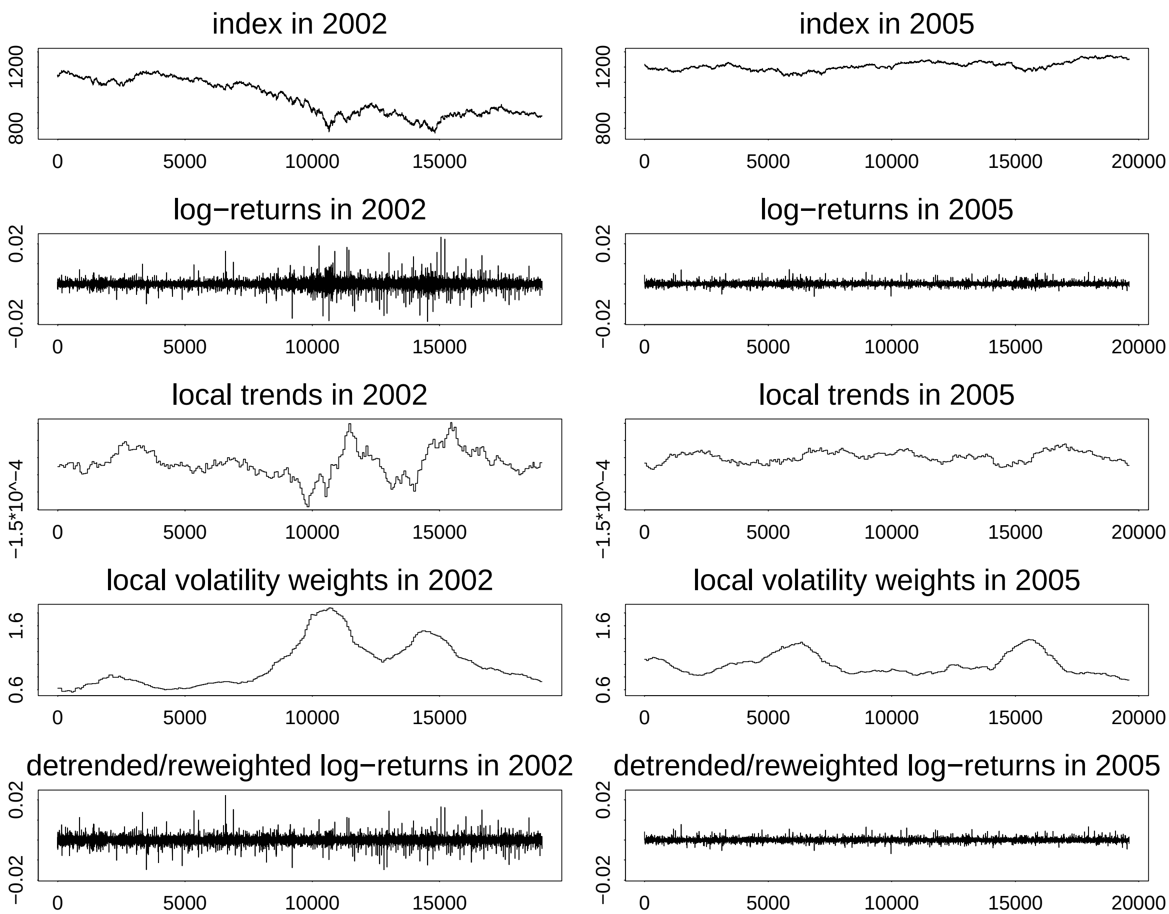

Figure 1: *5 minutely observations from S&P500 index data: log-returns, local trends, local volatility weights, and detrended and reweighted log-returns for years 2002 and 2005.*

$N(d) = 0$, since we take 5 minute data between 9:30 am and 4:00 pm. Note that $N(\cdot)$ can also be 0, when the NYSE opened later or closed earlier, since $N(\cdot)$ counts only missing values when trading really took place at that time. Although only very few observations are missing overall (less than 0.5%), we introduce, to act very precisely, the function $I : \{1, \ldots, T\} \to \mathbb{N}$, $i \mapsto I(i)$, which returns the number of 5-minute intervals elapsed before observation i. Usually $I(i) = 1$, and if, e.g. one observation is missing, $I(i) = 2$, and so on. Later, we will also need the more precise functions $I_k : \{1, \ldots, T\} \to \mathbb{N}$, $i \mapsto I_k(i)$, for $k \in \{9, 10, \ldots, 15\}$, which specify how many 5-minute intervals during trading hour k elapsed before observation i. E.g., if we have an observation at 9:55 am and the value at 10:00 am was deleted, so that the next observation i is from 10:05 am, we have $I(i) = 2$, $I_9(i) = I_{10}(i) = 1$, and $I_k(i) = 0$ for $k = 11, \ldots, 15$.

We now assume that the log-returns follow the model

$$y_i = m_{d(i)} + v_{d(i)} x_i, \qquad i = 1, \ldots, T,$$

where $m_{d(i)}$ represents a *local trend*, $v_{d(i)}$ a *local volatility weight*, and the x_i are detrended and locally reweighted log-returns. This approach takes irregularities of the stock prices and index data into account, which cannot be captured by the COGARCH model. Both local trends and volatility weights are assumed to be constant over trading days, and are estimated as follows. Using a fixed $M \in \mathbb{N}$, the local trends $m_{d(i)}$ are estimated as moving averages over $2M + 1$ trading days:

$$\widehat{m}_{d(i)} = \left[\sum_{d=d(i)-M}^{d(i)+M} (A(d) + N(d))\right]^{-1} \sum_{d=d(i)-M}^{d(i)+M} \sum_{\{j|d(j)=d\}} y_j, \qquad M < d(i) \leq D - M.$$

For the cases $d(i) \leq M$ and $d(i) > D - M$ we can again employ this formula by using data from the previous and following year, respectively, or else we can shrink the time frame of estimation. Note that the addition $A(d) + N(d)$ makes sense, since log-returns are additive.

Similarly, the volatility weights are estimated, for some $V \in \mathbb{N}$, by computing preliminary weights

$$\widehat{v}^*_{d(i)} = \left[\sum_{d=d(i)-V}^{d(i)+V} (A(d) + N(d))\right]^{-1} \sum_{d=d(i)-V}^{d(i)+V} \sum_{\{j|d(j)=d\}} |y_j - \widehat{m}_{d(j)}|, \quad V < d(i) \leq D - V,$$

and then by reweighting these according to

$$\widehat{v}_{d(i)} = \frac{\sum_{j=1}^{T} |y_j - \widehat{m}_{d(j)}| / \widehat{v}^*_{d(j)}}{\sum_{j=1}^{T} |y_j - \widehat{m}_{d(j)}|} \, \widehat{v}^*_{d(i)}.$$

This implies that for $\widehat{x}_i := (y_i - \widehat{m}_{d(i)}) / \widehat{v}_{d(i)}$ we have $\sum_{i=1}^{T} |\widehat{x}_i| = \sum_{i=1}^{T} |y_i - \widehat{m}_{d(i)}|$, so that the magnitude of the values is preserved.

In our analysis, we set $M = V = 10$, so that we use a time frame of 21 trading days to determine the local trend and volatility weight. This seems to be a reasonable choice since we usually get larger standard errors for the COGARCH parameter estimates for very large or very small values of M and V. In Figure 1, as an example, the third and fourth rows show the estimated local trends and estimated local volatility weights for the years 2002 and 2005. The fifth row shows the detrended and reweighted log-returns. Note that, for each row, the left figure for 2002 has the same scale as the right figure for 2005. We chose the years 2002 and 2005 for illustration since they apparently show quite different patterns; however, the cleaning and pre-processing procedure is the same for all years. We emphasize once more that both $m_{d(i)}$ and $v_{d(i)}$ depend only on the $d(i)$, so that all observations of the same day are reweighted by the same weight. This way we do not lose information about the dependence of the volatility on the exact trading time during the day. This dependence is accounted for in the following subsection.

3.3 Accounting for Trading Time by Transformation of Time

To take the possible impact of trading time on the volatility into account, we first conducted a standard regression analysis for the squared log-returns to check for explanatory variables having an influence on the volatility. We used indicator variables for the month January, for all weekdays from Monday to Friday, and for all trading hours $(9, 10, 11, \ldots, 15)$. For reasons of identifiability we had to remove one weekday and one trading hour (we chose Friday and hour 15), so that this sums up to a collection of 11 variables. The four most significant indicators always turned out to be hours 9, 12, 13, and 10, in this order. All these had p–values of less than 1%.

An easy way to account for these explanatory variables within the COGARCH context is to apply the COGARCH to a fictive business time axis. That means that we do not insert the explanatory variables into the COGARCH model itself, but use the PML method both to rescale the physical time axis and to estimate the COGARCH parameters, simultaneously. The idea is to replace Δt_i in Equations (2.3) and (2.4) by

$$\Delta\tau_i := \Delta t_i I(i) + h_9 I_9(i) + h_{10} I_{10}(i) + h_{12} I_{12}(i) + h_{13} I_{13}(i),$$

where h_9, h_{10}, h_{12}, and h_{13} are unknown parameters. Since we account for missing values within the functions I and I_k, respectively, Δt_i does not depend on i in our setup and serves just as the basic time unit. To get estimates nearly on an annual basis, we choose Δt_i as $1/(\sum I(j))$, since $\sum I(j) = T + \sum N(d) = \sum A(d) + N(d)$. However, since the time axis is transformed during the estimation procedure, values such as the exact annualized volatilities in each year have to be computed separately after the estimation procedure.

Of particular interest are the quantities $f_k := (h_k + \Delta t_i)/\Delta t_i$ for $k = 9, 10, 12, 13$. These report the factors by which the basic time unit has to be rescaled during a certain trading hour to get business time.

3.4 Results and Interpretation

Table 2 reports the PML estimates together with their corresponding standard errors. We first note that our estimates satisfy the stationarity condition $\eta > \varphi$. Moreover, the estimates for η and φ are very similar for all 10 years, with values around 0.20 and 0.10, respectively. Such stability is very reassuring as to the applicability of the model. As a consequence, the parameter β directly reflects the long-run volatility in each year, since the mean of the COGARCH variance equation can be expressed as $\beta/(\eta - \varphi)$, with $\eta - \varphi$ almost constant in our case. For example, from the plots in Figure 1 one can immediately see that in 2002 the overall mean volatility was much higher than in 2005. This conjecture is now confirmed by the estimates of β, which are around 0.0039 in 2002 and only 0.0007 in 2005.

Table 3 contains the factors which have to applied to the physical time axis to get business time. For example, the value 5.3057 of the estimate for f_9 in 1998 means that during this year usiness time was running at around 5.3 times faster than physical time between 9:30am and 10:00am, which reflects the high activity in the market after the opening of the exchange. Between 10am and 11am the activity decreased, but is still higher than on average. In general, during lunch time, between 12pm an 2pm, business time runs slower than physical time, usually the activity is only 50% to 70% of the average observed between 11am and 12pm as well as after 2pm.

4 Discussion

4.1 Extensions: COGARCH(p,q), ECOGARCH, Multivariate Models

There exist a couple of model extensions to the COGARCH model which we briefly mention in this section. For more details we refer to the corresponding papers.

Motivated by the generalization of the GARCH(1,1) to the GARCH(p,q) model, Brockwell, Chadraa, and Lindner [8] introduced the COGARCH(p,q) model. Here the volatility follows a CARMA (continuous-time ARMA) process, which is again driven by a Lévy process. As in the discrete-time case, this model displays a broader range of autocorrelation structures than those of the COGARCH(1,1) process.

Haug and Czado [20] introduce an exponential continuous time GARCH (ECOGARCH) process as analogue to the EGARCH(p,q) models. They investigate stationarity and moments and show an instantaneous leverage effect for the ECOGARCH(p,p) model. In a subsequent paper, Czado and Haug (2008) [11] derive a quasi-maximum likelihood estimation procedure for the ECOGARCH(1,1) model, in the case when it is driven by a compound Poisson process, assuming normally distributed jumps.

In Stelzer (2008a) [35] multivariate COGARCH(1,1) processes are introduced constituting a dynamical extension of normal mixture models and covering again such features as dependence of returns (but without autocorrelation), jumps, heavy tailed distributions, etc. As in the univariate case, the model has only one source of randomness, a single multivariate Lévy process. The time-varying covariance matrix is modelled as a stochastic process in the class of positive semi-definite matrices. The paper analyses the probabilistic properties and gives a sufficient condition for the existence of a stationary distribution for the stochastic covariance matrix process, and criteria ensuring the finiteness of moments.

As for the univariate COGARCH, the multivariate COGARCH can be extended to a multivariate ECOGARCH model, as is done in Haug and Stelzer (2008) [21].

Analoguously to the papers by Szimayer and Maller (2007) [37] and Maller et. al. (2008) [29], Stelzer (2008b) [36] generalises the first jump approximation of a pure jump Lévy process, which converges to the Lévy process in the Skorokhod topology in probability, to a multivariate setting and an infinite time horizon. Applying this result to multivariate ECOGARCH(1,1) processes, he shows that there exists a sequence of piecewise constant processes determined by multivariate EGARCH(1,1) processes in discrete time which converge in probability in the Skorokhod topology to the continuous-time process.

4.2 Other Theory

We pointed out in Section 2.2 the striking similarity between the theoretical formulations of the discrete-time GARCH and the COGARCH volatility equations. In relation to this, Kallsen and Vesenmayer (2008) [24] derive the infinitesimal generator of the bivariate Markov process representation of the COGARCH model and show that any COGARCH process can be represented as the limit in law of a sequence of GARCH(1,1) processes. The result of Maller, Müller, and Szimayer (2008) [29] is even stronger. They approximate the COGARCH with an embedded sequence of discrete time GARCH(1,1) models which converges to the continuous-time model in a strong sense (in probability, in the Skorokhod metric), as the discrete approximating grid grows finer. Whereas the diffusion limit in law established by Nelson (1990) [31] occurs from GARCH by aggregating its innovations, the COGARCH limit arising in Kallsen and Vesenmayer (2008) [24] and Maller et. al. (2008) [29] both occur when the innovations are randomly thinned. We sketched briefly the basic idea of the Maller et. al. (2008) [29] approximation in Section 2.3.

The question arises also, as to how strong this similiarity may be from a statistical point of view, and how this similarity might be measured mathematically. A sophisticated approach to measuring the similarity between two statistical models is Le Cam's framework of *statistical equivalence*. As was shown by Wang (2002) [38], the diffusion limit in law of the GARCH(1,1) established by Nelson (1990) [31] is *not* statistically equivalent to the approximating series of GARCH models. A recent paper by Buchmann and Müller (2008) investigates in detail the possibility of statistical equivalence between the GARCH and the COGARCH models. They show that, if the corresponding volatilities of the COGARCH process are unobservable, the limit experiment is again not equivalent to GARCH in "deficiency". If, however, in addition, full information about the volatility processes is available, then the limiting COGARCH experiment is in fact equivalent to the approximating sequence of GARCH models.

5 Summary and Conclusion

In Section 2 we summarised the literature on the theoretical properties of the COGARCH, on estimation procedures, and on some model extensions which have been proposed. In Section 3 we used the COGARCH model to analyse high-frequency data on the S&P500.

We emphasize once more that the COGARCH model incorporates the most important stylized features of financial data. Since it is a continuous-time model, it provides great flexibility in modeling different aspects of the data. While we are accustomed to seeing financial time-series at relatively low, or coarser, frequencies – such as the closing value of the index reported in nightly news broadcasts – it is relatively easy to forget that such values are merely one point on the high-frequency activity function. High-frequency data readily lends itself to analysis within a continuous time framework. Further, the COGARCH methodology demonstrated in this chapter is also appropriate to model the discontinuities which are a natural, but often ignored, feature of the data; as we have noted, discontinuities arise from breaks such as holidays but also the important break between the market's close and its opening the following morning.

However, we must be careful only to apply the COGARCH model to a stationary time series. This may require pre-processing the data as we did in Section 3.2, before the COGARCH model can be used. The PML method which we employed for fitting the COGARCH to the S&P500 data is applicable in general to irregularly spaced data. In this way we can very easily deal with missing data, or transform to another time scale, as needed to reflect some special properties of the data. In our case, we were able to account for the dependence of market activity on trading time. We have chosen to focus on modeling the volatility of the S&P500 due to its importance as a widely followed measure of the market. Its component stocks are large and liquid and, as a consequence, it is reasonable to believe that the behavior of the S&P500 will not be greatly affected by nonsynchronous trading, or nontrading, of its constituent securities. Nonetheless, the premises of COGARCH suggest that it appropriately accounts for such microstructure issues. Analysis of COGARCH in microstructure research for stocks where discontinuities arise due to non-trading offers interesting and potentially insightful research opportunities.

Although many aspects of the COGARCH have already been investigated, it remains an area of active research. We have demonstrated its implementation and applicability to an important financial time series. While we have used high-frequency data to demonstrate the potential of COGARCH, we do not wish to suggest that its only application is in microstructure research. We are confident that that the technique will become an important, and well-known tool, for research in financial economics.

References

[1] Abel, A.B. (1988). Stock Prices Under Time-Varying Dividend Risk. An Exact Solution in an Infinite-Horizon General Equilibrium Model. Journal of Monetary Economics 22, 375-393.

[2] Anderson, T.G., Benzoni, L., Lund, J. (2002) An Empirical Investigation of Continuous-Time Equity Return Models. Journal of Finance 57, 1239-1284.

[3] Ang, A., Hodrick, R.J., Xing, Y., Zhang, X. (2006) The cross-section of volatility and expected returns. Journal of Finance 61, 259-299.

[4] Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics, 93. Cambridge: Cambridge Univ. Press.

[5] Barsky, R.B. (1989) Why Don't the Prices of Stocks and Bonds Move Together? American Economic Review 79, 1132-1145.

[6] Bertoin, J. (1996). *Lévy Processes*. Cambridge: Cambridge Univ. Press.

[7] Bollerslev, T., Engle, R.F., and Nelson, D.B. (1995) ARCH models. In: *The Handbook of Econometrics, Volume 4.* R.F. Engle and D. McFadden, Eds. North Holland, Amsterdam.

[8] Brockwell, P.J., Chadraa, E. and Lindner, A. (2006). Continuous time GARCH processes. *Ann. Appl. Probab.* **16** 790–826.

[9] Buchmann, B., Müller, G. (2008). On the limit experiments of randomly thinned GARCH(1,1) in deficiency. Preprint, Technische Universität München.

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

[11] Czado, C. and Haug, S. (2008). Quasi maximum likelihood estimation and prediction in the compound Poisson ECOGARCH(1,1) model. Working paper. Technische Universität München.

[12] Durand, R.B., Lim, D., Zumwalt, J.K. (2007) Fear and the Fama-French Factors, working paper available at SSRN: http://ssrn.com/abstract=965587

[13] Durham, G.B. (2007) SV Models with Application to S&P500 Index Returns. Journal of Financial Economics 85, 822-856.

[14] Fama, E. F., French, K. R. (1993) Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-57.

[15] Engle, R.F. (1995) *ARCH: Selected Readings*. Oxford University Press, Oxford.

[16] Eraker, B., Johannes, M., Polson, N. (2003) The Impact of Jumps in Volatility and Returns. Journal of Finance 58, 1269-1300.

[17] Fasen, V., Klüppelberg, C., Lindner, A. (2005). Extremal behavior of stochastic volatility models. In: A. N. Shiryaev, M. d. .R. Grossinho, P. E. Oliviera, and M. L. Esquivel (Eds.) *Stochastic Finance*, pp. 107-155. New York: Springer.

[18] French, K., Schwert, G.W., Stambaugh, R. (1987) Expected Stock Returns and Volatility. Journal of Financial Economics 19, 3-30.

[19] Ghysels, E., Santa-Clara, P., Valkanov, R. (2005) There Is A Risk-Return Trade-Off After All. Journal of Financial Economics 76, 590-548.

[20] Haug, S. and Czado, C. (2007). An exponential continuous time GARCH process. *J. Appl. Probab.* **44**(4) 960–976.

[21] Haug, S. and Stelzer, R. (2008). Multivariate ECOGARCH processes - a multivariate stochastic volatility model with the leverage effect. Working paper, Technische Universität München.

[22] Haug, S., Klüppelberg, C., Lindner, A. and Zapp, M. (2007). Method of moments estimation in the COGARCH(1,1) model. *Econom. J.* **10** 320–341.

[23] Hsieh, D.A. (1991) Chaos and Nonlinear Dynamics: Application to Financial Markets. Journal of Finance 46, 1839-1877.

[24] Kallsen, J. and Vesenmayer, B. (2008). COGARCH as a continuous time limit of GARCH(1,1). *Stochastic Process. Appl.* to appear.

[25] Klüppelberg, C., Lindner, A. and Maller, R.A. (2004). A continuous time GARCH process driven by a Lévy process: stationarity and second order behaviour. *J. Appl. Probab.* **41** 601–622.

[26] Klüppelberg, C., Lindner, A. and Maller, R.A. (2006). Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. In: Kabanov, Yu., Lipster, R. and Stoyanov, J. (Eds.) *From Stochastic Calculus to Mathematical Finance. The Shiryaev Festschrift*. Berlin: Springer.

[27] Merton, R. (1980) On Estimating the Expected Return on the Market: An Exploratory Investigation. Journal of Financial Economics 8, 323-361.

[28] Mikosch, Th. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28** 1427–1451.

[29] Maller, R.A., Müller, G. and Szimayer, A. (2008). GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli* **14**(2) 519–542.

[30] Müller, G. (2007). MCMC estimation of the COGARCH(1,1) model. Preprint, Technische Universität München.

[31] Nelson, D.B. (1990). ARCH models as diffusion approximations. *J. Econometrics* **45** 7–38.

[32] Protter, P. (2005). *Stochastic Integration and Differential Equations*. 2nd ed. Heidelberg: Springer.

[33] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge Univ. Press.

[34] Sharpe, W. (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425-442.

[35] Stelzer, R. (2008a). Multivariate Continuous Time Lévy -Driven GARCH(1,1) Processes. Preprint, Technische Universität München.

[36] Stelzer, R. (2008b). First jump approximation of a multivariate Lévy driven SDE and an application to ECOGARCH processes. Preprint, Technische Universität München.

[37] Szimayer, A., Maller, R.A. (2007). Finite Approximation Schemes for Lévy Processes, and their Application to Optimal Stopping Problems. *Stochastic Process. Appl.* to appear.

[38] Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Statist.* **30** 754–783.

	β		η		φ	
1998	0.00215457	(0.00009185)	0.2035	(0.0099)	0.1002	(0.0075)
1999	0.00212128	(0.00007431)	0.2022	(0.0040)	0.1028	(0.0023)
2000	0.00312924	(0.00015303)	0.2032	(0.0052)	0.1019	(0.0032)
2001	0.00224318	(0.00010266)	0.1988	(0.0047)	0.1035	(0.0019)
2002	0.00385930	(0.00013885)	0.2067	(0.0068)	0.0918	(0.0054)
2003	0.00152757	(0.00011513)	0.2011	(0.0112)	0.1024	(0.0051)
2004	0.00076915	(0.00002222)	0.1979	(0.0037)	0.1048	(0.0025)
2005	0.00071658	(0.00002205)	0.2018	(0.0046)	0.1027	(0.0038)
2006	0.00069020	(0.00002447)	0.2039	(0.0054)	0.1009	(0.0049)
2007	0.00145146	(0.00008295)	0.2017	(0.0071)	0.0999	(0.0053)

	$h_9 \cdot 10^4$		$h_{10} \cdot 10^4$		$h_{12} \cdot 10^4$		$h_{13} \cdot 10^4$	
1998	2.2616	(0.1230)	0.1225	(0.0194)	-0.2587	(0.0081)	-0.2320	(0.0090)
1999	2.0057	(0.1038)	0.2080	(0.0213)	-0.2534	(0.0077)	-0.2290	(0.0085)
2000	1.9162	(0.1035)	0.2232	(0.0219)	-0.2420	(0.0082)	-0.1957	(0.0096)
2001	2.7368	(0.1571)	0.5309	(0.0349)	-0.2257	(0.0096)	-0.1637	(0.0113)
2002	2.7753	(0.1317)	0.4186	(0.0267)	-0.1412	(0.0106)	-0.1381	(0.0107)
2003	3.0822	(0.1526)	0.5732	(0.0329)	-0.1694	(0.0103)	-0.1970	(0.0096)
2004	2.5946	(0.1318)	0.2317	(0.0223)	-0.2085	(0.0091)	-0.1944	(0.0095)
2005	1.8003	(0.0988)	0.1923	(0.0198)	-0.1979	(0.0088)	-0.1504	(0.0106)
2006	2.1767	(0.1170)	0.2547	(0.0228)	-0.1667	(0.0103)	-0.1346	(0.0112)
2007	2.6795	(0.1377)	0.1301	(0.0193)	-0.1965	(0.0093)	-0.1766	(0.0099)

Table 2: *PML estimates and corresponding approximated standard errors of the parameters β, η, φ, h_9, h_{10}, h_{12}, and h_{13}. For notational convenience, the estimates and standard errors of h_9, h_{10}, h_{12}, h_{13} have been multiplied by 10000.*

	f_9	f_{10}	f_{12}	f_{13}
1998	5.3057	1.2332	0.5076	0.5583
1999	4.8966	1.4040	0.5077	0.5551
2000	4.7377	1.4353	0.5280	0.6183
2001	6.2005	2.0089	0.5712	0.6889
2002	6.4119	1.8164	0.7246	0.7307
2003	6.9899	2.1140	0.6709	0.6172
2004	6.0300	1.4491	0.5958	0.6231
2005	4.5319	1.3772	0.6117	0.7049
2006	5.2458	1.4967	0.6748	0.7374
2007	6.2170	1.2533	0.6174	0.6561

Table 3: *Factors that have to be applied to get from physical to business time scale, for trading hours 9, 10, 12, and 13, respectively.*