LUND UNIVERSITY

MASTER'S THESIS PROJECT

Return Models and Covariance Matrices

Author:
XIE Xiaolei

Supervisor:
Prof. Sven ÅBERG

June 26, 2014

Abstract

Return models and covariance matrices of return series have been studied. In particular, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and Stochastic Volatility (SV) models are compared with respect to their forecasting accuracy when applied to intraday return series. SV models are found to be considerably more accurate and more consistent in accuracy in forecasting.

Covariance matrices formed from Gaussian and GARCH return series, and in particular, return series auto-correlated as an AR(1) process, have been studied. In the case of Gaussian returns, the largest eigenvalue is found to approximately follow a gamma distribution also when the returns are auto-correlated. Expressions relating the mean and the variance of the asymptotic Gaussian distribution of the matrix elements are derived. In the case of GARCH returns, both the largest and the smallest eigenvalues of the covariance matrix are seen to increase with increasing auto-correlation. The matrix elements are found to follow Lévy distributions with different Lévy indices for the diagonal and the non-diagonal elements.

Localization of eigenvectors of covariance matrices of returns from GARCH processes has been investigated. It is found that the localization is reduced as the auto-correlation is increased. Quantitatively, the number of localized eigenvectors decreases approximately as a quadratic function with the auto-correlation strength, i.e. the autoregressive coefficient of the AR(1) process.

Acknowledgements

First of all, I thank my supervisor, Prof. Sven Åberg, for the many hours of mentoring, discussion, encouragement and guidance. Without them, this report will not exist. Then I wish to give my gratitude to the people in the division of mathematical physics. The casual discussions that I had with them truly inspired me, and the friendly and lively atmosphere that they create has helped bring peace and clarity to my mind.

Last but not least, I want to thank my wife, Wang Xiaodan, for her unconditional support and understanding.

Contents

Acronyms

ACF Auto-correlation Function. 11

AR Autoregressive. 10, 31, 33, 51

ARIMA Integrated Autoregressive Moving Average. 9, 12, 20, 28

ARMA Autoregressive Moving Average. 10, 20

CDF Cummulative Distribution Function. 26, 38

GARCH Generalized Autoregressive Conditional Heteroscedasticity. 1, 6–8, 10, 11, 16–20, 32, 37, 38, 46, 50, 51, 53, 54, 56, 58–61

IPR Inverse Participation Ratio. 44

MGF Moment Generating Function. 25, 26

MLE Maximum Likelihood Estimate. 12, 23, 26

PDF Probability Density Function. 21, 25, 26, 28, 31, 38, 44–48, 65, 67, 71, 72

PR Participation Ratio. 45, 48

SV Stochastic Volatility. 1, 5–8, 10, 11, 16, 17, 19, 20, 37, 50, 54, 56, 58, 60, 61

Chapter 1

Introduction

The mathematics of the financial market has always been a topic that arouses interest and imagination, and with no doubt, has been studied from many aspects and in many different ways.

Central to all these studies are the concepts of probability distributions and correlations. The values of stocks, futures, options, etc. are stochastic in nature and are governed by the laws of stochastic processes, which are expressed in terms of probability distributions and correlations.

Meanwhile, the observables of the market are prices, volumes, turn-over (the amount of money paid in a trade), names of the brokers, and time of the trades. These quantities don't make much sense by themselves but do reveal the probabilistic dynamics of the market when put together and turned into statistics.

The dynamics of an asset is affected by its own history as well as by the histories and current values of other assets in the market. The influence from the asset's own history is termed autocorrelations, i.e. correlations in time, while the influence from other assets are termed cross correlations.

Quite often, instead of an asset's price, the relative price change, i.e. the return, is studied. Analytically solvable models often assume Gaussian return distribution, although data suggest fatter tails [1]. By numerical methods, statistical features of historical returns data, such as fat tails, can be reproduced.

Statistical models built on historical data describe the time evolution of the aforementioned observables in terms of probability distributions, autocorrelations and cross correlations. They have been used to forecast future volatilities and help determine a pair price of a contingent claim, e.g. an option. Here the term "volatility" refers to the standard deviation of the distribution of returns, and measures how volatile the returns are. The volatility of an asset constitutes a major risk of investments on the asset and hence is very important.

[1] While the exact nature of the tails of the distribution of historical returns r is debatable, it is agreed $P(|r| > x) > ae^{-bx^2}$, for any positive constants a and b. Here $P(\cdot)$ denotes probability. We adopt this as the definition of "fat" tails.

In addition to the model of a single asset, the correlation between a group of assets is also of great interest. For example, in principle component analysis, one wishes to identify a number of factors that "drive" the price evolution of a group of assets in the sense that each of the assets' returns can be expressed as a linear combination of the factors' returns. In this scenario, the eigenvalues of the covariance matrix are the variances of the factors' returns and their corresponding orthogonalized eigenvectors give the composition of the factors, i.e. the coefficients with which the factors are contructed as a linear combination of the assets.

Therefore, in this thesis we also study the elements and the eigenvalues distribution, as well as the eigenvectors composition of the covariance matrix. When the matrix is constructed from returns with simple Gaussian distribution, the matrix is termed a Wishart matrix and has been studied extensively in the literature. If the returns have Lévy distributions, the matrix is termed Wishart-Lévy and has been studied to some extent, particularly regarding its eigenvalue distribution [1].

However, it is understood that real stock/index returns are much more complicated than a straight-forward Gaussian or Lévy distribution can describe — instead, one needs structured models. For this reason, we are particularly interested in a covariance matrix obtained for realistic return models. The so called GARCH(1,1) model is a realistic return model proven to have regularly varying tails [2]. So we study properties of eigenvalues and eigenvectors of such covariance matrices. Moreover, we also study how auto-correlations in the returns influence the aforementioned properties. Such auto-correlations, known as second-order auto-correlations decay exponentially but may still leave footprints in the covariance matrix.

This report is organized as follows: Chapter 2 reviews some of the most influential return model. Parameters of the models are fitted to a few intraday return series and the predictive power of the models is compared. A calculation of the unconditional distribution functions of SV models, especially in the case where the residual of the log-volatility and the innovation of the return are correlated normal variates, is also presented. In chapter 3 we investigate distributions of eigenvalues of the Wishart matrix, and study the influence of auto-correlated returns. In chapter 4 distributions of elements and eigenvalues of the covariance matrix of identically specified GARCH(1,1) series are studied. Finally, chapter 5 summarizes the results. Supplementary material is provided in the appendices.

Chapter 2

Return Models

In this chapter we review some of the discrete-time return models and fit them to intraday returns. The intention is to compare these models in terms of forecast accuracy and to understand their statistical properties. In the following we first describe these models briefly, then in section 2.1, section 2.2 and appendix A we describe how the GARCH and the SV models are fitted to intraday returns and compare their forecasts. In section 2.3 we calculate the unconditional distribution functions of the SV model.

1. **Gaussian Distribution**

   The justification of modeling return series as independent, identically distributed normal variates comes from imagining the price process S_t as a stochastic process with independent and normally distributed relative increments $dS_t/S_t = \sigma dw_t$, where w_t represents the path of a geometric Brownian motion and dw_t is normally distributed with mean 0 and variance dt.

   By adding a drift term μdt to allow a deterministic trend in the relative price change, one can express the price S_t as a stochastic differential equation:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dw_t$$

   By Itô's lemma, a stochastic differential equation for the logarithmic price $\ln S_t$ can be obtained

$$d(\ln S_t) = (\mu - \frac{1}{2}\sigma^2)dt + \sigma dw_t$$

   Solving this equation gives

$$\ln S_t - \ln S_0 = (\mu - \frac{1}{2}\sigma^2)dt + \sigma w_t$$

It follows from this equation that $\ln S(t) - \ln S_0$, where S_0 is the price at time 0, has Gaussian distribution with mean $(\mu - \sigma^2/2)t$ and variance $\sigma^2 t$ [3]. Therefore, in a discrete-time model, where the length of each time step Δt is fixed, the return $r_t = \ln S(t) - \ln S(t - \Delta t)$ is assumed to be Gaussian distributed and have mean and variance that are functions of Δt.

The picture depicted above is of course overly simplified, and the distribution of returns is not really Gaussian. Nevertheless, the assumption of Gaussian distributed returns underlies such important theories as Black and Scholes theory of option pricing. Thus, albeit inaccurate in describing very large returns, the Gaussian distribution as a return model shall not be forgotten. In the next few sections, we discuss some more realistic return models.

2. GARCH models

"GARCH" is the acronym for "Generalized Autoregressive Conditional Heteroscedasticity". A GARCH(p, q) model is defined by the following equation system [4]:

$$r_t = \mu_t + \epsilon_t \tag{2.1}$$

where r_t is the return as mentioned earlier; p and q are constant integers; μ_t is the mean process and considered a small constant for intraday returns. ϵ_t is called the innovation of the return and is a variate whose conditional distribution [1]is Gaussian and has variance σ_t^2:

$$\sigma_t^2 \quad = \quad \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

The quantity σ_t introduced here is termed the volatility, and its meaning and significance has been discussed in the introduction 1. In the model of geometric Brownian motion it is assumed constant and hence simply denoted σ.

It can be shown that the autocorrelation function $\varrho_n = \mathrm{corr}(\epsilon_t, \epsilon_{t-n})$ for $n > \max(p, q)$ of ϵ_t^2 is given by [5]:

$$\varrho_n \quad = \quad \sum_{i=1}^{p \vee q} (\alpha_i + \beta_i) \varrho_{n-i} \text{ for } n > p$$

where α_i with $i > q$ and β_i with $i > p$ are taken as zeros. $p \vee q$ denotes the maximum of p and q. From these equations, it is clear that the partial autocorrelation function cuts off at $\max(p, q)$.

Here we note that, in the GARCH model, the volatility σ_t is $\mathcal{F}_{t-1}$ measurable: given the history up to time $t - 1$, σ_t is a deterministic quantity.

[1] By conditional distribution we mean the distribution conditional on the variate's history: The conditional distribution at time t means the distribution conditional on the history up to time $t - 1$ in the discrete-time setting.

In contrast, SV models, which we discuss shortly, treat σ_t as a random variable even given the same history.

3. Stochastic Volatility Models

For the purpose of intraday returns, we specify the SV model as

$$\begin{aligned} r_{t,t-h} &= \ln S_t - \ln S_{t-h} \\ r_{t,t-h} &= \mu + \sigma_{t,t-h} b_t \end{aligned} \tag{2.2}$$

where S_t is the price of the asset at time t; $b_t \sim N(0,1)$; $r_{t,t-h}$ is the return over the time interval $[t-h, t]$. For simplicity, in the rest of this chapter we shall just write t for the subscript "t, t-h", since the time interval h is fixed for each time series and is a known constant.

Andersen et al proved the following in [6] (theorem 2):

$$r_t | \mathcal{F}_{t-h} \sim N(\int_0^h \mu_{t-h+s} ds, \int_0^h \sigma^2_{t-h+s} ds) \tag{2.3}$$

In plain words, conditional on the information up to time $t-h$, the distribution of r_t is Gaussian with integrated mean and variance. The variance of this conditional distribution, $\int_0^h \sigma^2_{t-h+s} ds$, can be approximated by [7, 6]:

$$\int_0^h \sigma^2_{t-h+s} ds = \sum_{k=1}^{n} \left(\ln S_{t-h+kh/n} - \ln S_{t-h+(k-1)h/n} \right)^2 \tag{2.4}$$

where n is a chosen constant. The square root of the right hand side of equation 2.4 is the realized volatility, call it $\hat{\sigma}_t$.

With the availability of transaction data, estimating conditional volatility using returns sampled at a higher frequency (realized volatility) gives superior accuracy and reliability. However, at which frequency the time series should be sampled (the choice of n) in order to give an unbiased and consistent estimate of the volatility is not a trivial question. Naively one would believe that the often the series is sampled, the better the estimate, but in fact, due to noise introduced by market micro-structure, there is an optimal sampling frequency, depending on h. While the method to determine this optimal frequency is a subject of debate (see for example [8]), it is not difficult to find a fairly satisfactory frequency in practice:

Keeping equation 2.3 in mind, one can simply try a few frequencies and compare the distribution of $(r_t - \mathbb{E}(r_t))/\hat{\sigma}_t$ with the standard Gaussian. If the two match, $\hat{\sigma}_t$ is a good approximation of

$$\sigma_t = \left(\int_0^h \sigma^2_{t-h+s} ds \right)^{1/2}$$

Once a good approximation of σ_t has been found, it is convenient to model $\ln \sigma_t$ so that the positivity of σ_t is implied by construction [9]. As is seen in later sections, Integrated Autoregressive Moving Average (ARIMA) models often serve well for this purpose. An ARIMA(p, d, q) model, where p, d, q are integers, is defined as [10]

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B)^d \ln \sigma_t \quad = \quad (1 - \sum_{i=1}^{q} \theta_i B^i) y_t$$

where B is the back-shift operator such that $Bx_t = x_{t-1}$ for any time series x_t. y_t are termed the residuals of the model, d is the order of integration, and p, q are orders of autoregression and moving average, respectively. ϕ_i and θ_i are constant parameters.

Quite often, the auto-correlation function of the time series in question manifests periodic patterns, thus seasonal components are added to the above model to account for the seasonality:

$$\begin{aligned}
&(1 - \sum_{i=1}^{P} \Phi_i B^{is})(1 - \sum_{i=1}^{p} \phi_i B^i)(1 - B^s)^D (1 - B)^d \ln \sigma_t \\
&= (1 - \sum_{i=1}^{Q} \Theta_i B^{is})(1 - \sum_{i=1}^{q} \theta_i B^i) y_t
\end{aligned}$$

where D, analogous to d in the non-seasonal model, is the order of seasonal integration; P, Q are orders of seasonal autoregression and moving average, respectively. Φ_i and Θ_i are constant parameters. In the most general situations, the seasonal and the non-seasonal components do not necessarily combine in the above multiplicative fashion, so the following model is also of interest:

$$(1 - \sum_{i=1}^{p+P} \phi_i B^i)(1 - B^s)^D (1 - B)^d \ln \sigma_t \quad = \quad (1 - \sum_{i=1}^{q+Q} \theta_i B^i) y_t$$

In all the above cases, if $d = 0$ and $D = 0$, the model does not involve integration and hence is called an Autoregressive Moving Average (ARMA)(p,q) model. An even simpler case is where $q = 0$ and $Q = 0$ in addition to $d = 0$ and $D = 0$. These conditions make the model a pure Autoregressive (AR) process, denoted AR(p).

To compare GARCH and SV models at the face of intraday returns, we study the *Nordea Bank* 15-minute returns during the period 2012/01/16 - 2012/04/20 in section 2.1, and *Volvo B* 30-minute returns during 2013/10/10 - 2014/04/04 in section 2.2. Another 4 intraday series are also studied and collected in appendix A. We have chosen these particular stocks because they have the largest trading volumes in the Swedish market and hence provide the largest amount of data for analysis. The time intervals of 15 and 30 minutes are chosen because they

are relatively short and hence give a large amount of data and yet they are not too short for noise of market micro-structure[2]to become a concern.

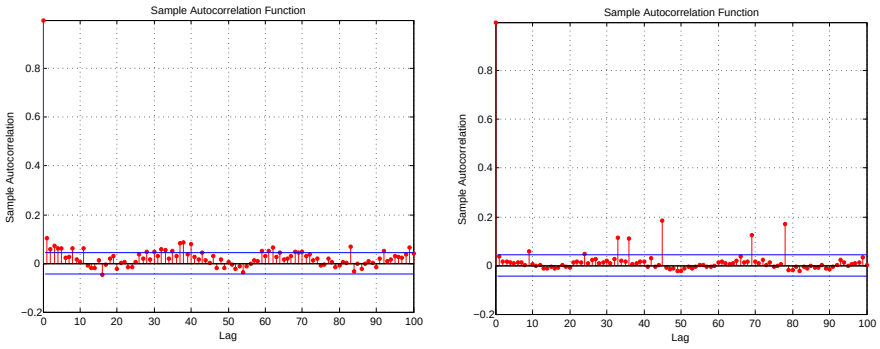
2.1 Case Study: Nordea 15-minute Returns

In this section we investigate the Nordea 15-minute returns sampled during the period 2012/01/16 - 2012/04/20. In total, these amount to 2022 returns. We use the first 80% (1617) for model estimation and the remaining 20% (405) for comparing with model forcasts. In section 2.1.1 we study the series with a GARCH model and in section 2.1.2 we study it with a SV model.

2.1.1 GARCH Model

When volatilities are auto-correlated or squared returns and volatilities are correlated, a GARCH(p, q) model may be appropriate for the return series under investigation. To find out whehter this is true in our case, we plot the Auto-correlation Function (ACF) of the squared returns. This is shown in figure 2.1(a). If the aforementioned features are absent from the series, the auto-correlations are expected to be Gaussian distributed with mean 0 and variance 1/T, and hence mostly reside within the confidence bounds set by $\pm 2/\sqrt{T}$ [4, 5].

Clearly this is not the case in figure 2.1(a) — the first 5 auto-correlations are rather significant. Moreover, figure 2.1(b) shows even more clearly that the conditional variances of the series are correlated. These observations suggest a GARCH(p, q) model can be appropriate.



(a) (b)

Figure 2.1: *2.1(a): Auto-correlations (ACF) of the squared returns; 2.1(b): Auto-correlations (ACF) of squared realized volatilities ($\hat{\sigma}_t^2$). The blue lines are confident bounds set at $\pm 2/\sqrt{T}$. T is the length of the time series.*

[2] Noise of market micro-structure is due to, most importantly, the difference between the bid and asked prices (bid-ask spread), and the discreteness of price changes.

Starting with a GARCH(1,1) model and taking advantage of the knowledge that the log-volatility $\ln \sigma_t$ has seasonality $s = 33$ (see figure 2.4(a)), we fit to the return series a GARCH(33, 33) model, limiting to lags 1 and 33 for both ARCH and GARCH parameters.

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_s \epsilon_{t-s}^2 + \beta_1 \sigma_{t-1}^2 + \beta_s \sigma_{t-s}^2 \end{aligned}$$

where the mean process of r_t, denoted μ_t earlier, is simplified to a mere constant μ owing to its smallness. by Maximum Likelihood Estimate (MLE), parameter values listed in table 2.1 are obtained.

Parameter	α_0	α_1	α_s	β_1	β_s
Value	4.7833×10^{-7}	0.1600	0.0667	0.6846	0.0342

Table 2.1: *GARCH model parameters*

2.1.2 Stochastic Volatility Model

For the Nordea Bank 15-minute returns under consideration, it can be verified that the square root of the sum of squared 30-second returns makes a good proxy for the volatility. This can be seen from the probability plot of $z_t = (r_t - \mathbb{E}(r_t))/\hat{\sigma}_t$ (figure 2.2), i.e. the quotient of the 15-minute returns over the volatility proxy.

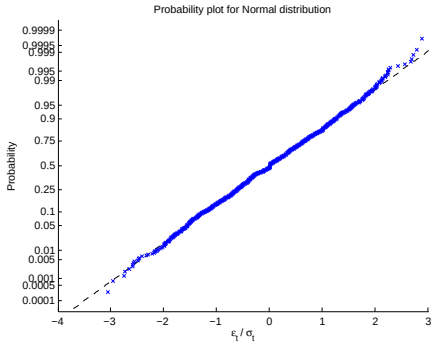


Figure 2.2: *Probability plot of $z_t = (r_t - \mathbb{E}(r_t))/\sigma_t$. ϵ_t are derived from Nordea Bank 15min returns while σ_t are realized volatilities calculated using 30s returns within each 15min interval. Horizontal axis: z_t; Vertical axis: CDF of z_t, arranged on such a scale that the CDF of the standard Gaussian is a straight line.*

Andersen and Bollerslev et al reported that, for the exchange rates between Deutch mark, yen and dollar, $\ln \sigma_t$ is Gaussian distributed [6]. This is, how-

ever, not the case for our modestly sized series. In fact, in our case, $\ln \sigma_t$ is right skewed (skewness 0.3342) and leptokurtic (kurtosis 6.1006). See figure 2.3 Moreover, the series of $\ln \sigma_t$ shows long-lasting and periodic autocorrelations

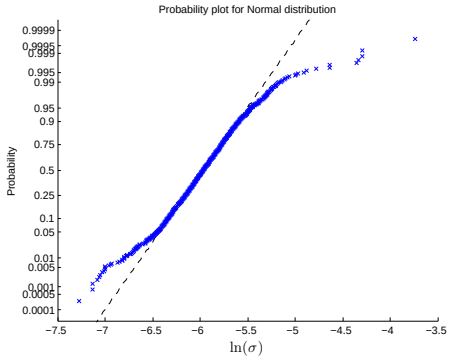


Figure 2.3: *Probability plot of Nordea 15min log-volatility* $\ln \sigma_t$ *unconditional distribution*

with an apparent period of 33 (see figure 2.4(a)). This suggests the series may be described by a seasonal ARIMA model. Thus we first simplify the series by differencing [10]:

$$w_t = (1 - B)(1 - B^s) \ln \sigma_t \tag{2.5}$$

where B is the back-shift operator[3] and $s = 33$ is the seasonality.

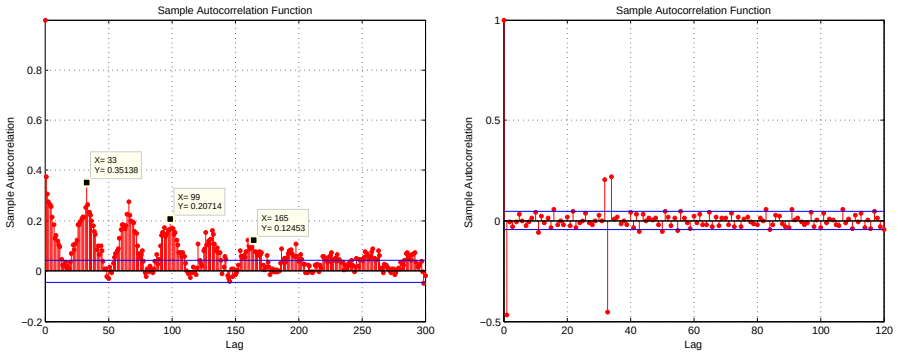
The autocorrelation function of the differenced process w_t, as shown in figure 2.4(b), clearly points to a seasonal moving-average model: There are only 4 non-zero autocorrelations in the plot, located at lags 1, 32, 33, 34, respectively; furthermore, the two at 32 and 34 are approximately equal. Thus we can write down the model as

$$w_t \quad = \quad (1 - \theta B)(1 - \Theta B^s) y_t \tag{2.6}$$

where θ and Θ are parameters to be determined and y_t is a noise process with constant variance σ_y^2 and mean 0. y_t is often refered to as the residuals.

The above seasonal moving average model has the following autocovariance

[3]For example, $B\, x_t = x_{t-1}$



(a) ACF of log-volatility ($\ln \sigma_t$) (b) ACF of differenced log-volatility (w_t)

Figure 2.4: *Auto-correlations of Nordea 15min log-volatility ($\ln \sigma_t$) and differenced log-volatility (w_t).*

structure [10]:

$$\begin{aligned} \gamma_0 &= \sigma_y^2(1+\theta^2)(1+\Theta^2) \\ \gamma_1 &= -\sigma_y^2\theta(1+\Theta^2) \\ \gamma_s &= -\sigma_y^2\Theta(1+\theta^2) \\ \gamma_{s+1} &= \gamma_{s-1} = \sigma_y^2\theta\Theta \end{aligned}$$

These equations together with the measured autocorrelations make possible an initial estimate of the parameters θ and Θ:

$$\begin{aligned} \varrho_{s+1}/\varrho_s &= \gamma_{s+1}/\gamma_s = -\frac{\theta}{1+\theta^2} \\ \varrho_{s+1}/\varrho_1 &= \gamma_{s+1}/\gamma_1 = -\frac{\Theta}{1+\Theta^2} \end{aligned}$$

Substituting in the measured values shown in table 2.2, we get

ϱ_1	ϱ_{s-1}	ϱ_s	ϱ_{s+1}
-0.4703	0.2053	-0.4564	0.2212

Table 2.2: *autocorrelations of differenced log-volatility (w_t)*

$$\begin{aligned} \theta &= 0.6890 \\ \Theta &= 0.6378 \end{aligned}$$

Among the two roots of each of the 2nd order equations in the above, we have chosen the one in the range $(-1, 1)$ so as to ensure invertibility of the model [10].

With an estimate of θ and Θ, one can then infer the noise process i.e. the residuals y_t:

$$y_t = w_t + \theta y_{t-1} + \Theta y_{t-s} - \theta\Theta y_{t-s-1} \tag{2.7}$$

where we substitute y_t $(t \leq 0)$ with their unconditional expectation 0.

In order to forecast the w_t process, and hence the return process itself, we must also know the distribution of y_t. Moreover, to properly estimate the parameters of the model in the sense of maximum likelihood, we are also in need of the distribution of y_t.

Figure 2.5 shows the normal probability plot of y_t. It is evident from this figure that y_t has fat tails. In addition, a simple calculation reveals that the

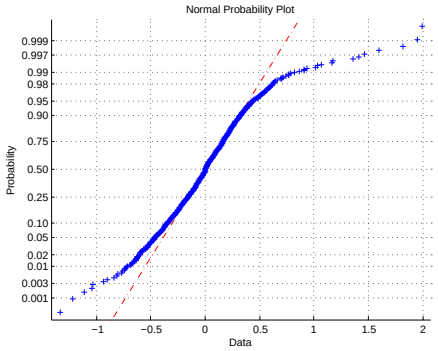


Figure 2.5: *Normal probability plot of residuals of log-volatility (y_t)*

distribution of y_t has skewness 0.2988 (shown in table 2.3). Based on this

mean	variance	skewness	kurtosis
0.0012	0.0935	0.2988	6.8691

Table 2.3: *Moments of log-volatility residuals (y_t)*

information, we find that y_t can be well described by a Johnson Su distribution. This is a flexible distribution that features fat tails and positive skewness. It is defined by the following transformation [11]:

$$y_t = \xi + \lambda \sinh \frac{z_t - \gamma}{\delta}$$

where $\gamma, \delta, \lambda, \xi$ are parameters to be determined and $z_t \sim N(0,1)$. The goodness of fitting is demonstrated in figure 2.6 by the empirical cummulative distribution function in comparison to the theoretical one.

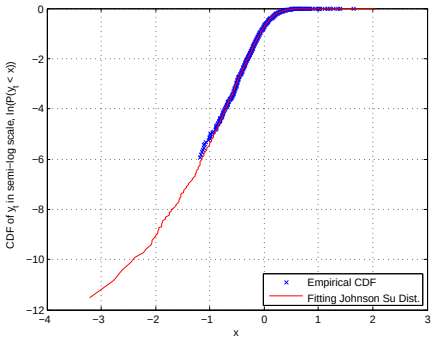


Figure 2.6: *Log-volatility residuals y_t fitted to a Johnson Su distribution. Horizontal: values of y_t, denoted x; Vertical:* $\ln(P(y_t < x))$.

The first 4 moments of the Johnson Su distribution are expressible in closed form in $\gamma, \delta, \lambda, \xi$ [11]. By matching the theoretical expressions of the moments with their measured values, and taking help from published tables [12], one can solve for the parameters $\gamma, \delta, \lambda, \xi$.

Under the assumption of i.i.d Johnson Su distributed residuals, the log-likelihood function of the parameters θ, Θ conditional on the sample w_t can be written as

$$L(\theta, \Theta) = -\frac{1}{2}\sum_{t=1}^{n} z_t^2 + n \ln \frac{\delta}{\lambda\sqrt{2\pi}} - \frac{1}{2}\sum_{t=1}^{n} \ln\left[1 + \left(\frac{y_t - \xi}{\lambda}\right)^2\right]$$

where y_t are inferred from w_t using eq.2.7 and z_t from y_t using

$$z_t = \delta \sinh^{-1} \frac{y - \xi}{\lambda} + \gamma$$

Note that $\gamma, \delta, \lambda, \xi$ are not really free parameters but rather are implied by θ and Θ: Once the latter have been chosen and the corresponding y_t inferred, the former are determined by the moments of y_t.

2.1.3 Comparison of the Forecasts

In this section we compare the one-step-ahead forecasts from the GARCH model and from the SV model. For this purpose, we compute the difference between a forecast $\ln \sigma_t^F$ and its measured counterpart, i.e. the realized volatility of the same period $\ln \hat{\sigma}_t$. As a reference, we also consider the results obtained by taking the mean of the realized volatilities of the first 80% of the data set as forecast for the volatilities of the remaining 20%. We call this naive forecast the "sample mean".

First of all, we look at the means and standard deviations of $\ln \sigma_t^F - \ln \hat{\sigma}_t$, which are listed in table 2.4. It is seen from table 2.4 that, on average, the SV

	SV	GARCH	Sample mean
$\mathbb{E}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	0.0040	-0.0008	-0.2210
$\text{std}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	0.2659	0.3011	0.2893

Table 2.4: *Mean and standard Deviation of the forecasts' distribution*

model over-estimates while GARCH under-estimates. In terms of the standard deviation of $\ln \sigma_t^F - \ln \hat{\sigma}_t$, the SV model wins with a small margin. In contrast, the sample mean forecast clearly under-estimates the log-volatilities to a large extent. So the efforts of building models have indeed led to more accurate forecasts.

Figure 2.7 compares the 3 kinds of forecasts by plotting the distribution function and the complementary distribution function of $\ln \sigma_t^F - \ln \hat{\sigma}_t$. Here one can see that the SV model yields a better quality of forecasts than does GARCH with respect to both under-estimates and over-estimates.

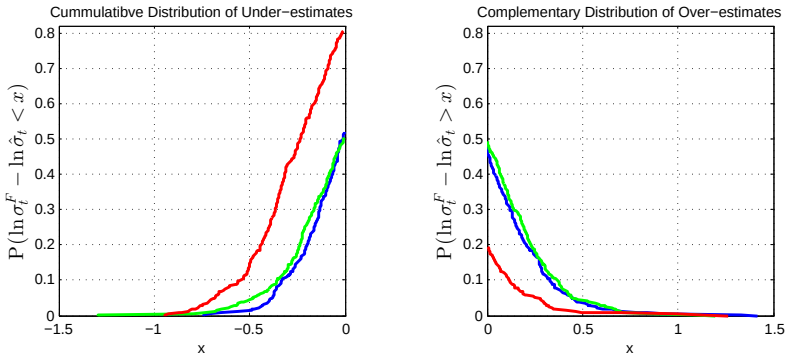


Figure 2.7: *Blue: SV forecasts; Green: GARCH forecasts; Red: sample mean forecasts. Left:* $P(\ln \sigma_t^F - \ln \hat{\sigma}_t < x)$*; Right:* $P(\ln \sigma_t^F - \ln \hat{\sigma}_t > x)$*. Horizontal:* x*.*

Another measure of the forecasts' quality can be the percentage of good forecasts, where the criterion of "good" is defined, respectively, as the forecast deviating less than 1%, 5%, or 10% from the corresponding realized log-volatility. Table 2.5 shows the respective percentage of the 3 kinds of forecasts. Again in this table it is seen that the SV model gives more accurate forecasts than does GARCH. For example, defining a "good" forecast as one that deviates less than 5% from the realized volatility, the probability of obtaining such a good forecast is 75% using the SV model, 71% using the GARCH model, and only 53% using the sample mean.

$\frac{\lvert \ln \sigma_t^F - \ln \hat{\sigma}_t \rvert}{\lvert \ln \hat{\sigma}_t \rvert}$	SV	GARCH	sample mean
1%	22%	14%	11%
5%	75%	71%	53%
10%	97%	94%	89%

Table 2.5: *The percentage of "good" forecasts when the criterion of being good is deviating no more than 1%, 5% or 10%.*

2.2 Case study: Volvo 30-minute Returns

In this section we model the log-volatility of *Volvo B* 30-minute returns during the period 2013/10/10 - 2014/04/04. This series contains 1884 log-volatilities computed using 2-minute returns in each 30-minute interval. Among them we use the first 1507 for model estimation and the last 377 for forecast and model verification.

The left plot of figure 2.8 shows the distribution of $(r_t - \mathbb{E}(r_t))/\sigma_t$. We observe in the figure a nice Gaussian variate, so we can be sure that the sum of squared 2-minute returns makes a good approximation to the variance of 30-minute returns in this particular case.

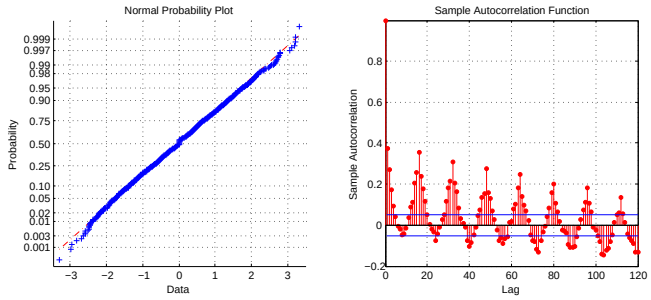


Figure 2.8: *Left: probability plot of* $(r_t - \mathbb{E}(r_t))/\sigma_t$*; Right: auto-correlations of* $\ln \sigma_t$

Guided by the auto-correlations of $\ln \sigma_t$ shown in the right plot of figure 2.8 we find the following model:

$$\begin{aligned} & (1-B)(1-B^s)\ln \sigma_t \\ = \; & (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)(1-\Theta B^s) y_t \end{aligned} \tag{2.8}$$

where $s = 16$ is the seasonality and is apparent from the auto-correlations of $\ln \sigma_t$. Fitting this model to the measured realized volatilities yields the parameter values listed in table 2.6. Forecasting using the estimated model parameters gives the forecast series $\ln \sigma_t^{\mathrm{SV}}$. To access the quality of the forecast, we also

Parameter	θ_1	θ_2	θ_3	θ_4	Θ	residual variance
Value	0.7305	0.0575	0.0574	0.0346	0.8324	0.1340

Table 2.6: *Volvo B log-volatility parameters*

estimate a GARCH model using the returns. The result is a GARCH(1, 1) model, whose parameter values are listed in table 2.7.

Parameter	α_0	α_1	β_1
Value	3.125×10^{-7}	0.05	0.90

Table 2.7: *GARCH(1, 1) model of Volvo B 30-minute returns*

The forecasts from SV, GARCH, and the sample mean are compared using the difference $\ln \sigma_t^F - \ln \hat{\sigma}_t$, where $\ln \sigma_t^F$ stands for the forecast. The distributions of this difference is plotted in figure 2.9; the mean and the standard deviation of the distributions are listed in table 2.8.

	SV	GARCH	Sample mean
$\mathbb{E}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	-0.0123	0.0242	-0.1055
$\mathrm{std}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	0.3261	0.4250	0.3708

Table 2.8: *Standard deviation of* $\ln \sigma_t^F - \ln \hat{\sigma}_t$

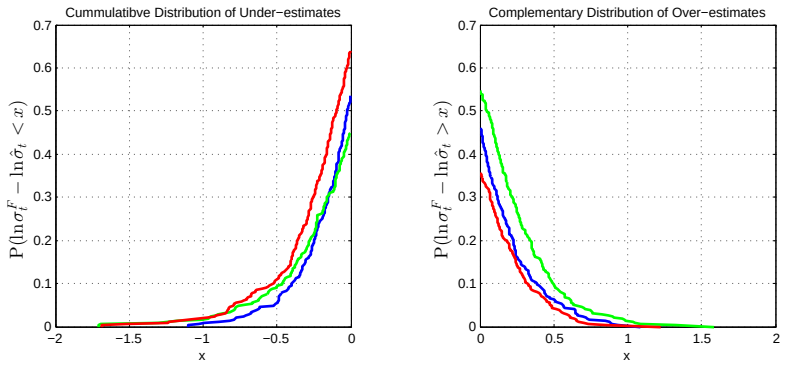


Figure 2.9: *Blue: SV forecasts; Green: GARCH forecasts; Red: sample mean forecasts. Left:* $P(\ln \sigma_t^F - \ln \hat{\sigma}_t < x)$*; Right:* $P(\ln \sigma_t^F - \ln \hat{\sigma}_t > x)$*. Horizontal:* x*.*

Figure 2.9 shows, as in the previous case of Nordea Bank 15-minute returns, the SV model performs the best, GARCH the second, and the sample mean the worst. However, when it comes to over-estimates, the sample mean appears to be the best estimator, while SV excels over GARCH. But a check of $\mathbb{E}(\ln \sigma_t)$ over

the sample for estimation (1507 data points) and over the sample for comparison (377 data points) reveals that the first sample has mean -6.3127 while the second has -6.2072. This increment explains the low probability of over-estimation when using the first sample mean as forecast.

Table 2.9 compares the fraction of "good" estimates as measured by $\frac{|\ln \sigma_t^F - \ln \hat{\sigma}_t|}{|\ln \hat{\sigma}_t|}$ being less than 1%, 5% and 10%. We see from the table that the SV model con-

$\frac{\lvert \ln \sigma_t^F - \ln \hat{\sigma}_t \rvert}{\lvert \ln \hat{\sigma}_t \rvert}$	SV	GARCH	Sample Mean
1%	22%	12%	14%
5%	72%	62%	66%
10%	92%	88%	90%

Table 2.9: *Fraction of "good" forecasts as defined by* $\frac{|\ln \sigma_t^F - \ln \hat{\sigma}_t|}{|\ln \hat{\sigma}_t|}$ *being less than 1%, 5% and 10%.*

sistently excels over the other two alternatives. In addition, it is also noted that the GARCH forecast is even worse than the sample mean. This is surprising but not totally unexpected — with only 3 parameters, the GARCH(1,1) model can only describe the most prominent auto-correlations in the volatility. When the volatility is influenced by relatively weak auto-correlations at several different time lags, the GARCH forecast cannot be expected to have good accuracy.

In this particular case, we see that the ARIMA model has 3 relatively small moving average coefficients, located at lags 2, 3, and 4 and evaluated to 0.06, 0.06, 0.03, suggesting a scattered auto-correlation structure, so the GARCH model cannot be expected to perform very well. In contrast, the Volvo 15-minute returns studied in section A.2 has more concentrated auto-correlations — 0.12 and 0.06 at lags 2 and 3 — thus the GARCH(1,1) model is also found to perform better and even marginally better than the SV model for the same series.

2.3 Unconditional Distribution Functions of SV Models

In this section we study the unconditional distribution function of the SV model specified as equation 2.2. As is discussed at the beginning of this chapter, $\ln \sigma_t$ can be described by an ARMA or ARIMA model, possibly with seasonal components. Here we note that all these models can be re-written as a moving average model, which is infinite in extent if autoregressive components are present:

$$\ln \sigma_t \quad = \quad y_t + \sum_{n=1}^{\infty} c_n y_{t-n} + \text{Const.}$$

Since the y_t are independent and identically distributed,

$$y_t + \sum_{n=1}^{\infty} c_n y_{t-n}$$

has Gaussian distribution by the central limit theorem, on condition that y_t for all t have finite second moment — this is what we assume in the rest of this section. It follows that the unconditional distribution of $\ln \sigma_t$ is the same as the distribution of $\bar{v} + v$ where $v \sim N(0, \sigma)$ and $\bar{v}$, σ are constants. Now we can state that the unconditional distribution of the returns r_t

$$\begin{aligned} r_t &= \mu + \sigma_t b_t \\ &= \mu + b_t \exp\left(y_t + \sum_{n=1}^{\infty} c_n y_{t-n} + \text{Const.} \right) \end{aligned}$$

is the same as

$$r = \mu + e^{\bar{v}+v} b \tag{2.9}$$

where $b \sim N(0, 1)$. For convenience, let $r' = e^v b$

In section 2.3.1 we first study the model in the relatively simple case when v and b are uncorrelated and $\mu = 0$. If this simplified version proves inadequate, one may resort to the general model studied in section 2.3.2.

2.3.1 The Simplified model

In the following we derive the unconditional Probability Density Function (PDF) of r', the de-meaned[4] and rescaled return. Denote this PDF $f_{r'}(x)$. Then the PDF of r is $e^{-\bar{v}} f_{r'}(e^{-\bar{v}} x)$. First we consider

$$\begin{aligned} P(r' < x) &= P(b < x e^{-v}) \\ f_{r'}(x) &= f_b(x e^{-v}) e^{-v} \end{aligned}$$

Averaging over all v, we get

$$\begin{aligned} f_{r'}(x) = &\int_{-\infty}^{\infty} dv (2\pi\sigma^2)^{-1/2} e^{-v^2/2\sigma^2} (2\pi)^{-1/2} \exp(-x^2 e^{-2v}/2) e^{-v} \\ = &\frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} dv \exp\left(-\frac{1}{2\sigma^2} v^2 - v - \frac{1}{2} x^2 e^{-2v} \right) \end{aligned} \tag{2.10}$$

The last part of the integrand, $e^{-x^2 e^{-2v}/2}$, is plotted in figure 2.10. We see from the figure that the landscape defined by this function, roughly speaking, consists of two levels, the lower one lying at height 0 and the higher one at height 1. The slope connecting the two levels is rather steep. It changes as a double exponential along lines of constant x and as e^{-x^2} along lines of constant v. Therefore we approximate this function as a 2-dimensional step function:

[4] By "de-mean" we mean subtracting from a random variable its expectation value.

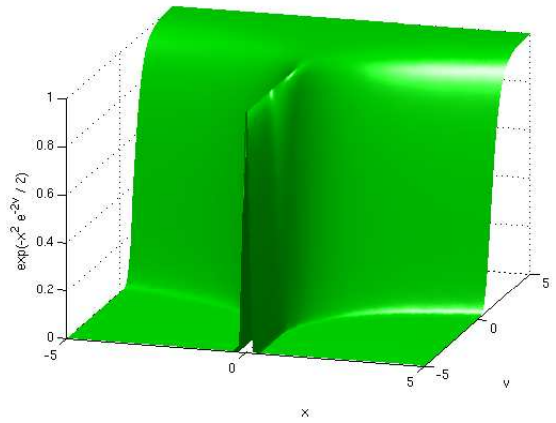


Figure 2.10: *Plot of* $\exp(-\frac{1}{2}x^2e^{-2v})$

$$\exp\left(-\frac{1}{2}x^2e^{-2v}\right) \approx \begin{cases} 0 & \text{if } v < \ln|x| - \frac{1}{2}\ln(2\ln 2) \\ 1 & \text{otherwise} \end{cases}$$

Here we note that $\exp\left(-\frac{1}{2}x^2e^{-2v}\right) = 1/2$ at $v = \ln|x| - \frac{1}{2}\ln(2\ln 2)$.

With this approximation we have

$$\begin{aligned} f_{r'}(x) &= \frac{1}{C}\frac{1}{2\pi\sigma}\int_{\ln(|x|/\sqrt{\ln 4})}^{\infty} dv \exp\left(-\frac{1}{2\sigma^2}v^2 - v\right) \\ &= \frac{1}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\ln\frac{|x|}{\sqrt{\ln 4}} + \frac{\sigma}{\sqrt{2}}\right) \end{aligned}$$

where $1/C$ has been added for the purpose of normalization.

At large x, we may use the asymptotic expansion of erfc to write

$$\begin{aligned} Cf_{r'}(x) &= \frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\mathrm{erfc}(\xi) \\ &= \frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\frac{e^{-\xi^2}}{\xi\sqrt{\pi}}\left[1 + \sum_{n=1}^{N}(-1)^n\frac{(2n-1)!!}{(2\xi^2)^n}\right] + O(\xi^{-2N-1}e^{-\xi^2}) \end{aligned}$$

where $\xi = \frac{1}{\sqrt{2}\sigma}\ln\frac{|x|}{\sqrt{\ln 4}} + \frac{\sigma}{\sqrt{2}}$. The slowest-decaying term is

$$f_{r,0}(x) = \frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\frac{e^{-\xi^2}}{\xi\sqrt{\pi}}$$

Let $\zeta = \frac{|x|}{\sqrt{\ln 4}}$. With a bit manipulation one obtains

$$f_{r,0}(x) = \frac{1}{\pi\sqrt{8}}\frac{1}{\left(\ln\zeta/\sigma\sqrt{2} + \sigma/\sqrt{2}\right)\zeta\zeta^{\ln\zeta/2\sigma^2}}$$

From the last equation one can see that, at any neighborhood of large x, $f_{r'}(x)$ may be approximated by $C/|x|^{\alpha}$, i.e. a power law. The normalization constant C is calculated in appendix B to be $C = \sqrt{\frac{2\ln 4}{\pi}}$. So, in summary, we can write:

$$\begin{aligned} f_r(x) &= e^{-\bar{v}} f_{r'}(e^{-\bar{v}}x) \\ &= \frac{e^{-\bar{v}}}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\mathrm{erfc}\left(\frac{1}{\sqrt{2}\sigma}\ln\frac{|e^{-\bar{v}}x|}{\sqrt{\ln 4}} + \frac{\sigma}{\sqrt{2}}\right) \end{aligned}$$

Using the same technique for integration as for normalization, the cummulative distribution function of r is found to be $F(x)$, which is the following:

1. if $x < 0$

$$\begin{aligned} F(x) &= \frac{\sqrt{\ln 4}}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\left[\frac{e^{-\bar{v}}x}{\sqrt{\ln 4}}\mathrm{erfc}\left(\frac{1}{\sigma\sqrt{2}}\ln\frac{-e^{-\bar{v}}x}{\sqrt{\ln 4}} + \frac{\sigma}{\sqrt{2}}\right)\right. \\ &\quad \left. + e^{-\sigma^2/2}\mathrm{erfc}\left(\frac{1}{\sigma\sqrt{2}}\ln\frac{-e^{-\bar{v}}x}{\sqrt{\ln 4}}\right)\right] \end{aligned}$$

2. if $x \geq 0$

$$\begin{aligned} F(x) &= \frac{1}{2} + \frac{\sqrt{\ln 4}}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\left[\frac{e^{-\bar{v}}x}{\sqrt{\ln 4}}\mathrm{erfc}\left(\frac{1}{\sigma\sqrt{2}}\ln\frac{e^{-\bar{v}}x}{\sqrt{\ln 4}} + \frac{\sigma}{\sqrt{2}}\right)\right. \\ &\quad \left. + e^{-\sigma^2/2}\mathrm{erfc}\left(-\frac{1}{\sigma\sqrt{2}}\ln\frac{e^{-\bar{v}}x}{\sqrt{\ln 4}}\right)\right] \end{aligned}$$

To verify the validity of the model, we fit the above probability density function to the de-meaned 30min returns of Volvo B [5] by means of MLE using the MATLAB function "mle". Then for the parameters σ and $\bar{v}$ we get

$$\begin{aligned} \sigma &= 0.6355 \\ \bar{v} &= -6.0308 \end{aligned}$$

Then we plot $P(r' > x)$ of the model against its empirical counterpart on a log-log scale, as shown in figure 2.11.

In general figure 2.11 shows a good fit, but a closer look reveals that deviations are significant in the regions $r \in (0, \sigma_r)$ and $r \in (2\sigma_r, 3\sigma_r)$, where σ_r stands for the empirical standard deviation of the de-meaned returns. These observations are shown in greater details in figure 2.12.

Moreover, the inefficiency of the model is also manifest in the skewness of the data. For the Volvo 30min returns, the data has a skewness of 0.2419, but our model is strictly symmetric, since the return x appears in equation 2.10 only as x^2. Hence the above model needs to be improved to accommodate the non-zero skewness as well as to account for the discrepancies shown in figure 2.12. This is the subject of the next section.

[5] By "de-meaned returns" we mean the quantity $r_t - \langle r_t \rangle$, where r_t are the measured returns and $\langle r_t \rangle$ is the sample mean. The data set covers the transaction records of Volvo B on the OMX market (Stockholm) between 2013-10-10 and 2014-03-12. The returns are computed using 1-minute mean prices.

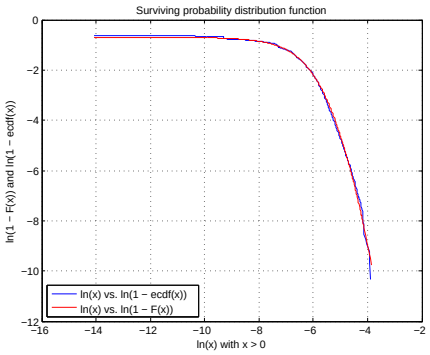
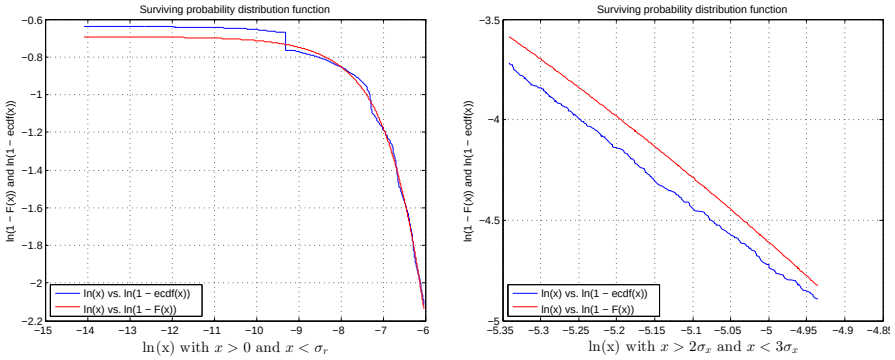


Figure 2.11: $P(r' > x)$ *for* $x > 0$. *Blue: empirical probabilities. Red: probabilities predicted by the model.*



(a) $\ln(P(r > x))$ with $x \in (0, \sigma_r)$ (b) $\ln(P(r > x))$ with $x \in (2\sigma_r, 3\sigma_r)$

Figure 2.12: *Surviving probabilities. Blue: empirical values of* $\ln(P(r > x))$. *Red: Predicted values of* $\ln(P(r > x))$.

2.3.2 The General model

It has long been hypothesized in the liturature that skewness is the result of price-volatility correlation (for example [13]). Therefore, the most apparent modification is to allow v, the zero-mean Gaussian variate in the log-volatility, and b, the zero-mean Gaussian variate in the innovation of the return, to be

correlated. For convenience we decompose $v \sim N(0,\sigma)$ as $v = \sigma a$ and assume

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim N(0,\Sigma)$$

with a covariance matrix

$$\Sigma \quad = \quad \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}$$

where $|\psi| < 1$. Then we rewrite equation 2.9 as

$$\begin{aligned} r &= \mu + e^{\bar{v}} r' \\ r' &= e^{\sigma a} b \end{aligned} \tag{2.11}$$

We then get

$$\begin{aligned} & P(r' < x) \\ = \quad & P(b < e^{-a\sigma} x) \\ = \quad & \frac{1}{2\pi |\det(\Sigma)|^{1/2}} \int_{-\infty}^{\infty} da \int_{-\infty}^{e^{-a\sigma}x} db \exp\left[-\frac{1}{2}(a,b)\Sigma^{-1} \begin{pmatrix} a \\ b \end{pmatrix} \right] \end{aligned}$$

After some manipulations we get

$$\begin{aligned} & P(r' < x) \\ & = \frac{1}{2\pi\sqrt{1-\psi^2}} \int_{-\infty}^{\infty} da e^{-a^2/2} \int_{-\infty}^{e^{-a\sigma}x} db \exp\left[-\frac{(b-a\psi)^2}{2(1-\psi^2)} \right] \\ & = \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-a^2/2} \mathrm{erfc} \frac{a\psi - e^{-a\sigma}x}{\sqrt{2(1-\psi^2)}} da \end{aligned} \tag{2.12}$$

Differentiating with respect to x yields the PDF

$$\begin{aligned} f_{r'}(x) \quad &= \quad \frac{1}{2\pi\sqrt{1-\psi^2}} \times \\ & \qquad \int_{-\infty}^{\infty} da \exp\left[-\frac{a^2 + 2\sigma(1-\psi^2)a - 2a\psi e^{-a\sigma}x + e^{-2a\sigma}x^2}{2(1-\psi^2)} \right] \end{aligned} \tag{2.13}$$

$$\begin{aligned} f_r(x) \quad &= \quad \frac{dr'}{dr} f_{r'}[e^{-\bar{v}}(x-\mu)] \\ &= \quad e^{-\bar{v}} f_{r'}[e^{-\bar{v}}(x-\mu)] \end{aligned} \tag{2.14}$$

Unlike 2.10 where a relatively simple approximation can be found and thus leads to an analytic result of the integral, no such approximation has been found by the author for the integral 2.13. However, the Moment Generating Function (MGF) of PDF 2.13 and 2.14 are easy enough to find and give the moments about the origin in closed form. Then by matching the analytic expressions of

the moments with their statistical values from the sample, the parameters σ, ψ and $\bar{v}$ corresponding to the sample can be obtained.

The MGF of 2.14 can be found as follows:

$$\begin{aligned} & M_r(t) \\ = \; & \mathbb{E}(e^{tr}) \\ = \; & \frac{1}{2\pi\sqrt{1-\psi^2}} \int_{-\infty}^{\infty} da \exp\left(-\frac{a^2}{2} - \sigma a\right) \int_{-\infty}^{\infty} dx \exp\left[t(e^{\bar{v}}x + \mu) - \frac{(a\psi - e^{-a\sigma}x)^2}{2(1-\psi^2)}\right] \\ = \; & \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} da \exp\left[-\frac{a^2}{2} + \frac{1}{2}e^{2a\sigma+2\bar{v}}(1-\psi^2)t^2 + a\psi e^{a\sigma+\bar{v}}t + \mu t\right] \end{aligned}$$

With the MGF, the first 4 moments of r can be computed:

$$\begin{aligned} \mathbb{E}(r) &= \mu + e^{\bar{v}+\frac{\sigma^2}{2}}\sigma\psi \\ \mathbb{E}(r^2) &= \mu^2 + 2e^{\bar{v}+\frac{\sigma^2}{2}}\mu\sigma\psi + e^{2\left(\bar{v}+\sigma^2\right)}\left(1+4\sigma^2\psi^2\right) \\ \mathbb{E}(r^3) &= \mu^3 + 3e^{\bar{v}+\frac{\sigma^2}{2}}\mu^2\sigma\psi + 9e^{3\bar{v}+\frac{9\sigma^2}{2}}\sigma\psi\left(1+3\sigma^2\psi^2\right) + 3e^{2\left(\bar{v}+\sigma^2\right)}\mu\left(1+4\sigma^2\psi^2\right) \\ \mathbb{E}(r^4) &= \mu^4 + 4e^{\bar{v}+\frac{\sigma^2}{2}}\mu^3\sigma\psi + 36e^{3\bar{v}+\frac{9\sigma^2}{2}}\mu\sigma\psi\left(1+3\sigma^2\psi^2\right) \\ &\quad + 6e^{2\left(\bar{v}+\sigma^2\right)}\mu^2\left(1+4\sigma^2\psi^2\right) + e^{4\bar{v}+8\sigma^2}\left(3+96\sigma^2\psi^2+256\sigma^4\psi^4\right) \end{aligned} \tag{2.15}$$

These equations are rather complicated and directly solving them to obtain the parameters μ, σ, ψ and $\bar{v}$ is infeasible. However, in practice, $\mathbb{E}(r)$ is often very small — so, to get a rough estimate of the parameters, we may set $\mu = 0$ in the above equations and, with a bit of manipulation, find the following equation for $\psi\sigma$:

$$\frac{\mathbb{E}(r^3)\mathbb{E}(r^4)^{-3/8}}{\mathbb{E}(r^2)^{3/4}} = \frac{9\sigma\psi(1+3\sigma^2\psi^2)}{(1+4\sigma^2\psi^2)^{3/4}} \frac{1}{(3+96\sigma^2\psi^2+256\sigma^4\psi^4)^{3/8}} \tag{2.16}$$

This equation can be solved numerically for a given sample to yield an estimate for $\sigma\psi$, which in turn can be substituted in equations 2.15, where μ has been set to 0, to give estimates for all the 4 parameters. These estimates then serve as initial values in a numerical solution to 2.15 where μ is kept as a free variable. This full solution can now be used as the initial estimate in an MLE procedure.

It is particularly important in our situation to have a good initial estimate, because, as has been shown earlier, the PDF and the Cummulative Distribution Function (CDF) cannot be obtained in closed forms and consequently have to be evaluated by numerical integration. This is a rather costly procedure especially in the context of MLE. So the computation of the moments under the assumption of $\mu = 0$ is worthwhile.

Following the aforementioned procedure, i.e. computing the initial estimate by matching the moments and then refining the estimate by MLE, we obtain the parameter values for a number of return series, including the Volvo B 30-minute returns described in section 2.3.1. Table 2.10 and 2.11 present the obtained

parameter values and the resulting moments, respectively. Figure 2.13, 2.14, and 2.15 compare the empirical distribution functions with the analytic distribution functions evaluated at these parameters.

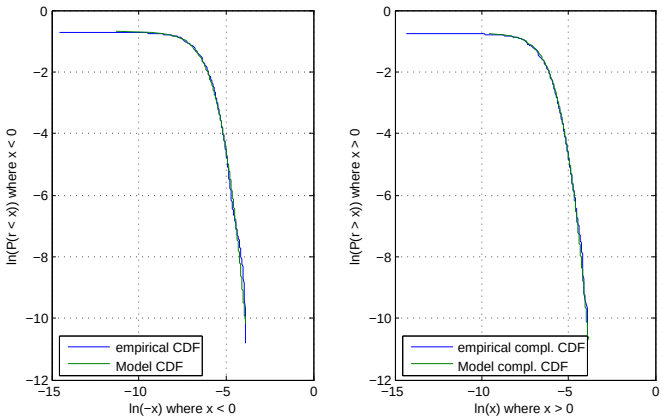


Figure 2.13: *Volvo B 30min returns' unconditional distribution fit to the SV model. The time series runs from 2013/10/10 to 2014/03/12. Left:* $P(r < x)$ *with* $x < 0$*; Right:* $P(r > x)$ *with* $x > 0$*. Blue: empirical CDF obtained from data; Green: model CDF computed using equation 2.12 with* x *replaced by* $(x - \mu)e^{-\bar{v}}$*. Both plots are on log-log scale.*

	ψ	σ	$\bar{v}$	μ
Volvo B	-1.6×10^{-2}	4.7×10^{-1}	-6.2	-5.3×10^{-5}
Nordea Bank	1.7×10^{-2}	4.3×10^{-1}	-6.4	7.5×10^{-5}
Ericsson B	1.8×10^{-2}	4.5×10^{-1}	-6.3	-9.3×10^{-5}

Table 2.10: *Parameter Values of Selected Assets' returns. The time series are 30-minute returns and run from 2013/10/10 to 2014/03/12.*

Clearly these figures show a fairly good match of the empirical and the analytic distribution functions. However, we also see that the skewness of the returns of Volvo B and Ericsson B have rather different values when computed from the sample and from the model. This could be the consequence of the limited sample size or deficiencies in the estimation procedure described above. We leave these issues to future studies.

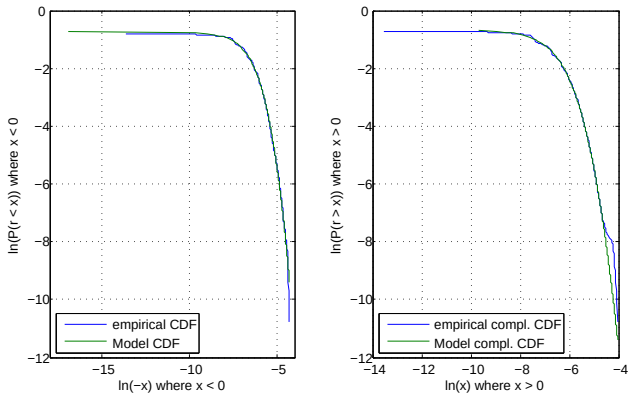


Figure 2.14: *Nordea Bank 30min returns' unconditional distribution fit to the SV model. The time series runs from 2013/10/10 to 2014/03/12. Left:* $P(r < x)$ *with* $x < 0$*; Right:* $P(r > x)$ *with* $x > 0$*. Blue: empirical CDF obtained from data; Green: model CDF computed using equation 2.12 with* x *replaced by* $(x - \mu)e^{-\bar{v}}$*. Both plots are on log-log scale.*

	mean	std	skewness	kurtosis
Volvo B	-8.7×10^{-5}	2.5×10^{-3}	-2.7×10^{-2}	7.3
	-6.9×10^{-5}	2.5×10^{-3}	-7.3×10^{-2}	7.2
Nordea Bank	7.5×10^{-5}	2.0×10^{-3}	6.9×10^{-2}	6.7
	8.9×10^{-5}	2.0×10^{-3}	6.7×10^{-2}	6.2
Ericsson B	-8.3×10^{-5}	2.3×10^{-3}	2.7×10^{-1}	7.7
	-7.6×10^{-5}	2.3×10^{-3}	7.9×10^{-2}	6.7

Table 2.11: *Moments of Selected Assets' returns. For each return series, the 1st row contains the sample moments, while the 2nd constains those computed with MLE parameters. The time series are 30-minute returns and run from 2013/10/10 to 2014/03/12.*

2.3.3 Relation to Conditional Distribution Functions

In the last section we have shown that the unconditional distribution of the returns are skewed, which implies the zero-mean Gaussian variate a in the log-volatility is correlated to the zero-mean Gaussian variate b in the return (c.f. equation 2.9).

In the context of conditional distributions and forecast, this correlation translates to the correlation between the residual of the log-volatility, denoted y_t in section 2.1 and section A.2, and b_t. Now let us consider the forecast function

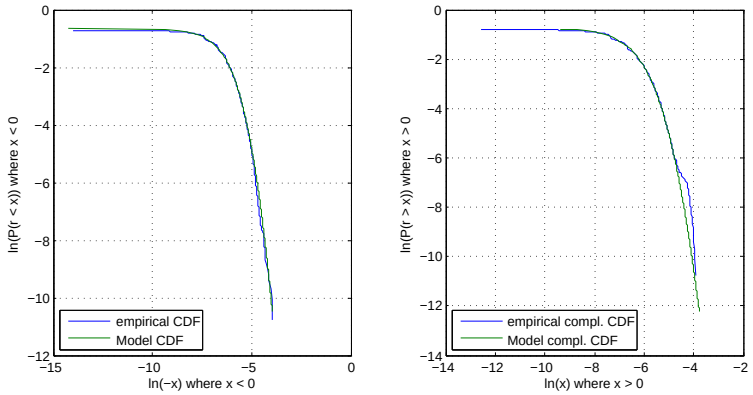


Figure 2.15: *Ericsson B 30min returns' unconditional distribution fit to the SV model. The time series runs from 2013/10/10 to 2014/03/12. Left:* $P(r < x)$ *with* $x < 0$*; Right:* $P(r > x)$ *with* $x > 0$*. Blue: empirical CDF obtained from data; Green: model CDF computed using equation 2.12 with* x *replaced by* $(x - \mu)e^{-\bar{v}}$*. Both plots are on log-log scale.*

of an ARIMA model:

$$\ln \sigma_t \quad = \quad y_t + \sum_{i=1}^{P} \phi_i \ln \sigma_{t-i} - \sum_{i=1}^{Q} \theta_i y_{t-i}$$

where $\phi_i < 2$ if the model involves integration, and $\phi_i < 1$ otherwise. Comparing this equation with equation 2.11, one immediately realizes that the conditional PDF of r_t

$$\begin{aligned} r_t &= \mu + \sigma_t b_t \\ &= \mu + \exp\left(y_t + \sum_{i=1}^{P} \phi_i \ln \sigma_{t-i} - \sum_{i=1}^{Q} \theta_i y_{t-i} \right) b_t \end{aligned}$$

is given by equation 2.14 and its moments given by the equations 2.15 if one makes the following substitutions:

$$\begin{aligned} \sigma &\rightarrow \mathrm{std}(y_t) \\ \bar{v} &\rightarrow \sum_{i=1}^{P} \phi_i \ln \sigma_{t-i} - \sum_{i=1}^{Q} \theta_i y_{t-i} \\ \psi &\rightarrow \mathrm{corr}(y_t, b_t) \end{aligned}$$

For the Nordea series considered in section 2.1, $\text{corr}(y_t, b_t)$ is found to be 3.56×10^{-2}; while for the Volvo series considered in section A.2, $\text{corr}(y_t, b_t)$ is found to be -7.3×10^{-3}. These values are comparable to those in table 2.10, where ψ is 1.7×10^{-2} for the Nordea series and -1.6×10^{-2} for the Volvo series. The similarity in these values provides some evidence about the validity of the model.

Chapter 3

Covariance Matrix of Gaussian Returns

In chapter 2 we have studied GARCH and SV models and seen their power of forecasting future volatilities. However, we have not considered the important fact that a financial market comprises many assets and the volatilities of these assets are correlated to each other in a complicated manner. Practically useful volatility forecasts require good understanding of these correlations.

In the literature, covariance matrices of Gaussian and Lévy distributed returns have been studied (see e.g. [1, 14, 15]). However, as is seen in chapter 2, real-world returns are not described by any particular distribution but rather by stochastic processes that account for auto-correlations in the returns and the volatilities.

Therefore, the focus of this and the next chapter is on covariance matrices of realistic return series, and especially covariance matrices in the case when the return series have considerable auto-correlations. In particular, we study covariance matrices of GARCH (1,1) return series in chapter 4 and show the influence of auto-correlations on these matrices. But before that, it is useful to first understand the influence of auto-correlations on covariance matrices of Gaussian return series. When auto-correlations are absent, these matrices are called Wishart matrices and have been studied extensively. The results we obtain in this chapter will provide a reference to the studies in chapter 4.

In section 3.1 we discuss how the distributions of the matrix elements are affected by autocorrelations, and in section 3.2 we investigate the distribution of the eigenvalues.

3.1 Distribution of the Matrix Elements

In this section we study the distribution of the elements of a covariance matrix $C = \mathbf{R}\mathbf{R}'/T$, where

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1T} \\ r_{21} & r_{22} & \cdots & r_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N22} & \cdots & r_{NT} \end{pmatrix}$$

and $\mathbf{R}'$ denotes the transpose of $\mathbf{R}$. In words, an element r_{it} of $\mathbf{R}$ is the return of asset i at time t, with $i = 1, \cdots, N$ and $t = 1, \cdots, T$. If each column of $\mathbf{R}$ follows a zero-mean Gaussian distribution, i.e.

$$\begin{pmatrix} r_{1t} \\ \vdots \\ r_{Nt} \end{pmatrix} \sim N(0, \mathbf{\Sigma})$$

for all $t = 1, \cdots, T$, and none of the return series is auto-correlated, i.e. $\mathrm{corr}(r_{it}, r_{i,t'}) = 0$ for all $i = 1, \cdots, N$ and $t \neq t'$, then $\mathbf{R}\mathbf{R}'$ is a Wishart matrix whose probability density function is well known [16].

When auto-correlations are indeed present in the returns, $\mathbf{R}\mathbf{R}'$ no longer follows the Wishart distribution. However, the joint distribution function of its elements can be expressed in terms of the Wishart PDF and the auto-correlations. We show this in appendix C. Also, in appendix D we derive an approximate expression for the asymptotic distribution of these matrix elements, assuming r_{it} is an AR(1) process [1], i.e. $r_{it} = \phi r_{i,t-1} + a_{it}$, and

$$\begin{pmatrix} a_{1t} \\ \vdots \\ a_{Nt} \end{pmatrix} \sim N(0, \mathbf{\Sigma})$$

where

$$\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

[1] It is straight forward to derive the auto-correlation function ϱ_k of an AR(1) process with autoregressive coefficient ϕ:

$$\begin{aligned} \varrho_k &= \mathrm{corr}(r_{it}, r_{i,t-k}) \\ &= \phi \varrho_{k-1} \\ &= \phi^k \end{aligned}$$

Let τ denote the time lag at which $\varrho_\tau = 1/2$, then it follows $\tau = -\ln 2 / \ln \phi$.

where $-1 < \rho < 1$ is a constant parameter describing the correlation between a_{it} and a_{jt} — so, by construction, we assume such correlations are constant across all asset pairs and over all time.

The elements C_{ij} $(i \neq j)$ of the covariance matrix, $C = \mathbf{R}\mathbf{R}'/T$, are found to be normally distributed with mean

$$\mu'_X = \frac{\sigma^2}{\sqrt{2\pi}(1-\phi^2)(1-\rho^2)^{1/4}} \left[P^{-3/2}_{-1/2}(-\rho) - P^{-3/2}_{-1/2}(\rho) \right] \tag{3.1}$$

and variance

$$\begin{aligned} \sigma'^2_X &= \frac{1}{(1-\phi^2)^2} \left[\sum_{t=1}^{T} \sum_{k=1}^{t-1} 2 \left(\frac{\phi^k}{T} \right)^2 \sigma^6 + \sum_{t=1}^{T} \frac{\sigma^4 (1-\rho^2)^2}{T^2} v^2(\rho) \right] \\ &= \frac{2\sigma^6}{T(1-\phi^2)^2} \left[\frac{\phi^2}{1-\phi^2} - \frac{\phi^2(1-\phi^{2T})}{T(1-\phi^2)} \right] + \frac{\sigma^4(1-\rho^2)^2 v^2(\rho)}{T(1-\phi^2)^2} \\ &\approx \frac{2\sigma^6\phi^2}{T(1-\phi^2)^3} + \frac{\sigma^4(1-\rho^2)^2 v^2(\rho)}{T(1-\phi^2)^2} \end{aligned} \tag{3.2}$$

where $P^{\mu}_{\nu}(\cdot)$ is Ferrer's function of the first kind. See appendix D. For the definition of Ferrer's function, see equation D.4.

Equation 3.1 tells that, if two return series i and j are not correlated, i.e. $\rho = 0$, auto-correlation in the returns does not introduce a bias into the estimation of their covariance, i.e. μ'_X, since the difference between the two Ferrer's functions evaluate to 0 in equation 3.1; if, however, the return series are indeed correlated, auto-correlation in the returns rescales the covariance through a multiplicative factor $1/(1-\phi^2)$.

In addition, equation 3.2 tells that auto-correlation in the returns always makes the covariance estimation more noisy — auto-correlation not only rescales the variance of the no-autocorrelation estimation by $1/(1-\phi^2)^2$ but even adds an extra term $\frac{2\sigma^6\phi^2}{T(1-\phi^2)^3}$.

For the diagonal elements of the covariance matrix C we have

$$\begin{aligned} \mathbb{E}(C_{ii}) &= \frac{1}{T} \left[\sum_{k=0}^{t-1} \phi^{2k} \sigma^2 \right] \\ &= \frac{\sigma^2}{(1-\phi^2)T} \left[T - \frac{\phi^2(1-\phi^{2T})}{1-\phi^2} \right] \\ &\approx \frac{\sigma^2}{1-\phi^2} \left[1 - \frac{\phi^2}{T(1-\phi^2)} \right] \end{aligned} \tag{3.3}$$

and

$$\begin{aligned}
\mathrm{var}(C_{ii}) &= \sum_{t=1}^{T}\left[\sum_{k=0}^{t-1}\frac{\phi^{4k}\sigma^4}{T^2}2+\sum_{k,l=0}^{t-1}\frac{\phi^{2(k+l)}}{T^2}\sigma^6\right] \\
&= \sum_{t=1}^{T}\left[\frac{2\sigma^4}{T^2}\frac{1-\phi^{4t}}{1-\phi^4}+\frac{\sigma^6}{T^2}\left(\frac{1-\phi^{2t}}{1-\phi^2}\right)^2\right] \\
&= \frac{2\sigma^4}{T(1-\phi^4)}-\frac{2\sigma^4\phi^4(1-\phi^{4T})}{T^2(1-\phi^4)^2}+ \\
&\quad \frac{\sigma^6}{T(1-\phi^2)^2}-\frac{2\sigma^6\phi^2(1-\phi^{2T})}{T^2(1-\phi^2)^3}+\frac{\sigma^6\phi^4(1-\phi^{4T})}{T^2(1-\phi^2)^2(1-\phi^4)} \\
&\approx \frac{2\sigma^4}{T(1-\phi^4)}+\frac{\sigma^6}{T(1-\phi^2)^2} \qquad (3.4)
\end{aligned}$$

From equation 3.3 we see that auto-correlation in the returns increases the variance of the return series; and from equation 3.4 we see that the variance of that variance estimation is also increased by auto-correlations. Moreover, we note that $\mathrm{var}(C_{ii})$ scales with T approximately as $1/T$, similar to the behavior of $\mathrm{var}(C_{ij})$. This is to be compared with the case of GARCH returns discussed in chapter 4.

3.2 Distribution of the Eigenvalues

For a Wishart matrix $\mathbf{RR}'$, theoretical results are available for the eigenvalue distribution, and we summarize them in appendix E. In short, the joint probability density function of the eigenvalues is given by equation E.1 when neither auto-correlation nor cross-correlation is present in $\mathbf{R}$. Moreover, the mariginal distribution of the largest eigenvalue is approximated by a gamma distribution [14].

However, as detailed in the derivation leading to equation C.2, the distribution of $\mathbf{RR}'$ is not Wishart when the columns of $\mathbf{R}$ are correlated (auto-correlation). Deriving the eigenvalue distribution analytically in this case is beyond the scope of this thesis. Instead, we resort to numerical methods.

As before we consider the AR(1) process:

$$\mathbf{r}_t = \phi\mathbf{r}_{t-1}+\mathbf{a}_t$$

where $\mathbf{a}_t \sim N(0,\mathbf{I})$, i.e. the elements of $\mathbf{a}_t$ are independent Gaussian random variable with zero mean and unit variance. Now we investigate how the eigenvalue distribution depends on the auto-correlation strength parameter ϕ. Figure 3.1 shows the results of the simulation. It is clear from the figure that the maximum eigenvalue moves consistently to the right as the value of ϕ increases, and, as shown in figure 3.2, the minimum eigenvalue also increases with ϕ.

Chiani showed that the marginal distribution of the maximum eigenvalue (λ_1) is approximately gamma when neither cross-correlation nor auto-correlation

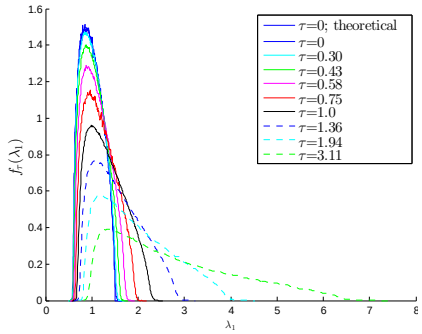


Figure 3.1: *Eigenvalue distribution with correlation time τ ranging from 0 to 3. The 1st blue line, which is shown as stairs, is the theoretical eigenvalue distribution according to the Marcenko-Pastur law (see eq.E.2). In the simulation we have chosen $q = N/T = 50/1000 = 0.05$ and the standard deviation of the returns $\sigma = 1$. For each value of τ we generate 2000 instances of $N \times T$ random matrix R, and compute C as $C = RR'/T$. Hence each curve in the figure is constructed from 2000 sets of eigenvalues. The correspondance between the correlation time τ and the auto-regressive coefficient ϕ is $\tau = -\ln 2/\ln \phi$.*

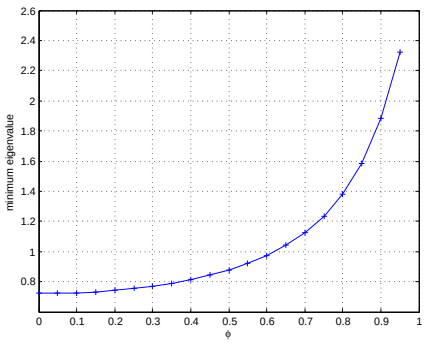


Figure 3.2: *The minimum eigenvalue versus auto-correlation strength ϕ. For each value of ϕ 2000 random matrices are generated and their eigenvalues are calculated. The minimum eigenvalue of each random matrix is noted and the mean of the 2000 such minimum eigenvalues are plotted against the chosen value of ϕ. 20 values of ϕ are included in the plot, ranging from 0 to 0.95 with step size 0.05.*

is present[14]. So we compare in figure 3.3 the empirical cumulative distribution function (CDF) of the maximum eigenvalue with the CDF of a gamma distribution. The cases where auto-correlations are present ($\phi > 0$) have also

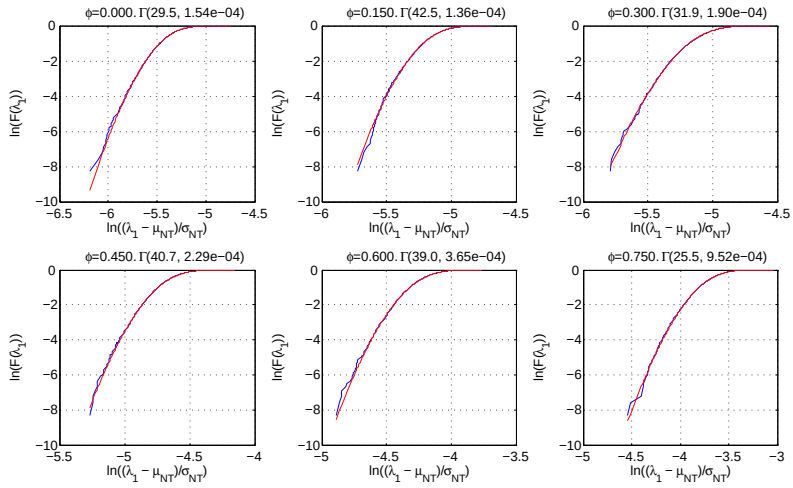


Figure 3.3: *Cummulative distribution function (CDF) of the maximum eigenvalue (λ_1) from numerical simulations (blue lines) are compared to the fitted gamma distribution (red lines). Each pair of CDFs correspond to a fixed autocorrelation strength (ϕ). The parameters k and θ of the gamma distribution are fit to data by matching the 2nd and the 3rd moments of the gamma distribution to the corresponding moments of the empirical distribution. Then the parameter α in equation E.3 is chosen to be $\alpha = k\theta - \mathbb{E}\left(\frac{\lambda_1 - \mu_{NT}}{\sigma_{NT}}\right)$. The curves are plotted on log-log scale.*

been included. It is seen in the figure that gamma distributions with different parameters fit fairly well in all cases. So we conclude that a gamma distribution not only approximates the maximum eigenvalue distribution at the absence of autocorrelations but does so even at the *presence* of autocorrelations.

Since the maximum eigenvalue distribution is approximated by a gamma distribution characterized by parameters k, θ, and α [2], the influence of the autocorrelations can be characterized by the dependence of k, θ, and α on ϕ. While these dependences are rather intricate, good support can be found in the data for the following approximate relation:

$$k\theta \quad = \quad a \tan^2 \frac{\pi\phi}{2} + b \tan \frac{\pi\phi}{2} + c \tag{3.5}$$

Here we note that $k\theta$ is the mean of the gamma distribution. To verify this

[2] The mean, variance and skewness of the gamma distribution are given by

$$\begin{aligned} \text{mean} &= k\theta \\ \text{variance} &= k\theta^2 \\ \text{skewness} &= 2/\sqrt{k} \end{aligned}$$

relation, we first fit a 2nd order polynomial and obtain the coefficients a, b, c; then for each data point $k_n\theta_n$ we solve the quadratic equation

$$a\tan^2\frac{\pi\phi'_n}{2} + b\tan\frac{\pi\phi'_n}{2} + c - k_n\theta_n \quad = \quad 0 \tag{3.6}$$

for $\tan\frac{\pi\phi'_n}{2}$. If relation 3.5 is a good approximation, a close match between $\tan\frac{\pi\phi'_n}{2}$ and $\tan\frac{\pi\phi_n}{2}$ is expected. From figure 3.4 one can see this is indeed the case.

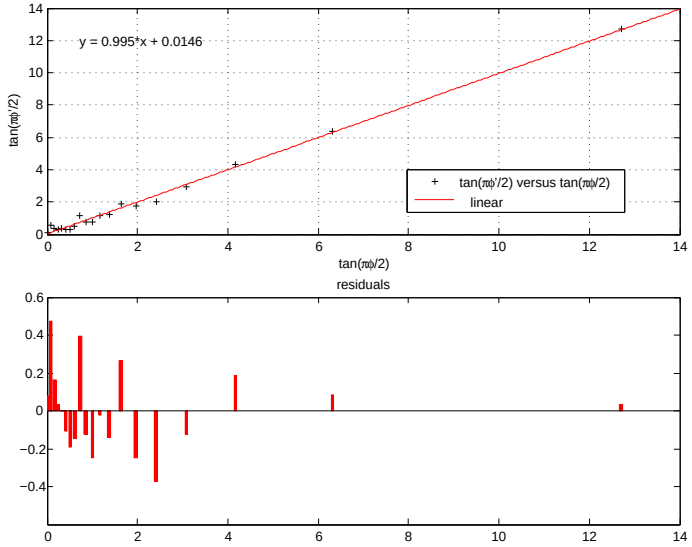


Figure 3.4: *Upper plot:* $\tan\frac{\pi\phi'_n}{2}$ *against* $\tan\frac{\pi\phi_n}{2}$. *The fitted line has equation* $y_n = 0.995\tan\frac{\pi\phi_n}{2} + 0.0146$. *Lower plot: Residuals of the linear fit, i.e.* $\tan\frac{\pi\phi'_n}{2} - y_n$. *20 values of* ϕ *are included in the plot, ranging from 0 to 0.95 with step size 0.05.*

Chapter 4

Covariance Matrix of GARCH(1,1) Returns

GARCH models and particularly GARCH(1,1) models are widely used to model financial return series of various time scales, ranging from daily and monthly returns that have been studied extensively in the literature to intraday returns that we have selectively investigated in chapter 2. One particularly nice feature of GARCH(1,1) models is that they have regularly varying tails (power-law tails) [9, 2] even when the innovations (denoted z_t in the following text) are normally distributed — something not shared by other classes of models (e.g. not by SV models, as shown in section 2.3) but well documented for realistic returns data by empirical studies [17, 13, 18]. However, what this tail behavior implies for the covariance matrix is much less understood, especially when the return series of the covariance matrix are auto-correlated.

So in this chapter we consider the covariance matrix of N identically specified, possibly auto-correlated GARCH(1, 1) processes:

$$\begin{aligned} r_{it} &= \phi r_{i,t-1} + \epsilon_{it} \\ \epsilon_{it} &= \sigma_{it} z_{it} \end{aligned} \tag{4.1}$$

where $i = 1, 2, ..., N$; $t = 1, 2, ..., T$; z_{it} is independent, identically distributed, and

$$\sigma_{it}^2 = \alpha_0 + \alpha_1 z_{i,t-1}^2 + \beta_1 \sigma_{i,t-1}^2$$

Mikosch and Starica showed in [2] that a GARCH(1,1) process satisfying

$$\begin{aligned} \alpha_0 &> 0 \\ \mathbb{E} \ln(\alpha_1 Z^2 + \beta_1) &< 0 \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[(\alpha_1 Z^2 + \beta_1)^{p/2}] &\geq 1 \\ \mathbb{E}|Z|^p \ln|Z| &\leq \infty \end{aligned}$$

for some $p > 0$, is stationary and has regularly varying tails. The tail exponent α is determined by:

$$\mathbb{E}[(\alpha_1 Z^2 + \beta_1)^{\alpha/2}] = 1 \tag{4.2}$$

Here Z is a random variable that has the same distribution as z_{it}. In our simulations described hereafter, z_{it} and Z have standard Gaussian distribution. In this section and the next, we first study situations where no auto-correlations are present among the returns, i.e. $\phi = 0$; then in section 4.2 we look at how auto-correlations change the picture.

With regularly varying tails, the eigenvalue distribution of a covariance matrix built from GARCH(1,1) returns is expected to differ from the Marcenko-Pastur law discussed in section E. In figure 4.1 we simulate N=50 independent GARCH(1,1) returns series, each with identical parameters, namely $\alpha_0 = 2.3 \times 10^{-6}$, $\alpha_1 = 0.15$, $\beta_1 = 0.84$, $\phi = 0$ and T=8×10^4 time steps, then we build the covariance matrix as

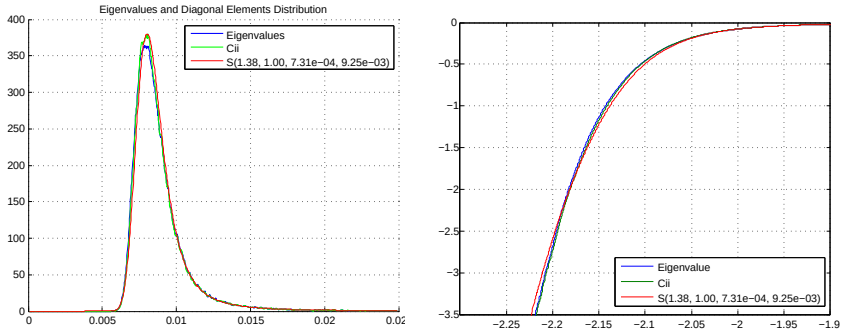
$$\mathbf{C} = \frac{1}{T^{2/\alpha}} \mathbf{R}\mathbf{R}'$$

where $\mathbf{R}$ is an $N \times T$ matrix, whose elements r_{it} are specified by equation 4.1 with $\phi = 0$. The normalization factor $\frac{1}{T^{2/\alpha}}$ has been chosen such that the eigenvalue distribution is independent of T in the limit $T \to \infty$ [1, 19].

The PDF of the eigenvalue distribution of C is plotted in figure 4.1(a) and CDF of the distribution is plotted on log-log scale in 4.1(b). Also plotted in the same figure is the distribution of the diagonal elements of $\mathbf{C}$. It is clear from the figure that the two PDFs coincide, implying $\mathbf{C}$ is diagonal. This is further confirmed by figure 4.2 which shows the distribution of the non-diagonal elements of C. One can see the non-diagonal elements are distributed symmetrically around 0 with a very small width in comparison to the distribution of the diagonal elements — in fact, 1 order of magnitude smaller (2.14×10^{-5} v.s. 7.31×10^{-4}). Hence $\mathbf{C}$ is very close to a diagonal matrix.

Figure 4.1 also shows the two curves are well fitted by an α-stable distribution. An estimate of the Lévy index α of the stable distribution is also obtained via fitting, $\alpha \approx 1.38$, as shown in figure 4.1. This is really an expected result for the diagonal elements. Using $\alpha_1 = 0.15$, $\beta_1 = 0.84$, which are the values used for simulating the GARCH returns, one can obtain, by solving equation 4.2, $\alpha = 2.96$. Then according to Mikosch and Starica [2]

$$\begin{aligned} P(|r_t| > x) &\sim \frac{\mathbb{E}(z^\alpha) c_0}{x^\alpha} \\ P(|r_t|^2 > x) &\sim \frac{\mathbb{E}(z^\alpha) c_0}{x^{\alpha/2}} \end{aligned}$$



(a) (b)

Figure 4.1: *4.1(a): Eigenvalues and Diagonal elements' distribution of a covariance matrix built from independent GARCH return series. Blue: PDF of eigenvalues; Green: PDF of diagonal elements; Red: PDF of a α-stable distribution fitted to the diagonal elements. 4.1(b): CDF of the same quantities in 10-based log-log scale.*

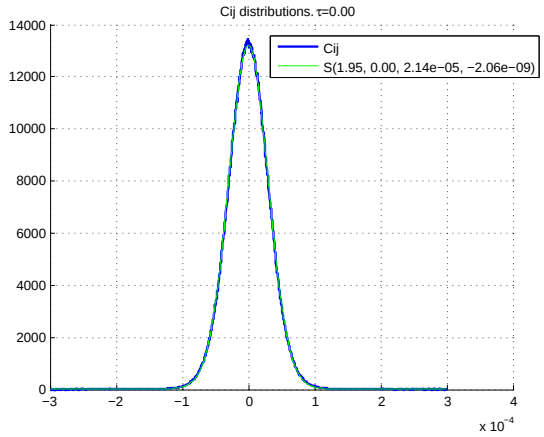


Figure 4.2: *Distribution of the non-diagonal elements of the covariance matrix. Blue: PDF of the non-diagonal elements; Green: α-stable distribution fitted to the non-diagonal elements' PDF.*

for some constant c_0. Now that $P(|r_t|^2 > x)$ has power-law tail behavior with power $\alpha/2 < 2$, one can deduce

$$\sum_{t=1}^{T} r_t^2 \xrightarrow{d} S(\alpha/2, 1, \gamma, \mu) \text{ as } T \to \infty \tag{4.3}$$

where $\xrightarrow{d}$ denotes convergence in distribution, and $S(\alpha/2, 1, \gamma, \mu)$ denotes an α-stable distribution with parameters $(\alpha/2, 1, \gamma, \mu)$. Here $\alpha/2$ is the Lévy index, 1 is the asymmetry, γ is the scaling parameter and μ is the mean value of the distribution. Asymmetry being 1 means a random variable so distributed only takes positive values [20, 21].

The mean μ in equation 4.3 is given by

$$\begin{aligned} \mu &= T\mathbb{E}(|r_t|^2) \\ &= T\frac{\alpha_0}{1-\alpha_1-\beta_1} \end{aligned}$$

where we have used the result $\mathbb{E}(|r_t|^2) = \alpha_0/(1-\alpha_1-\beta_1)$ [4]. The scaling parameter γ in 4.3 is determined by the limit [20]

$$\lim_{T\to\infty} \frac{T\mathbb{E}(z^{\alpha/2})c_0}{\gamma^{\alpha/2}} = C_{\alpha/2}$$

where

$$\begin{aligned} C_{\alpha/2} &= \left(\int_0^\infty \frac{\sin x}{x^{\alpha/2}} dx\right)^{-1} \\ &\approx \frac{1}{\sqrt{2\pi}} \end{aligned}$$

Therefore

$$\begin{aligned} \gamma^{\alpha/2} &= \sqrt{2\pi} T\mathbb{E}\left(|z|^{\alpha/2}\right) c_0 \\ \gamma &= (2\pi)^{1/\alpha} T^{2/\alpha} \left(\mathbb{E}|z|^{\alpha/2}\right)^{2/\alpha} c_0^{2/\alpha} \end{aligned}$$

Here we note that an α-stable distribution $S(\alpha, \beta, \gamma, \mu)$ has characteristic function [18]

$$\varphi(k;\alpha,\beta,\gamma,\mu) = \exp\left[i\mu k - \gamma^\alpha |k|^\alpha \left(1 - i\beta\frac{k}{|k|}\tan\frac{\pi\alpha}{2}\right)\right] \text{ for } \alpha \neq 1$$

from which we see $\varphi(ak;\alpha,\beta,\gamma,\mu) = \varphi(k;\alpha,\beta,a\gamma,a\mu)$, implying that, if $x \sim S(\alpha,\beta,\gamma,\mu)$, then $ax \sim S(\alpha,\beta,a\gamma,a\mu)$.

Now that

$$\sum_{t=1}^{T} r_t^2 \xrightarrow{d} S(\alpha/2, 1, \gamma, \mu) \text{ as } T \to \infty$$

we have

$$\begin{aligned} C_{ii} &= \frac{1}{T^{2/\alpha}} \sum_{t=1}^{T} r_{it}^2 \qquad (4.4) \\ &\xrightarrow{d} S(\alpha/2, 1, \gamma_D, \mu_D) \text{ as } T \to \infty \qquad (4.5) \end{aligned}$$

where

$$\begin{aligned} \gamma_D &= (2\pi)^{1/\alpha}\mathbb{E}\left(|z|^{\alpha/2}\right)^{2/\alpha} c_0^{2/\alpha} \\ \mu_D &= \frac{\alpha_0}{1-\alpha_1-\beta_1}T^{1-2/\alpha} \end{aligned} \tag{4.6}$$

So the diagonal elements of the covariance matrix converge to an α-stable distribution with Lévy index $\alpha/2 \approx 1.48$. This is comparable to the index value 1.38 obtained by fitting. Considering the slow convergene of regularly varying tails, this is a reasonably good match.

Now we look at the distribution of the non-diagonal elements. Figure 4.2 shows that an α-stable distribution fits rather well, and additionally, figure 4.3 shows that the distribution of these non-diagonal elements has fat tails, supporting an α-stable distribution.

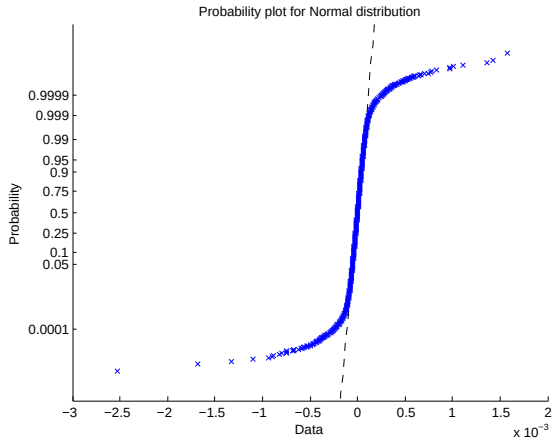


Figure 4.3: *Probability plot of the non-diagonal elements. Blue: The accumulative probability function (CDF) of the non-diagonal elements. Black, dashed: CDF of the Gaussian distribution that has the same mean and variance as the sample. The graph is arranged on such a scale that the Gaussian CDF is a straight line.*

The parameters of the fitted α-stable distribution $S(\alpha', \beta', \gamma', \mu')$ are obtained in the procedure of fitting. The results have been shown in the legend of figure 4.2. In table 4.1 we list them with a higher precision. Since r_{it} and r_{jt} are independent of each other, β' and μ' are expected to be 0 — but with a finite T, some deviation from 0 is not surprising.

Now that the parameters' values have been obtained, the way they scale

α'	β'	γ'	μ'
1.9453	0.0018	2.1381×10^{-5}	2.0637×10^{-9}

Table 4.1: *Parameters of the non-diagonal elements' distribution*

with T, i.e. the length of the return series, can be deduced. Consider

$$T^{2/\alpha} C_{ij} = \sum_{t=1}^{T} r_{it} r_{jt} \xrightarrow{d} S(\alpha', 0, \gamma', 0) \tag{4.7}$$

where the width γ' is determined by [20],

$$\begin{aligned} \lim_{T\to\infty} \frac{TC}{\gamma'^{\alpha'}} &= C_{\alpha'} \\ C_{\alpha'} &= \left(\int_0^\infty \frac{\sin x}{x^{\alpha'}} dx \right)^{-1} \end{aligned}$$

Here the constant C is such that

$$P(|r_{it} r_{jt}| > x) \sim \frac{C}{x^{\alpha'}} \text{ as } x \to \infty$$

So we have

$$\gamma' = \left(\frac{CT}{C_{\alpha'}} \right)^{1/\alpha'}$$

Divide throughout equation 4.7 by $T^{2/\alpha}$ then gives

$$C_{ij} = \frac{1}{T^{2/\alpha}} \sum_{t=1}^{T} r_{it} r_{jt} \xrightarrow{d} S(\alpha', 0, \gamma_N, 0) \tag{4.8}$$

where

$$\gamma_N = \left(\frac{C}{C_{\alpha'}} \right)^{1/\alpha'} T^{1/\alpha' - 2/\alpha}$$

So we see that distribution of the non-diagonal elements has a width that scales with T as $T^{1/\alpha'-2/\alpha} \approx T^{-1/6}$, while the width of the diagonal elements' distribution, as shown in equation 4.6, does not scale with T. This is to be compared with the Wishart case where the return series have Gaussian distribution, and hence the asymptotic distributions of both the diagonal and the non-diagonal elements of the covariance matrix are Gaussian with a variance that scales as $1/T$ (c.f. 3.1).

4.1 Implications of Finite Number of Observations

In the last section we have discussed the limiting situation where the return series of the covariance matrix have an infinite number of observations ($T \to \infty$), and in the simulation studies we have generated a large number of observations for each return series, namely $T = 8 \times 10^4$. However, in practice, one often does not have such a large number of observations available. Therefore, it is useful to investigate situations where T only has a modest size.

Figure 4.4 shows the distributions of the diagonal elements as well as the eigenvalues when the number of return series (N) is fixed at 250 and the number of observations in each series (T) is gradually increased from 3000 to 6000. It is

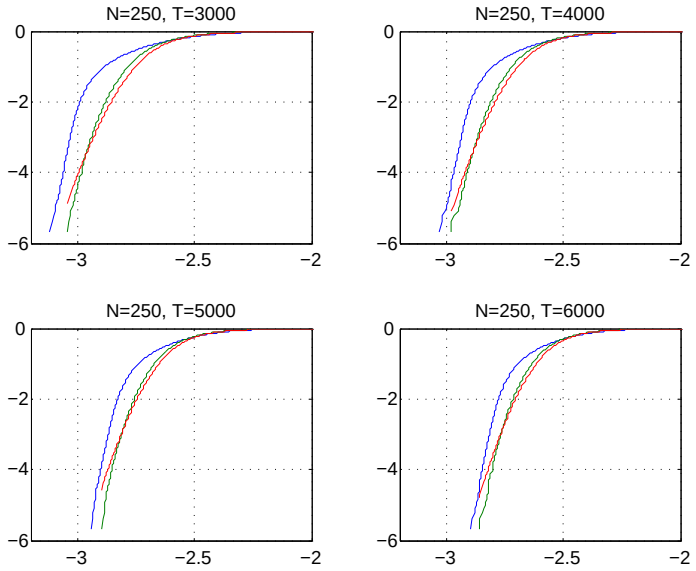


Figure 4.4: *The Diagonal elements' and the eigenvalues' distributions with modest T. Blue: eigenvalues' CDF; Green: Diagonal elements' CDF; Red: CDF of the α-stable distribution fitted to the diagonal elements. All the curves are drawn on 10-based log-log scale.*

seen from the figure that, compared to the earlier case where $T = 8 \times 10^4$, the eigenvalue distribution and the diagonal elements' distribution do not coincide as well but differ rather siginificantly for small values. For large values of the diagonal elements and the eigenvalues, the two do coincide and comply with the

fitted α-stable distribution. Moreover, the difference between the eigenvalues' distribution and the diagonal elements' distribution, as well as that between the diagonal elements' distribution and the fitting α-stable distribution, is also seen to diminish as T increases.

The convergence of the eigenvalues (λ) towards the diagonal elements C_{ii} as $\lambda \to \infty$ is really an anticipated result. For convenience, we order the diagonal elements so that

$$C_{11} < C_{22} < \cdots < C_{NN}$$

where, as before, N is the dimension of the covariance matrix C. Then, as $x \to \infty$,

$$\begin{aligned} P(C_{ii} > x) &\sim \frac{c}{x^{\alpha/2}} \\ f(C_{ii}) &\sim \frac{c}{x^{\alpha/2+1}} \end{aligned}$$

where $f(\cdot)$ denotes the PDF of the diagonal elements and c is some constant. Thus, at the limit $C_{ii} \to \infty$, the distance between two adjacent diagonal elements C_{ii} and $C_{i+1,i+1}$ can be expressed as

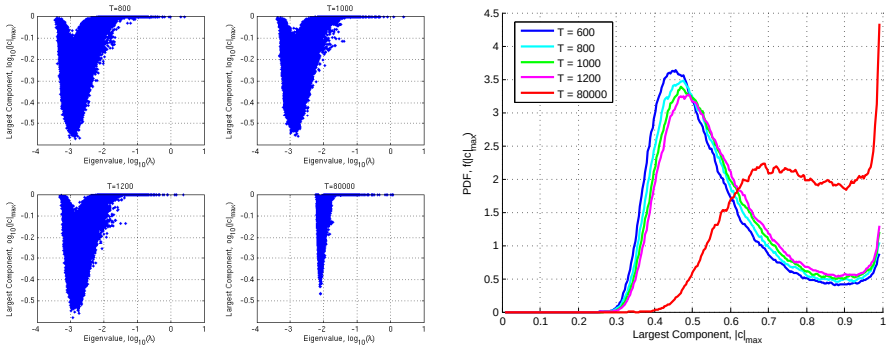
$$\begin{aligned} \frac{1}{N(C_{i+1,i+1} - C_{ii})} &= f(C_{ii}) \\ C_{i+1,i+1} - C_{ii} &\sim \frac{C_{ii}^{\alpha/2+1}}{cN} \end{aligned}$$

Thus as $C_{ii} \to \infty$, $C_{i+1,i+1} - C_{ii} \to \infty$ while $C_{ij} \to 0$ $(i \neq j)$. Now that the spacing between adjacent diagonal elements become wider and the non-diagonal elements become smaller, the eigenstates considered as a mixture of the basis states, become more and more localized to a prominent basis state. This localization can be measured by the size of the component in each eigenvector that has the largest absolute value($|c|_{\max}$), provided that the eigenvectors have been normalized.

Figure 4.5(a) shows how $|c|_{\max}$ changes in response to increasing λ. It is seen that as T increases, localization of the eigenstates proceeds from those with very large eigenvalues towards those with relatively smaller eigenvalues. At the same time, the minimum of $|c|_{\max}$ increases and so does the minimum of the eigenvalues. The last point here is further illustrated in figure 4.5(b) where the PDF of $|c|_{\max}$ is plotted. We see in this figure that an increased value of T leads to advancement of $\min(|c|_{\max})$ to larger values as well as to an increased proportion of large $|c|_{\max}$. The mean of $|c|_{\max}$ has apparently been increased too.

Another informative quantity that measures the localization is the "Inverse Participation Ratio" (IPR). For a given normalized eigenvector $\mathbf{c}_i = (c_{1,i}, c_{2,i}, ..., c_{N,i})$, the Inverse Participation Ratio (IPR) is defined as [22]

$$\text{IPR}(\mathbf{c}_i) = \sum_{k=1}^{N} c_{k,i}^4 \tag{4.9}$$

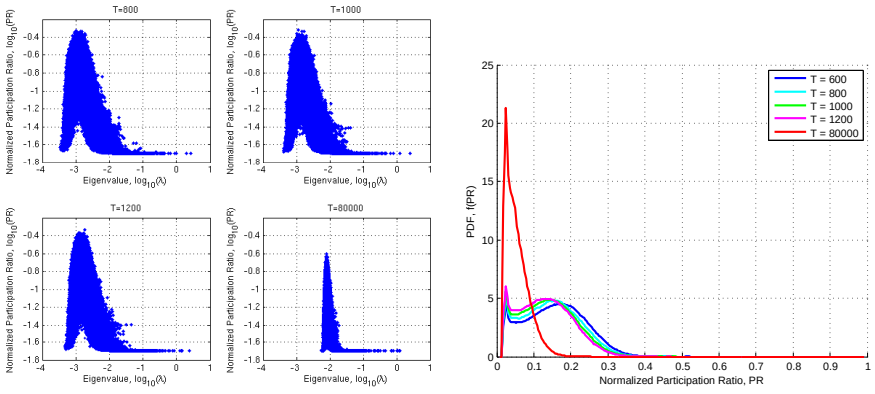


(a) (b)

Figure 4.5: *4.5(a): Horizontal: eigenvalues ($\log\lambda$); Vertical:* $\log|c|_{max}$ *of the corresponding eigenvector. From left to right and 1st to 2nd row, T has values 800, 1000, 1200, 80000. 4.5(b): PDF of the largest component ($|c|_{max}$) of each eigenvector. In both plots the number of returns (N) is 50.*

Figure 4.6(a) shows $\frac{1}{N\cdot \mathrm{IPR}(\mathbf{c}_i)}$ in correspondence to the eigenvalues. This quantity is sometimes termed the normalized Participation Ratio (PR) and measures the proportion of basis vectors that contribute considerably to the eigenvector in question. From this figure we see that, for all values of T, if an eigenvalue is larger than $10^{-1.5} \approx 0.03$, its corresponding participation ratio is less than $10^{-1.6} = 2.5\%$, meaning less than $50 \times 0.025 = 1.26$ basis vectors contribute — each of the corresponding eigenvectors is localized to a single basis vector and hence the distribution of such large eigenvalues is the same as the diagonal elements' distribution.

Figure 4.6(b) shows the PDF of the normalized PR. We see that as T increases, the distribution of PR is compressed towards 0, suggesting increased localization of the eigenvectors. In conclusion, localization of the eigenvectors, which implies coincident eigenvalue and diagonal elements' distributions, begins with those associated to large eigenvalues. Increased observation points lead to increased localization and hence increased coincident sections of the eigenvalue and diagonal elements' distributions. However, this increment with T is slow, because the diagonal elements mean μ_D increases only as a fractional power of T, namely $T^{1-2/\alpha}$, and the non-diagonal elements' variance decreases only as a fractional power too, namely $T^{1/\alpha'-2/\alpha}$. These have been detailed in equations 4.6 and 4.8.



(a) (b)

Figure 4.6: *4.6(a): Normalized participation ratio (PR) versus eigenvalue (λ). Horizontal:* $\log \lambda$*; Vertical:* $\log PR$ *of the corresponding eigenvector. From left to right and 1st to 2nd row, T has values 800, 1000, 1200, 80000. 4.6(b): PDF of the normalized PR. In both plots, the number of returns (N) is 50.*

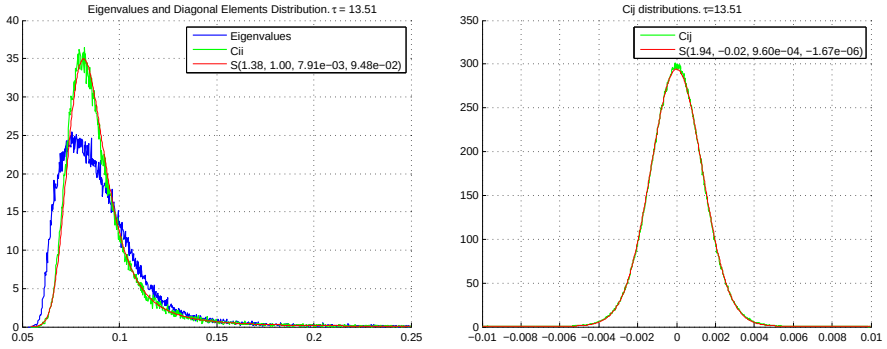
4.2 Influence of auto-correlations

In the previous two sections we have studied situations where $\phi = 0$ in the specification 4.1, i.e. no auto-correlation is in the returns. In this section we investigate how auto-correlations change the picture.

Figure 4.7(a) shows the eigenvalue as well as the diagonal elements' distribution when $\phi = 0.95$, i.e. $\tau = 13.51$. The values of N and T are 50 and 8×10^4 as before. The GARCH(1,1) parameters are also unchanged, namely $\alpha_0 = 2.3 \times 10^{-6}$, $\alpha_1 = 0.15$, and $\beta_1 = 0.84$. Figure 4.7(b) shows the non-diagonal elements' distribution in the same setup. Included in these plots are $50 \times 2000 = 1 \times 10^5$ eigenvalues and diagonal elements, as well as $\binom{50}{2} \times 2000 = 2,450,000$ non-diagonal elements. These data come from 2000 simulated matrices.

From figure 4.7(a) we see that, as auto-correlations become significant, the distribution of the eigenvalues no longer coincides with the diagonal elements' distribution — instead it becomes wider and fatter on the tails. We also notice that the widths of both the diagonal and the non-diagonal elements' PDF's have increased. In figure 4.1 we see that, when no auto-correlation is present, the PDF of the diagonal elements has width (γ) 7.31×10^{-4}, while in figure 4.7(a) we see that the width has become 7.91×10^{-3} as τ becomes 13.51. Similarly the non-diagonal elements' PDF has width 2.14×10^{-5} when $\tau = 0$, as shown in figure 4.2, and this width becomes 9.60×10^{-4} when $\tau = 13.51$, as shown in figure 4.7(b).

Figure 4.8 shows the eigenvalues' distribution corresponding to a range of ϕ values. The number of eigenvalues in each curve is the same as in figure 4.7(a).



(a) (b)

Figure 4.7: *4.7(a): Eigenvalues' and diagonal elements' distribution when* $\phi = 0.95$*, i.e.* $\tau = 13.51$*; 4.7(b): Non-diagonal elements' distribution in the same situation.*

From this figure one can see that, as auto-correlation strengthens,

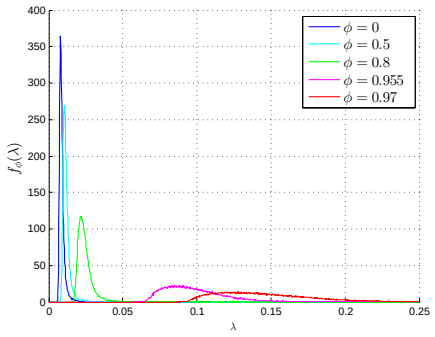


Figure 4.8: *Eigenvalue distribution of covariance matrix built from auto-correlated GARCH processes.* $\phi = 0, 0.5, 0.8, 0.955, 0.97$ *correspond to correlation time* $\tau = 0, 1.00, 3.11, 15.05, 22.76$*.*

- the PDF of the eigenvalue distribution flattens and widens;
- the minimum as well as the maximum eigenvalues increase.

To find out more about this series of deformation, we first look at how the largest component and the normalized participation ratio of the eigenvectors change as ϕ takes on larger values. Figure 4.9 shows the largest eigenvector component $|c|_{\max}$ in correspondence to the eigenvalue. Apparently, as auto-correlation strengthens, the eigenvectors' composition fractures, leading to a

reduced degree of localization and even reduced certainty of localization — for a fixed eigenvalue, $|c|_{\max}$ now varies in a larger range than it does with smaller ϕ. The same story of reduced localization is also evident from the plot

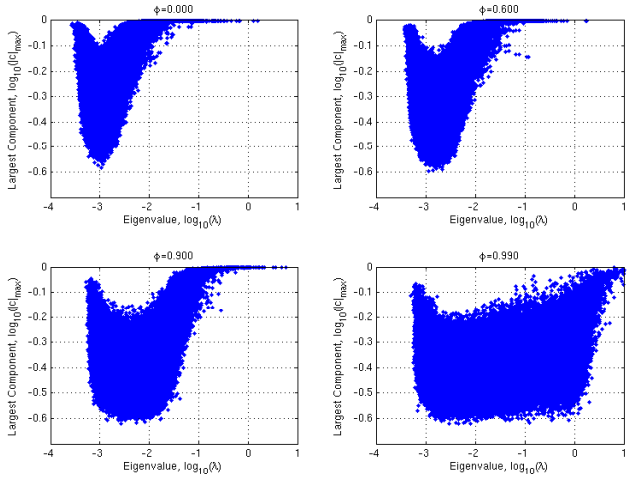
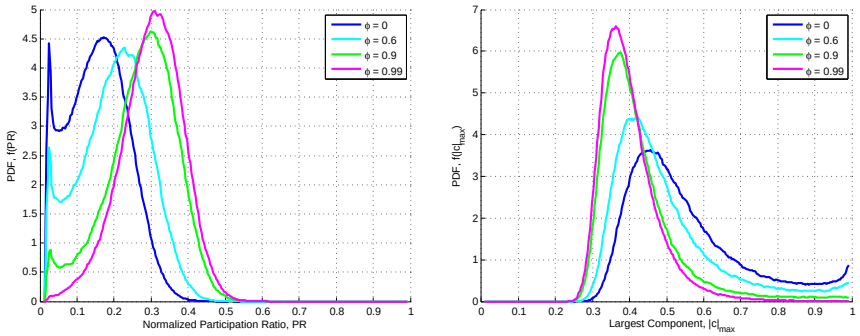


Figure 4.9: *Eigenvectors' largest component versus eigenvalue, for 4 auto-correlation strengths ϕ = 0, 0.6, 0.9, and 0.99. The number of returns (N) is 50.*

of the normalized PR, shown in figure 4.10(a), and from the PDF of $|c|_{\max}$ shown in figure 4.10(b). It is seen in 4.10(a) that the peak at the left of the plot, representing the group of localized eigenvectors, falls with increased auto-correlation, and essentially disappears when ϕ reaches the extreme value 0.99. Figure 4.10(b) shows the proportion of large $|c|_{\max}$ values is severely reduced and the mean of $|c|_{\max}$ is pushed to smaller values by increased auto-correlation.

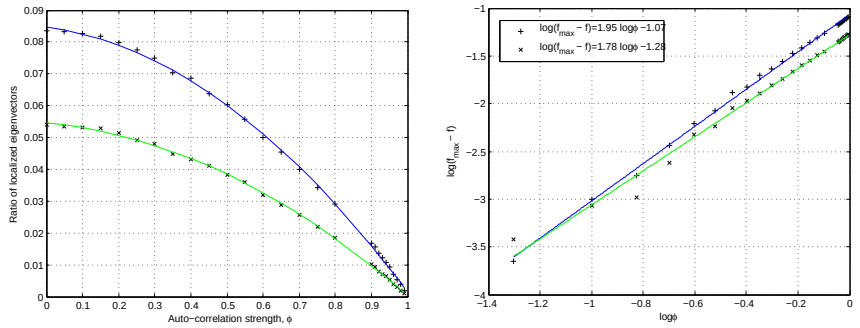
It is also useful to look at how the fraction of localized eigenvectors changes with the auto-correlation. For definiteness, we classify an eigenvector as being localized when (1) its number of participating basis vectors is less than 2, or (2) the largest of its components' absolute values is larger than 0.9.

Figure 4.11(a) shows how the ratio of localized eigenvectors depends on the auto-correlation strength ϕ. In either way of classification, the ratio falls with ϕ in accordance with a power law, the power exponent lying a bit below 2. This is further confirmed in plot 4.11(b), where the ratios are plotted versus ϕ on log-log scale.



(a) (b)

Figure 4.10: *4.10(a): PDF of the normalized participation ratio (PR). 4.10(b): PDF of the largest component of the eigenvectors. The number of returns (N) is 50. ϕ values of 0, 0.6000, 0.9000, 0.9900 correspond to correlation time τ = 0, 1.3569, 6.5788, 68.9676*



(a) (b)

Figure 4.11: *4.11(a): Ratio of localized eigenvectors versus the auto-correlation strength ϕ. Black "+": ratio of localized eigenvectors as measured by the number of participating basis vectors being lower than 2; Black "x": ratio of localized eigenvectors as measured by the largest component being larger than 0.9. Blue curve: quadradic function fitted to "+". Green curve: quadratic function fitted to "x". There are 27 data points in the plot, corresponding to 27 ϕ values: 0 to 0.8 with step size 0.05, and 0.9 to 0.99 with step size 0.01. The corresponding values of the correlation time τ range from 0 to 69. 4.11(b): $\ln(f_{max} - f)$ is plotted versus $\ln \phi$, where f stands for the ratio of localized eigenvectors. The fitted curves are linear.*

Chapter 5

Results

Realistic dynamic models for relative price changes (returns) of financial assets have been studied. In particular, the GARCH and SV models have been utilized and compared in their forecast accuracies by case studying of 15- and 30-minute returns of Nordea, Volvo and Ericsson. The following are the conclusions from this investigation:

- In all the 7 studied series, the log-volatility $\ln \sigma_t$ is well described by a seasonally integrated moving average model. Long memory in these series is accounted for by the compounded difference operator $(1 - B)(1 - B^s)$ where B denotes the back-shift operator. s stands for seasonality, which is 33 in the cases of 15-minute returns and 16 in the cases of 30-minute returns.

- An SV model generally yields more accurate forecasts than does the GARCH model for the same series. This is certainly well expected, considering that the SV model incorporates much more data than does GARCH. However, it must be noted that the validity of the aforementioned comparison is underlain by the accuracy of realized volatility as a proxy to the true conditional volatility.

- SV models perform more consistently in terms of forecast accuracy than does GARCH, as shown in appendix A.5.

Regarding the covariance matrix of Gaussian return series, we have arrived at the following results:

- Auto-correlations in the returns rescale the covariance of two series by a factor of $1/(1 - \phi^2)$, where $\phi = \mathrm{corr}(r_t, r_{t-1})$. See eq. 3.1. Moreover, they rescale the variance of each series by $\frac{1}{(1-\phi^2)} \left[1 - \frac{\phi^2}{T(1-\phi^2)}\right]$. See eq. 3.3.

- Auto-correlations in the returns increase the variance of the covariance and also the variance of the variance. See eq. 3.2 and 3.4.

- The largest eigenvalue obeys approximately a gamma distribution even when the returns are auto-correlated. It was shown by Chiani in 2012 that the largest eigenvalue obeyed gamma distribution when the returns were *not* auto-correlated [14]. In addition, we find that the mean of the gamma distribution is approximately quadratic in $\tan \frac{\pi\phi}{2}$.

For the covariance matrix of GARCH(1,1) series, our contributions are the following:

- The diagonal and non-diagonal matrix elements both have Lévy distributions but with different Lévy indices.
- The power-law tails of GARCH(1,1) returns lead to a group of localized eigenvectors that correspond to large eigenvalues.
- Auto-correlations in the returns reduce the localization of the eigenvectors. The fraction of localized eigenvectors decreases approximately as a quadratic function of $\phi = \text{corr}(r_t, r_{t-1})$. See figure 4.11(a) and 4.11(b).

In summary, auto-correlations in the returns create illusory cross-correlations. To assess the true cross-correlations among the assets in a market, one has to adopt a model for each of the return series, infer the residuals and then assess the cross-correlations among the residuals instead (c.f. chapter 2).

However, in an efficient market, auto-correlations in the returns are necessarily as weak as indistinguishable from measurement errors so that exploitable arbitrage opportunities do not exist. Hence auto-correlations cannot be completely eliminated by taking residuals of the returns. When estimating the covariance matrix and its eigenvalues and eigenvectors, the illusory effects caused by auto-correlations must be considered as an inherent source of uncertainty.

Chapter 6

Outlook

Continuing from the obtained results, areas of future research may include, first of all, a covariance matrix formed from SV models with power-law tails, which may be obtained by assuming, for example, Student's t distribution or Normal Inverse Gaussian distribution for the return's innovation.

Another interesting area could be models that treat the conditional covariance matrix as an inherent part of the model specification rather than treating it as an inferred quantity, which is the approach taken by the current work. Predecessors on this path are the multivariate GARCH and SV models [9]. To have sufficient flexibility, many of these models involve a large number of parameters, which make it hard to fit them to data and increase the chances of model mis-specification. A model that strikes a good balance between flexibility and complexity will be of great interest. These approaches may well lead to more accurate volatility forecasts and improved estimation of the uncertainties introduced by auto-correlations in the returns.

Analytically deriving the eigenvalue and eigenvector distributions of a covariance matrix formed from return series described by a realistic model will be very challenging but interesting. It is also interesting to see whether the methods developed for Wishart matrices, for example the holonomic gradient method [23], can be applied to the aforementioned covariance matrices. Advancement in these areas will undoubtedly find applications in e.g. principle component analysis and portfolio management.

Chapter 7

Self-reflection

First of all, during the thesis project I learned a lot about models of time series in finance, such as GARCH, and SV models, as well as traditional time series analysis, for example, ARIMA models. In addition, a study of continuous-time models gave me an opportunity to increase my knowledge of stochastic differential equations, probability theory, and statistics. I am a beginner, but nonetheless find it comfortable to read books and articles on these subjects after the thesis project.

Moreover, I have introduced myself to the random matrix theory, multivariate analysis, and extreme value theory. These became relevant when I worked on the covariance matrix of returns described by Gaussian or GARCH models. I am still a long way from being able to do serious proofs or derivations in these areas, but I know what to read and can understand the basics.

In addition, I have become better at Matlab programming and dealing with databases, and have grown more confident in numerical analysis.

Appendix A

Case Study of Some Intraday Series

A.1 Nordea 30-minute Returns

In this section we study the volatility of Nordea Bank 30-minute returns in the period 2013/10/10 - 2014/04/04. The realized volatilities that approximate the volatilities of these 30-minute returns are computed using 1-minute returns. As in the previous cases, this choice of 1-minute returns is confirmed by the normality of the quotient $(r_t - \mathbb{E}(r_t))/\sigma_t$, which is shown in figure A.1. The

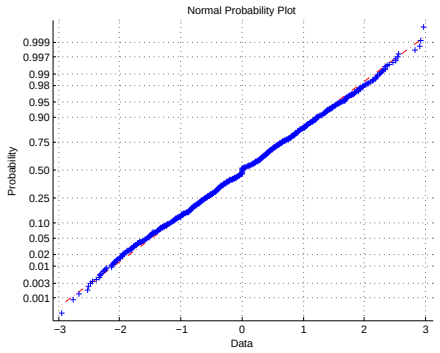
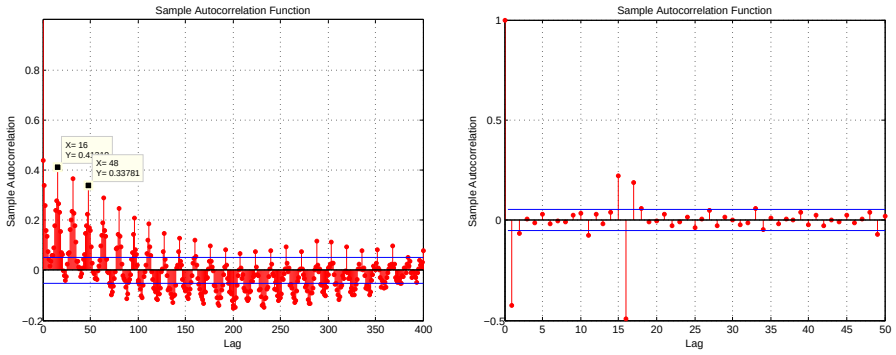


Figure A.1: *Normal probability plot of* $(r_t - \mathbb{E}(r_t))/\sigma_t$

auto-correlations of $\ln \sigma_t$ is shown in figure A.2(a), where we see an apparent seasonality of 16. By differencing we get $w_t = (1 - B)(1 - B^s) \ln \sigma_t$ where

$s = 16$. Its auto-correlations are shown in figure A.2(b). Once again, this



(a) (b)

Figure A.2: *A.2(a): Auto-correlations (ACF) of* $\ln \sigma_t$*; A.2(b): Auto-correlations (ACF) of* $w_t = (1-B)(1-B^s)\ln \sigma_t$.

auto-correlation structure points to a seasonal moving average model:

$$w_t \quad = \quad (1-\theta B)(1-\Theta B^s)y_t$$

where y_t is assumed to have Gaussian distribution, neglecting slight non-normality as before. Maximum likelihood estimation gives the parameter values listed in table A.1. To fit a GARCH model to the same series, we assume Gaussian inno-

Parameter	θ	Θ	var(y_t)
Value	0.7612	0.8036	0.0736

Table A.1: *Parameter values of the Seasonal Moving Average model*

vations (c.f. equation 2.1). The result is a GARCH(1,1) model. Its parameters are listed in table A.2.

Parameter	α_0	α_1	β_1
Value	2.3×10^{-7}	0.05	0.9

Table A.2: *Parameter values of GARCH(1,1) model fitted to Nordea 30-minute returns.*

As before, we compare the accuracies of the GARCH and the stochastic volatility model by comparing their one-ste-ahead forecasts. We compute the

difference between their forecasts and the realized volatilities and then look at the statistics of these difference values. Firstly we show the mean and the standard deviation of the differences in table A.3. It is seen in this table that the

	SV	GARCH	Sample mean
$\mathbb{E}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	-0.0047	0.0130	0.0584
$\text{std}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	0.2602	0.3194	0.3093

Table A.3: *Mean and standard deviation of the 3 kinds of forecasts*

SV forecasts have a mean whose absolute value is just above 1/3 of that of the GARCH forecasts. The standard deviation of the SV forecasts is also smaller than that of GARCH. In addition, we check the distribution of $\ln \sigma_t^F - \ln \hat{\sigma}_t$, and the percentage of "good" forecasts with respect to different criteria of goodness. These are shown in figure A.3 and table A.4, respectively. It is clear that the

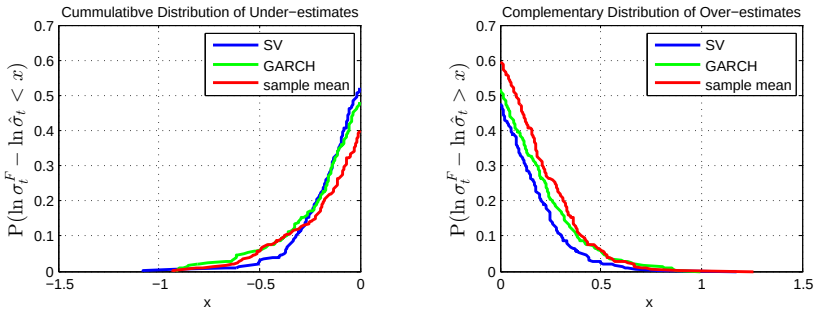


Figure A.3: *Blue: SV forecasts; Green: GARCH forecasts; Red: sample mean forecasts. Left:* $P(\ln \sigma_t^F - \ln \hat{\sigma}_t < x)$*; Right:* $P(\ln \sigma_t^F - \ln \hat{\sigma}_t > x)$*. Horizontal:* x*.*

$\frac{\lvert \ln \sigma_t^F - \ln \hat{\sigma}_t \rvert}{\lvert \ln \hat{\sigma}_t \rvert}$	SV	GARCH	Sample Mean
1%	18%	18%	15%
5%	81%	71%	69%
10%	98%	93%	95%

Table A.4: *Percentage of "good" forecasts as defined by deviating nore more than 1%, 5% and 10% from the measured realized volatility.*

SV forecasts are considerably more accurate.

A.2 Volvo 15-minute returns in 2013/14

In this section we study the volatility of Volvo B 15-minute returns during the period 2013/10/10 - 2014/04/04. The subject of modeling is the realized volatility computed as the square root of the sum of squared 1-minute returns. The normal probability plot of the quotient $(r_t - \mathbb{E}(r_t))/\sigma_t$ is shown in figure A.4, from which one can see it is normally distributed, confirming the choice of 1-minute returns for computing the realized volatility. The auto-correlations of

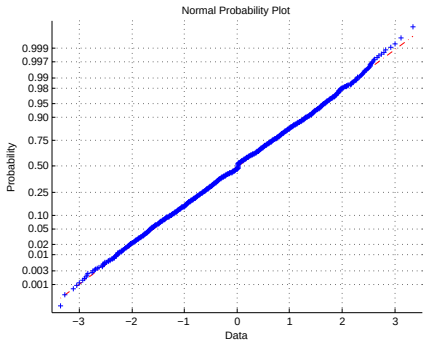
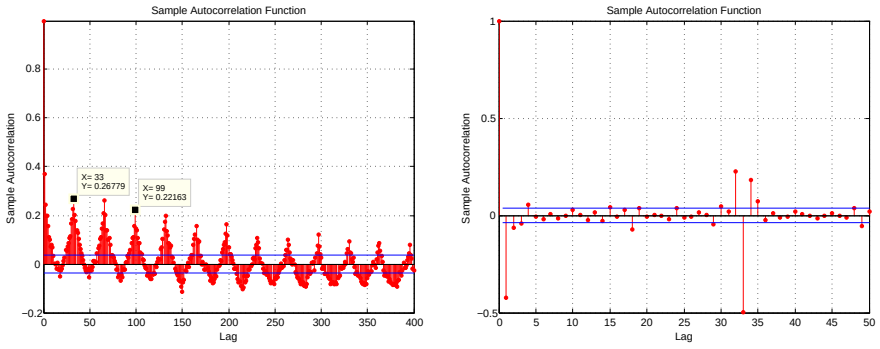


Figure A.4: *Normal probability plot of* $(r_t - \mathbb{E}(r_t))/\sigma_t$

$\ln \sigma_t$ and $w_t = (1-B)(1-B^s)\ln \sigma_t$ are plotted in figure A.5(a) and A.5(b). The former clearly shows the seasonality $s = 33$ in the auto-correlations of $\ln \sigma_t$.



(a) (b)

Figure A.5: *Left: auto-correlations of* $\ln \sigma_t$*; Right: auto-correlations of* $w_t = (1 - B)(1 - B^s)\ln \sigma_t$*.*

Based on the auto-correlations of w_t, a seasonal moving average model is

estimated:

$$(1-B)(1-B^s)\ln\sigma_t \quad = \quad (1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)(1-\Theta B^s)y_t$$

By maximum likelihood estimation, the parameter values listed in table A.5 are obtained.

Parameter	θ_1	θ_2	θ_3	Θ	$\mathrm{var}(y_t)$
Value	0.7491	0.1170	0.0571	0.8646	0.1245

Table A.5: *Parameter values of the stochastoc volatility (SV) model.*

A GARCH model (c.f. equation 2.1) assuming Gaussian innovations is also fitted to the same series. The parameter values are listed in table A.6:

Parameter	α_0	α_1	β_1
Value	6.54×10^{-7}	0.1565	0.6445

Table A.6: *Parameter values of the GARCH model.*

Performing one-step-ahead forecasts using both models gives the series $\ln\sigma_t^F - \ln\hat{\sigma}_t$, where $\ln\hat{\sigma}_t$ are the measured realized volatilities. The mean and standard deviation of $\ln\sigma_t^F - \ln\hat{\sigma}_t$ are shown in table A.7. We see that both the SV and

	SV	GARCH	Sample mean
$\mathbb{E}(\ln\sigma_t^F - \ln\hat{\sigma}_t)$	0.0035	0.0313	-0.1448
$\mathrm{std}(\ln\sigma_t^F - \ln\hat{\sigma}_t)$	0.3356	0.3894	0.3826

Table A.7: *Mean and standard deviation of the 3 kinds of forecasts*

the GARCH model give biased forecasts, but the SV forecasts are only 1/10 as biased as those of GARCH (0.0035 vs. 0.0313). In addition, the standard deviation of the SV forecasts is smaller too. These results are confirmed by the distribution of $\ln\sigma_t^F - \ln\hat{\sigma}_t$ and the fraction of "good" forecasts, which are shown in figure A.6 and table A.8, respectively.

$\frac{\lvert\ln\sigma_t^F - \ln\hat{\sigma}_t\rvert}{\lvert\ln\hat{\sigma}_t\rvert}$	SV	GARCH	Sample Mean
1%	58%	60%	39%
5%	84%	84%	66%
10%	97%	95%	91%

Table A.8: *Percentage of "good" forecasts as defined by deviating nore more than 1%, 5% and 10% from the measured realized volatility.*

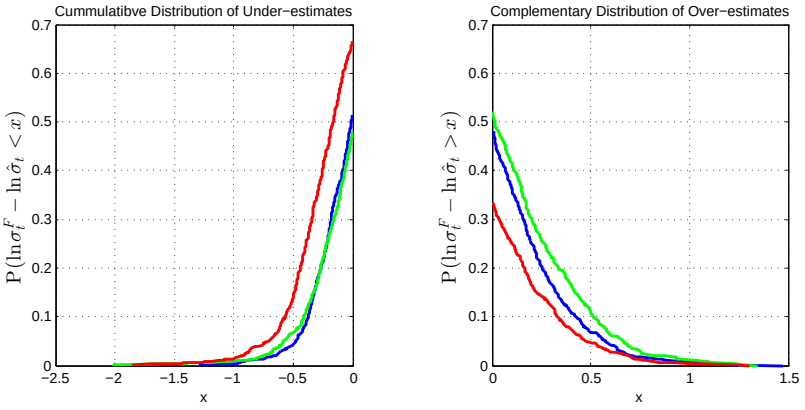


Figure A.6: *Blue: SV forecasts; Green: GARCH forecasts; Red: sample mean forecasts. Left:* $P(\ln\sigma_t^F - \ln\hat{\sigma}_t < x)$*; Right:* $P(\ln\sigma_t^F - \ln\hat{\sigma}_t > x)$*. Horizontal:* x*.*

A.3 Ericsson 15-minute Returns

In this section we model the volatility of *Ericsson B* 15-minute returns during the period 2013/10/10 - 2014/04/04. Using the same method as with other series, we first find the right higher-frequency returns that best approximate the volatility of the 15-minute returns. These turn out to be 50-second returns, as figure A.7 shows.

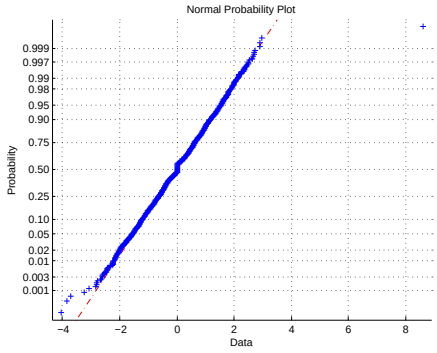


Figure A.7: *Normal probability plot of* $(r_t - \mathbb{E}(r_t))/\sigma_t$*.* σ_t^2 *is computed as the sum of squared 50-second returns.*

Following the procedure described in previous sections, an ARIMA model is found for this series:

$$(1-B)(1-B^s)\ln\sigma_t \quad = \quad (1-\theta_1 B-\theta_2 B^2-\theta_3 B^3)(1-\Theta B^s)y_t$$

where the seasonality s is 33 and the other parameter values are estimated to be those listed in table A.9.

Parameter	θ_1	θ_2	θ_3	Θ	$\mathrm{var}(y_t)$
Value	0.8078	0.0454	0.0943	0.8798	0.1242

Table A.9: *Ericsson B log-volatility parameters*

Then a GARCH model is also found with parameter values listed in table A.10:

Parameter	α_0	α_1	α_s	β_s
Value	2.4181×10^{-7}	0.1513	0.1409	0.6027

Table A.10: *GARCH model of Volvo B 30-minute returns*

The forecasts from both models are compared as follows: Table A.11 shows the mean and standard deviation of the 3 kinds of forecasts: Consistent with

	SV	GARCH	Sample mean
$\mathbb{E}(\ln\sigma_t^F-\ln\hat{\sigma}_t)$	0.0035	0.0600	0.1116
$\mathrm{std}(\ln\sigma_t^F-\ln\hat{\sigma}_t)$	0.3278	0.3667	0.3590

Table A.11: *Standard deviation of* $\ln\sigma_t^F-\ln\hat{\sigma}_t$

previous results, the bias introduced by the SV model is significantly smaller than that introduced by GARCH. Figure A.8 shows the distribution of the difference between a forecast and its corresponding measured realized volatility, i.e. $\ln\sigma_t^F-\ln\hat{\sigma}_t$; table A.12 compares the percentage of "good" forecasts using the 3 alternatives.

$\frac{\lvert\ln\sigma_t^F-\ln\hat{\sigma}_t\rvert}{\lvert\ln\hat{\sigma}_t\rvert}$	SV	GARCH	Sample Mean
1%	17%	17%	16%
5%	73%	69%	69%
10%	96%	93%	94%

Table A.12: *Fraction of "good" forecasts as defined by* $\frac{\lvert\ln\sigma_t^F-\ln\hat{\sigma}_t\rvert}{\lvert\ln\hat{\sigma}_t\rvert}$ *being less than 1%, 5% and 10%.*

It is clear from figure A.8 and table A.12 that the SV model out-performs GARCH. We see that the SV model yields considerably higher fractions of good forecasts by all criteria, and in particular, gives much few under-estimates.

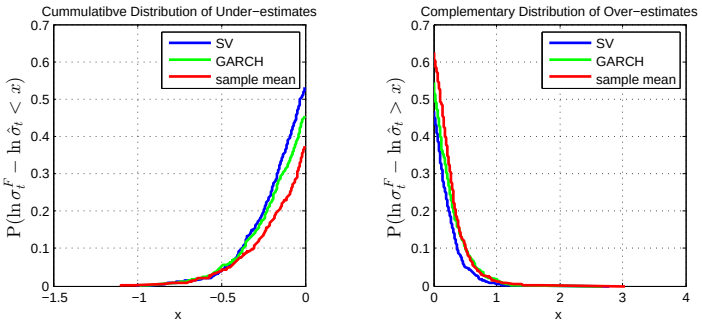


Figure A.8: *Blue: SV forecasts; Green: GARCH forecasts; Red: sample mean forecasts. Left:* $P(\ln\sigma_t^F - \ln\hat{\sigma}_t < x)$*; Right:* $P(\ln\sigma_t^F - \ln\hat{\sigma}_t > x)$*. Horizontal:* x*.*

A.4 Ericsson 30-minute Returns

In this section we look at the 30-minute returns of Ericsson B during the period 2013/10/10 - 2014/04/04. Since the methods are the same as with other series, we shall only present the results here.

First of all, the volatility of this series is found to be well approximated by realized volatilities computed from 75-second returns. The normal probability plot is shown in figure A.9. The following ARIMA model is found for the log-volatility $\ln\sigma_t$:

$$(1-B)(1-B^s)\ln\sigma_t \quad = \quad (1-\theta_1 B-\theta_2 B^2)(1-\Theta B^s)y_t$$

where the seasonality $s = 16$. The parameter values are estimated to be those listed in table A.13: A GARCH model is also found with parameters listed in

Parameter	θ_1	θ_2	Θ	var(y_t)
Value	0.6842	0.2470	0.8391	0.0918

Table A.13: *Ericsson B log-volatility parameters*

table A.14.

Parameter	α_0	α_1	α_s	β_s
Value	7.4856×10^{-7}	0.0532	0.1506	0.6778

Table A.14: *GARCH(1, 1) model of Ericsson B 30-minute returns*

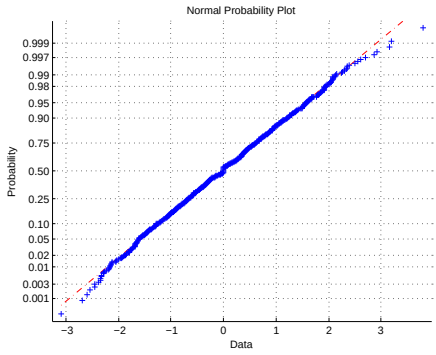


Figure A.9: *Normal probability plot of* $(r_t - \mathbb{E}(r_t))/\sigma_t$. σ_t^2 *is computed as the sum of squared 75-second returns.*

Table A.15 shows the mean and standard deviation of the difference series $\ln \sigma_t^F - \ln \hat{\sigma}_t$, where $\ln \sigma_t^F$ is the forecast log-volatility and $\ln \hat{\sigma}_t$ is the measured realized volatility. Apparently, the GARCH forecasts have a fairly large bias

	SV	GARCH	Sample mean
$\mathbb{E}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	0.0004	0.0526	0.1039
$\text{std}(\ln \sigma_t^F - \ln \hat{\sigma}_t)$	0.2519	0.3019	0.2963

Table A.15: *Standard deviation of* $\ln \sigma_t^F - \ln \hat{\sigma}_t$

compared with those of the SV model. Turning to figure A.10 and table A.16, we also see the SV model performs consistently better — a higher percentage of "good" forecasts are delivered and, while it under-estimates to around the same extent as does GARCH, it certainly over-estimates a lot less. These are confirmative to what we have observed for the other series.

$\frac{\lvert \ln \sigma_t^F - \ln \hat{\sigma}_t \rvert}{\lvert \ln \hat{\sigma}_t \rvert}$	SV	GARCH	Sample Mean
1%	22%	16%	16%
5%	78%	69%	69%
10%	98%	97%	96%

Table A.16: *Fraction of "good" forecasts as defined by* $\frac{\lvert \ln \sigma_t^F - \ln \hat{\sigma}_t \rvert}{\lvert \ln \hat{\sigma}_t \rvert}$ *being less than 1%, 5% and 10%.*

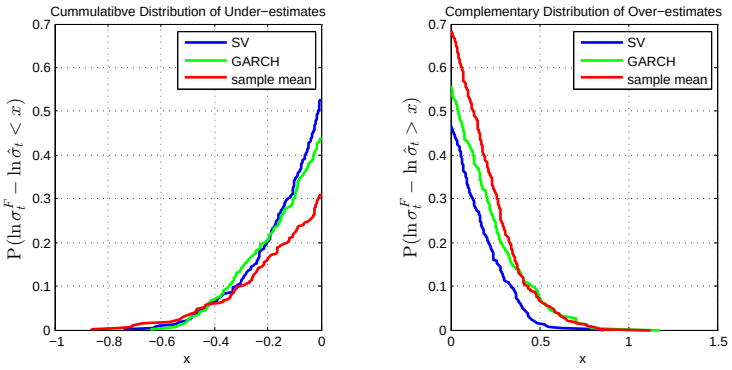


Figure A.10: *Blue: SV forecasts; Green: GARCH forecasts; Red: sample mean forecasts. Left:* $P(\ln\sigma_t^F - \ln\hat{\sigma}_t < x)$*; Right:* $P(\ln\sigma_t^F - \ln\hat{\sigma}_t > x)$*. Horizontal:* x*.*

A.5 Volatility of Forecast accuracy

In this section we compare the volatility of the accuracy of GARCH and SV models' forecasts. Table A.17 lists the the percentage of forecasts that deviate less than 5% from the corresponding realized volatilities, and also gives the standard deviation of the 6 numbers. It is seen that the forecasts of SV models have standard deviation 0.0432 while those of GARCH have 0.0722. Clearly the accuracy of SV forecasts varies considerably less than does that of GARCH forecasts. In other words, a SV model performs more consistently around a level of accuracy.

	Volvo 15m	Volvo 30m	Ericsson 15m	Ericsson 30m
SV	84%	72%	73%	78%
GARCH	84%	62%	69%	69%
	Nordea 1 15m	Nordea 2 30m	standard deviation	
SV	75%	81%	0.0432	
GARCH	71%	71%	0.0722	

Table A.17: *Comparison of the percentage of forecasts that lie within 5% of the corresponding realized volatilities. Nordea 1 refers to the Nordea data set of 2012/01/16 - 2012/04/20 (see section 2.1) and Nordea 2 refers to the data set of 2013/10/10 - 2014/04/04 (see section A.1).*

Appendix B

Normalization Constant in the Unconditional PDF of the Symmetric SV model

In the following we work out the normalization constant C used in section 2.3.1:

$$\begin{aligned}\int_{-\infty}^{\infty} f_{r'}(x)dx &= \frac{1}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\int_{-\infty}^{\infty}\operatorname{erfc}\left(\frac{1}{\sqrt{2}\sigma}\ln\frac{|x|}{\sqrt{\ln 4}}+\frac{\sigma}{\sqrt{2}}\right)dx \\ &= \frac{2}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\int_{0}^{\infty}\operatorname{erfc}\left(\frac{1}{\sqrt{2}\sigma}\ln\frac{x}{\sqrt{\ln 4}}+\frac{\sigma}{\sqrt{2}}\right)dx\end{aligned}$$

Let

$$\begin{aligned}a &= \frac{1}{\sigma\sqrt{2}} \\ b &= -\frac{1}{2\sigma\sqrt{2}}\ln\ln 4+\frac{\sigma}{\sqrt{2}} \\ y &= a\ln|x|+b\end{aligned}$$

Then

$$\begin{aligned}
\int_{-\infty}^{\infty} f_{r'}(x)dx &= \frac{2}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\int_{-\infty}^{\infty} dy \frac{e^{(y-b)/a}}{a}\mathrm{erfc}(y) \\
&= \frac{2}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\, e^{(y-b)/a}\mathrm{erfc}(y)\Big|_{y=-\infty}^{\infty} + \frac{2}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}}\int_{-\infty}^{\infty} dy \frac{2}{\sqrt{\pi}} e^{(y-b)/a} e^{-y^2} \\
&= \frac{2}{C}\frac{e^{\sigma^2/2}}{\sqrt{8\pi}} 2e^{\ln\ln 4/2 - \sigma^2/2} \\
C &= \sqrt{\frac{2\ln 4}{\pi}} \\
&\approx 0.9394
\end{aligned}$$

Appendix C

PDF of the covariance matrix of auto-correlated Gaussian Returns

When autocorrelations exist among the columns of the Gaussian returns matrix $\mathbf{R}$, the distribution of the covariance matrix $\mathbf{C} = \mathbf{R}\mathbf{R}'$ is no longer Wishart but is nonethelss closely related to it. In the following we consider the situation where $r_{i,t}$ can be represented as a vector auto-regressive process. Specifically, suppose

$$\begin{aligned} \mathbf{r}_t &= \sum_{k=1}^{p} \phi_k \mathbf{r}_{t-k} + \mathbf{a}_t \\ \mathbf{a}_t &= \mathbf{r}_t - \sum_{k=1}^{p} \phi_k \mathbf{r}_{t-k} \end{aligned}$$

where $\mathbf{a}_t = (a_{1,t}, a_{2,t}, \cdots, a_{N,t})' \sim N(0, \mathbf{\Sigma})$ and comprise the columes of $\mathbf{A}$; $\mathbf{r}_t$ are the columes of $\mathbf{R}$. Here $N(0, \mathbf{\Sigma})$ denotes the multivarate normal distribution with covariance matrix $\mathbf{\Sigma}$. The last equation can be written in matrix form

$$\mathbf{A} = \mathbf{R}\mathbf{M}$$

For example, in the case of an AR(1) process

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,T} \\ \vdots & \ddots & \vdots & \\ a_{N,1} & a_{N,2} & \cdots & a_{N,T} \end{pmatrix} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,T} \\ \vdots & \ddots & \vdots & \\ r_{N,1} & r_{N,2} & \cdots & r_{N,T} \end{pmatrix} \begin{pmatrix} 1 & -\phi & & & \\ & 1 & -\phi & & \\ & & \ddots & \ddots & \\ & & & 1 & -\phi \\ & & & & 1 \end{pmatrix}$$

Let $\mathbf{Q}\mathbf{R} = \mathbf{R}\mathbf{M}$, then $\mathbf{Q} = \mathbf{R}\mathbf{M}\mathbf{R}^{-1}$. Since the set of $\{r_{ij}\}$ for which $\det\mathbf{R} = 0$ has probability zero, a matrix $\mathbf{Q}$ satisfying the above equation almost surely

exists. Thus $\mathbf{A} = \mathbf{RM} = \mathbf{QR}$ and $\mathbf{AA'} = \mathbf{QRR'Q'}$ Since the columns of A are not auto-correlated, $\mathbf{AA'} \sim W(\mathbf{\Sigma}, T)$. Then $\mathbf{RR'} \sim W(\mathbf{Q^{-1}\Sigma Q'^{-1}}, T)$ follows from [16], section 7.3.3. Now we observe the Wishart PDF

$$f_W(\mathbf{X}) = \frac{|\det X|^{(T-N-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\,\mathbf{\Sigma^{-1}X}\right)}{2^{NT/2}\pi^{N(N-1)/4}|\det \mathbf{\Sigma}|^{T/2}\prod_{i=1}^{N}\Gamma\left[(n+1-i)/2\right]} \tag{C.1}$$

where, assuming $\mathbf{X} = \mathbf{UU'}$, $\mathbf{\Sigma}$ is the covariance matrix of the columns of $\mathbf{U}$. we notice in C.1 that $\mathbf{\Sigma}$ enters $f_W(\cdot)$ through $\mathrm{tr}\,(\mathbf{\Sigma^{-1}RR'})$ and $\det \mathbf{\Sigma}$. $\mathbf{RR'}$ enters through $\mathrm{tr}\,\mathbf{\Sigma^{-1}RR'}$ and $\det \mathbf{RR'}$. Clearly

$$\begin{aligned} \det \mathbf{Q^{-1}\Sigma Q'^{-1}} &= \det \frac{\mathbf{\Sigma}}{(\det \mathbf{Q})^{\mathbf{2}}} \\ &= \frac{\det \mathbf{\Sigma}}{(\det \mathbf{M})^2} \\ &= \det \mathbf{\Sigma} \end{aligned}$$

As to $\mathrm{tr}\,\mathbf{\Sigma^{-1}C}$, we have

$$\begin{aligned} & \mathrm{tr}\,\left[(\mathbf{Q^{-1}\Sigma Q'^{-1}})^{\mathbf{-1}}\mathbf{RR'}\right] \\ = \;& \mathrm{tr}\,\left[\mathbf{Q'\Sigma^{-1}QRR'}\right] \\ = \;& \mathrm{tr}\,\left[\mathbf{R'^{-1}M'R'\Sigma^{-1}RMR^{-1}RR'}\right] \\ = \;& \mathrm{tr}\,\left[\mathbf{\Sigma^{-1}RM(RM)'}\right] \\ = \;& \mathrm{tr}\,\left[\mathbf{\Sigma^{-1}AA'}\right] \end{aligned}$$

Moreover, $\det \mathbf{RR'} = \det \mathbf{AM^{-1}M'^{-1}A'} = \det \mathbf{AA'}$. Thus if we use $f_R(\cdot)$ to denote the joint probability density of the entries of $\mathbf{RR'}$ and $f_A(\cdot)$ to denote that of $\mathbf{AA'}$, we can write

$$f_R(\mathbf{RR'}) = f_A(\mathbf{RMM'R'}) \tag{C.2}$$

Substituting for f_A the Wishart PDF C.1, we get the PDF of $\mathbf{RR'}$.

Appendix D

Asymptotic Distributions of the Elements of a Covariance Matrix of Autocorrelated Gaussian Returns

In this appendix we derive the approximate distributions of the elements of a Covariance Matrix of autocorrelated Gaussian returns.

In the following we denote the diagonal elements as C_{ii} and the non-diagonal elements as C_{ij} $(i \neq j)$. We assume that $\mathrm{var}(a_{i,t}) = \sigma^2$ for any i and t, and $a_{i,t}$ are not autocorrelated, i.e. $\mathrm{corr}(a_{i,t}, a_{i,t'}) = 0$ for any i, t and t'. Then we can write

$$\begin{aligned} r_{i,t}r_{j,t} &= [\phi r_{i,t-1} + a_{i,t}]\,[\phi r_{j,t-1} + a_{j,t}] \\ C_{ij} &= \frac{1}{T}\sum_{t=1}^{T} r_{i,t}r_{j,t} \\ &= \phi^2 \frac{1}{T}\sum_{t=1}^{T} r_{i,t-1}r_{j,t-1} + \phi\frac{1}{T}\sum_{t=1}^{T}(r_{i,t-1}a_{j,t} + r_{j,t-1}a_{i,t}) + \frac{1}{T}\sum_{t=1}^{T} a_{i,t}a_{j,t} \end{aligned}$$

We note that the two sums $\sum_{t=1}^{T} r_{i,t}r_{j,t}$ and $\sum_{t=1}^{T} r_{i,t-1}r_{j,t-1}$ only differ by the first and the last addend, which is negligible for sufficiently large T. Thus we have

$$(1-\phi^2)C_{ij} \approx \phi\frac{1}{T}\sum_{t=1}^{T}(r_{i,t-1}a_{j,t} + r_{j,t-1}a_{i,t}) + \frac{1}{T}\sum_{t=1}^{T} a_{i,t}a_{j,t} \qquad \text{(D.1)}$$

Now we write the AR(1) process $r_{i,t}$ as an infinite moving-average process:

$$\begin{aligned} r_{i,t} &= \phi r_{i,t-1} + a_{i,t} \\ (1-\phi B) r_{i,t} &= a_{i,t} \end{aligned}$$

where B is the back-shift operator. Then it follows from the above equation

$$\begin{aligned} r_{i,t} &= \frac{1}{1-\phi B} a_{i,t} \\ &= \sum_{k=0}^{\infty} \phi^k B^k a_{i,t} \\ &= \sum_{k=0}^{\infty} \phi^k a_{i,t-k} \end{aligned}$$

where it is left implicit that $a_{i,t}$ with $t \leq 0$ is zero (in words, this implies that the $r_{i,t}$ process is not affected at all by events before $t = 1$).

Substituting this into eq.D.1 for $r_{i,t-1}$ yields

$$\begin{aligned} (1-\phi^2) C_{ij} &\approx \frac{1}{T} \sum_{t=1}^{T} \sum_{k=0}^{\infty} \phi^{k+1} (a_{i,t-k-1} a_{j,t} + a_{j,t-k-1} a_{i,t}) + \frac{1}{T} \sum_{t=1}^{T} a_{i,t} a_{j,t} \\ &= \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{t-1} \phi^{k} (a_{i,t-k} a_{j,t} + a_{j,t-k} a_{i,t}) + \frac{1}{T} \sum_{t=1}^{T} a_{i,t} a_{j,t} \end{aligned} \tag{D.2}$$

Given two Gaussian random variables x and y with zero mean and covariance matrix

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

The PDF of xy can be found by considering $P(xy < z)$:

$$\begin{aligned} P(xy < z) &= \left(\int_0^{\infty} dx \int_{-\infty}^{z/x} dy + \int_{-\infty}^{0} dx \int_{z/x}^{\infty} dy \right) \frac{1}{2\pi\sigma\sqrt{1-\rho^2}} \\ &\quad \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2\sigma^2(1-\rho^2)} \right] \\ f(z;\sigma,\rho) &= \frac{d}{dz} P(xy < z) \\ &= \frac{1}{\pi\sigma\sqrt{1-\rho^2}} \exp\left[\frac{\rho z}{\sigma^2(1-\rho^2)} \right] K_0 \left[\frac{|z|}{\sigma^2(1-\rho^2)} \right] \end{aligned}$$

where $K_n(z')$ is the modified Bessel function of the second kind. It is worth taking note that, when $\rho \neq 0$, $f(z;\sigma,\rho)$ is not symmetric with respect to z. As a result, if $\rho > 0$, the mean of $f(z;\sigma,\rho)$ is positive, and vice versa.

Secondly, because $|\rho| < 1$ and

$$K_0(z) \sim \sqrt{\frac{\pi}{2z}} e^{-z}$$

as $z \to \infty$ [24], all the moments of $f(z;\sigma,\rho)$ are finite, implying the applicability of the Lyapunov central limit theorem provided that T is large, which is very often the case and what we assume here.

With this in mind, we observe that ϕ only affects the first sum in D.2. If $a_{i,t-k}$ and $a_{j,t}$ are not correlated for non-zero k, ϕ will not affect the mean of C_{ij}. Furthermore, it is also clear from equation D.2 that the variance of C_{ij} is always increased by a non-zero ϕ, regardless of the sign of ϕ. We compute this increment in the following.

In light of the above expression for $f(z;\sigma,\rho)$, we rewrite equation D.2 as

$$\begin{aligned}(1-\phi^2)C_{ij} \approx & \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{t-1}\phi^k(a_{i,t-k}a_{j,t} + a_{j,t-k}a_{i,t}) \\ & + \frac{1}{T}\sum_{t=1}^{T}\sigma^2(1-\rho^2)\frac{a_{i,t}}{\sigma\sqrt{1-\rho^2}}\frac{a_{j,t}}{\sigma\sqrt{1-\rho^2}}\end{aligned} \tag{D.3}$$

In addition, we assume

$$\text{corr}(a_{i,t-k}, a_{j,t}) = \begin{cases} 1 & \text{if } i = j \text{ and } k = 0 \\ \rho & -1 < \rho < 1. \text{ if } i \neq j \text{ and } k = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, because $a_{i,t-k}$ and $a_{j,t}$ with $i \neq j$ and $k > 0$ are not correlated, the mean of $a_{i,t-k}a_{j,t}$ is 0 ($f(z;\sigma,0)$ is symmetric), and the variance of it can be found to be σ^6 using formula (10.43.19) of [25]:

$$\int_0^\infty dt K_\nu(t) t^{\mu-1} = 2^{\mu-2}\Gamma\left(\frac{\mu+\nu}{2}\right)\Gamma\left(\frac{\mu-\nu}{2}\right)$$

On the other hand, $a_{i,t}/\sigma\sqrt{1-\rho^2}$ and $a_{j,t}/\sigma\sqrt{1-\rho^2}$ have variance $1/(1-\rho^2)$ and are correlated - $\text{corr}(a_{i,t}, a_{j,t}) = \rho$. The mean of $a_{i,t}a_{j,t}/\sigma^2(1-\rho^2)$ can be found using formula (10.43.22) of [25], given that $-1 < \rho < 1$:

$$\begin{aligned}\int_0^\infty t^{\mu-1}e^{-at}K_\nu(t)dt &= (\pi/2)^{1/2}\Gamma(\mu+\nu)\Gamma(\mu-\nu)(1-a^2)^{-\mu/2+1/4} \times \\ & \qquad P^{-\mu+1/2}_{\nu-1/2}(a)\end{aligned}$$

where $P^\mu_\nu(\cdot)$ is Ferrers function of the first kind[1]. The result is

$$\mathbb{E}\left[\frac{a_{i,t}a_{j,t}}{\sigma^2(1-\rho^2)}\right] = \frac{1}{\sqrt{2\pi}(1-\rho^2)^{5/4}}\left[P^{-3/2}_{-1/2}(-\rho) - P^{-3/2}_{-1/2}(\rho)\right]$$

[1] Ferrers function of the first kind is defined through the hypergeometric functon $F(a,b;c;z)$ [25]:

$$\mathsf{P}^\mu_\nu(x) = \left(\frac{1+x}{1-x}\right)^{\mu/2}\mathbf{F}\left(\nu+1,-\nu;1-\mu;\tfrac{1}{2}-\tfrac{1}{2}x\right). \tag{D.4}$$

Similarly, the variance of $a_{i,t}a_{j,t}/\sigma^2(1-\rho^2)$ is found to be

$$\begin{aligned} v^2(\rho) &= \frac{4}{\sqrt{2\pi}(1-\rho^2)^{7/4}}\left[P^{-5/2}_{-1/2}(\rho)+P^{-5/2}_{-1/2}(-\rho)\right] \\ &\quad -\mathbb{E}^2\left[\frac{a_{i,t}a_{j,t}}{\sigma^2(1-\rho^2)}\right] \end{aligned}$$

Now we can apply the Lyapunov central limit theorem [26] to the sum in equation D.3 and write down the asymptotic Gaussian distribution of C_{ij}:

$$C_{ij} \sim N(\mu'_X, \sigma'_X)$$

where

$$\mu'_X = \frac{\sigma^2}{\sqrt{2\pi}(1-\phi^2)(1-\rho^2)^{1/4}}\left[P^{-3/2}_{-1/2}(-\rho)-P^{-3/2}_{-1/2}(\rho)\right] \tag{D.5}$$

$$\begin{aligned} \sigma'^2_X &= \frac{1}{(1-\phi^2)^2}\left[\sum_{t=1}^{T}\sum_{k=1}^{t-1}2\left(\frac{\phi^k}{T}\right)^2\sigma^6+\sum_{t=1}^{T}\frac{\sigma^4(1-\rho^2)^2}{T^2}v^2(\rho)\right] \\ &= \frac{2\sigma^6}{T(1-\phi^2)^2}\left[\frac{\phi^2}{1-\phi^2}-\frac{\phi^2(1-\phi^{2T})}{T(1-\phi^2)}\right]+\frac{\sigma^4(1-\rho^2)^2v^2(\rho)}{T(1-\phi^2)^2} \\ &\approx \frac{2\sigma^6\phi^2}{T(1-\phi^2)^3}+\frac{\sigma^4(1-\rho^2)^2v^2(\rho)}{T(1-\phi^2)^2} \end{aligned} \tag{D.6}$$

We may apply a similar treatment to the diagonal elements of the covariance matrix, which we denote as C_{ii} here:

$$\begin{aligned} C_{ii} &= \frac{1}{T}\sum_{t=1}^{T}r_{it}^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{l=0}^{t-1}\phi^l a_{i,t-l}\sum_{k=0}^{t-1}\phi^k a_{i,t-k} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{k=0}^{t-1}\phi^{2k}a^2_{i,t-k}+\sum_{k,l=0}^{t-1}\phi^{k+l}a_{i,t-k}a_{i,t-l}\right] \end{aligned}$$

By the Lyapunov central limit theorem under the assumption of large T, the asymptotic distribution of C_{ii} is Gaussian, the mean and variance being

$$\begin{aligned} \mathbb{E}(C_{ii}) &= \frac{1}{T}\left[\sum_{k=0}^{t-1}\phi^{2k}\sigma^2\right] \\ &= \frac{\sigma^2}{(1-\phi^2)T}\left[T-\frac{\phi^2(1-\phi^{2T})}{1-\phi^2}\right] \\ &\approx \frac{\sigma^2}{1-\phi^2}\left[1-\frac{\phi^2}{T}\right] \end{aligned} \tag{D.7}$$

and

$$\begin{aligned}
\mathrm{var}(C_{ii}) &= \sum_{t=1}^{T}\left[\sum_{k=0}^{t-1}\frac{\phi^{4k}\sigma^4}{T^2}2+\sum_{k,l=0}^{t-1}\frac{\phi^{2(k+l)}}{T^2}\sigma^6\right] \\
&= \sum_{t=1}^{T}\left[\frac{2\sigma^4}{T^2}\frac{1-\phi^{4t}}{1-\phi^4}+\frac{\sigma^6}{T^2}\left(\frac{1-\phi^{2t}}{1-\phi^2}\right)^2\right] \\
&= \frac{2\sigma^4}{T(1-\phi^4)}-\frac{2\sigma^4\phi^4(1-\phi^{4T})}{T^2(1-\phi^4)^2}+ \\
&\quad \frac{\sigma^6}{T(1-\phi^2)^2}-\frac{2\sigma^6\phi^2(1-\phi^{2T})}{T^2(1-\phi^2)^3}+\frac{\sigma^6\phi^4(1-\phi^{4T})}{T^2(1-\phi^2)^2(1-\phi^4)} \\
&\approx \frac{2\sigma^4}{T(1-\phi^4)}+\frac{\sigma^6}{T(1-\phi^2)^2} \qquad \text{(D.8)}
\end{aligned}$$

Appendix E

Eigenvalues Distribution of Wishart Matrix

In this section we summarize the analytic results regarding the eigenvalue distribution of a Wishart matrix $\mathbf{R}\mathbf{R}'$. In the simplest case where the elements of $\mathbf{R}$ are all independent of each other, i.e. neither auto-correlation nor cross-correlation exists, we have [14]

$$f(x_1, \cdots, x_N) = K \prod_{i=1}^{N} e^{-x_i/2} x_i^{(T-N-1)/2} \prod_{i<j}^{N} (x_i - x_j) \tag{E.1}$$

where the eigenvalues have been indexed in descending order: $x_1 \geq x_2 \geq \cdots \geq x_N$. The normalization constant K is given by

$$K = \frac{\pi^{N^2/2}}{2^{NT/2} \Gamma_N(T/2) \Gamma_N(N/2)}$$

where the function $\Gamma_m(a)$ is defined as

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{k=1}^{m} \Gamma\left(a - \frac{k-1}{2}\right)$$

If the order of the eigenvalues is ignored, the PDF of their distribution is given by the Marcenko-Pastur law [18]

$$f(x) = \frac{1}{2\pi\sigma^2 q} \frac{\sqrt{(x_2 - x)(x - x_1)}}{x} \tag{E.2}$$

where it is assumed $r_{it} \sim N(0, \sigma^2)$ and

$$\begin{aligned} q &= \lim_{N,T\to\infty} \frac{N}{T} \\ x_1 &= \sigma^2(1-\sqrt{q})^2 \\ x_2 &= \sigma^2(1+\sqrt{q})^2 \end{aligned}$$

It is imposed that $q = N/T < 1$.

Moreover, K. Johnsson [27] and I. Johnstone [28] showed that, at the *absence* of autocorrelations and in the asymptotic limit $N, T \to \infty$, $N/T \to q < \infty$, the maximum eigenvalue λ_1 follows the Tracy-Widom distribution (denote $\mathscr{TW}_1$ here) when properly relocated and rescaled:

$$\frac{\lambda_1 - \mu_{NT}}{\sigma_{NT}} \sim \mathscr{TW}_1$$

where μ_{NT} and σ_{NT} are given by

$$\begin{aligned} \mu_{NT} &= \left(\sqrt{N-1/2} + \sqrt{T-1/2}\right)^2 \\ \sigma_{NT} &= \sqrt{\mu_{NT}} \left(\frac{1}{\sqrt{N-1/2}} + \frac{1}{\sqrt{T-1/2}}\right)^{1/3} \end{aligned}$$

The $\mathscr{TW}_1$ distribution has the following cummulative distribution function (CDF) [14]

$$F_1(x) = \exp\left[-\frac{1}{2}\int_x^\infty dy \left(q(y) + (y-x)q^2(y)\right)\right]$$

where $q(y)$ is defined as the solution to the Painlevé II differential equation

$$q''(y) = yq(y) + 2q^3(y)$$

which is unique when imposing the condition

$$q(y) \sim \mathrm{Ai}(y) \text{ as } y \to \infty$$

Then Marco Chiani showed recently that the $\mathscr{TW}_1$ distribution can be well approximated by a gamma distribution based on his proof that the exact distribution of the maximum eigenvalue is a mixture of gamma distributions. Specifically,

$$\frac{\lambda_1 - \mu_{NT}}{\sigma_{NT}} + \alpha \sim \mathscr{G}(k, \theta) \tag{E.3}$$

where $\mathscr{G}(k, \theta)$ denotes the Gamma distribution with parameters k and θ [14]. The PDF of the gamma distribution is given by

$$f_\gamma(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}$$

Moreover, the first 3 moments of the distribution are simple:

$$\begin{aligned} \text{mean} &= k\theta \\ \text{variance} &= k\theta^2 \\ \text{skewness} &= 2/\sqrt{k} \end{aligned}$$

Bibliography

[1] M. Politi, E. Scalas, and G. Germano. Spectral densities of wishart-lévy free stable random matrices. *Eur. Phys. J. B*, 73:13–22, 2010.

[2] Thomas Mikosch and Catalin Starica. Limit theory for the sample autocorrelations and extremes of a garch (1, 1) process. *Annals of Statistics*, pages 1427–1451, 2000.

[3] Bernt Øksendal. *Stochastic Differential Equations*. Springer, 5 edition, 2000.

[4] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307327, 1986.

[5] Tim Bollerslev. On the correlation structure for the generalized autoregressive conditional heteroskedastic processes. *Journal of Time Series Analysis*, 9(2):121131, 1987.

[6] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Economitrica*, 71:529–626, 2003.

[7] Philip E. Protter. *Stochastic integration and differential equations*. Springer, 2 edition, 2005.

[8] Yacine Ait-Sahalia, Per A. Mykland, and Lan Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18(2), 2005.

[9] Torben G. Andersen, Richard A. Davis, Jens-Peter Kreiß, and Thomas Mikosch. *Handbook of Financial Time Series*. Springer, 2009.

[10] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, 3 edition, 1994.

[11] J. Shang and P. Tadikamalla. Modelling financial series distributions: A versatile data fitting approach. *International Journal of Theoretical & Applied Finance*, 7:231–251, 2004.

[12] N. L. Johnson. Tables to facilitate fitting su frequency curves. *Biometrika*, 52(3/4):547–558, 1965.

[13] Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press, 2nd edition, 2003.

[14] M. Chiani. Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution. *ArXiv e-prints*, September 2012.

[15] Steve Lalley. Gaussian and wishart ensembles: Eigenvalue densities, 2013.

[16] T. W. Anderson. *An Introduction to multivariate statistical analysis*. John Wiley & Sons. Inc., 3 edition, 2003.

[17] R. N. Mantegna and H. E. Stanley. *An introduction to Econophysics*. Cambridge University Press, Cambridge, 2000.

[18] Thomas Guhr. *Econophysics*. Lunds University, Lund, Sweden, 2007.

[19] P. Cizeau and Jean-Philippe Bouchaud. Theory of lévy matrices. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 50(3):1810–1822, October 1994.

[20] Alex Bilik. Heavy-tail central limit theorem. Lecture notes from University of California, San Diego, 2008.

[21] P. Embrechts, C. Klüppelberg, and T.Mikosch. *Modelling Extreme Events*. Springer-Verlag, 1997.

[22] Sven Åberg. Spectral density of autocorrelated wishartlévy matrices. *Journal of Physics A: Mathematical and Theoretical*, 46(34), 2013.

[23] Hiroki Hashiguchi, Yasuhide Numata, Nobuki Takayama, and Akimichi Takemura. Holonomic gradient method for the distribution function of the largest root of a wishart matrix. *Journal of Multivariate Analysis*, 117, January 2012.

[24] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010. Print companion to [25].

[25] NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.6 of 2013-05-06. Online companion to [24].

[26] Patrick Billingsley. *Probability and Measure*. John Wiley & sons, 3 edition, 1995.

[27] K. Johnsson. Shape fluctuations and random matrices. *Communications in Mathematical Physics*, 209:437–476, 2000.

[28] I. Johnstone. On the distribution of the largest eigenvalue in principle component analysis. *The annals of statistics*, 29(2):259–327, 2001.