



Norwegian University of
Science and Technology

Modelling Correlated Financial Time Series

Agnar Ask Ellingsen

Master of Science in Physics and Mathematics

Submission date: June 2011

Supervisor: Håvard Rue, MATH

Problem Description

The aim of this thesis is to study financial time series and how their correlations change over time and during crisis situations, and to come up with a good statistical model to describe these. The model should also be able to give realistic realisations of the studied financial time series.

Assignment given: 25. January 2011

Supervisor: Håvard Rue, MATH

Preface

This thesis was written at the Department of Mathematical Sciences, Faculty of Information Technology, Mathematics and Electrical Engineering at the Norwegian University of Science and Technology. The course is called *TMA4905 Statistics, Master Thesis* and gives 30 out of a total of 30 credits in a semester, and was written over a period of 20 weeks. All the coding in this thesis is done in the open source statistical software R.

I would like to give thanks to my supervisor Håvard Rue, a professor at the Department of Mathematical Sciences, for help and instructions, and Egil Husby and Tormod Teig at Western Bulk for their help and expertise.

Trondheim June 17, 2011
Agnar Ask Ellingsen

Abstract

Western Bulk is a shipping company that is interested in generating realistic realisations of five different price series; two freight rates, one interest rate and two oil rates. The realisations should keep the internal correlation-structure. There should also be possible to introduce a crisis-regime where all the prices are correlated with a downward trend similar to what was observed in 2008.

The prices series modelling is approached through a state-space method. The volatility is modelled by a multivariate stochastic volatility model on the compounded returns. The multivariate model decomposes the five dimensional correlation matrix into fifteen univariate models; five for the volatilities and ten for the correlation between each of the series. Only the two freight rates and the two oil rates are correlated, but the correlation-structure is not strong enough to create series in line with what is historically observed. A stochastic trend is therefore also introduced. The stochastic trend is a local linear trend model, where the underlying state-space variable follows an AR(1) model.

To get good simulations on a long time span both simulating directly from the model of the underlying state-space time series as well as bootstrapping them are compared. For the stochastic volatility models with a trend there is little difference between bootstrapping and simulating, except for the interest rate which tends to diverge when simulated. For the stochastic trend bootstrapping gives more stable results, most notably on realisations without crisis-regimes.

Contents

| | | |
|-----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | The data set | 3 |
| 3 | Autoregressive models | 7 |
| 3.1 | Univariate AR models | 7 |
| 3.2 | Vector Autoregressive models | 7 |
| 3.3 | Fitting our data to a VAR model | 8 |
| 4 | Bayesian Inference | 11 |
| 5 | Markov Chain Monte Carlo | 14 |
| 5.1 | Markov Chains | 14 |
| 5.2 | The Metropolis-Hastings algorithm and the Gibbs sampler | 15 |
| 6 | Bootstrap | 20 |
| 6.1 | Introduction to bootstrapping | 20 |
| 6.2 | Bootstrapping time series | 20 |
| 7 | Stochastic volatility model | 24 |
| 7.1 | Deriving the standard model | 24 |
| 7.2 | Simulating from the standard model | 27 |
| 7.3 | Extending the model to encompass high volatility regimes | 31 |
| 7.4 | Testing the model on generated data | 34 |
| 8 | Stochastic trend model | 38 |
| 8.1 | Deriving the standard model | 38 |
| 8.2 | Simulating from the standard model | 38 |
| 8.3 | Extending the model to encompass regimes with drop in prices | 39 |
| 8.4 | Testing the model on generated data | 41 |
| 9 | Implementating the method on our data | 44 |
| 9.1 | Implementing the stochastic volatility model on our data | 44 |
| 9.2 | Implementing the stochastic trend model on our data | 49 |
| 10 | Bootstrapping the series and simulating from the models | 53 |
| 10.1 | How to bootstrap and simulate the volatility | 53 |
| 10.2 | Comparing with the stochastic volatility model | 53 |
| 10.3 | How to bootstrap and simulate the trend | 61 |
| 10.4 | Comparing with the trend model | 63 |

| | |
|---|-----------|
| 10.5 Combining the models | 67 |
| 11 Reviewing the model and the results | 70 |
| 11.1 About the model | 70 |
| 11.2 Improvements and further work | 71 |
| Bibliography | 73 |
| A Probability Distributions | 75 |
| B Stochastic volatility plots | 76 |
| B.1 Generated data | 76 |
| B.2 Real data | 79 |

1 Introduction

Western Bulk is a leading dry bulk shipping company with offices in key spots all over the world. As a company they are risk aware and take a responsive approach to changes in the market. Western Bulk has Value-at-Risk (VaR) as its most important risk measure, but also apply stress tests and Cash-flow-at-Risk (CfaR) measures. Western Bulk has good models to calculate VaR, but wants to improve its methodology for long term CfaR. To do this it needs a model able to simulate correlated marked rates. More specifically it needs a model to simulate correlated freight rates, oil rates and interest rates.

The prediction of future values of price series based on past observed prices is not possible by basic market hypothesis (see for example Wilmott, Howison and Dewynne (1995)). However, it should not be impossible to generate realisations of such price series, that share some of the same characteristics as past observed prices. Given a portfolio of several asset prices there are several ways to find if and how price movements in one asset correlates with the others, see for example McAleer, Asai and Yu (2004) for a review of some of the most popular models. Western Bulk wants a model that keep the internal correlation-structure of the historical prices, as well as having the possibility to introduce a crisis-regime, where all prices plummet.

One of the simplest ways to describe the data is to fit them to some sort of multivariate time series (e.g. a vector auto-regressive (VAR) model). A more complicated but flexible way to view time series than the ordinary Box-Jenkins approach, is the state-space approach (see Durbin and Koopman (2001) for a thorough review of the subject). The main idea of the state-space method is dividing the series in different parts; a trend term and a variance (or volatility) term and choose the model most suited for the problem at hand to examine the different parts.

A popular approach to describe time-varying volatility is the GARCH models (or M-GARCH in the multivariate case), but several papers indicate that stochastic volatility models outperforms GARCH models (see for example Kim, Shephard and Chib (1998)), so we will focus on the multivariate stochastic volatility model as suggested by Plataniotis and Dellaportas (2010). A stochastic volatility model is a state-space model, where we cannot observe the volatility directly. Since we cannot observe it directly, we need to draw inference about it through its probability distribution. We will also be working in a Bayesian setting where the parameters have their own probability distributions, and this will influence the probability distribution of the volatility states. If we call all the parameters for ρ , the unobservable volatility states X and our observations for Y , the probability distribution we draw inference about is $P(\rho, X|Y)$. This is called the posterior distribution, and can be very complicated to draw inference from directly. We

will therefore use a procedure called Markov Chain Monte Carlo to simulate it. Markov Chain Monte Carlo is procedure where you design a Markov Chain with the wanted distribution as its limiting distribution and iterate through the chain until you reach (arbitrarily close to) it.

We model the trend with a slightly changed local linear trend model as described in Durbin and Koopman (2001). This is also a state-space model, so we need to use Markov Chain Monte Carlo in a similar way as with the stochastic volatility. But since Western Bulk wants the model to be able to replicate the effects the financial crisis in the late 2000s had on the different price series, we have extended the models to encompass different regimes. The volatility model will have an increase in volatility during a crisis-regime, and the trend model will have a downwards slope during a crisis-regime. As Western Bulk is interested in series over 2-3 years, but simulating directly from a model over such a long time can lead to high variance, we also want to compare simulating from the model directly with bootstrapping from the state-spaces.

The thesis is outlined as follows: section 2 will introduce the data sets, and discuss some of the properties of the series. Section 3 will quickly introduce the autoregressive models, and try to fit the data to a VAR model. Section 4, 5 and 6 will introduce the theory needed for defining and simulating from our volatility and trend model; namely Bayesian statistics, Markov Chain Monte Carlo and Bootstrap. Section 7 and 8 introduce the multivariate stochastic volatility model and the trend model. And finally in section 9 and 10 we fit our data to the models and simulate from them, comparing bootstrap with simulating from the model directly on large time scales (3 years \approx 700 trading days). We finish with some concluding remarks and suggestions for further work in section 11.

2 The data set

The data sets we use in this thesis are five different daily observed price series. Two freight rates, TC Average BSI and TC Average BPI, where BSI denotes the supramax index price and BPI the panamax index price. The supramax bulk carriers are the largest ships in the handysize class with cargo capacity of 50 000 to 60 000 metric tonnes, often with self-loading capabilities. The panamax class ships are the largest ships that fulfil the size requirements to be able to pass through the Panama canal. The third observed price series we have is the three months USD LIBOR rates. LIBOR is an acronym for London InterBank Offered Rate, and denotes the interest rates at which banks borrow unsecured funds from other banks in the London market. The two last price series are oil prices. One denotes the bunker fuel the ships runs on (Singapore 380), and the other is just the price of crude oil (ICE Brent).

The data are given from the 4th January 2005 until the 25th January 2011 for all series except the crude oil which is given from the 28th April 2006, and there is one missing observation for the bunker fuel data on the 1st of August 2005. The series can be seen plotted against the series in the same segment (freight rates and oil rates together) in Figure 2.1. It clearly is a close connection between the two freight prices, but there is also one between the oil prices. In Figure 2.2 we have scaled down both oil prices to the same level and plotted them against each other. An observation worth noting is while the prices across the different segments might seem independent, they still share one characteristic; due to the financial crisis in the late 2000s all series had a steep fall in their prices.

With financial time series it can often be rewarding to look at the compounded return series. The compounded return is defined as $r_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$, where S_t is the observed price at time t . Where time series of prices often have a trend, the return series do not, and they often capture how the volatility of the series behave. More information on this can be found in the Stochastic Volatility section. The return series for our observed time series can be seen in Figure 2.3. Just as we can observe from the price series directly the BPI is more volatile than the BSI, and the oil prices have about the same magnitude on their volatility. The LIBOR rates are a bit different from the others; they are hardly volatile at all except at certain times when a lot happens to the interest rates. One of the most simple models to describe the volatility of a financial asset is to assume that the returns are normally distributed. We can easily see that the observed returns are not from a stationary distribution; the volatility is larger at the end of 2008 for all of the series, and the LIBOR rates have periods of low and periods of high volatility throughout the series. It can be interesting to find the kurtosis and the skewness for the different series to test if the volatility still can come from a normal distribution in some form, for example through autoregressive time series. One test to see if the data

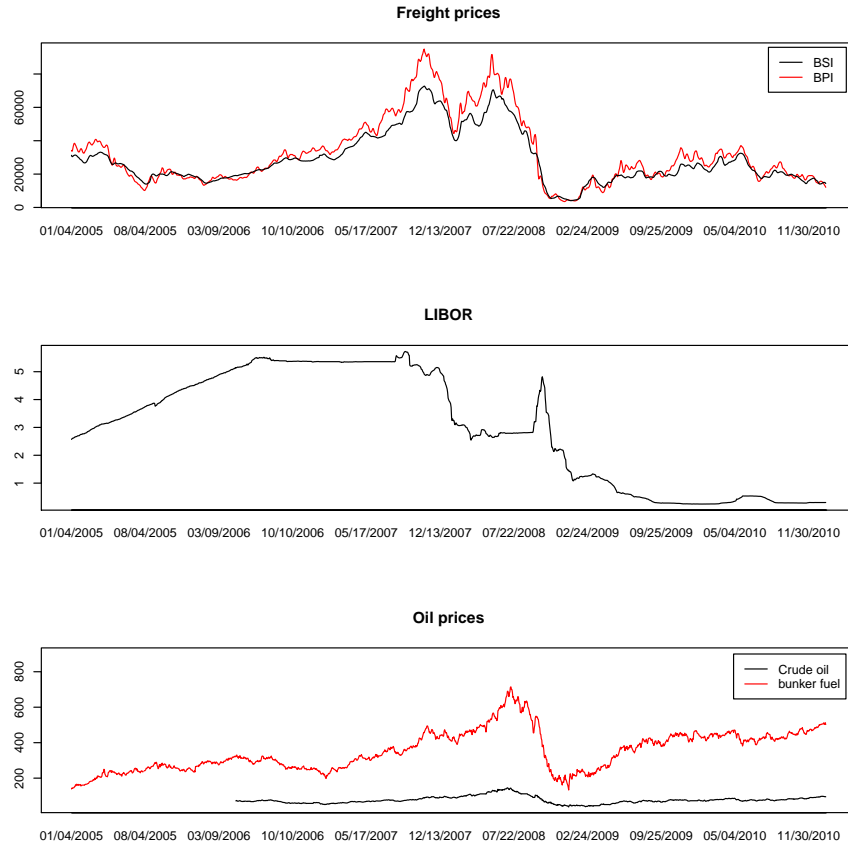


Figure 2.1: The observed price series.

is identically normal distributed, which uses the skewness (an indication of the symmetry of the underlying distribution) and the excess kurtosis (an indication of the heaviness of the tails of the underlying distribution), is the Jarque-Bera test introduced in Jarque and Bera (1987). The test statistic is $JB = \frac{n}{6} \left(S^2 + \frac{1}{4} K^2 \right)$, where S is the sample skewness defined as $\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}$ and K is the sample excess kurtosis defined as $\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3$. Here $\hat{\mu}_i = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^i$, when $x = (x_1, \dots, x_n)$ is the vector of observations. Under the hypothesis of normality both the skewness and excess kurtosis should be zero, and the test statistic JB χ^2 -distributed with 2 degrees of freedom. It should be a less than 0.5% probability that the test statistics are greater than 10.597 under the hypothesis of normality, and we should discard it if we get a value close to this magnitude or larger. In Table 1 we can see the test statistics and both skewness and kurtosis for all of the five series, and it leaves us

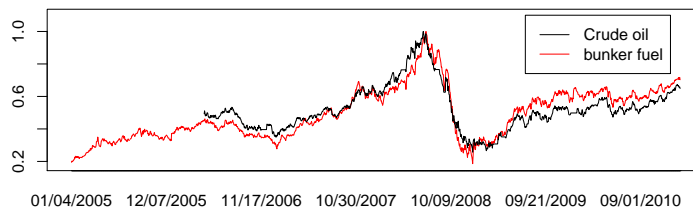


Figure 2.2: Scaled bunker fuel and crude oil prices compared against each other.

| | JB | S | K |
|--------|-------|-------|------|
| BSI | 23947 | .658 | 19.4 |
| BPI | 2335 | -.469 | 6.0 |
| LIBOR | 66905 | 1.85 | 32.4 |
| CRUDE | 2211 | .369 | 6.7 |
| BUNKER | 5875 | .754 | 9.5 |

Table 1: JB -test statistic with skewness and kurtosis for the whole return series.

with little doubt that the data are not normally distributed. It was not surprising that the LIBOR returns were the ones with the largest deviation from the normal-hypothesis, considering they are fairly stationary until there is an interest rate correction due to external effects, which gives a large change in the returns. The other series are clearly not homoscedastic with major spikes in volatility during the financial crisis. If we run the JB test on the data up to where the crisis really kicks in (the 4th of September 2008), we get better results as we can see in Table 2 (if we ignore the LIBOR rates). It should be noted that even though the data itself is not normally distributed, we cannot say they do not come from a series with normally distributed innovations. After a couple of simulated AR(1) models of 1000 observation with $\phi = 0.99$ and $N(0, 1)$ distributed innovations, one of them returned a $JB = 228$ with $S = -0.799$ and $K = 1.7$. We will not assume the returns to be iid normal, but for simplicity we will still be working on models that in some way incorporate the normal distribution, like the autoregressive (AR) model.

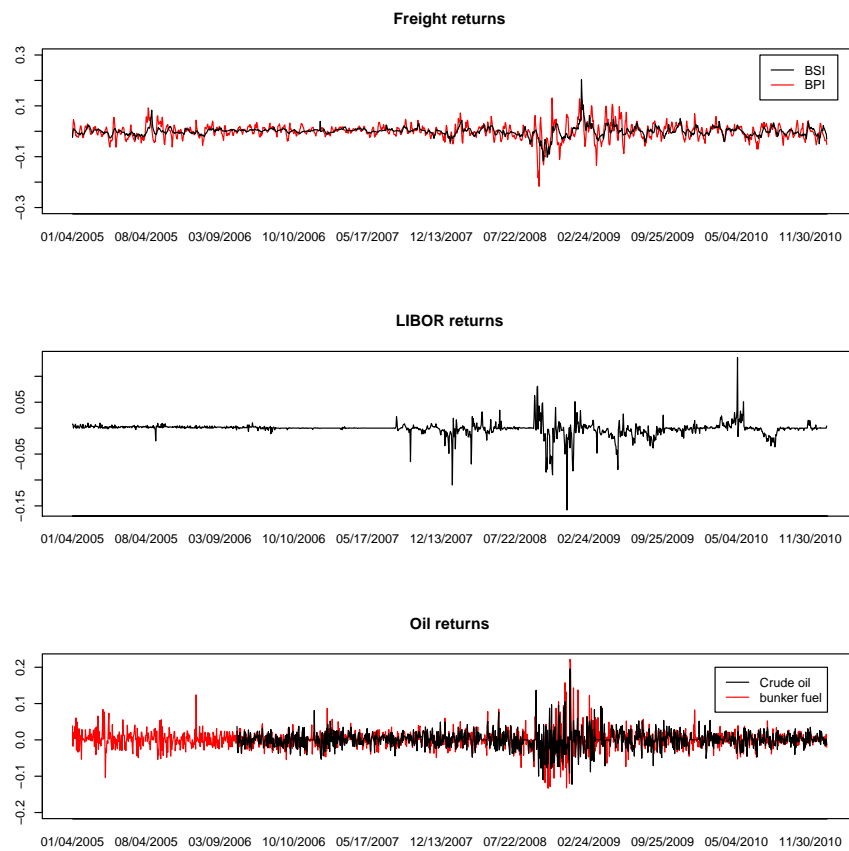


Figure 2.3: The observed return series.

| | JB | S | K |
|--------|--------|-------|------|
| BSI | 1417 | .504 | 6.0 |
| BPI | 52 | .133 | 1.1 |
| LIBOR | 282142 | -6.58 | 84.8 |
| CRUDE | 17 | .160 | 0.8 |
| BUNKER | 285 | .307 | 2.7 |

Table 2: JB -test statistic with skewness and kurtosis for the return series up to the 4th of September 2008.

3 Autoregressive models

Autoregressive (AR) models are popular models to capture the dynamics of several different types of time series. The idea behind these models is that the value at a given time is dependent on past values and an innovation-term which gives the models their stochastic behaviour.

3.1 Univariate AR models

In the univariate case we only consider the time series of a single asset x_t , where $t \in [1, T]$. The p -order autoregressive model, $AR(p)$, gives the k -th value of x_t as a weighted sum of its p earlier values and a random innovation term. More mathematically the model can be described as

$$\dot{x}_k = \dot{x}_{k-1}\phi_1 + \dots + \dot{x}_{k-p}\phi_p + \epsilon_k = \sum_{i=1}^p \phi_i \dot{x}_{k-i} + \epsilon_k; \quad k > p,$$

where ϵ_k is the random innovation term (often from a zero mean normal distribution), ϕ_i the weights and $\dot{x}_k = x_k - \nu$. Here we use ν to denote the mean value of the stationary process. When we use the term stationary process, we mean a second order weakly stationary process, also called covariance stationary. The criteria for such processes are that the mean value is stationary, the variance is stationary and the covariance between two values in time is dependent on the time difference alone. This can be written as

$$\begin{aligned} E[x_t] &= \nu_t = \nu \\ Var[x_t] &= \sigma_t^2 = \sigma^2 \\ Cov[x_{t_1}, x_{t_2}] &= \gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k) = \gamma_{t_2 - t_1}. \end{aligned}$$

Two realisations of a stationary $AR(1)$ model with different values for their autoregressive coefficients ϕ can be seen in Figure 3.1. As we can see, when the value of ϕ is large low values tend to follow low values, and high values tend to follow high ones. This effect is called clustering and will be used later in this thesis.

For a deeper explanation of stationary time series and more on the theory from this section the reader is referred to Wei (2006).

3.2 Vector Autoregressive models

The vector autoregressive (VAR) models describe the dynamics of the time series of several assets. In the univariate case the value of an asset at a given time is only dependent on the earlier values of the same asset, but in the n dimensional

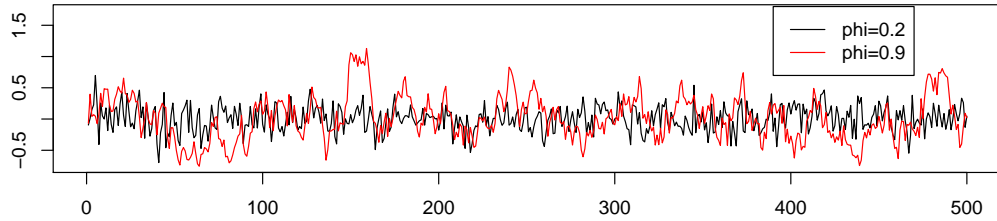


Figure 3.1: Two realisations of AR(1) models with different ϕ , but with the same standard deviation of 0.2.

VAR models a vector at a given time is dependent on the earlier vectors. In other words, the value of one asset in this vector at a given time is not necessarily only dependent of the earlier values of the same asset, but also the earlier values of the other assets in the vector. The p order VAR models can be written as

$$\dot{x}_t = \sum_{i=1}^p \Phi_i \dot{x}_{t-i} + \epsilon_t; \quad t > p,$$

where \dot{x}_t now is vector of assets, ϵ_t a vector of innovations from an n -dimensional multivariate distribution and Φ_i is an $n \times n$ matrix of weights. Again we consider process that are stationary in the same sense as in the univariate case. So the mean vector is stationary and the covariance matrix between two vectors in time is only dependent on the time difference.

3.3 Fitting our data to a VAR model

In R there is a package that can be installed called `dse`, written by Gilbert (2011), which can fit data to a VAR model and simulated realisations from that model. We use the data from the 28th of April 2006 and onwards to get a complete data set, and we let the algorithm decide the autoregressive order. This might lead to over-fitting, but it will give us a good idea on the strengths and weaknesses of this model. We plot the realisations we get in the same way we plotted the original data in Figure 2.1, and a typical realisation can be seen in Figure 3.2. Some of the problems we get if we simulate from this model are captured well in this figure. Prices often tend to get negative, and even though the freight rates followed each other in a way similar to what can be observed, the oil prices did not. There seems to be no connection between the bunker price and the crude oil

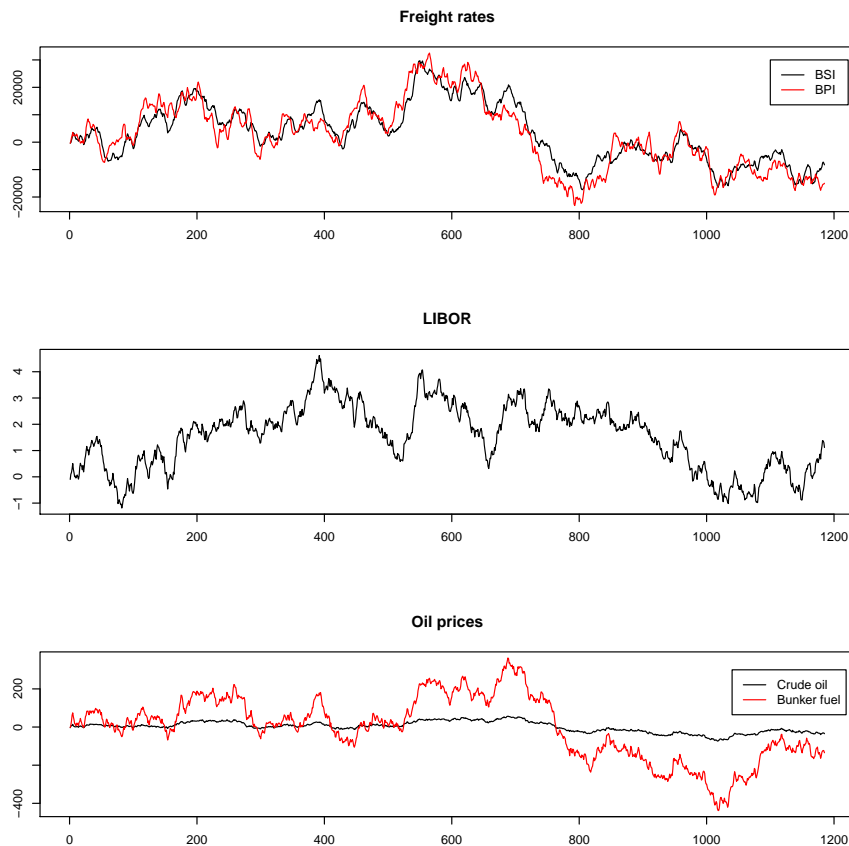


Figure 3.2: One realisation from the fitted VAR model.

price, and the bunker price often tends to be significantly lower than the crude price over long periods, which is not realistic considering that the bunker is just refined crude oil.

One way to overcome those two problems is to model the return series in the VAR framework, and from a given start value S_0 create the series by setting $S_t = S_{t-1} \exp(r_t)$. Now the series cannot become negative (given that $S_0 \geq 0$), and since the oil prices have similar sized return series, the VAR model might be able to capture the relation between bunker rates and crude rates better. It turns out that those two problems are indeed solved when we model the returns rather than the prices directly. But this model is not without its flaws either. An extreme version of these flaws can be seen in Figure 3.3. The prices can tend to become unrealistically large. Since we model on a log scale, high values drawn in sequence will cause the model to return too high prices. The largest observed price for the

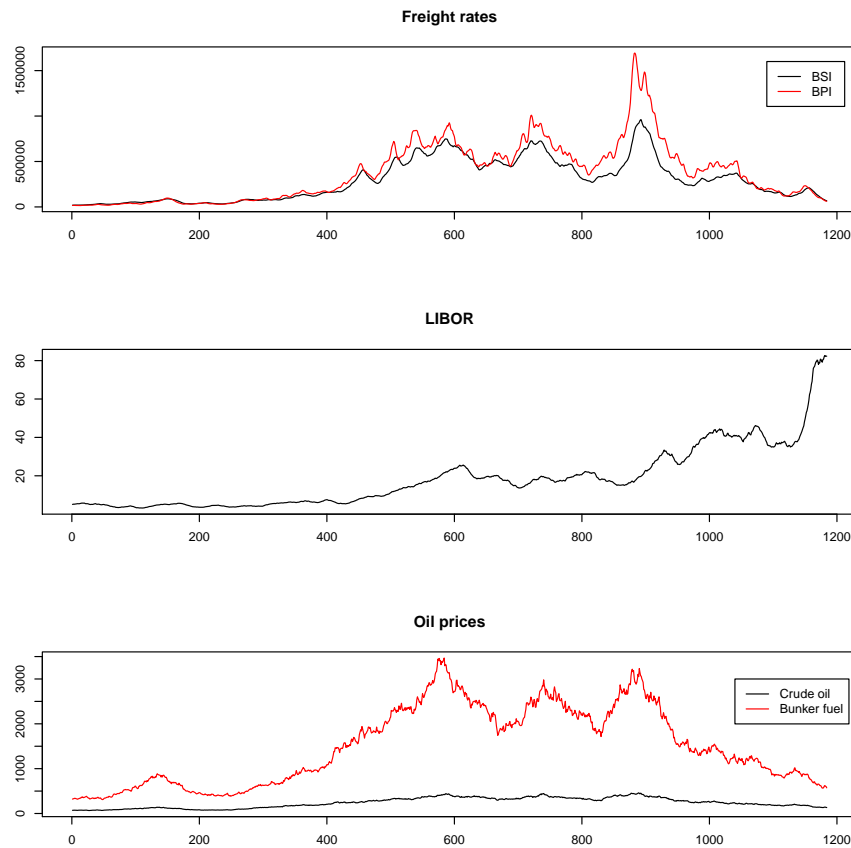


Figure 3.3: One realisation from the fitted VAR model where we first modelled the returns, and worked backwards to find the price series.

BPI in the six year history is less than 100 000, but in the realisation in Figure 3.3 it is larger 1 500 000. Also due to the non-autoregressive characteristic of the LIBOR return series with its sudden large spikes, the volatility is overestimated and we end up with too large values much too often.

Even though some of these problems might be possible to overcome, we also would like to be able to implement an external crisis which would cause all the prices to plummet. We will therefore use a more complicated model than the simple VAR in this thesis. To be able to use this more complicated model we need knowledge about Bayesian statistics and Markov Chain Monte Carlo.

4 Bayesian Inference

The Bayesian approach to statistics is philosophically different from the classical (also called frequentist) one, and there used to be much controversy as to which was right. Nowadays the strengths and weaknesses of both approaches are understood, and practitioners can use the one most suited to their problem. The theory and examples in this section come from Gamerman and Lopes (2006), and a more extensive discussion can be found there.

Both cases were developed in the presence of observations x whose value is initially uncertain, but is described by its probability density function $p(x|\rho)$, where ρ are all the parameters of the probability function of the observations (for example the mean and the variance of a normal distribution). When the values of ρ are of interest, the researcher often already has some information about what these values are. Often it would make sense if the researcher could use this information in the analysis of the problem, and this is where the Bayesian and frequentist approaches differ. For the frequentist the values of ρ are fixed constants and information about them can only be gained through observations. For a Bayesian however the parameters ρ have their own probability distribution. What kind of distributions the parameters have together with their parameters (the parameters of the parameters ρ are called the hyperparameters) are decided by the practitioner based on the information he/she has prior to observing x . The probability distribution of ρ , $p(\rho)$, is therefore called the prior distribution.

To draw inference about the problem and the value of ρ , the practitioner is interested in the probability distribution of ρ given the observed quantities, x . This distribution, $p(\rho|x)$, is called the posterior distribution. This distribution is given by the functions $p(x|\rho)$ (called the likelihood by Bayesians) and $p(\rho)$ through what is called Bayes' Theorem

$$p(\rho|x) = \frac{p(x|\rho)p(\rho)}{p(x)},$$

where $p(x) = \int p(x|\rho)p(\rho)d\rho$. It should be noted that in $p(\rho|x)$ x is just a constant and $p(x)$ is just the normalizing constant, as ρ is the variable.

An example to show the difference between the frequentist and the Bayesian approach is shown below.

Example Given a problem where the observations are (iid) normal distributed, and we want to draw inference about the mean when the variance is known. Or in other words $x = (x_1, \dots, x_n)$ and $f(x_i|\mu) \sim N(\mu, \sigma^2)$ where σ^2 is known.

Frequentist We expect that the frequentist approach is well known to the reader, so we will only sketch it out one of the possible ways to find

a good statistic of μ . The body of literature and methods available to the frequentist are massive and the interested reader can see Casella and Berger (2002) for a thorough explanation. From experience we want to check the distribution of the sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ so we look at its moment generating function $M_{\bar{x}}(t) = [M_x(t/n)]^n = \left[\exp \left\{ \mu \frac{t}{n} + \frac{\sigma^2 (t/n)^2}{2} \right\} \right]^n = \exp \left\{ \mu t + \frac{(\sigma^2/n)t^2}{2} \right\}$ which is the moment generating function of $N(\mu, \sigma^2/n)$. So \bar{x} is a good estimator of μ (actually it is the minimum-variance unbiased estimator), and usually found through maximum likelihood rather than moment generating functions as we did here.

Bayesian From the Bayesian perspective we need to put a prior on the parameter of interest, and let's assume we have some prior knowledge about the problem which gives us $p(\mu) \sim N(\mu_0, \sigma_0^2)$. To overemphasise what is already said; the hyperparameters μ_0 and σ_0 are known constants we can use because we have information about μ prior to observations (μ_0 is what we believe μ to be, and σ_0 is how sure we are where low values means more sure than high ones). The likelihood function (as a function of μ) can be written as

$$\begin{aligned} f(x|\mu) &= \prod_{i=1}^n f(x_i|\mu) \propto \exp \left\{ \frac{-\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right\} \\ &= \exp \left\{ \frac{-(n\mu^2 - 2\mu \sum_{i=1}^n x_i + (\sum_{i=1}^n x_i)^2)}{2\sigma^2} \right\} \\ &\propto \exp \left\{ \frac{-(\mu - \sum_{i=1}^n x_i/n)^2}{2\sigma^2/n} \right\}. \end{aligned}$$

We use a lot *proportional to* because we know that when the prior and likelihood are proper distributions, the posterior will be a proper distribution. And if we recognise the functional form of the posterior as a known distribution, we know what its normalising constant is (for a normal distribution with mean μ and variance σ^2 the normalising constant is $\frac{1}{\sqrt{2\pi\sigma^2}}$), so we do not need to concern ourselves with constants along the way. The posterior distribution is then

$$\begin{aligned} p(\mu|x) &\propto p(x|\mu)p(\mu) \propto \exp \left\{ \frac{-(\mu - \bar{x})^2}{2\sigma^2/n} \right\} \exp \left\{ \frac{-(\mu - \mu_0)^2}{2\sigma_0^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \frac{\mu^2 \sigma_0^2 - 2\mu \bar{x} \sigma_0^2 + \mu^2 \sigma^2/n - 2\mu \mu_0 \sigma^2/n}{\sigma^2 \sigma_0^2/n} \right\} \\ &= \exp \left\{ -\frac{1}{2} \frac{\sigma_0^2 + \sigma^2/n}{\sigma^2 \sigma_0^2/n} \left(\mu^2 - 2\mu \frac{\bar{x} \sigma_0^2 + \mu_0 \sigma^2/n}{\sigma_0^2 + \sigma^2/n} \right) \right\}, \end{aligned}$$

which we can recognise as the normal distribution and after observing closely we can recognise the mean $\hat{\mu}$ and variance $\hat{\sigma}^2$. Here $\hat{\sigma}^2 = (\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2})^{-1}$ and $\hat{\mu} = (\frac{\bar{x}}{\sigma^2/n} + \frac{\mu_0}{\sigma_0^2})\hat{\sigma}^2$. This looks more complicated than the frequentist approach, and it is. But it is more robust against poorly chosen or few observations if one is certain about the prior. If one is very uncertain about the prior (that is σ_0^2 tends to infinity) the posterior distribution becomes $N(\bar{x}, \sigma^2/n)$, that is the Bayesian will use the same estimator as the frequentist with the same level of certainty. One should still note the philosophical difference still exist; for the frequentist μ is an unobservable constant and the expected value of \bar{x} , whereas for the Bayesian μ is a random quantity with \bar{x} as its expectation.

As the example above shows, which prior you choose is very important. The reason the normal distribution was chosen as a prior for μ when the likelihood also had a normal distribution was obviously not a coincidence when we see that the posterior then also got a normal distribution. It is not given that for any prior and likelihood distribution, the posterior will be easy to distinguish. It is not even given that it will be a known distribution. But given some likelihood distributions an experienced practitioner can choose the prior such that the posterior distribution is in the same family of distributions as the prior. Those are called conjugate distributions. There is obviously a dilemma here. Do you choose a conjugate prior to make the computations simpler, or do you chose a different prior, one you think will more clearly express the information you have on the problem. This balance between tractability and realism is a choice the practitioner has to choose, and will depend on the problem at hand.

Later we will use two types of conjugate priors. The normal prior for the expected value in a normal distribution when the variance is known, which gives us a normal posterior distribution. And a gamma distribution for the precision in a normal distribution with known expectation will give a gamma posterior distribution (the precision τ is defined as the inverse of the variance in a normal distribution; $\tau = \frac{1}{\sigma^2}$, and is called the precision because as opposed to the variance a high precision implies that an observation from this distribution will be close to its theoretical mean).

5 Markov Chain Monte Carlo

The motivation for developing an MCMC scheme is the ability to draw random values from an arbitrary probability distribution. This is achieved by using some properties of Markov Chains together with cleverly chosen transition probabilities. First we explain what a Markov chain is, and some of its properties important for our work. We have not included proofs in this section, and interested readers can look at Gamerman and Lopes (2006), the book the theory and examples in this section is taken from. The second part uses the theory of Markov chains to draw samples from general distributions through what is called Metropolis-Hastings algorithms.

5.1 Markov Chains

A Markov Chain is a special kind of stochastic process $\{\theta^{(k)} : k \in K\}$ with state space S and index set K . Generally speaking a Markov Chain is a process where given the present state, past and future states are independent. This can be expressed more mathematically as

$$Pr(\theta^{(k+1)} = y | \theta^{(k)} = x, \theta^{(k-1)} = x_{k-1}, \dots, \theta^{(0)} = x_0) = Pr(\theta^{(k+1)} = y | \theta^{(k)} = x)$$

for all $x_0, \dots, x_{k-1}, x, y \in S$ in the discrete case and

$$Pr(\theta^{(k+1)} \in A | \theta^{(k)} = x, \theta^{(k-1)} \in A_{k-1}, \dots, \theta^{(0)} \in A_0) = Pr(\theta^{(k+1)} \in A | \theta^{(k)} = x) \quad (1)$$

for all $A_0, \dots, A_{k-1}, A \subseteq S$ and $x \in S$ in the continuous case. When this chain is independent of k , it is called homogeneous. Then for all $x \in S$, the transition kernel $P(x, \cdot)$ is a probability distribution over S in the continuous case, or $P(x, y)$ a transition probability in the discrete case. Although we will mostly use the notation for the discrete case from here on, one could exchange the state y with the state set A and it would for the most hold true in the continuous case.

The probability of going from x to y in m steps is written as $P^m(x, y)$, and T_y (if it exist) is the number of transitions it takes before hitting y for the first time after starting in y or said more precisely

$$T_y = \{\#n > 0 : \theta^{(n)} = y | \theta^{(0)} = y, \theta^{(k)} \neq y, \forall k \in \{1, \dots, n-1\}\}.$$

A state y is said to be transient when $P^m(y, y) < 1$ for every m , recurrent if $P^m(y, y) = 1$ for some m (possibly infinite), and positive recurrent if $E(T_y) < \infty$. A set $R \subseteq S$ is said to be irreducible if for all $x, y \in R$, $P^m(y, x) = 1$ for some m . A Markov Chain is said to be irreducible if S itself is irreducible.

A very important definition is that of a stationary distribution π . A distribution

π is said to be a stationary distribution of a chain with transition probabilities $P(x, y)$ if

$$\sum_{x \in S} \pi(x)P(x, y) = \pi(y), \quad \forall y \in S. \quad (2)$$

Once the chain reaches a stage where π is the distribution of the chain, it will retain this distribution for all subsequent stages.

Some chains will, given enough time, enter π by themselves, but to find those that do we need to introduce the concept of periodicity. The period of a state x , denoted by d_x , is the largest common divisor of the set

$$\{n \geq 1 : P^n(x, x) > 0\},$$

i.e. d_x is the smallest number of steps you can, with a positive probability, return to x with. It is obvious that if $P(x, x) > 0$ (that there is a positive probability to stay in state x) $d_x = 1$, and the chain is then called aperiodic. For an irreducible chain, all states have the same period. It can then be shown that an irreducible, positive recurrent, aperiodic chain with stationary distribution π has the property

$$\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y)$$

for all $x, y \in S$. Or said differently: given enough time, the chain is bound to reach its stationary distribution.

Although it is said earlier that the continuous case is completely analogue to the discrete, it should be noted that continuous chains need the slightly stronger notion of Harris recurrent rather than positive recurrent and sums needs to be exchanged with integrals like $\int_S \pi(dx)P(x, dy) = P(x, dy)$ in (2).

The last important feature we need to know about Markov Chains, is the concept of reversible chains. A reversible chain is a chain satisfying

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad (3)$$

for all $x, y \in S$. Reversible chains are useful because if there is a distribution π satisfying equation (3) for an irreducible chain, then it is a positive recurrent (or Harris-recurrent), reversible chain with π as its stationary distribution. This feature is paramount to MCMC and the Metropolis-Hastings algorithm.

5.2 The Metropolis-Hastings algorithm and the Gibbs sampler

The Metropolis-Hastings (MH) algorithm is a method of creating transition kernels, $P(x, y)$, for chains that have our wanted distribution as its stationary one. It uses the properties of a reversible chain given by equation (3).

The main idea behind MH is that the transition kernel $P(x, y)$ can be divided into two parts, one arbitrary transition kernel $Q(x, y)$ called the proposal kernel, and an acceptance probability $\alpha(x, y)$ such that

$$P(x, y) = Q(x, y)\alpha(x, y), \quad \text{if } x \neq y.$$

In other words, the kernel defines a density $P(x, \cdot)$ for every possible value of the parameter different from x . There is also a positive probability for the chain to remain in state x given by

$$P(x, x) = 1 - \int Q(x, y)\alpha(x, y)dy.$$

This gives us a general expression for the transition from state x into a subset A of the parameter space

$$P(x, A) = \int_A Q(x, y)\alpha(x, y)dy + I(x \in A) \left[1 - \int Q(x, y)\alpha(x, y)dy \right]. \quad (4)$$

Algorithms based on the transition kernel in (4) and with acceptance probability given by

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right\} \quad (5)$$

are called MH-algorithms.

If we just insert $Q(x, y)\alpha(x, y)$ instead of $P(x, y)$ in equation (3) we get

$$\pi(x)Q(x, y) \min \left\{ 1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right\} = \pi(y)Q(y, x) \min \left\{ 1, \frac{\pi(x)Q(x, y)}{\pi(y)Q(y, x)} \right\},$$

which we can see is equal. We can then see that chains with this kernel have π as limiting distributions by the properties of reversible chains.

We have not specified how to choose $Q(x, y)$, but it turns out that any irreducible aperiodic proposal kernel works as long as the acceptance probability is positive for every choice of x and y . In this paper will mostly use a random walk proposal kernel, which is defined as $Q(x, \cdot) \sim N(x, \sigma^2)$, that is, a normal distribution with the present state as its mean. This is obviously a irreducible and aperiodic proposal kernel on the real line.

That the wanted distribution appears in the acceptance probability does not become a problem. In practice we can't sample from π and we might not know its normalising constant, but we only need the general form because it appears as a fraction in (5). The MH-algorithm with a random walk proposal distribution can be summarized in 3 steps

1. Draw $\theta^{(i+1)}$ from $Q(\theta^{(i)}, \cdot)$.

2. Move to state $\theta^{(i+1)}$ with probability $\alpha(\theta^{(i)}, \theta^{(i+1)})$.
3. Change the counter i to $i + 1$ and go to step 1.

For our random walk proposal kernel there is a parameter that has not been specified, σ^2 . The variance of Q does not have one correct value, but needs to be tuned in each case. The rule of thumb is that we choose a σ^2 to get an acceptance rate of about 20%-50%. One more strength of the random walk proposal, other than its easy implementation, is that it is symmetric. In other words $Q(x, y) = Q(y, x)$, so the kernel-part disappears from equation (5).

How many iterations are needed depends on the problem, but very simply we can say we run the code until convergence. The values we have before convergence is reached is called the burn-in and are discarded, and all the values after convergence are supposed to be from the distribution of interest and can be used to find the mean value of the parameters, their medians etcetera. How to check for convergence is a complicated field, but in our case we can usually check the values of the parameters and assume convergence is reached when all parameters seem to stabilise about a certain mean.

MH example In pharmacology studies it is common to specify concentration levels of substances introduced in a system by non-linear equation on the form $f(\psi, x) = \psi_1 + \frac{\psi_2 x}{\psi_3 + x}$, where $\psi = (\psi_1, \psi_2, \psi_3)$ and x is the explanatory variable. Let us assume that y_i gives the velocity observation of a reaction when the concentration is x_i by the relation $y_i = f(\psi, x_i) + \epsilon_i$, $i = 1, \dots, n$, where the observation errors ϵ_i are iid normally distributed with zero mean and variance σ^2 . Let us for simplicity also assume that $(\psi_1, \psi_2, \sigma^2) = (50, 170, 126)$ i.e. known and that $\psi_3 = \theta$ is the parameter of interest. We put a $N(0, 100)$ prior on θ , and the posterior distribution becomes

$$p(\theta | \dots) \propto p(\theta)p(y|\theta) = f_N(\theta|0, 100) \prod_{i=1}^n f_N(y_i|50 + 170x_i/(\theta + x_i), 126).$$

We decide to use a random walk proposal kernel $Q(\theta, \phi)$ with tuning parameter $\sigma^2 = 0.01$, such that the next proposed step $\phi \sim N(\theta, 0.01)$. In each step we accept ϕ with probability $\alpha = \min \left\{ 1, \frac{p(\phi|\dots)}{p(\theta|\dots)} \right\}$, and keep θ with probability $1 - \alpha$. We would typically run this code for some iterations, look at the acceptance rate, and decide if we want to keep the 0.01 tuning variance in the kernel. If we are pleased we run the code until convergence. Let us say we run the code for 10 000 iterations and convergence is reached after 2000. We could then estimate the correct value of θ by using the mean of the last 8000 values the MH-algorithm gave us.

There is also a different special case of a MH-algorithm we will use; the Gibbs-sampler. The main motivation for using the Gibbs-sampler is that it has an acceptance probability of 1.

Earlier we have not burdened ourselves with deciding if the states x and y are scalars or vectors, but in practice we often have vectors and our algorithms can in each iteration stepwise update the state vector. This is an important point when explaining the Gibbs sampler. There are some cases when it is possible to sample from $\pi(x_i|x_{-i})$ even though it is impossible to sample from $\pi(x)$ itself. Here x_i denote the i -th part of the state vector x , and x_{-i} denote the state vector when the i -th part is removed. The distribution $\pi(x_i|x_{-i})$ is called the full conditional distribution for x_i and the Gibbs-sampler uses this distribution as its transition kernel. If we were to use the Gibbs-sampler to update a state vector x we could at the j -th iteration sample the first element as $x_1^j \sim \pi(x_1|x_{-1}^{j-1})$ and the second $x_2^j \sim \pi(x_2|x_1^j, x_{-\{1,2\}}^{j-1})$ and so on till all the elements were updated.

The reason that the Gibbs-sampler has an acceptance probability of 1 is obvious when we realise that $Q(x, y) = \pi(y_i|x_{-i})$ and $\pi(y) = \pi(y_i, x_{-i}) = \pi(y_i|x_{-i})\pi(x_{-i})$ and insert this fact into equation (5). Remember that here we have that $y = (x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)$.

Gibbs example Consider a sample of observations $y = (y_1, \dots, y_n)$ from a Poisson distribution. We know that at some point in time m , somewhere in $1, \dots, n$, the mean of the Poisson distribution changes and remains unchanged. So we have, given m and the parameters, $y_i | (\lambda, m) \sim \text{Pois}(\lambda)$, $i = 1, \dots, m$ and $y_i | (\phi, m) \sim \text{Pois}(\phi)$, $i = m+1, \dots, n$. From knowledge we have of the problem we choose to put gamma distributions as priors on both λ and ϕ , and a uniform prior on m , so $\lambda \sim \Gamma(\alpha, \beta)$, $\phi \sim \Gamma(\gamma, \delta)$ and $m \sim U[1, n]$. If we write out the posterior density we get

$$\begin{aligned} p(\lambda, \phi, m|y) &\propto p(y|\lambda, \phi, m)p(\lambda)p(\phi)p(m) \\ &= \left[\prod_{i=1}^m f_P(y_i|\lambda) \prod_{i=m+1}^n f_P(y_i|\phi) \right] f_\Gamma(\lambda|\alpha, \beta) f_\Gamma(\phi|\gamma, \delta) \frac{1}{n} \\ &\propto \left[\prod_{i=1}^m e^{-\lambda} \lambda^{y_i} \prod_{i=m+1}^n e^{-\phi} \phi^{y_i} \right] \lambda^{\alpha-1} e^{-\beta\lambda} \phi^{\gamma-1} e^{-\delta\phi} \\ &\propto \lambda^{\alpha+(\sum_{i=1}^m y_i)-1} e^{-(\beta+m)\lambda} \phi^{\gamma+(\sum_{i=m+1}^n y_i)-1} e^{-(\delta+n-m)\phi}. \end{aligned}$$

From experience we can recognise the full conditional distributions for both λ and ϕ ; $p(\lambda|\dots) \sim \Gamma(\alpha + \sum_{i=1}^m y_i, \beta + m)$ and $p(\phi|\dots) \sim \Gamma(\gamma + \sum_{i=m+1}^n y_i, \delta + n - m)$. For m we get a distribution we do not recognise, so we have to update that parameter with a MH-algorithm. So the procedure becomes as follows: draw λ from its gamma distribution where we use the last value of

m as the correct value, then draw ϕ from its gamma distribution where we again use the last value we have of m as the correct, then we draw m from its MH-algorithm with the newly drawn λ and ϕ as correct values. This is repeated until all the parameters reach convergence.

A MCMC scheme can often consist of a state space where, given a state vector x , a full conditional distribution can be found for some of its parts and we will use a Gibbs-sampler to update those. We will use more general MH-algorithms to create a sampler for the rest of the parts where a full conditionals can not be directly drawn from, just as we did in the Gibbs-example.

Although we have said MCMC enables us to draw from general distributions, it actually draws from distributions that only asymptotically equals π . The difference between these distributions goes towards 0 as the number of iterations increases. For a deeper discussion see Gamerman and Lopes (2006).

6 Bootstrap

6.1 Introduction to bootstrapping

The name bootstrap comes from a story of Baron von Münchhausen. He was supposed to have fallen to the bottom of a deep lake, but rescued himself by pulling himself up by his own bootstraps. The analogy is that you can solve a problem, seemingly magically, without any external help. The theory in this section is taken from Efron and Tibshirani (1994). The bootstrap principle was developed as a way to ascertain the standard error of an estimated parameter $\hat{\theta}$, calculated from a vector of observations $x = (x_1, \dots, x_n)$ from an unknown probability distribution F . So $\hat{\theta} = s(x)$ is an estimate of the correct statistic $\theta = t(F)$ (if we consider the sample mean as an estimate of the expected value we have $s(x) = \frac{1}{n} \sum_i x_i$ and $t(F) = \int x dF$), but we might have no concept of how good this estimate is. The non-parametric bootstrap principle tries to solve this when we have no more than the vector of observations. It uses the empirical distribution \hat{F} , which puts a probability $\frac{1}{n}$ on all the observations (x_1, \dots, x_n) , and draws a sample from that with replacement. E.g. say we have a sample $x = (x_1, x_2, x_3)$, \hat{F} would then equal $(1/3, 1/3, 1/3)$ and a random sample, denoted a bootstrap sample, $x^* = (x_1^*, x_2^*, x_3^*)$ could be (x_1, x_3, x_1) . Now you can create a bootstrap replica of $\hat{\theta}$, $\hat{\theta}^* = s(x^*)$, by using the bootstrap sample x^* . To ascertain the error of $\hat{\theta}$ by using bootstrap, we would create B bootstrap replicas of $\hat{\theta}$ and find the standard error of this; if the error is large it is not a “good“ (or certain) estimate of θ and vice versa. This can be summarised in three steps

1. Draw B independent bootstrap samples of length n , x^{*1}, \dots, x^{*B} , from $x = (x_1, \dots, x_n)$.
2. Find the bootstrap replicas of $\hat{\theta}$, $\hat{\theta}^*(b) = s(x^{*b})$ for $b = 1, \dots, B$.
3. Calculate the standard error of the replicas $\left\{ \sum_{b=1}^B [\hat{\theta}^*(b) - \bar{\theta}^*]^2 / (B - 1) \right\}^{1/2}$, where $\bar{\theta}^* = \sum_{b=1}^B \hat{\theta}^*(b) / B$.

6.2 Bootstrapping time series

Bootstrap have been used for financial time series (see Ruiz and Pascual (2002) for a review of literature on some of the applications), but when we want to bootstrap time series, there are things we need to consider which we did not consider in the introduction above. Most importantly we need to capture the correlation-structure the data exhibit over time. If we were to create bootstrap replicas without taking the correlation-structure in consideration, the resulting series would only be noise.

There are several ways to bootstrap time series; parametric, semi-parametric and non-parametric. A parametric approach would be to fit the data to a model, for example an AR(p)-model, and make bootstrap replicas of the residuals of the fitted model. This approach would generate time series that exhibit the same attributes as the original, given that the chosen model was a good one. If you are not sure about the underlying distribution of the time series, you should use a more general approach. One of the most general approaches is what is called block bootstrap. What you do is instead of resampling all of the observations, you divide the data in m blocks, and resample those. The problem is now how to choose the size of the blocks. You want large enough sizes to keep the properties of the original dataset, but small enough to not simply resample the observed history. There are also two different block bootstraps, one where the data do not overlap, and one where they do. Let us assume the observed time series $x = (z_1, \dots, z_m)$, where z_1 is the first block z_2 the second and so on. Each block have length l , i.e. $z_1 = (x_1, \dots, x_l)$, so the non-overlapping bootstrap would give us $z_2 = (x_{l+1}, \dots, x_{2l})$ and the overlapping would give us $z_2 = (x_2, \dots, x_{l+1})$. Which of the overlapping or non-overlapping is optimal to use is not clear, but Lahiri (1999) shows that the overlapping bootstrap have the smallest variance and might be more efficient, so we will use the overlapping bootstrap (also called moving block bootstrap). One other thing to be conscious about when using block bootstrap is trend in the data. If there is a trend in the data the resulting bootstrap resampling would not capture this well, as Figure 6.1 shows. In the figure we have a linear trend so it might be possible to translate each bootstrap sample such that every endpoint are connected, but this will not work on general trend-structures and we will therefore not use this on the observed trend of our data directly when we try to model the trend later. A semi-parametric form which sometimes outperforms the moving block bootstraps is called the sieve bootstrap (a Ph.D. that deals with some of these issues is Smeeke (2009)), but it is more complicated so the comparison between the moving block and the sieve bootstrap is left for further study.

Even though we will be working on models without trend we can still get problems if we just connect the blocks directly after each other. In our case it turns out that we work on series which sometimes have large volatilities, and if we are on a volatility-top and the next one starts at a low value we can get jagged series where the blocks are apparent. There are several ways to overcome this, and E. Carlstein et al. (1998) does this by defining an algorithm specifying which blocks that comes after one another. The way we choose to fix this is to use an AR(1)-model as a “glue” between blocks. Given ϕ and σ we can generate a short AR series that have the last value of the previous block as its first value, and the first value of the new block as its last. Since a AR(1) series can be seen as a draw from a multivariate

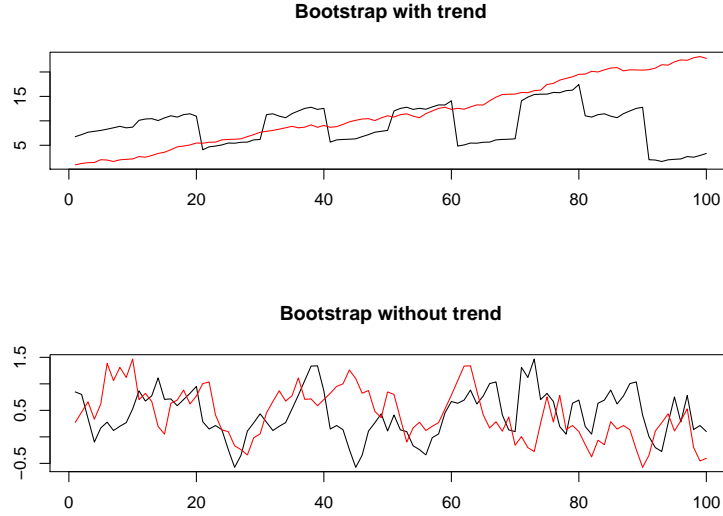


Figure 6.1: Bootstrap replicas of time series with and without trends (original series in red).

normal distribution (see for example Rue and Held (2005)), we can use theory from multivariate statistics to get the distribution of $(x_2, \dots, x_{n-1} | x_1, x_n)$. The following theory is taken from Rencher and Schaalje (2008).

When we have $f(x_1, \dots, x_n) \sim N(0, \Sigma)$ we can write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix},$$

where Σ_{11} is 2×2 , $\Sigma_{12} = \Sigma_{21}^T$ is $(n-2) \times 2$ and Σ_{22} is $(n-2) \times (n-2)$. If we denote $X_1 = (x_1, x_2)$ and $X_2 = (x_3, \dots, x_n)$ we get that $(X_2 | X_1)$ is normally distributed with $E(X_2 | X_1) = \Sigma_{12} \Sigma_{11}^{-1} X_1$ and $\text{var}(X_2 | X_1) = \Sigma_{22} - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{21}$. We will show how we use this more clearly in an example where $x = (x_1, \dots, x_5)$.

The joint distribution of these observations can be written as $f(x_1, \dots, x_5) \sim N(0, \Sigma)$, where

$$\Sigma = \frac{\sigma^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 \\ \phi & 1 & \phi & \phi^2 & \phi^3 \\ \phi^2 & \phi & 1 & \phi & \phi^2 \\ \phi^3 & \phi^2 & \phi & 1 & \phi \\ \phi^4 & \phi^3 & \phi^2 & \phi & 1 \end{pmatrix}$$

We want to find $f(x_2, x_3, x_4 | x_1 = a, x_5 = b)$, so we need the same kind of block-structure described for Σ_{11} , Σ_{22} and Σ_{12} . We observe that $f(x_1, x_5, x_2, x_3, x_4) \sim$

$N(0, \tilde{\Sigma})$, where

$$\tilde{\Sigma} = \frac{\sigma^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi^4 & \phi & \phi^2 & \phi^3 \\ \phi^4 & 1 & \phi^3 & \phi^2 & \phi \\ \phi & \phi^3 & 1 & \phi & \phi^2 \\ \phi^2 & \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi & \phi^2 & \phi & 1 \end{pmatrix}$$

So $\tilde{\Sigma}_{11} = \begin{pmatrix} 1 & \phi^4 \\ \phi^4 & 1 \end{pmatrix}$, $\tilde{\Sigma}_{22} = \begin{pmatrix} 1 & \phi & \phi^2 \\ \phi & 1 & \phi \\ \phi^2 & \phi & 1 \end{pmatrix}$ and $\tilde{\Sigma}_{12} = \begin{pmatrix} \phi & \phi^3 \\ \phi^2 & \phi^2 \\ \phi^3 & \phi \end{pmatrix}$. We now have that $f(x_2, x_3, x_4 | x_1 = a, x_5 = b)$ is normally distributed with expected value $E(x_2, x_3, x_4 | x_1 = a, x_5 = b) = \tilde{\Sigma}_{12} \tilde{\Sigma}_{11}^{-1} \begin{pmatrix} a \\ b \end{pmatrix}$ and variance $\text{var}(x_2, x_3, x_4 | x_1 = a, x_5 = b) = \tilde{\Sigma}_{22} - \tilde{\Sigma}_{12} \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{12}^T$. We can use this procedure to generate AR(1) series that are n long, which can act as “glue” between bootstrap blocks when we have the values of σ and ϕ , by sampling from a multivariate normal distribution with the parameters indicated above.

7 Stochastic volatility model

7.1 Deriving the standard model

The univariate log-stochastic volatility model can be used to model time-varying volatility, and was first proposed in Taylor (1982). It can be derived from an Itô-process under certain assumptions. When an equity price is written as $S(t)$, an Itô-process can be written as

$$dS(t) = \mu(S(t), t)S(t)dt + \sigma(S(t), t)S(t)dW(t),$$

where $W(t)$ is a Brownian motion. Since we are only concerned with the volatility-structure of the model now, we discard the drift term $\mu(S(t), t)$. We can say a stochastic volatility model is a product process between two different and independent shocks (or innovations), $\sigma(S(t), t)$ and $W(t)$, where one models the volatility process and the other models unpredictable information about the asset. To capture the volatility-clustering observed in practice we assume the volatility model to follow an autoregressive process, and thus write the model in logarithmic prices as

$$\begin{aligned} d \ln S(t) &= \sigma(t)dB_1(t) \\ d \ln \sigma^2(t) &= \alpha\beta + \alpha \ln \sigma^2(t)dt + \sigma dB_2(t). \end{aligned} \tag{6}$$

Here B_1 and B_2 are two uncorrelated Brownian motions, and σ is the volatility of the log-volatility-structure not to be mistaken with volatility-structure $\sigma(t)$ itself. For simulation and computational purposes it needs to be discretised, and we use the Euler discretisation. If we first denote the continuously compounded returns as $\ln(\frac{S(t+1)}{S(t)}) = r_t$ and $\ln(\sigma_t^2) = h_t$ our model can be written as

$$\begin{aligned} r_t &= \exp\left\{\frac{h_t}{2}\right\}\epsilon_t \\ h_t &= \nu + \phi(h_{t-1} - \nu) + \sigma\eta_t, \end{aligned} \tag{7}$$

where ϵ_t and η_t are i.i.d. $N(0, 1)$ variables for $t = 1, \dots, T$, $\nu = \beta$ and $\phi = 1 + \alpha$. So in other words r_t given all other processes is normal distributed with 0 mean, and variance $\exp\{h_t/2\}$, and h_t given the state at all other times, h_{-t} , and all its parameters is normal distributed with mean $\nu + \phi(h_{t-1} - \nu)$ and variance σ^2 .

We will, however, work on a 5-dimensional time series, so we need a multivariate stochastic volatility model. There are several ways to specify a multivariate stochastic volatility model (see McAleer et al. (2004)), but our way to move from the univariate to multivariate setting follows the work of Plataniotis and Dellaportas (2010). In the multivariate case we do not have a time-varying variance, but rather a time-varying covariance matrix. In other words we do not model the

σ_t in $N(0, \sigma_t^2)$, but the Σ_t in $r_t \sim N_5(0, \Sigma_t)$ when r_t is a 5-dimensional vector of observations. We will estimate the model parameters using a MCMC-scheme. It should be noted that a MCMC-scheme is not feasible when the number of dimensions grows sufficiently large, say above 10, as we would need to estimate $O(N^2)$ latent paths for an N -dimensional problem. For larger dimensions we would rather use the INLA approach introduced by Rue, Martino and Chopin (2009).

So we need to estimate the T number of covariance matrices Σ_t . A major challenge is that a covariance matrix has to be positive-definite, so we need a state-space model that keeps the constraint of positive-definiteness and gives us volatility clustering. To ensure positive-definiteness we will look at the spectral decomposition of the covariance matrices, $\Sigma_t = U_t \Lambda_t U_t^T$. If we focus on these time dependent matrices, rather than Σ_t directly, and keep all the eigenvalues positive and the eigenvector matrices orthonormal, the covariance matrix will be positive definite. An N -dimensional covariance matrix consist of $\frac{N(N+1)}{2}$ unique parameters due to symmetry, and the spectral decomposition puts N of them in the diagonal matrix and $\frac{N(N-1)}{2}$ in the orthonormal matrices.

Because orthonormal matrices can be seen as a series of rotations and/or reflections, an orthonormal matrix can be written as a product of $\frac{N(N-1)}{2}$ rotation matrices (here called Givens matrices) $U_t = \prod_{i=1}^{N(N-1)/2} G_i(\theta_{i,t})$. The idea is that any multidimensional rotation can be seen as several pairwise rotations in a plane, as shown by Hoffman, Raffenetti and Ruedenberg (1972), and the number of ways to divide N values as pairs of two are $\frac{N(N-1)}{2}$. We can then let the different $\theta_{i,t}$ vary without worrying about keeping the orthonormality constraints we would need if we work on U_t directly, as every Givens matrix is orthonormal by design.

So for each point in time, t , we have a 5-dimensional eigenvalue matrix Λ_t and 10 5-dimensional Givens-matrices $G_i(\theta_{i,t})$ each with only one variable; $\theta_{i,t}$ for $i = 1, \dots, 10$. This gives us the possibility to decompose the covariance matrix to 15 variables, $\lambda_{1,t}, \dots, \lambda_{5,t}$ and $\theta_{1,t}, \dots, \theta_{10,t}$. This shows that the complexity increases drastically with increased dimensionality, from 1 variable in the 1-dimensional case to 15 in the 5-dimensional case.

In our multivariate model where $r_t = (r_{1,t}, \dots, r_{5,t})^T$, we now have a model that gives us the volatility of each asset (when we in this thesis use the term *the volatility of $r_{i,t}$* we will most often mean $\lambda_{i,t}$) and the correlation between them (when we use the term *correlation* we mean θ_t , and the correlation between $r_{1,t}$ and $r_{2,t}$ is given through $\theta_{1,t}$ and $r_{1,t}$ and $r_{3,t}$ through $\theta_{2,t}$, etcetera). When we write this

out completely we get

$$\Lambda_t = \begin{pmatrix} \lambda_{1,t} & & & & \\ & \ddots & & & \\ & & \lambda_{5,t} & & \end{pmatrix}, \text{ and}$$

$$G_{1,t} = \begin{pmatrix} \cos(\theta_{1,t}) & -\sin(\theta_{1,t}) & & & \\ \sin(\theta_{1,t}) & \cos(\theta_{1,t}) & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}$$

$$G_{2,t} = \begin{pmatrix} \cos(\theta_{2,t}) & -\sin(\theta_{2,t}) & & & \\ & 1 & & & \\ \sin(\theta_{2,t}) & & \cos(\theta_{2,t}) & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}$$

$$\vdots$$

$$G_{10,t} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \cos(\theta_{10,t}) & -\sin(\theta_{10,t}) \\ & & & \sin(\theta_{10,t}) & \cos(\theta_{10,t}) \end{pmatrix}.$$

In the univariate case we worked on $\ln(\sigma_t)$ to ensure a positive variance. Analogous with this we need positive eigenvalues in the multivariate case for the covariance matrix to be positive-definite, and to ensure that the Givens matrices are unique we also need the angles to be between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. So instead of using $\lambda_{i,t}$ and $\theta_{j,t}$ directly we use $h_{i,t} = \ln(\lambda_{i,t})$ and $\delta_{j,t} = \ln\left(\frac{\frac{\pi}{2} + \theta_{j,t}}{\frac{\pi}{2} - \theta_{j,t}}\right)$. We can use about any time dependent model we want on $h_{i,t}$ and $\delta_{j,t}$ and the resulting covariance matrix Σ_t we would get when we transformed it back would be positive-definite. For parsimonious reasons we would like to use a simple model that can capture volatility clustering, and like before we will use an AR(1) model to achieve this. The resulting model seen below looks much like the univariate one although its derivation was more complicated;

$$\begin{aligned} h_{i,t} &= \nu_{h_i} + \phi_{h_i}(h_{i,t-1} - \nu_{h_i}) + \sigma_{h_i}\eta_{h_i,t} \\ \delta_{j,t} &= \nu_{\delta_j} + \phi_{\delta_j}(\delta_{j,t-1} - \nu_{\delta_j}) + \sigma_{\delta_j}\eta_{\delta_j,t}, \end{aligned} \tag{8}$$

for $i = 1, \dots, 5$ and $j = 1, \dots, 10$.

7.2 Simulating from the standard model

Because we are in a Bayesian setting we need to find the posterior distribution for the parameters and the state-space variables (h_t and δ_t). When we look at (8) we see that every state variable are dependent on their former value, except of course the first value; $h_{i,1}$ and $\delta_{j,1}$. And at the same time, every value will have a dependency in their next value, except the last value; $h_{i,T}$ and $\delta_{j,T}$. This will obviously affect the full conditional distribution for these values. We see that the full conditional distribution for $h_t = (h_{1,t}, \dots, h_{5,t})$, $t \in [2, T-1]$ is

$$\begin{aligned}
 p(h_t | \dots) &= p(r_t | h_t, \dots) p(h_t | \rho_{h_1}, \dots, \rho_{h_5}, h_{t\pm 1}) \\
 &\propto \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^5 h_{i,t} + r_t^T \Sigma_t^{-1} r_t \right]\right\} \\
 &\quad \exp\left\{-\sum_{i=1}^5 \frac{\tau_{h_i}}{2} [(h_{i,t+1} - \nu_{h_i} - \phi_{h_i}(h_{i,t} - \nu_{h_i}))^2 + (h_{i,t} - \nu_{h_i} - \phi_{h_i}(h_{i,t-1} - \nu_{h_i}))^2]\right\},
 \end{aligned} \tag{9}$$

here we use $\tau = \sigma^{-2}$ called the precision rather than the variance directly. The reason for this is because of the full conditional for that parameter then has a well known distribution as we will see later. The full conditional for h_T is not so different from the one above. You simply drop the h_{t+1} -part and end up with

$$\begin{aligned}
 p(h_T | \dots) &\propto \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^5 h_{i,T} + r_T^T \Sigma_T^{-1} r_T \right]\right\} \\
 &\quad \exp\left\{-\sum_{i=1}^5 \frac{\tau_{h_i}}{2} [(h_{i,T} - \nu_{h_i} - \phi_{h_i}(h_{i,T-1} - \nu_{h_i}))^2]\right\}
 \end{aligned}$$

as the full conditional. For h_1 it is slightly different. We want a dependency through the former value, but there are no former value for $t = 1$, so we use the limiting distribution of h_t instead. The limiting distribution is the distribution the process tends towards given an infinite amount of time. The limiting distribution can be found from (8) by substitution, by using that $\phi \in (-1, 1)$ and the fact that the variance of h_t is homoscedastic i.e. do not change over time;

$$\begin{aligned}
 \mathbb{E}(h_t) &= \nu + \phi(\mathbb{E}(h_{t-1}) - \nu) = \nu + \phi((\nu + \phi(\mathbb{E}(h_{t-2}) - \nu)) - \nu) \\
 &= \dots = \nu + \phi^n(\mathbb{E}(h_{t-n}) - \nu) \approx \nu \\
 \text{Var}(h_t) &= \phi^2 \text{Var}(h_{t-1}) + \sigma^2 = \phi^2 \text{Var}(h_t) + \sigma^2 \\
 \Rightarrow \text{Var}(h_t) &= \frac{\sigma^2}{1 - \phi^2} = \frac{1}{\tau(1 - \phi^2)}.
 \end{aligned}$$

So we end up with a full conditional that looks like this:

$$p(h_1|\dots) \propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^5 h_{i,1} + r_1^T \Sigma_1^{-1} r_1\right]\right\} \\ \exp\left\{-\sum_{i=1}^5 \frac{\tau_{h_i}}{2} [(h_{i,2} - \nu_{h_i} - \phi_{h_i}(h_{i,1} - \nu_{h_i}))^2 + (1 - \phi^2)(h_{i,1} - \nu)^2]\right\}.$$

The full conditional for $\delta_t = (\delta_{1,t}, \dots, \delta_{10,t})$, $t \in [2, T - 1]$ is

$$p(\delta_t|\dots) = p(r_t|\delta_t, \dots)p(\delta_t|\rho_{\delta_1}, \dots, \rho_{\delta_{10}}, \delta_{t\pm 1}) \\ \propto \exp\left\{-\frac{1}{2}[r_t^T \Sigma_t^{-1} r_t]\right\} \\ \exp\left\{-\sum_{j=1}^{10} \frac{\tau_{\delta_j}}{2} [(\delta_{j,t+1} - \nu_{\delta_j} - \phi_{\delta_j}(\delta_{j,t} - \nu_{\delta_j}))^2 + (\delta_{j,t} - \nu_{\delta_j} - \phi_{\delta_j}(\delta_{j,t-1} - \nu_{\delta_j}))^2]\right\}. \quad (10)$$

And the full conditionals for δ_1 and δ_T are completely analogous to that of h_1 and h_T .

Here $\rho.$ is the parameters for the variable given by its index (e.g. ρ_{h_1} is τ_{h_1} , ν_{h_1} and ϕ_{h_1} , and so on) and r_t denote the 5-dimensional observation vector at time t . The likelihood term in the full conditional is

$$r_t^T \Sigma_t^{-1} r_t = r_t^T (U_t \Lambda_t U_t^T)^{-1} r_t = r_t^T U_t \Lambda_t^{-1} U_t^T r_t,$$

which our computer calculates through matrix multiplication.

To find the full conditional distribution of the parameters, we need to put priors on ν , ϕ and σ^2 . We do not need to specify which of the state-space variables the parameter belongs to, because they all have the same distribution. We want to make sure that ϕ is between -1 and 1 and that σ^2 is always positive (although we will be working on the precision $\tau = \sigma^{-2}$). To simplify the algorithms we choose conjugate priors on ν and τ , and we end up with the following priors

$$\nu \sim N(\mu_\nu, \sigma_\nu^2) \\ \frac{\phi + 1}{2} \sim \beta(\omega, \gamma) \\ \tau \sim \Gamma(\alpha, \beta).$$

Here N , β and Γ denote respectively the normal, beta and gamma distributions. The parametrisations of the different probability distributions can be seen in Appendix A. The parameters of these priors are simply hyperparameters whose values reflect what we know about the processes beforehand.

It should be noted that even though we here find the full conditionals of the parameters for some $h_{i,t}$ (for simplicity denoted h_t), we can get the full conditionals for the parameters of $\delta_{j,t}$ by exchanging h_t with $\delta_{j,t}$.

We chose conditional priors for both ν and τ , and we therefore know that the full conditionals for those have the same distributions, but with different parameters. Looking at the full conditional of τ we see that

$$\begin{aligned} p(\tau|\cdots) &\propto p(\tau|\alpha, \beta) \prod_{t=1}^T p(h_t|\nu, \tau, \phi, h_{-t}) \\ &\propto \tau^{\alpha-1} \exp\{-\beta\tau\} \times \tau^{T/2} \exp\left\{-\frac{\tau}{2}\left[(1-\phi^2)(h_1-\nu)^2 + \sum_{t=2}^T (h_t-\nu-\phi(h_{t-1}-\nu))^2\right]\right\} \\ &= \tau^{T/2+\alpha-1} \exp\left\{-\tau\left(\beta + \frac{1}{2}\left[(1-\phi^2)(h_1-\nu)^2 + \sum_{t=2}^T (h_t-\nu-\phi(h_{t-1}-\nu))^2\right]\right)\right\}, \end{aligned}$$

where h_{-t} denote all $h_i, i = 1, \dots, T$ except h_t . We see that this is the Gamma distribution with parameters $\tilde{\alpha}$ and $\tilde{\beta}$, where

$$\begin{aligned} \tilde{\alpha} &= \frac{T}{2} + \alpha \\ \tilde{\beta} &= \beta + \frac{1}{2}\left[(1-\phi^2)(h_1-\nu)^2 + \sum_{t=2}^T (h_t-\nu-\phi(h_{t-1}-\nu))^2\right]. \end{aligned}$$

To find the parameters of the full conditional distribution of ν we can in the same way write out the full expression, reorder the parameters so that we recognise the form of the normal distribution, and find our new expectation and variance from that. There is, however, very easy to make mistakes when reordering the parameters and we might therefore consider a different approach (just look at the Bayesian part of the example in the Bayesian Inference-section). Since we know that the full conditional is a normal distribution and we only want to know the expectation and variance, we can derive these algebraically. This procedure will work on all normal distributions, and we will only need to consider the quadratic equation in the exponent. If we denote the mean-value we wish to find as $\tilde{\mu}$ and the variance as $\tilde{\sigma}^2$ this can easily be rewritten as $f(\nu) = \frac{-(\nu-\tilde{\mu})^2}{2\tilde{\sigma}^2} = \frac{1}{2}C\nu^2 - B\nu + A$, for some A , B and C . If we evaluate f in three different points, say $\nu = -1$, $\nu = 0$ and $\nu = 1$ and call those values $f(-1) = p$, $f(0) = q$ and $f(1) = r$ respectively, we could use those to find which values A , B and C must have. Simple algebra shows us that $C = p + r - 2q$, $B = \frac{1}{2}(p - r)$ and $A = q$. From the definition of A , B and C it is can be seen that $\tilde{\sigma}^2 = \frac{-1}{C}$ and $\tilde{\mu} = \frac{B}{C}$. To find the new parameters we simply have to write out the full conditional of ν , evaluate it in three points namely -1 , 0 and 1 , and use these values to calculate the new parameters. The

full conditional of ν is

$$\begin{aligned} p(\nu|\cdots) &\propto p(\nu|\mu_\nu, \sigma_\nu^2) \prod_{t=1}^T p(h_t|\nu, \tau, \phi, h_{-t}) \\ &\propto \exp\left\{-\frac{\tau}{2}[(1-\phi^2)(h_1-\nu)^2 + \sum_{t=2}^T (h_t-\nu-\phi(h_{t-1}-\nu))^2] - \frac{\tau_\nu}{2}(\nu-\mu_\nu)^2\right\}, \end{aligned}$$

where τ_ν , according to former notation, is defined to be $\tau_\nu = \frac{1}{\sigma_\nu^2}$. So the new parameters are

$$\begin{aligned} \tilde{\mu} &= \frac{p-r}{2(p+r-2q)} \\ \tilde{\sigma}^2 &= \frac{-1}{p+r-2q}, \end{aligned}$$

where p, q and r is as defined above.

The full conditional for ϕ is

$$\begin{aligned} p(\phi|\cdots) &\propto p(\phi|\omega, \gamma) \prod_{t=1}^T p(h_t|\nu, \tau, \phi, h_{-t}) \\ &\propto \sqrt{1-\phi^2} \exp\left\{-\frac{\tau}{2}[(1-\phi^2)(h_1-\nu)^2 + \sum_{t=2}^T (h_t-\nu-\phi(h_{t-1}-\nu))^2]\right\} \\ &\quad \left(\frac{\phi+1}{2}\right)^{\omega-1} \left(1-\frac{\phi+1}{2}\right)^{\gamma-1}. \end{aligned}$$

This is not a distribution we know or can draw directly from.

In our simulation we sample the different ν and τ with a Gibbs sampler, and the ϕ , $h_{i,t}$ and $\delta_{j,t}$ using a Metropolis-Hastings algorithm with a random walk proposal distribution.

It should be noted that in our model we need to assume that δ_t has little variation about its mean to be able to draw inference (a small σ_{δ_j} or a high τ_{δ_j}). Later we will see that this indeed holds true for real data.

To check for convergence we will look at the parameter evolution, that is look at the values of the parameters in each iteration of the code and see how they diverge or converge about a certain value. To draw inference about the underlying model it can also be useful to look at what is called the filtered series. The filtered series is the mean value of each $h_{i,t}$ and $\delta_{j,t}$ over all the iterations after convergence is reached. If we denote the filtered series by a hat, this becomes $\hat{h}_{i,t} = \sum_{k=a}^K h_{i,t}^{(k)}$, for $h_{i,t}$ when $h_{i,t}^{(k)}$ is its value in the k th iteration, a is when convergence is reached and K is the total number of iterations. $\hat{\delta}_{j,t}$ is found in the same way.

7.3 Extending the model to encompass high volatility regimes

When we observe the return series in Figure 2.3 and couple it with our knowledge of the financial situation at the time, it seems as though the volatility of the returns, and perhaps even the correlation between them, increases during the financial crisis and the following recession. We would therefore like our model to be able to change parameters during such a crisis, and return to its original form after the crisis has passed. In other words, we want a model where the parameters have different values depending on an underlying regime. We will use the same approach as Hamilton (1988) did for US Treasury bills and as Hamilton (1989) did for US GNP to measure economic recessions. Those papers are mostly concerned with finding and filtering out such regimes in long data series, but in our case we only have one regime and we can assume knowledge of when it starts and stops. So we will focus mostly on the way those papers model how the time series change during different regimes. In our case we will assume there exist two different regime states; one ordinary regime and one crisis regime. At a certain time t we can say the series are in regime state s_t where $s_t = 0$ if we are in the ordinary regime at time t and $s_t = 1$ if we are in the crisis. Hamilton (1988) lets both the mean value ν and the variance σ^2 of the time series to be functions of the states. Since we do not have that much data in both regimes and because we would like to keep the model as simple as possible, we choose to only let the most important parameter to be a function of s_t , namely $\nu(s_t)$. $\nu(s_t)$ is defined as follows $\nu(s_t) = \nu_0 + \nu_1 s_t$, so in the ordinary state the mean is ν_0 , but during a crisis it will increase with ν_1 . We will expect ν_1 to be positive, especially for the different $h_{i,t}$, but we will allow for negative values as well. To simplify the notation we will simply call $\nu(s_t)$ for ν_t , so the mean value of $h_{2,t}$ would be $\nu_{h_{2,t}}$. It should be clear when we use ν_1 to denote $\nu(s_1)$ and when we use it to denote ν_1 in $\nu(s_t) = \nu_0 + \nu_1 s_t$ by our usage. When we implement this regime ν_t in our code, there are not much we need to change. The full conditional distributions for $h_{i,t}$, $\delta_{j,t}$ and the parameters τ and ϕ need to exchange the values of ν with ν_t where they appear. For example the full conditional of τ becomes $\tau \sim \Gamma(\tilde{\alpha}, \tilde{\beta})$ where $\tilde{\alpha} = \frac{T}{2} + \alpha$ and $\tilde{\beta} = \beta + \frac{1}{2}[(1 - \phi^2)(h_1 - \nu_1)^2 + \sum_{t=2}^T (h_t - \nu_t - \phi(h_{t-1} - \nu_{t-1}))^2]$. What obviously changes slightly more is the full conditional distribution of ν_t . Since we now have two parameters ν_0 and ν_1 we get a full conditional for each of them. We first need to decide the prior distribution. For the same reasons as in the standard model, we choose conjugate priors; i.e. the normal distribution for both of the parameters. We then get

$$\begin{aligned}\nu_0 &\sim N(\mu_{\nu_0}, \tau_{\nu_0}) \\ \nu_1 &\sim N(\mu_{\nu_1}, \tau_{\nu_1}).\end{aligned}$$

The full conditional distribution of ν_0 is then

$$\begin{aligned} p(\nu_0 | \dots) &\propto p(\nu_0 | \mu_{\nu_0}, \tau_{\nu_0}) \prod_{t=1}^T p(h_t | \nu_0, \nu_1, s_{t,t-1}, \tau, \phi, h_{-t}) \\ &\propto \exp\left\{-\frac{\tau}{2}[(1 - \phi^2)(h_1 - (\nu_0 + \nu_1 s_1))^2 + \right. \\ &\quad \left. \sum_{t=2}^T (h_t - (\nu_0 + \nu_1 s_t) - \phi(h_{t-1} - (\nu_0 + \nu_1 s_{t-1})))^2] - \frac{\tau_{\nu_0}}{2}(\nu_0 - \mu_{\nu_0})^2\right\}, \end{aligned}$$

and the full conditional of ν_1 is

$$\begin{aligned} p(\nu_1 | \dots) &\propto p(\nu_1 | \mu_{\nu_1}, \tau_{\nu_1}) I(s_1) \prod_{t=1}^T p(h_t | \nu_0, \nu_1, s_{t,t-1}, \tau, \phi, h_{-t}) I(s_t | s_{t-1}) \\ &\propto \exp\left\{-\frac{\tau}{2}[(1 - \phi^2)(h_1 - (\nu_0 + \nu_1 s_1))^2 I(s_1) + \right. \\ &\quad \left. \sum_{t=2}^T (h_t - (\nu_0 + \nu_1 s_t) - \phi(h_{t-1} - (\nu_0 + \nu_1 s_{t-1})))^2 I(s_t | s_{t-1}) \right. \\ &\quad \left. - \frac{\tau_{\nu_1}}{2}(\nu_1 - \mu_{\nu_1})^2\right\}, \end{aligned}$$

where $I(s_1)$ is an indicator variable which equals one if $s_1 = 1$ and $I(s_t | s_{t-1})$ is an indicator variable which equals one if either $s_t = 1$ or $s_{t-1} = 1$. We need these indicator variables because ν_1 only influences the data series during crisis (and on the border of the crisis) not outside of it. Since we used conjugate priors, we know the full conditional for both ν_0 and ν_1 have the normal distribution, and the parameters can be found by using algebra just as we did in the standard model. When we start our algorithm we want the state-space variables to mix properly, and therefore we will do what was done in Kim et al. (1998) and only use the expected values given by the hyperparameters for the first 3000 iterations as the correct value of the parameters. I.e. we only draw the state-space variables for the 3000 first iterations, and we will draw them with parameter values given by their hyperparameters ($\nu = \mu_\nu$, $\phi = 2\frac{\omega}{\gamma + \omega} - 1$ and so on), and then we will draw both the state-space variables and the parameters for the subsequent iterations. A pseudo-code of how our MCMC algorithm turns out can be seen below in Algorithm 1.

Algorithm 1: MCMC algorithm for the stochastic volatility model

Data: 5 dimensional vector of observations r_t , $t = [1, \dots, T]$ and the number of runs

Result: The filtered series of h_t and δ_t , and the parameter evolutions
begin

Set the values of the hyperparameters;

Define variables to hold the values of the parameters, h_t , δ_t and $r\Sigma_0 r^T$, and add initial values to them;

Set the regime-indicating variable s_t and the tuning variables σ_ϕ , σ_h and σ_δ ;

for i in 2 to number of runs **do**

foreach $h_{1,t}, \dots, h_{5,t}$ and $\delta_{1,t}, \dots, \delta_{10,t}$ **do**

if $i < 3000$ **then**

$$\tau = \frac{\alpha}{\beta};$$

$$\nu_0 = \mu_{\nu_0};$$

$$\nu_1 = \mu_{\nu_1};$$

$$\phi = 2 \frac{\omega}{\gamma + \omega} - 1$$

else

 Draw their $\tau \sim \Gamma(\tilde{\alpha}, \tilde{\beta})$;

 Draw their $\nu_0 \sim N(\tilde{\mu}_0, \tilde{\sigma}_0^2)$;

 Draw their $\nu_1 \sim N(\tilde{\mu}_1, \tilde{\sigma}_1^2)$;

 Draw their $\hat{\phi} \sim N(\phi, \sigma_\phi)$ and set $\phi = \hat{\phi}$ with probability

$$\alpha = \min\left\{\frac{p(\hat{\phi}|\dots)}{p(\phi|\dots)}, 1\right\}$$

foreach h_t where $t = [1, \dots, T]$ **do**

 Draw $\hat{h}_t \sim N(h_t, \sigma_h)$ and set $h_t = \hat{h}_t$ with probability

$$\alpha = \min\left\{\frac{p(\hat{h}_t|\dots)}{p(h_t|\dots)}, 1\right\}$$

foreach δ_t where $t = [1, \dots, T]$ **do**

 Draw $\hat{\delta}_t \sim N(\delta_t, \sigma_\delta)$ and set $\delta_t = \hat{\delta}_t$ with probability

$$\alpha = \min\left\{\frac{p(\hat{\delta}_t|\dots)}{p(\delta_t|\dots)}, 1\right\}$$

Calculate and return the mean value of h_t^i and δ_t^i for every i after convergence is reached, and return the parameter evolutions

end

7.4 Testing the model on generated data

To see how this model works it can be interesting to see how it is able to replicate the parameters and the state-space variables of series generated by us, and compare the estimated values against their theoretical counterparts. The parameters we choose can be seen in Table 3. For simplicity we choose the same parameters on almost all the h_t -series (with an exception of a lower mean of the volatility and a decrease in volatility during crisis on $h_{3,t}$ just to see if it effects the algorithm), and on all the δ_t -series. We choose a smaller variance for δ because these should

| | σ^2 | ν_0 | ν_1 | ϕ |
|----------|------------|---------|---------|--------|
| h_{-3} | 0.09 | -8 | 2 | 0.9 |
| h_3 | 0.09 | -9 | -1 | 0.9 |
| δ | 0.0025 | 0 | 0.5 | 0.95 |

Table 3: True parameter values of the generated series.

in realistic cases be quite small, and we choose a higher autoregressive coefficient ϕ for δ because those series in general change values slower than h do. In our MCMC- scheme we decide to put quite general priors i.e. choose hyperparameters that give relatively large variances, except for ν_1 which we force to be quite small. We also use the same hyperparameters on both h and δ for ϕ , but different for the rest of the parameters to capture what we know beforehand (i.e. lower variance for the δ -series and different expected values for h_t and δ_t).

$$\begin{array}{l}
 h \mid \mu_{\nu_0} = -5 \quad \sigma_{\nu_0}^2 = 10^2 \quad \mu_{\nu_1} = 2 \quad \sigma_{\nu_1}^2 = 0.5 \quad \omega = 30 \quad \gamma = 1 \quad \alpha_h = 100 \quad \beta_h = 10 \\
 \delta \mid \mu_{\nu_0} = 0 \quad \sigma_{\nu_0}^2 = 5 \quad \mu_{\nu_1} = 0.8 \quad \sigma_{\nu_1}^2 = 0.3^2 \quad \omega = 30 \quad \gamma = 1 \quad \alpha_\delta = 400 \quad \beta_\delta = 1
 \end{array}$$

We let our code run for some (>5000) iterations with several values of the tuning-parameter in our MH-algorithms (when we draw ϕ , h_t and δ_t), and decide to choose 0.1, 0.5 and 0.03 respectively. We now let the code run for 25 000 iterations and look at the parameter evolutions and the filtered series. The output shows that our code is much better at drawing inference about the volatility processes, i.e. $h_{i,t}$, than the angles. In Table 4 we can see the average parameter values after discarding the first 15 000 iterations with their standard deviations in parenthesis. The parameters of h_t are close to their correct values, except ν_1 which is too low (in absolute value) for all the cases. The parameters of δ_t are also quite good. The problems being that the estimates of ν_1 are not significantly larger than zero and ϕ is much too low for both $\delta_{t,1}$ and $\delta_{t,3}$. It seems like the code has a difficulty with finding what is the cause of the increase in volatility; the regime or just the high values of ϕ . One way to check the validity of the code, is by running it on

| | σ^2 | ν_0 | ν_1 | ϕ |
|---------------|-------------------|------------------|----------------|----------------|
| h_1 | 0.1(0.0081) | -7.81(0.112) | 0.869(0.251) | 0.915(0.0183) |
| h_2 | 0.106(0.00945) | -7.87(0.0972) | 0.815(0.218) | 0.898(0.0192) |
| h_3 | 0.0954(0.01) | -9.15(0.0956) | -0.364(0.201) | 0.903(0.0186) |
| h_4 | 0.103(0.0101) | -7.93(0.0965) | 0.942(0.258) | 0.896(0.0253) |
| h_5 | 0.107(0.00969) | -7.8(0.134) | 0.691(0.265) | 0.929(0.0158) |
| δ_1 | 0.00249(0.00012) | -0.0296(0.0228) | 0.031(0.0281) | 0.741(0.057) |
| δ_2 | 0.00252(0.000126) | 0.0361(0.0383) | 0.0479(0.0478) | 0.96(0.00945) |
| δ_3 | 0.00248(0.000123) | 0.0294(0.0127) | -0.034(0.0243) | 0.789(0.0634) |
| δ_4 | 0.00254(0.000133) | -0.00585(0.0211) | 0.0367(0.0402) | 0.93(0.0154) |
| δ_5 | 0.00253(0.000133) | 0.0519(0.0288) | 0.0309(0.0452) | 0.95(0.0111) |
| δ_6 | 0.0025(0.000128) | 0.00878(0.027) | 0.0573(0.0454) | 0.943(0.0124) |
| δ_7 | 0.00252(0.000121) | 0.0835(0.0478) | 0.0385(0.05) | 0.969(0.00785) |
| δ_8 | 0.00251(0.000115) | 0.0633(0.0356) | 0.0458(0.048) | 0.958(0.0111) |
| δ_9 | 0.00248(0.000128) | 0.0352(0.0788) | 0.0278(0.0482) | 0.977(0.00732) |
| δ_{10} | 0.00251(0.00014) | 0.0831(0.0326) | 0.0489(0.0436) | 0.923(0.0145) |

Table 4: Mean parameter values of generated state-space series with their standard deviations in parenthesis.

a data generated with negative ϕ -values. When we do this, we get much better estimates for ν_1 . Incidentally, when we get better estimates for ν_1 we also get better estimates for ν_0 . This indicate that the estimate of ν_0 over-compensate for the too low estimate of ν_1 , but relatively little compared to its actual error. So a practitioner should look at the filtered series and couple that with what he/she knows about the external regimes, and reflect this in the prior. If there seems to be an increase during the crisis-regime we can try to see how much, and choose a prior with a high mean and low variance to force the algorithm to give higher estimates of ν_1 . It is therefore interesting to see if the resulting filtered series reflect the “true” series well. The best and worst cases of h_t and δ_t can be seen in Figure 7.1. It is very easy to see that our algorithm is much better at finding out how the volatility behaves rather than the correlation. Which of the series our algorithm have problems distinguishing can actually easily be seen in Table 4; the δ_t with the low ϕ is basically only white noise.

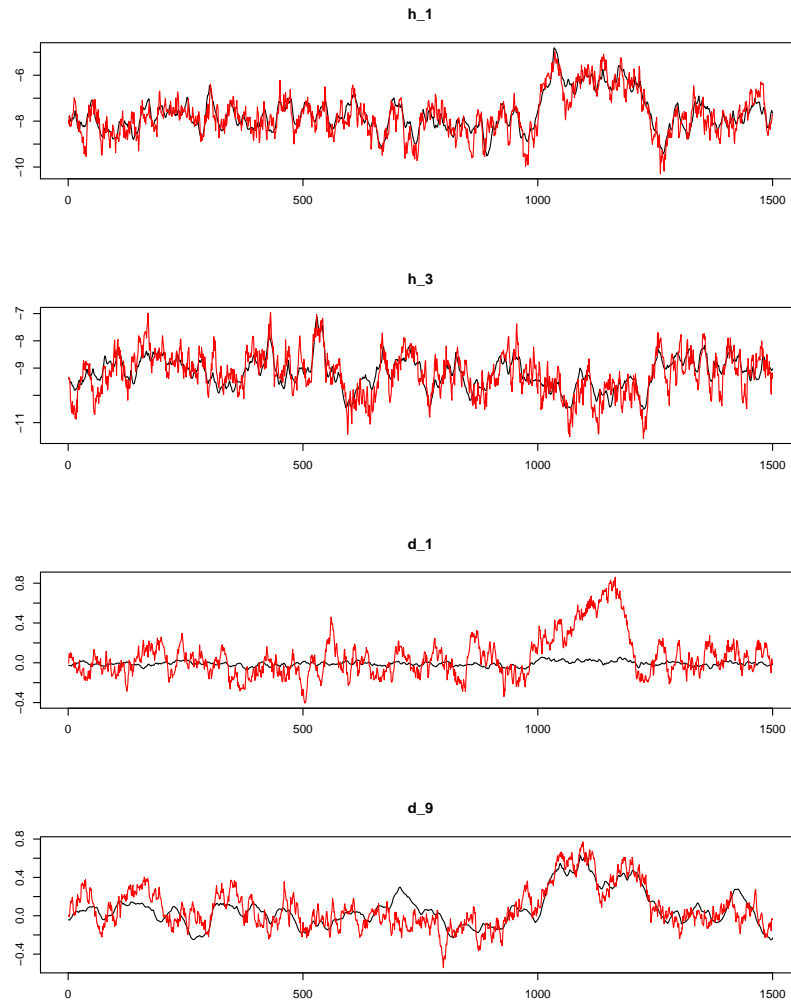


Figure 7.1: Filtered series of $h_{t,1}$, $h_{t,3}$, $\delta_{t,1}$ and $\delta_{t,9}$ with the correct series in red.

We have large correlation-structures between all the series, so the algorithm might have problems distinguishing between the correlation between series 1 and 3 directly, and that which comes from 1 and 2 and then 2 and 3. In practise the data will not be as correlated as it was here, but it is interesting to see how the algorithm cope with a complicated situation. It can anyway seem as though our algorithm underestimate the size of the correlation. All the filtered series can be seen in Appendix B.1.

When we perform this procedure on the real data later, we are concerned if the normality assumption holds between the volatility series and the observed prices. Remember that we assume $r_t \sim N(0, \Sigma_t)$ and we have that $\Sigma_t = U_t \Lambda_t U_t^T$. Then, by the property of the multivariate normal distribution, we should have that $\Lambda_t^{-1/2} U_t^T r_t \sim N(0, I)$, i.e. the scaled returns are standard normally distributed. Since we have the filtered series of both h_t and δ_t , which gives us Λ and U , we can use the Jarque-Bera test to check for normality. Since we now have filtered series we know come from the normal distribution, we can hope the test give reasonable results. We can see the values for the filtered series (\hat{h}_t and $\hat{\delta}_t$) compared to the generated series (h_t and δ_t) in Table 5. As we can clearly see all the generated

| | Generated series | | | Filtered series | | |
|----------|------------------|----------|----------|-----------------|----------|----------|
| | <i>JB</i> | <i>S</i> | <i>K</i> | <i>JB</i> | <i>S</i> | <i>K</i> |
| Series 1 | 2.79 | -0.0627 | -0.17 | 21.4 | -0.0823 | -0.562 |
| Series 2 | 0.702 | -0.0498 | 0.0362 | 10.8 | -0.0197 | -0.413 |
| Series 3 | 0.525 | 0.00173 | -0.0916 | 9.27 | 0.0328 | -0.379 |
| Series 4 | 2.03 | 0.0119 | -0.179 | 17.8 | -0.0195 | -0.533 |
| Series 5 | 2.12 | -0.0606 | 0.139 | 10.8 | -0.098 | -0.367 |

Table 5: *JB* test statistic with skewness *S* and kurtosis *K* for the scaled returns.

series all pass the test (not surprisingly since the observations are distributed by the generated series directly), but almost none of the filtered does. If we exchange the $\hat{\delta}_{1,t}$ and $\hat{\delta}_{3,t}$ with their generated counterparts, we would still reject three of the series (remember that *JB* should be less than 10.597). Even if we use all the generated δ_t with the filtered \hat{h}_t , we would still reject Series 1 under the normal hypothesis. This might be because the filtered series are smoothed versions of the series they estimate, since they are averaged over more than 10 000 simulations. It is worth noting that in the cases with real data we test the normal distribution against distributions with heavier tails, and the filtered series all gave lighter tails. Should we in the real case get heavy tails, e.g. $K > 0.6$, it would give us reasons to doubt the normality assumptions.

8 Stochastic trend model

8.1 Deriving the standard model

In our state-space time series approach we decide to use a model called the local linear trend model, to capture the trend-structure. The model is found in Durbin and Koopman (2001). In the basic form this model can be written for a time series of observations r_t

$$\begin{aligned} r_t &= \mu_t + \sigma_\epsilon \epsilon_t \\ \mu_{t+1} &= \mu_t + \chi_t + \sigma_\xi \xi_{t+1} \\ \chi_{t+1} &= \chi_t + \sigma_\zeta \zeta_{t+1}, \end{aligned}$$

where ϵ_t , ξ_{t+1} and ζ_{t+1} all are $N(0, 1)$ distributed error-terms. We will not use this simple model. First of all, our volatility-structure is much more complicated than a simple normal distributed error, and we use the stochastic volatility model to capture that. Secondly, it turns out that the underlying χ_t model does not follow the simple random walk the model above implies.

When we implemented the above model we got parameter estimations as well as filtered series just as in the stochastic volatility model. The filtered series captures what the underlying state-series should look like to give the observed series, and is not very sensitive to which type of model that links the observations to those states. So when we get a filtered series that is situated about zero, and spikes (or volatility) is clustered, we assume that the series have a autoregressive representation rather than a simple random walk. So we choose to let χ_t follow an AR(1) model with a zero mean value ($\nu = 0$ to use the notation from the stochastic volatility model), and if we for simplicity discard the stochastic volatility and assume the observed series is simply the trend series, we get

$$\begin{aligned} \mu_{t+1} &= \mu_t + \chi_t + \sigma_\xi \xi_{t+1} \\ \chi_{t+1} &= \phi \chi_t + \sigma_\zeta \zeta_{t+1}. \end{aligned}$$

8.2 Simulating from the standard model

When we want to simulate from this model we first need to estimate the parameters. We will use the same procedures as with the stochastic volatility model: put priors on the parameters and draw from the posterior (or actually full conditional) distributions. The model is not very different from the stochastic volatility so the full conditional distributions for the parameters will also be similar to those we found earlier. Both μ_t and χ_t given its parameters and earlier values are normally distributed. τ_ξ and τ_ζ with gamma priors have gamma posteriors due to the property of conditional distributions, and to ensure that ϕ is between $(-1, 1)$ we put

a $\beta(\gamma, \omega)$ prior on $\frac{\phi+1}{2}$. Since we again use the precision rather than the variance, we get

$$\begin{aligned} p(\tau_\xi | \dots) &\propto p(\tau_\xi | \alpha_{xi}, \beta_{xi}) p(\mu_2 | \tau_\xi, \mu_1, \chi_1) \cdots p(\mu_T | \tau_\xi, \mu_{T-1}, \chi_{T-1}) \\ &\propto \tau_\xi^{\alpha_\xi - 1} \exp(-\beta_\xi \tau_\xi) \times \tau_\xi^{\frac{(T-1)}{2}} \exp\left(-\frac{\tau_\xi}{2} \sum_{t=2}^T (\mu_t - (\mu_{t-1} + \chi_{t-1}))^2\right), \end{aligned}$$

so $p(\tau_\xi | \dots) \sim \Gamma(\tilde{\alpha}_\xi, \tilde{\beta}_\xi)$ where $\tilde{\alpha}_\xi = \alpha_\xi + \frac{T-1}{2}$ and $\tilde{\beta}_\xi = \beta_\xi + \frac{1}{2} \sum_{t=2}^T (\mu_t - (\mu_{t-1} + \chi_{t-1}))^2$. For simplicity we assume that we have observed μ_1 without error to overcome the border problem where $t = 1$.

Due to the similarity of χ_t to that of $h_{i,t}$ and $\delta_{j,t}$ the full conditional distributions of the parameters will be very similar to what was found for the stochastic volatility model. Without any need to go into further detail, we have that $p(\tau_\zeta | \dots) \sim \Gamma(\tilde{\alpha}_\zeta, \tilde{\beta}_\zeta)$, where $\tilde{\alpha}_\zeta = \alpha_\zeta + \frac{T}{2}$ and $\tilde{\beta}_\zeta = \beta_\zeta + \frac{1}{2} [(1 - \phi^2)\chi_1^2 + \sum_{t=2}^T (\chi_t - \phi\chi_{t-1})^2]$. We also have that $p(\phi | \dots) \propto \sqrt{1 - \phi^2} \exp\left\{-\frac{\tau_\zeta}{2} [(1 - \phi^2)\chi_1^2 + \sum_{t=2}^T (\chi_t - \phi\chi_{t-1})^2]\right\} \left(\frac{\phi+1}{2}\right)^{\omega-1} (1 - \frac{\phi+1}{2})^{\gamma-1}$.

It should not be too difficult to see that the full conditional of χ_t is given by

$$\begin{aligned} p(\chi_t | \dots) &= p(\chi_t | \chi_{t-1}, \phi, \tau_\zeta) p(\chi_{t+1} | \chi_t, \phi, \tau_\zeta) p(\mu_{t+1} | \mu_t, \chi_t, \tau_\xi) \\ &\propto \exp\left\{-\frac{\tau_\zeta}{2} [(\chi_t - \phi\chi_{t-1})^2 + (\chi_{t+1} - \phi\chi_t)^2] - \frac{\tau_\xi}{2} (\mu_{t+1} - (\mu_t + \chi_t))^2\right\}. \end{aligned}$$

With a bit of experience we can see that this is normally distributed, and we could reorder the parameters to find out what the new expectation and variance is. However, since we in the stochastic volatility model described an algorithm which uses algebra to find these parameters, we can use that directly here as well. We have, with an exception of ϕ , full conditionals we can draw directly from. We therefore use a Gibbs-sampler that stepwise updates each parameter and state space in order, except for ϕ which is updated with the same type of MH-algorithm with a random walk proposal distribution as the ϕ parameter in the stochastic volatility model.

8.3 Extending the model to encompass regimes with drop in prices

When we observe how the prices behaved in the aftermath of the financial crisis, there were obviously a steep drop in all the price series. We would like to implement a way to force the prices to fall when we enter a certain regime, and normalise when we exit that regime. There might be several ways to implement this, but the way we choose is to enforce a negative expected value on χ_t during the crisis regime. The way the local linear trend model works is that the trend at a certain

point in time is the sum of the last value, a random noise term and a hidden state space term χ_t . When this state space term is positive the trend is likely to increase, and when it is negative it is likely to decrease. So if we during a crisis regime force χ_t to be negative, the trend is likely to decrease throughout this regime. We can use the same notation as we did in the stochastic volatility section and introduce a time dependent parameter $\chi_0(s_t)$, where again we have $\chi_0(s_1) = 0$ during the normal regime and $\chi_0(s_t) = \chi_0$ during the crisis regime. Our new model for χ_{t+1} then becomes

$$\chi_{t+1} = \phi\chi_t - \chi_0 \cdot s_{t+1} + \sigma_\zeta \zeta_{t+1},$$

where $s_t = 0$ in the normal regime and $s_t = 1$ in the crisis regime. The full conditional distributions for ϕ , τ_ζ and χ_t needs to be updated by exchanging every appearance of $\phi\chi_{t-1}$ with $(\phi\chi_{t-1} - \chi_0(s_t))$. If we choose a normal prior for χ_0 , $N(\mu_{\chi_0}, \sigma_{\chi_0}^2)$, the full conditional distribution of χ_0 becomes

$$\begin{aligned} p(\chi_0 | \dots) &= p(\chi_0) p(\chi_1 | \chi_0(s_1), \phi, \tau_\zeta) \prod_{t=2}^T p(\chi_t | \chi_{t-1}, \chi_0(s_t), \phi, \tau_\zeta) \\ &\propto \exp\left\{ -\frac{\tau_{\chi_0}}{2} (\chi_0 - \mu_{\chi_0})^2 - \frac{\tau_\zeta}{2} [(1 - \phi^2)(\chi_1 - \chi_0 s_1)^2 I(s_1) \right. \\ &\quad \left. + \sum_{t=2}^T (\chi_t - (\phi\chi_{t-1} - \chi_0 s_t))^2 I(s_t)] \right\}, \end{aligned}$$

where again $I(s_t)$ is an indicator variable that equals 1 if $s_t = 1$ and 0 otherwise. Again, this is a normal distribution and the parameters of this distribution can be found in the same way all the other parameters of normal distributions are found in this thesis.

A pseudo-code of how our MCMC algorithm to find χ_t and its parameters can be seen in Algorithm 2 below.

Algorithm 2: MCMC algorithm for the stochastic trend model

Data: 5 dimensional vector of observations μ_t , $t = [1, \dots, T]$ and the number of runs

Result: The filtered series of χ_t and the parameter evolutions

begin

Set the values of the hyperparameters;

Define variables to hold the values of the parameters and χ_t , and add initial values to them;

Set the regime-indicating variable s_t and the tuning variable σ_ϕ ;

for i in 2 to number of runs **do**

 Draw $\tau_\zeta \sim \Gamma(\tilde{\alpha}_\zeta, \tilde{\beta}_\zeta)$;

 Draw $\tau_\xi \sim \Gamma(\tilde{\alpha}_\xi, \tilde{\beta}_\xi)$;

 Draw $\chi_0 \sim N(\tilde{\mu}_0, \tilde{\sigma}_0^2)$;

 Draw $\hat{\phi} \sim N(\phi, \sigma_\phi)$ and set $\phi = \hat{\phi}$ with probability

$\alpha = \min\{\frac{p(\hat{\phi}|\dots)}{p(\phi|\dots)}, 1\}$ **foreach** χ_t where $t = [1, \dots, T]$ **do**

 Draw $\chi_t \sim N(\mu_{\chi_t}, \sigma_{\chi_t})$

Calculate and return the mean value of χ_t^i for every i after convergence is reached and return the parameter evolutions;

end

8.4 Testing the model on generated data

To see what kind of trends this model produces and how well our algorithm replicates the state space and parameters, it can be interesting to test our model on generated data where we know the correct values. The chosen parameter values can be seen below, and the generated trend can be seen in Figure 8.1.

$$\frac{\sigma_\xi \quad \sigma_\zeta \quad \phi \quad \chi_0}{26 \quad 200 \quad 0.9 \quad 150}$$

With a much lower variance for μ_t than χ_t our generated μ_t -series is much smoother, and looks more like a real trend. This is why we choose a σ_ξ lower than σ_ζ . The prior parameters we choose are $\alpha_\xi = 0.001$, $\beta_\xi = 0.1$, $\alpha_\zeta = 0.0001$, $\beta_\zeta = 0.1$, $\omega = 30$, $\gamma = 1$, $\mu_\chi = 200$ and $\sigma_\chi = 100$. We let our code run for some iterations to decide the tuning parameter for the MH-algorithm for ϕ , and end up with $\sigma_\phi = 0.1$. We can run the code for as little as 5000 iterations, and still get reasonable convergence. The estimated parameter estimations can be seen in the table below with their standard deviations in parenthesis.

$$\frac{\hat{\sigma}_\xi \quad \hat{\sigma}_\zeta \quad \hat{\phi} \quad \hat{\chi}_0}{17.7(18.3) \quad 202(40) \quad 0.892(0.0109) \quad 155(19.3)}$$

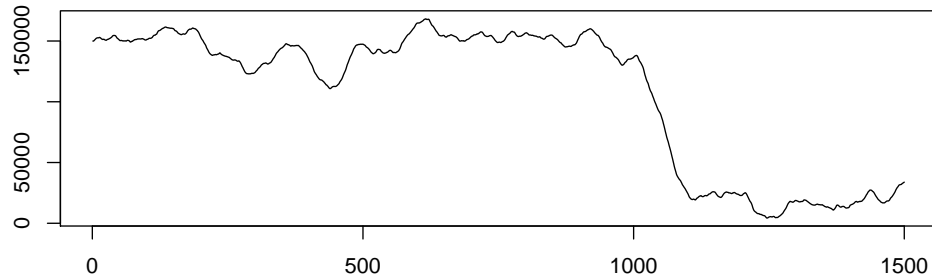


Figure 8.1: Generated trend series with a crisis regime from $t = 1000$ to $t = 1100$.

The variance of the estimated σ -parameters have large variances, and the estimated value of σ_ξ was a bit low, as also can be seen in Figure 8.3 which shows the parameter evolutions. As the figure shows there is a link between the variance of

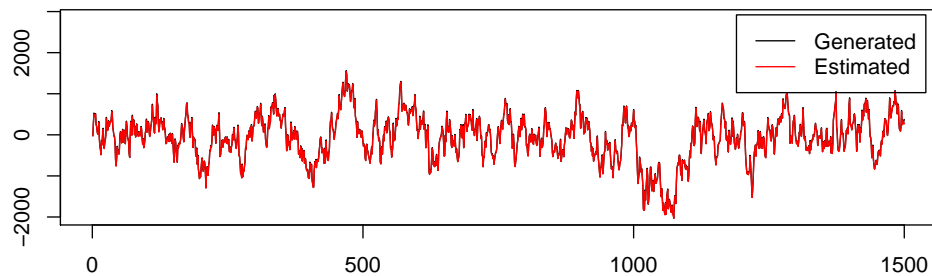


Figure 8.2: Filtered series of χ compared to the generated.

μ and χ , when one goes up the other one goes down. The filtered series of χ_t is really good, and it is impossible to see any differences as Figure 8.2 shows.

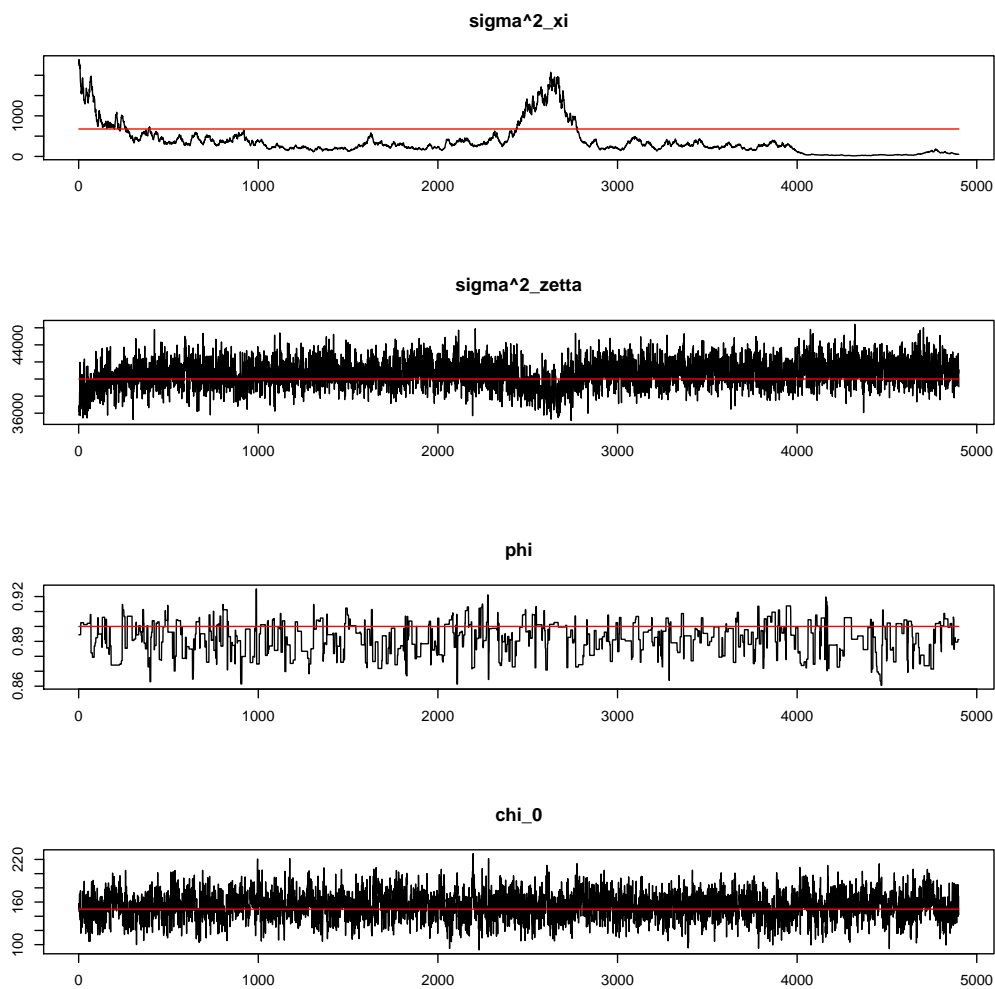


Figure 8.3: The parameter evolutions (with the first 100 values discarded). The red line shows the correct value.

9 Implementating the method on our data

First we implement our stochastic volatility model on the compounded return series of our data, then we implement the trend model on the observation series. In both cases we show the estimated parameters in tables, and the relevant filtered series in figures.

9.1 Implementing the stochastic volatility model on our data

To be able to use our stochastic volatility model, we need a fully specified data set without any NA-values. We could change our code such that when there are an NA-part in the data at time t , we would discard the likelihood contribution from the data. I.e. $r_t^T \Sigma_t r_t = 0$ if one of the $r_{i,t} = NA$ for $i = 1, \dots, 5$. But since we are missing all crude data up to the 28th of April 2006 and the set is fully specified from there on out, we only use that part of the data.

From some test runs and prior knowledge of stochastic volatility modelling of financial time series, we choose the following priors for our algorithm

$$\begin{array}{l|cccccccc}
 h & \mu_{\nu_0} = -8 & \sigma_{\nu_0}^2 = 10^2 & \mu_{\nu_1} = 2 & \sigma_{\nu_1}^2 = 0.5 & \omega = 1 & \gamma = 0.1 & \alpha = 100 & \beta = 10 \\
 \delta_1 & \mu_{\nu_0} = 0.5 & \sigma_{\nu_0}^2 = 5^2 & \mu_{\nu_1} = 0 & \sigma_{\nu_1}^2 = 1 & \omega = 1 & \gamma = 0.1 & \alpha = 100 & \beta = 0.1 \\
 \delta_{10} & \mu_{\nu_0} = 0.8 & \sigma_{\nu_0}^2 = 5^2 & \mu_{\nu_1} = 0 & \sigma_{\nu_1}^2 = 1 & \omega = 1 & \gamma = 0.1 & \alpha = 100 & \beta = 0.1 \\
 \delta_{2:9} & \mu_{\nu_0} = 0 & \sigma_{\nu_0}^2 = 5^2 & \mu_{\nu_1} = 0 & \sigma_{\nu_1}^2 = 1 & \omega = 1 & \gamma = 0.1 & \alpha = 100 & \beta = 0.1
 \end{array}$$

The reason we choose a larger μ_{ν_0} for $\delta_{1,t}$ and $\delta_{10,t}$ is because test-runs and observing the series have shown us that the freight rates and oil rates are generally correlated. We have also chosen γ and ω differently from the test case. The reason for this is because when series are uncorrelated, the ϕ -values of δ tends to be low and we use the same priors for the ϕ of the h -series as well as the δ -series. When γ and ω are so low, the resulting prior-variance is low, i.e. it is a more uncertain prior. But it has the same expected value as when $\gamma = 1$ and $\omega = 10$, so when we mix the series in the beginning of the algorithm, there is no difference. We hope this will lead us to capture the series that are correlated, and those that are not. As with the generated data we run the algorithm some times to decide the tuning parameters, and we end up with $\sigma_\phi = 0.1$, $\sigma_h = 0.5$ and $\sigma_\delta = 0.02$. We then let our code run for 25 000 iterations, and use the parameter evolutions to try to ascertain when we have reached convergence. We have reached convergence on most of the parameters after 15 000 iterations, and those parameters that do not converge are connected to the δ -series that are basically noise. The estimated values of the parameters after the 15 000 iterations can be seen in Table 6. From the value of ϕ it seems as though only the correlation series δ_1 and δ_{10} can be

| | σ^2 | ν_0 | ν_1 | ϕ |
|---------------|----------------------|-------------------|-------------------|----------------|
| h_1 | 0.135(0.0155) | -9.74(0.289) | 0.73(0.301) | 0.959(0.0111) |
| h_2 | 0.11(0.0115) | -7.8(0.219) | 0.784(0.291) | 0.948(0.0142) |
| h_3 | 0.312(0.0347) | -12.4(2.97) | 0.974(0.38) | 0.99(0.00536) |
| h_4 | 0.0995(0.00903) | -8.68(0.15) | 0.629(0.256) | 0.929(0.0161) |
| h_5 | 0.108(0.0109) | -7.57(0.0932) | 1.15(0.191) | 0.872(0.0297) |
| δ_1 | 0.00126(0.00019) | 0.413(0.798) | 0.00177(0.0339) | 0.986(0.00715) |
| δ_2 | 0.000944(9.28e - 05) | -0.0016(0.00617) | -0.0272(0.0159) | 0.535(0.163) |
| δ_3 | 0.000993(9.64e - 05) | 0.00561(0.00804) | -0.0347(0.011) | 0.556(0.214) |
| δ_4 | 0.001(0.000101) | -0.00664(0.0125) | 0.0155(0.0138) | 0.292(0.23) |
| δ_5 | 0.00086(7.36e - 05) | -0.00115(0.00374) | -0.00606(0.0152) | 0.299(0.155) |
| δ_6 | 0.00102(9.38e - 05) | -0.0152(0.0114) | -0.0108(0.0266) | 0.571(0.191) |
| δ_7 | 0.00101(9.5e - 05) | -0.0187(0.00794) | -0.000353(0.0202) | 0.732(0.133) |
| δ_8 | 0.000804(6.43e - 05) | 0.00166(0.00354) | 0.0159(0.0155) | 0.0686(0.209) |
| δ_9 | 0.000779(6.67e - 05) | 0.00206(0.00352) | -0.0198(0.0133) | 0.0159(0.117) |
| δ_{10} | 0.00103(0.000118) | 0.818(0.0225) | -0.019(0.0221) | 0.876(0.0244) |

Table 6: Mean parameter values of the state-space variables for the observed series with their standard deviations in parenthesis.

assumed significantly different from zero. If we observe the plotted filtered series, that strengthens this assumption. In Figure 9.1 and 9.2 we see respectively the filtered volatilities and the filtered non-zero correlation series. In Appendix B.2 all the correlation-series can be seen. Those non-zero correlation series are respectively the correlation between BSI and BPI, and between crude oil and bunker oil.

The volatility series h_t all have an increase during the financial crisis and we are not surprised when we observe which series that are the most volatile within their segment, when considering the observed prices and returns. The correlation series are somewhat more interesting. It seems the correlation between BSI and BPI is the most time-varying, and it is strictly positive. I.e those series have positive correlation all the time. The correlation between crude and bunker oil is the one that changes the least over time and we could say it is almost constant. It is also the series that have the largest ν_0 , so we would say the oil rates are constantly and highly correlated. If we look at this from the outside the correlation-structure does make sense. It is logical that the freight rates are correlated, and it is also easy to see that in the price history. It is also logical that the crude oil and its refined version, the bunker oil, are almost constantly correlated. What these plot don't capture is that sometimes when we run our algorithm there appears a slight correlation between the BSI freight-rate and the different oil-prices. That the oil-prices

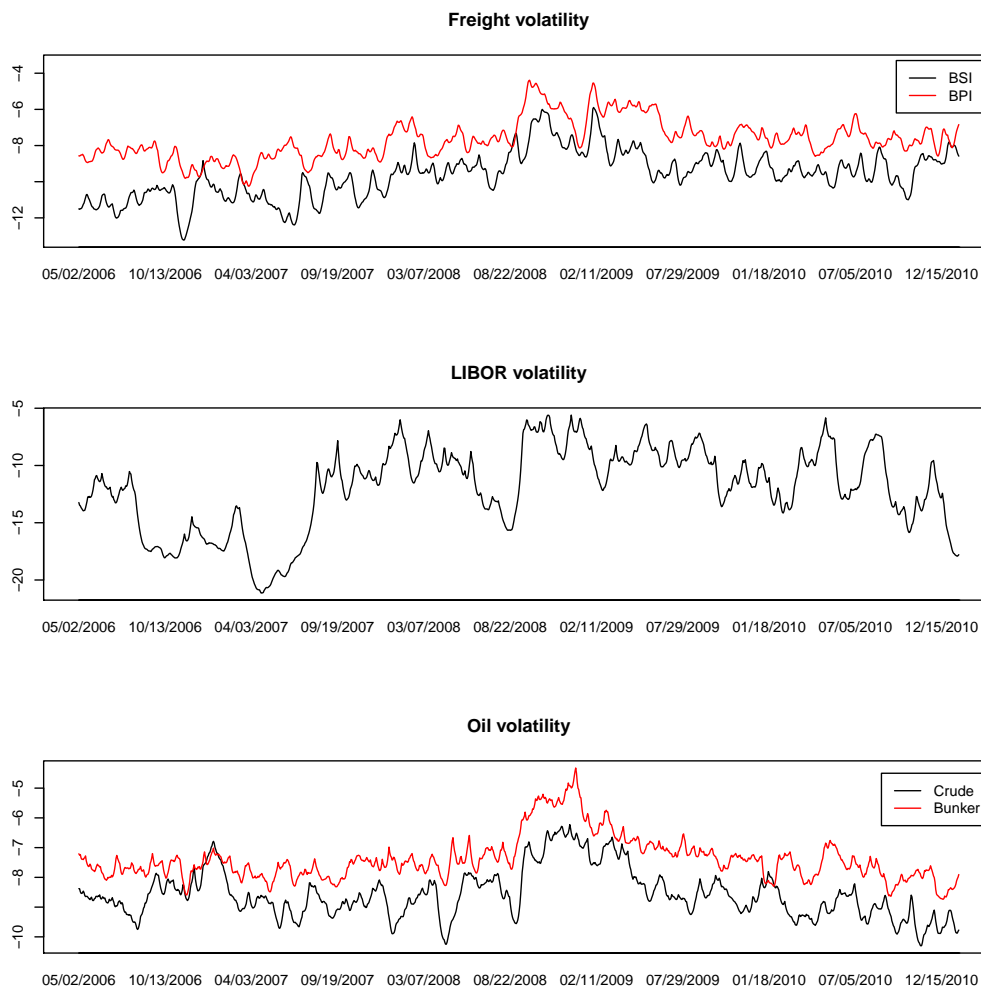
could affect the freight-rates is not unthinkable considering that bunker-prices is one of the largest expenses in shipping, but they don't seem to affect the short term price-structure of each other in any way worth mentioning.

It can also be interesting to see if the Jarque-Bera test indicates that the observed prices can come from a normal distribution after we try to remove the time dependency in the returns, as we did in the generated data case. As before we take the test on the scaled returns, $\Lambda_t^{-1/2} U_t^T r_t$, which under the stochastic volatility assumption should be identically normal distributed. The results can be seen in Table 7. The LIBOR rate are not surprisingly the one that deviated the most

| | <i>JB</i> | <i>S</i> | <i>K</i> |
|--------|-----------|----------|----------|
| BSI | 10.4 | -0.0385 | -0.454 |
| BPI | 44.5 | -0.0575 | -0.943 |
| LIBOR | 256000 | 3.4 | 71.8 |
| Crude | 4.55 | -0.0475 | -0.289 |
| Bunker | 3.04 | 0.0458 | -0.231 |

Table 7: *JB* test statistic with skewness *S* and kurtosis *K* for the scaled returns.

from the normality assumption, due to its heavy tails and unsymmetrical distribution. Neither the BSI or BPI can be said to come from a normal distribution, but the tails are too light rather than too heavy. If we plot the series they look more like autoregressive time series, and if we fit the transformed series of BSI to an AR(1)-model with the built-in R-function, it returns a model with $\phi = 0.764$ and $\sigma = 0.617$ and for BPI it returns a model with a even larger ϕ and lower σ . We should not be too surprised by this after looking at the returns of BSI and BPI. They do not change value “fast enough” to come from a normal distribution. There seems to be some memory, whereas the oil-rates have more the jagged returns that could hint of the normal distribution, and the *JB*-test confirms that the normal assumption holds for those rates.

Figure 9.1: Filtered series of h_t .

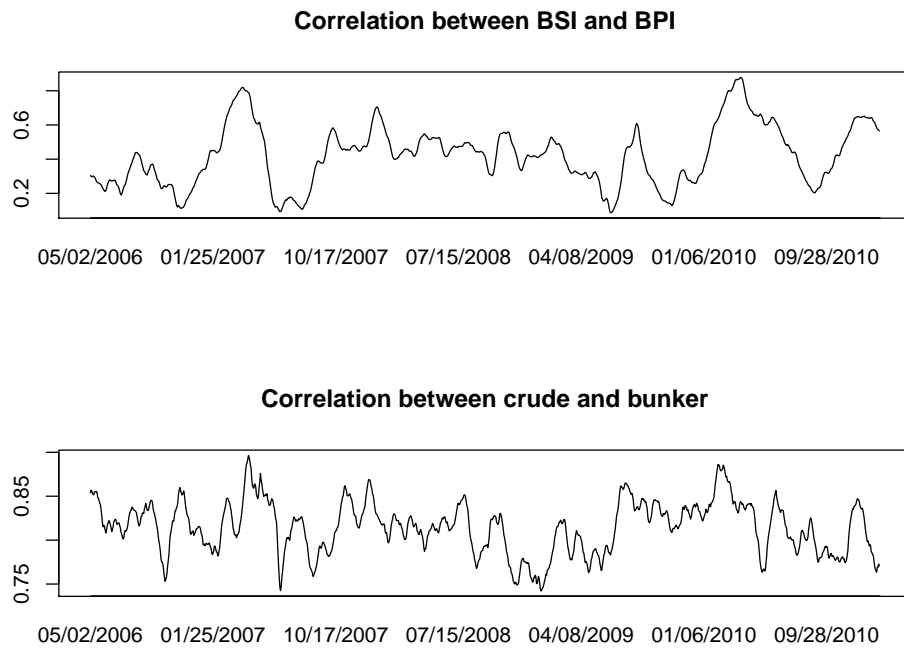


Figure 9.2: Filtered series of the non-zero δ_t .

9.2 Implementing the stochastic trend model on our data

As mentioned earlier in the thesis, the behaviour of the BSI and BPI are clearly linked, as are the behaviour of crude and bunker oil. One way to view this, and the way we choose in this thesis, is to assume that BSI and BPI have the same underlying trend and just their different volatility-structure give rise to the observed differences. Similarly we assume that crude and bunker oil have the same underlying trend, although on different scales. We assume no interaction between the trends of oil, freight and interest rates, except during an external crisis. In our model we assume that we observe the trend series directly, so this leaves us with the problem of which of the series BSI or BPI to choose as the correct trend. We simply choose the sample mean of the series $(BPI+BSI)/2$. We could do the same for the oil rates. Scale the series, find the mean of the scaled series and use that. Again we face the problem that one series is larger than the other, and since we have more data for the bunker oil series, we choose this as the correct trend. Since the LIBOR rate is assumed to be independent of the other series, the observed series is used as the observed trend directly.

Since there obviously are different scales between the different trends, we choose different priors for each as can be seen in the table below.

| | α_ξ | β_ξ | α_ζ | β_ζ | μ_χ | σ_χ^2 |
|---------|--------------|-------------|----------------|---------------|------------|-----------------|
| Freight | 0.001 | 0.1 | 0.0000001 | 0.001 | 200 | 100 |
| LIBOR | 1000000000 | 100000 | 1000 | 10 | 0.1 | 1 |
| Bunker | 1000 | 100 | 0.1 | 0.1 | 10 | 100 |
| Crude | 10000 | 100 | 0.1 | 0.1 | 10 | 100 |

The reason we also need to find the underlying χ_t process for the crude data, is simply to know how much to scale down the later simulations from the bunker model.

We run our code on the data to get the parameter estimations. The mean value of the parameters after convergence is reached is given in the table below with the parenthesis indicating their standard deviations.

| | $\hat{\sigma}_\xi$ | $\hat{\sigma}_\zeta$ | $\hat{\phi}$ | $\hat{\chi}_0$ |
|---------|----------------------|----------------------|-----------------|------------------|
| Freight | 2.38(3.13) | 278(53.1) | 0.851(0.0133) | 179(9.32) |
| LIBOR | 0.01($5.62e - 05$) | 0.078(0.012) | 0.575(0.0586) | 0.00519(0.00693) |
| Bunker | 0.316(0.0563) | 8.5(1.63) | 0.0674(0.0253) | 2.82(0.748) |
| Crude | 0.1(0.01) | 1.86(0.379) | -0.0316(0.0283) | 0.0348(0.163) |

The filtered series for χ_t with the corresponding trend can be seen in Figure 9.3, 9.4 and 9.5 for the freight, LIBOR and bunker series respectively.

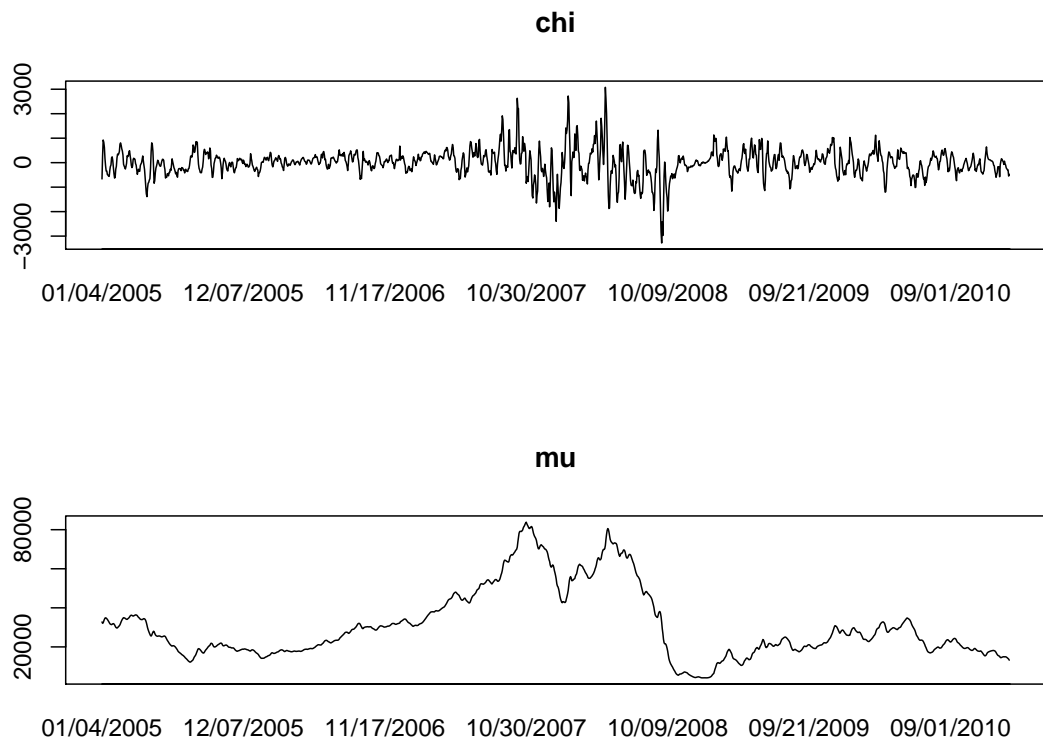


Figure 9.3: Filtered series of χ compared to the trend for the freight series.

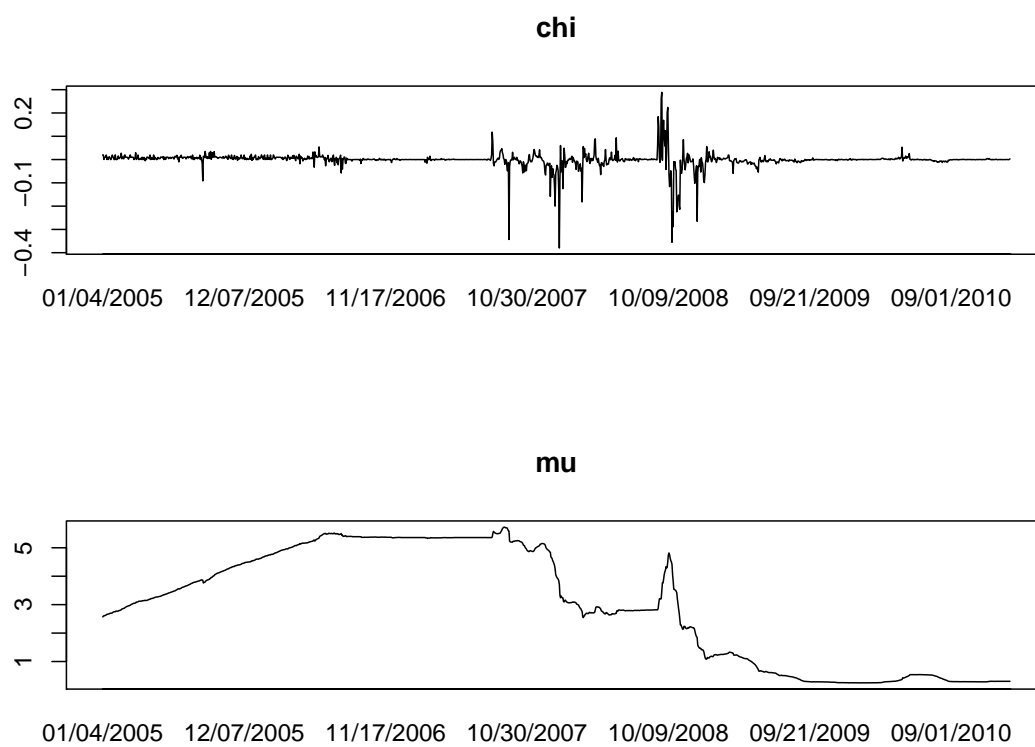


Figure 9.4: Filtered series of χ compared to the trend for the LIBOR series.

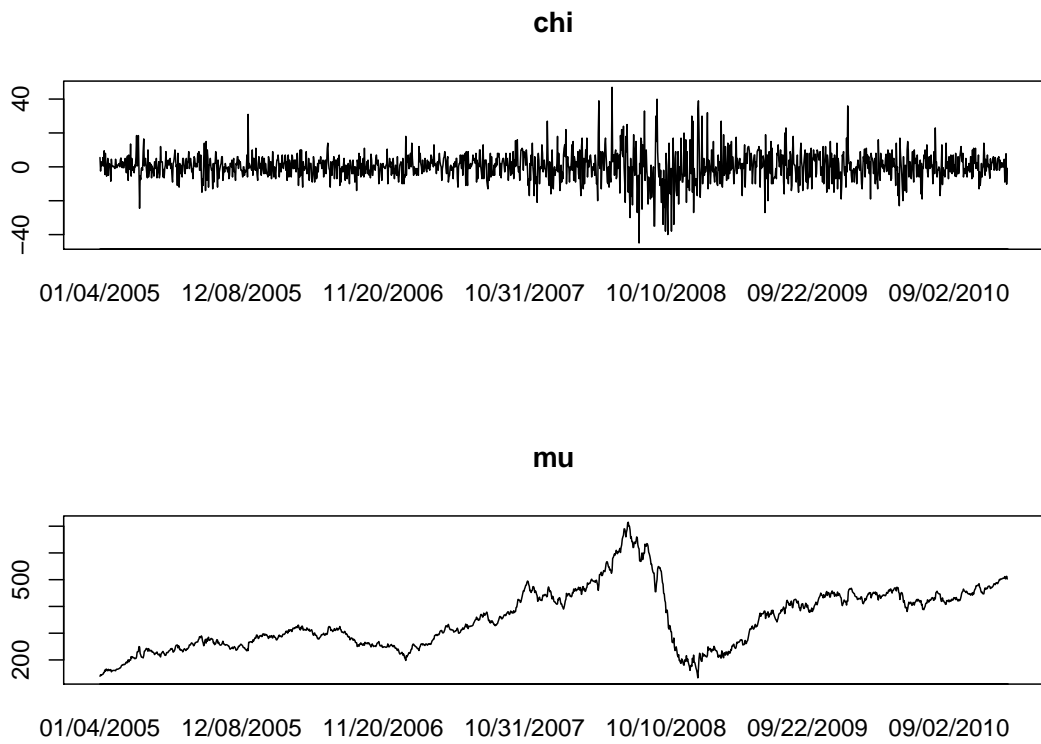


Figure 9.5: Filtered series of χ compared to the trend for the bunker series.

10 Bootstrapping the series and simulating from the models

10.1 How to bootstrap and simulate the volatility

When simulating the volatility from the stochastic volatility model, we first have to generate time series for the state-space variables h_t and δ_t by using the parameters estimated earlier. If the filtered series and estimated parameters indicate that some of the δ_t -series are just noise, we will assume they are constantly 0. In other words we will only estimate $\delta_{1,t}$ and $\delta_{10,t}$. When we have simulated a realisation of the state-space variables, we will use those to generate U_t and Λ_t and consequently Σ_t . We can now draw a return series, r_t , from the normal distribution $r_t \sim N(0, \Sigma_t)$. When bootstrapping the series, we basically use the standard moving block bootstrap. The exceptions being that we do not sample from the data we assume to be in the crisis-regime when we are outside of that regime, and that we “glue” the blocks in the different series together with an AR(1)-model where we use the same parameters as when we simulate the series. In the trend-model we introduce methods in the bootstrap samples to get more diversity, but we will not use those here. Since we bootstrap the variance of a multivariate normally distributed variable, the differences we introduce in the trend bootstrapping would not influence the variables here to a significant degree. However, what we will do here is to draw the same block numbers for all of the series. Say we divide our series in 1000 blocks and our bootstrap sample are the blocks (4, 252, 675, 353, 1, 994), we would first use the 4th block of $h_{1,t}$, the 4th block of $h_{2,t}$ and so on. The reason we do not sample the blocks of the different series independently of each other is to exploit possible correlations between the series which are not reflected in our model. Our model is defined under the assumption that there are no excess correlation between the series, i.e. the error-term in $h_{1,t}$ is uncorrelated with $h_{2,t}$ and all the other series. Bootstrap, however, is used in cases where we do not know the underlying model that well, so we will incorporate the possibility of such correlations in our moving block bootstrap.

10.2 Comparing with the stochastic volatility model

To simulate from the stochastic volatility model we chose overlapping block-bootstrap with block-size 90. The choice of block-size comes from simulating 1000 realisations of the different series h_t and δ_t , with block-size 10, 20, \dots , 100, calculating r_t and choosing the size that gave the smallest mean standard deviation in each time step. I.e. smallest $\frac{1}{1000} \sum_{i=1}^{1000} [\sum_{k=1}^K |r_{j,k}^i|]$, when $j=1, \dots, 5$. We chose this because we want a model with as small variance as possible. As mentioned above, we will

only bootstrap sample from the crisis-regime when we are in such a regime, and we will then simulate from the start of that regime and onwards. This will give us volatility series in crisis regimes equal to what we have found by our filter, but since this is only the variance of our returns, we will not get simulated series equal to what we have observed. A greater limitation is that we have a limit on how long our regime can last. This will be the same for the trend-model later, and how severe these limitations are for application purposes must be for the practitioner to decide. Series found by an r_t drawn from a bootstrap sample and simulated from our model can be seen in Figure 10.1 both with and without a crisis-regime. It can be difficult to ascertain the differences between these two procedures by only looking at one realisation, so the average of 1000 draws without a regime-change, but with the maximum and minimum of each of the series can be seen in Figure 10.2 As the figure clearly shows, the LIBOR rate diverges when we simulate from the model. Even though it might seem as though the LIBOR model is highly unstable, this is a little unfair. Should we view the 95% largest values at each time-step from the 1000 realisations the largest would only be 1.857, and for the 98% largest values the largest one is 10.45. Since the ϕ -value of $h_{3,t}$ (i.e of the LIBOR volatility) is so high (0.99), given enough realisations it is bound to reach extreme values, both large and small. One of the properties of the mean-statistic is that it is sensitive to outliers, and since we model on the log-scale our model is bounded below, so it is not surprising that the mean value of the LIBOR series get so high considering that the maximum value is almost 5 000 000. If we plot the median, a much more stable statistic, of the simulated LIBOR rather than the mean, the result get similar to what we observe for the freight series (see Figure 10.3). The solid lines are quite similar, but the envelopes for the simulated series increase more than the bootstrapped, indicating that those series have larger variance. So except for the few times where the simulated LIBOR-rate diverges, the values of the LIBOR realisation stay almost constant. Even the 75% largest values in each bootstrapped sample does not increase above 0.05 of its initial value. This is because of the low initial value (0.30438); a single (and non-extreme) realisation when the LIBOR initial value was 4 for example, had a maximum value of 6.5 and a minimum of 3.5 over 722 days.

We can also compare the series when the volatilities have been simulated from model against the bootstrapped ones when we also have a regime-change. In Figure 10.4 we see the average of 1000 realisations over 722 days, when we have a regime-change at 400 and 500 (i.e a crisis-regime that lasts from 400 until 500). Again the LIBOR series where the volatilities are simulated from a model diverges. Should we again use the median of those series, rather than the mean, we would reach the same conclusion as we did above. The resulting model would then be similar to what we observe for the freight series (that is, similar to the freight se-

ries with the crisis-regime, not the ones without). As we can see the bootstrapped values have a larger increase in variance during the regime. This only confirms what we observed on the generated data in the stochastic volatility section. Our algorithm underestimate the value of ν_1 , so a practitioner should view the filtered series and choose a ν_1 that would better fit the increase in volatilities we observe there. It is also clear that since the volatility just increases during a crisis-regime, the prices might just as well rise as fall. This is not a result we would like to get when we simulate a crisis similar to the one we had in 2008.

The properties we wanted from our model was prices that are highly correlated within each segment and possibly across segments, and a possibility to enforce a regime-change that gave a drop in prices across all segments. We have not found any definite correlation across segments, and as Figure 10.5 shows, the correlation-structure is not strong enough to give realisations similar to the historical observations. This, in addition to the price-drop property, lead us to use the defined stochastic trend-structure.

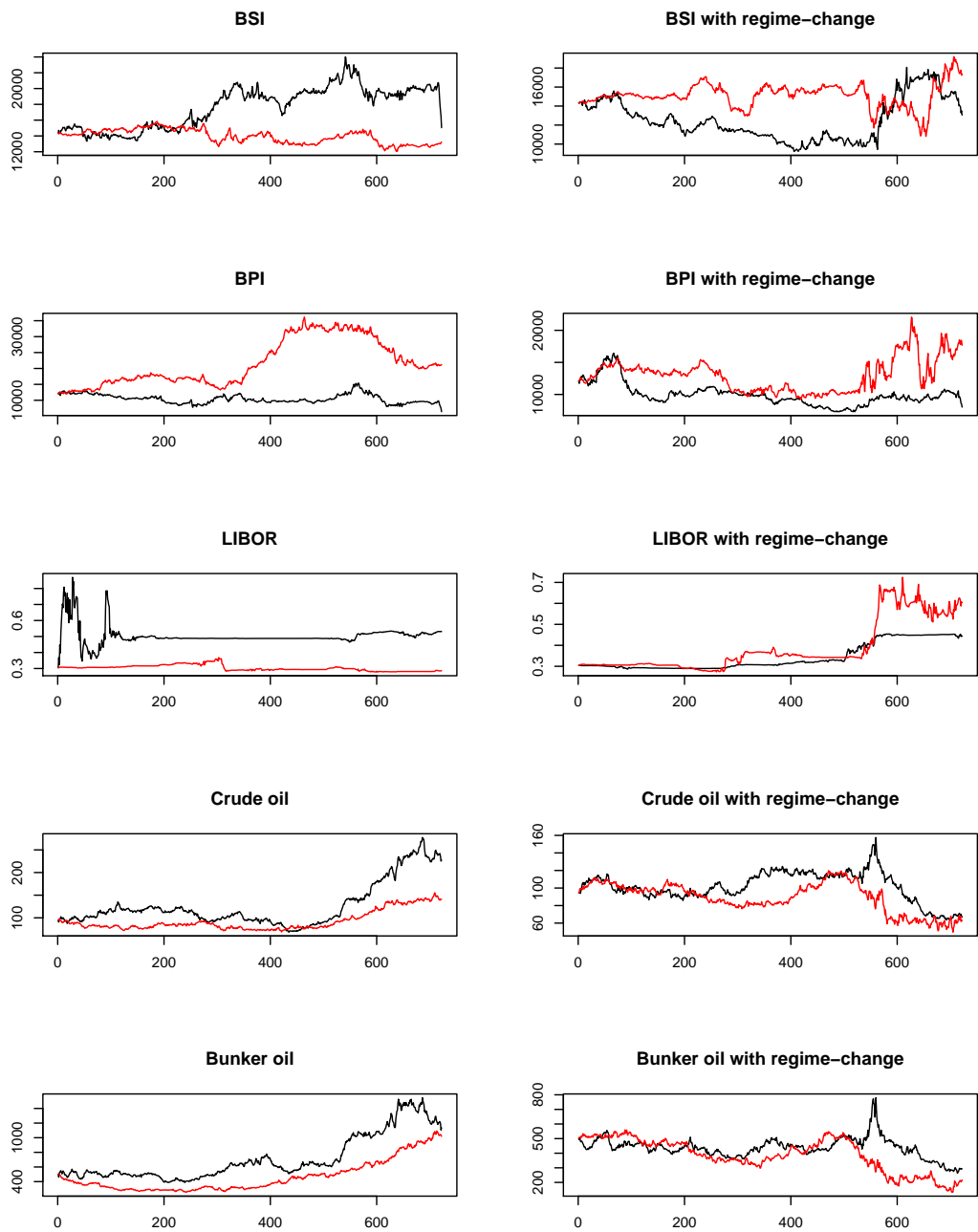


Figure 10.1: A single realisation, where black is simulated from the model and red is the bootstrap equivalent.

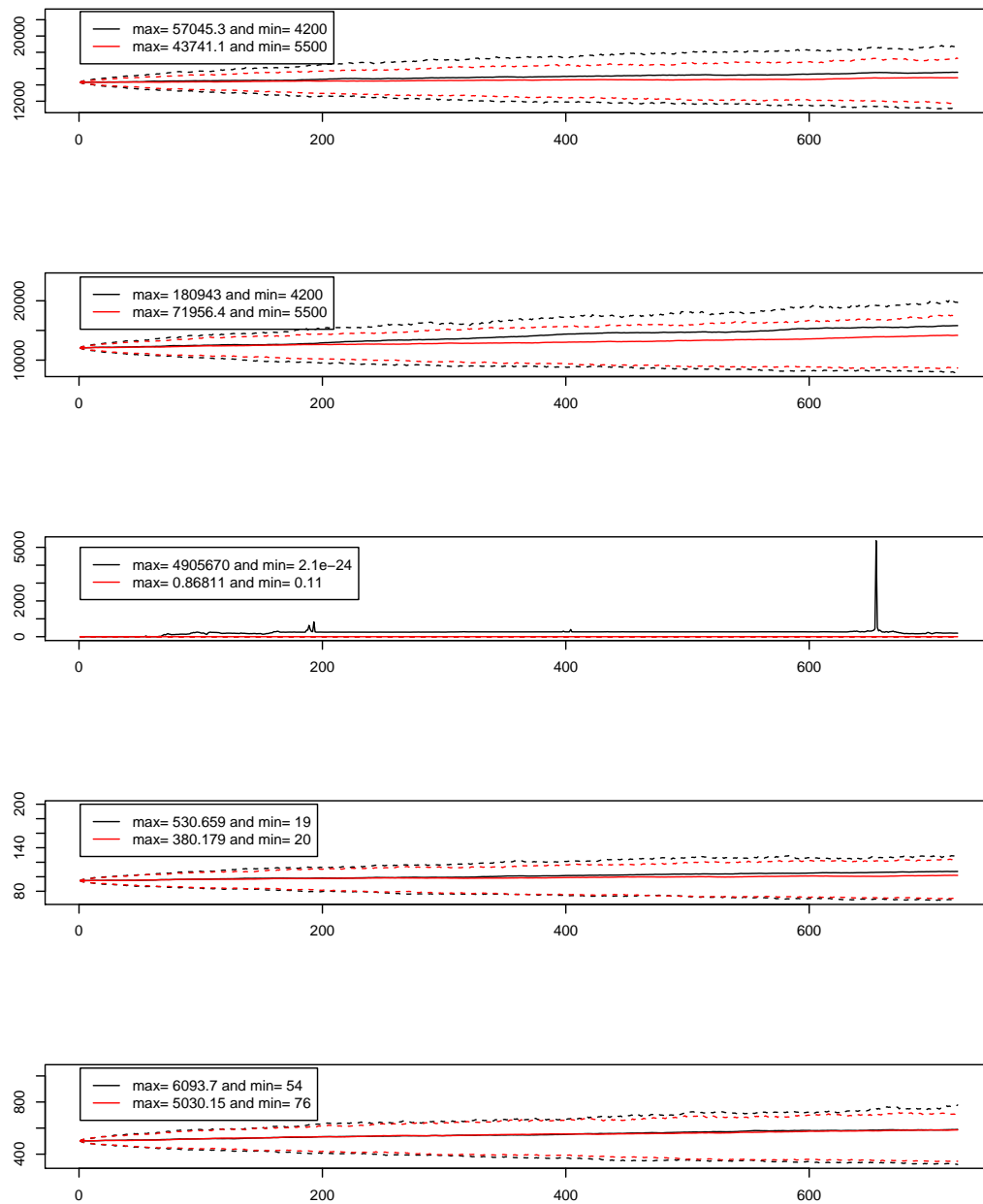


Figure 10.2: The average series over 1000 runs, where the volatility of the black is simulated from the model and red is the bootstrap equivalent, with 75% envelopes (dotted lines).

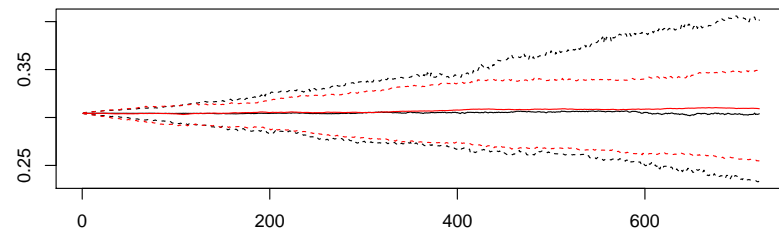


Figure 10.3: The median of the LIBOR series with volatilities simulated from the model (black) and the mean of its bootstrap equivalent (red), with 75% envelopes.

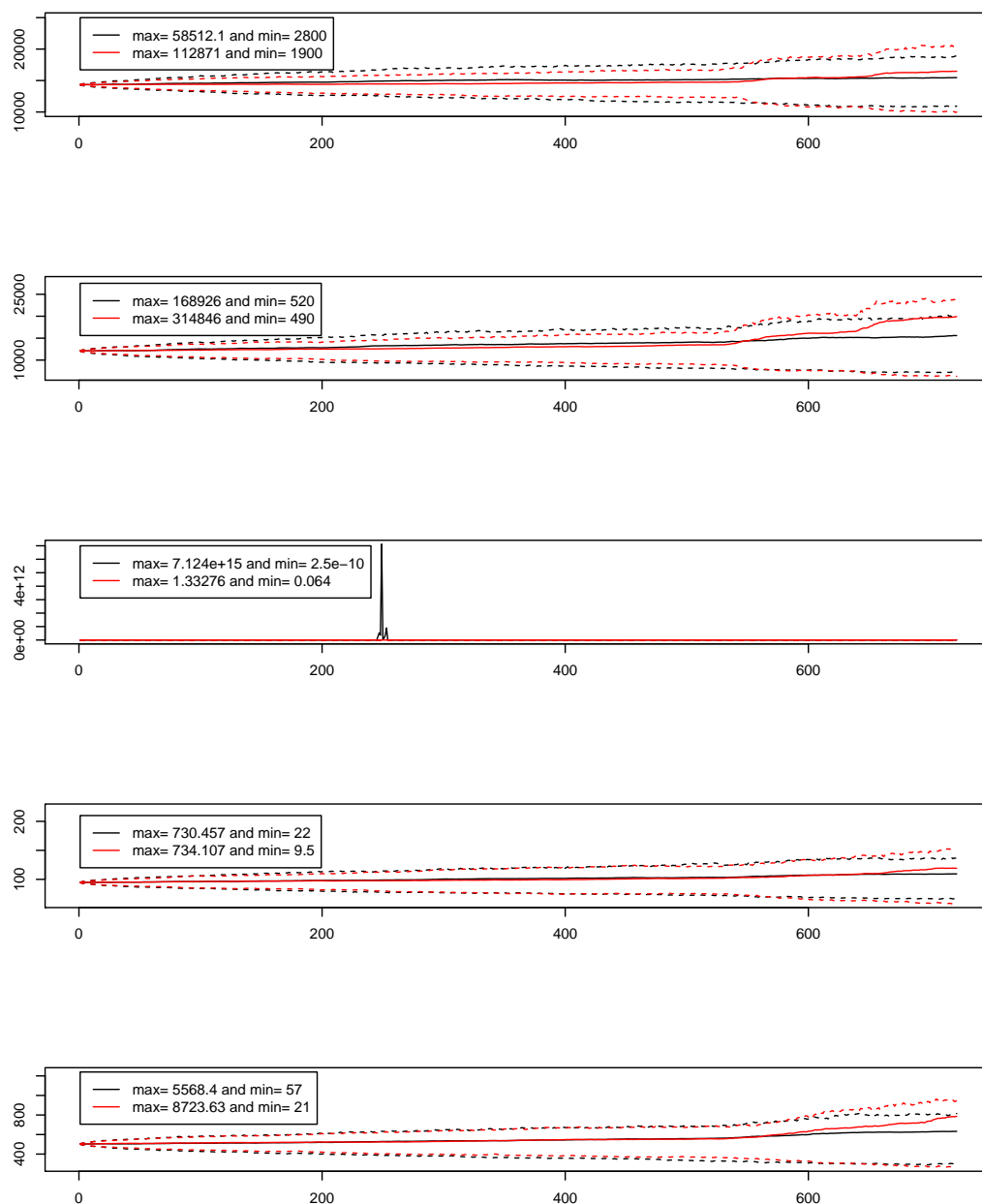


Figure 10.4: The average series over 1000 runs, where the volatility of the black is simulated from the model and red is the bootstrap equivalent, with 75% envelopes (dotted lines).

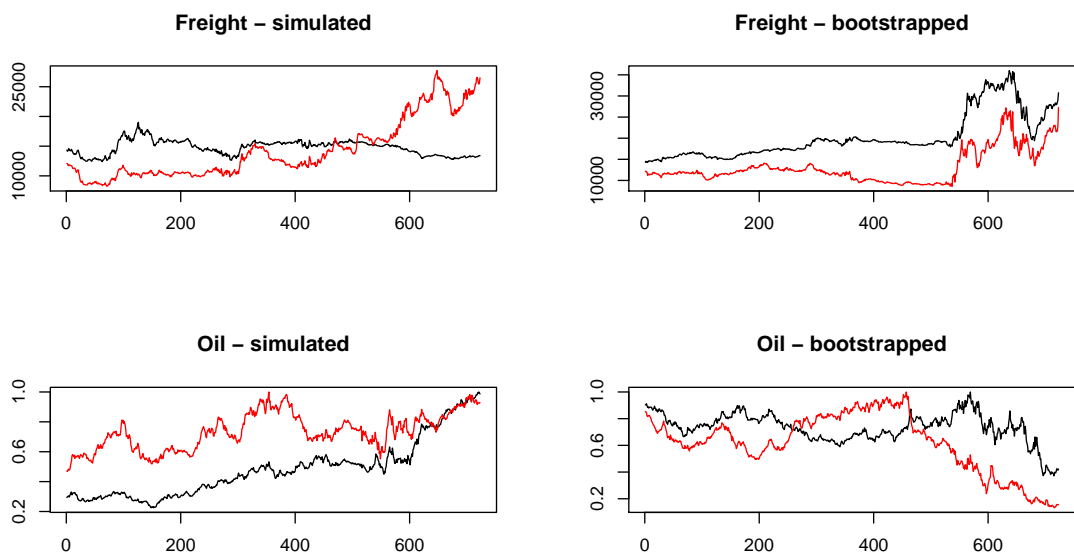


Figure 10.5: A realisation showing the correlated freight rates and the correlated (and scaled) oil rates compared against each other.

10.3 How to bootstrap and simulate the trend

When simulating the trend of financial time series there are some pitfalls one needs to be careful about, and the biggest of those is keeping the trend positive. Since the price series will not be negative (or at least so rarely that we can assume it never happens) and the prices are simulated about the trend, we need to keep it positive. There are several ways to do this, and it is difficult to say which is best. One simple way would be to simulate the whole series, discard the simulation if we get a negative value, and keep going until we get a trend with only positive values. The problem with such an algorithm is that we might weight series that have an upwards trend too much, and on average get too many series that takes large values. An other way would be to simulate in smaller blocks, check in each block if we have a negative value and discard the ones that do. Again we might get a bias for larger series, but not in the same extent as the first. The problem here is with the bootstrap case. There might be a bootstrap block that ends with a value such that no other bootstrap block can give a positive value, and the aforementioned algorithm would iterate through the possible values indefinitely without finding an acceptable block. The way we chose to overcome the problem in this thesis is by noting that a time series $\dot{x}_t = x_t - \nu$ for $t = 1, \dots, T$ and $-\dot{x}_t$ have the same properties when \dot{x}_t is stationary or at least without a trend. So we simulate χ_t in blocks χ_t^i , if $\mu(\chi_t^i) < 0$ in a block we choose either χ_t^i or $-\chi_t^i$ depending on $\max(\min(\mu(\chi_t^i), \mu(-\chi_t^i)))$. This will ensure that we do not have trends that tends upwards like the first case, and we will not have the problem with an infinite loop as in the second case. The trend can however get negative, although it is not negative for a “long time“. But since the trend should not be negative we fixate a lower limit μ_l , say 0.1, and return $\max(\mu_l, \mu_t)$. This might lead to parts where the trend is linear on μ_l for short times, but since we will simulate the volatility about the trend later, the final series will not be a line. This might not be a perfect solution, but it solves the problems without adding too much upward bias to the trends, and it will also allow us to increase the downward slope in the crisis-regime, should we want to, without worrying about the price going really negative. They will simply reach the lower limit, and stay there until the regime changes again. There is an additional change we introduce to the bootstrap sampling. We do not want to simply replicate history, so we want to add additional variability to our samples. As mentioned above, in our case there are little noticeable difference between $\chi_{1:T}$ and $-\chi_{1:T}$, the flipped series. There are actually also not much difference between $\chi_{1:T}$ and the reversed series $\chi_{T:1}$, and by extension neither between $\chi_{1:T}$ and $-\chi_{T:1}$. This can be observed in Figure 10.6, where we have plotted a block from the freight series with its flipped, reversed and flipped and reversed series. Had we not known which was the correct one, guessing would have been difficult, even after looking at Figure 9.3. And all the versions have the same

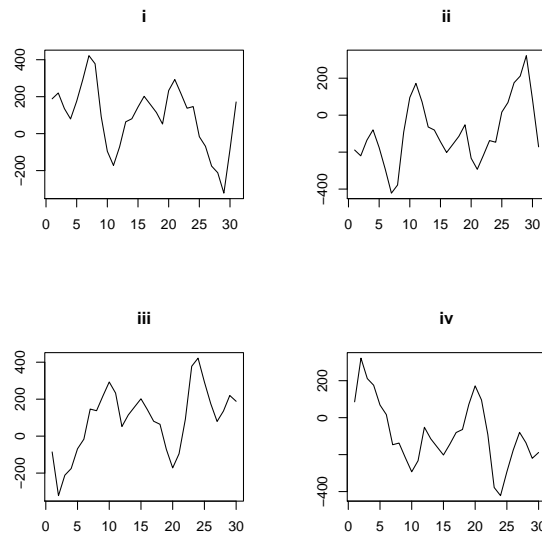


Figure 10.6: A freight χ_t -block with its ordinary, flipped, reversed and flipped and reversed (respectively i, ii, iii and iv) versions.

characteristics, when thinking of clustering effect and mean (had the mean value of the series been different from 0, we would have had to subtract it from the block before transforming, and added it back afterwards). To add additional variability we therefore use one of these four transformations when we bootstrap χ . Now we have the possibility to draw steep increases in the LIBOR rate, something we couldn't when we only simulated from the historical data directly since its history only have gentle upwards trend, but a steep downwards (outside of the crisis-regime). Because we do not want to draw blocks from the crisis-regime when we are not in the regime, we draw from the blocks up to the regime starts, and from the regime ends.

So to simulate from the model we first simulate an n long χ_t -block χ^i (for example $n = 30$) and calculate $\mu(\chi^i)$ and $\mu(-\chi^i)$. If $\mu(\chi^i) < 0$ choose the μ with the largest minimum-value, and repeat the procedure with a new block until the simulated series is long enough, then return $\max(\mu_t, \mu_t)$. It is a little more complicated with the bootstrap sample, but the algorithm is sketched in Algorithm 3 (when we in the algorithm use the ceiling delimiter, $\lceil x \rceil$, it denotes the integer value equal or greater than x , so $\lceil 3.6 \rceil = 4$).

Algorithm 3: Bootstrap algorithm for the stochastic trend model

Data: The filtered series of χ_t with estimated parameters, the length of the output series K and regime indicator s_t

Result: Estimated μ_t for $t = 1, \dots, K$

begin

Divide the series (excluded the crisis-regime) in blocks of length l ;

Draw $\lceil K/l \rceil$ numbers J from 1 to #blocks with replacement **for** j in J

do

while $s_t == 0$ **do**

Either flip, reverse, both or do nothing with the block χ^j with probability 0.25;

Generate the “glue” between the transformed series and the last block, add it to the transformed series and call the result $\hat{\chi}^j$;

Calculate both $\mu(\hat{\chi}^j)$ and $\mu(-\hat{\chi}^j)$ **if** $\mu(\hat{\chi}^j) < 0$ **then**

└ Add $\max(\min(\mu(\hat{\chi}^j), \mu(-\hat{\chi}^j)))$ to μ_t

else

└ Add $\mu(\hat{\chi}^j)$ to μ_t

while $s_t == 1$ **do**

Choose χ from the beginning of the crisis-regime and onwards;

└ Calculate $\mu(\chi)$ and add it to μ_t

Return $\max(\mu_l, \mu_t)$;

end

When we generate the AR-model to glue the different blocks together, we use the same parameters, ϕ and σ , as we use when we simulate the series to get as smooth a transition as possible.

10.4 Comparing with the trend model

To bootstrap from the trend series, we use overlapping block-bootstrap with blocks of 80 from χ_t for freight and LIBOR and 90 for bunker. The choice of block-size comes from simulating 1000 realisations with block-size 10, 20, \dots , 100, and choosing the one with smallest drift. To achieve this we choose the one that have the lowest absolute mean deviance from the start value $\frac{1}{1000} \sum_{i=1}^{1000} [\sum_{k=1}^K |\mu_k^i - \mu_0|]$, when μ^i is the i th trend realisation and μ_0 its start value. We choose this because we want our model to be as close to the initial value as possible on average (i.e. not diverge away from the initial value).

As mentioned above we sample from the series outside of the crisis, and during a crisis regime we only sample from the observed crisis. This will make every simulated crisis-regime identical to what was observed in 2008, and there will also

be a limitation on the length of the crisis-regime. It cannot be significantly longer than what have been observed. A bootstrap draw and a simulation, both with and without a crisis-regime, can be seen in Figure 10.7. It can be difficult to

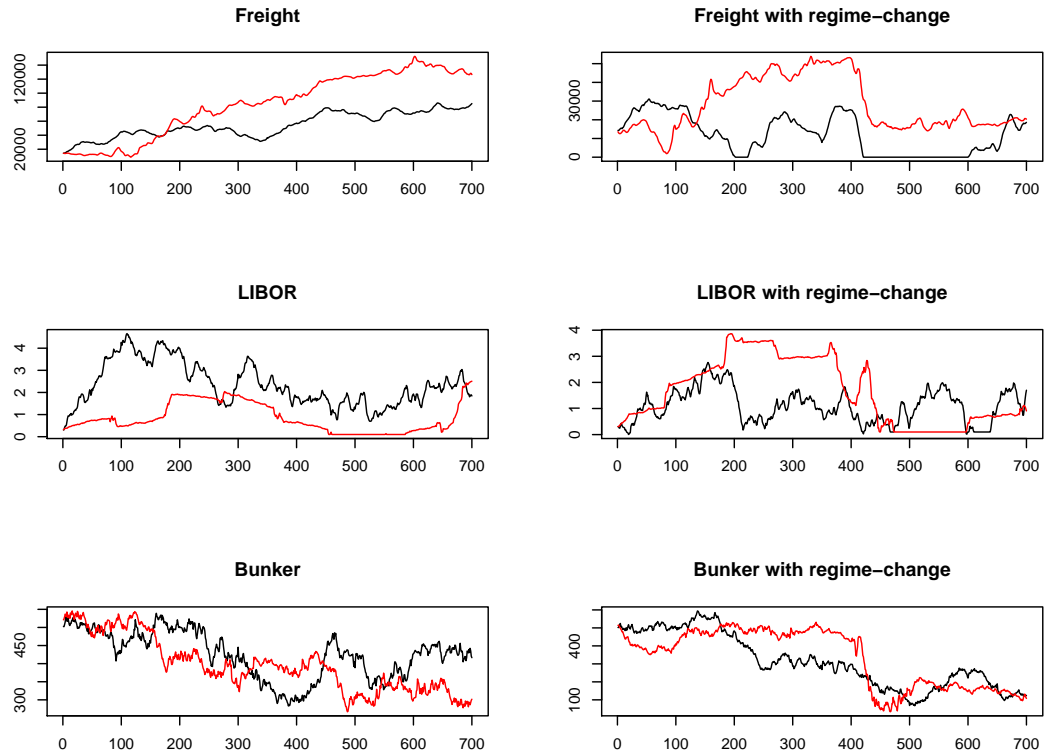


Figure 10.7: A single realisation, where black is simulated from the model and red is the bootstrap equivalent. The models with regime-changes have a crisis-regime from $t = 400$ to $t = 500$.

ascertain what the big differences between these two models are by only looking at one realisation, so in Figure 10.8 we can see the average of 1000 draws with the minimum and maximum value of the series (that is the smallest minimum and largest maximum of the 1000 series, not of the average series). Here we have also added the 75% largest generated value at each point in time and the 25% lowest, to see how the range including 50% of all the observations looks like for the different procedures. The main difference here is obviously that the bootstrap variant have a smaller variance, and therefore does not rise as fast as the simulation model does over the time-span. And, except from the LIBOR series, the range is quite similar for the two other series. The average value in the model should actually

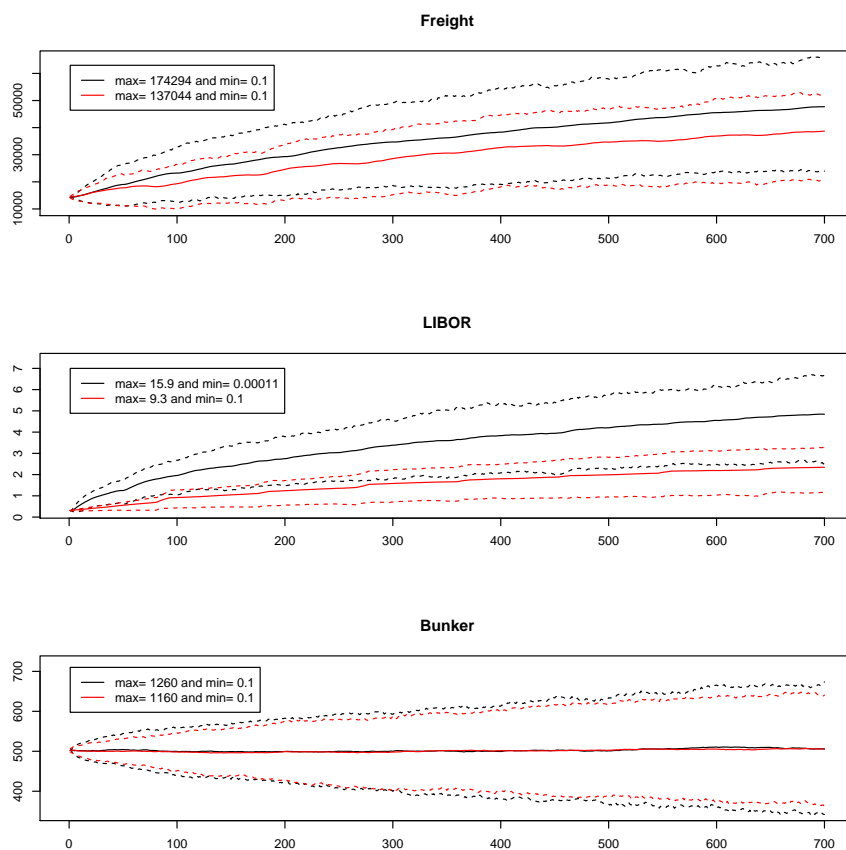


Figure 10.8: The average trend series over 1000 runs, where black is simulated from the model and red is the bootstrap equivalent, with 75% envelopes (dotted lines).

neither rise nor fall much since χ_t is situated around zero. However, since the series is bounded below by zero and not bounded above, it will on average rise. The larger the variance, the more often it will take large values, and the steeper the average rise will be (i.e. the trend of the trend). It should also be noted that the initial value will influence this. A small value (relative to the series) will cause the model to start closer to its “floor“, and therefore give the trend an upwards sloping trend whereas a large initial value will not do so in the same extent. This is observed in the bunker series, where the large initial value keeps the average trend about its initial value. In fact, should we start with a relative low initial bunker value, say 100, the resulting average series would have a slightly upwards trend and look very much like what we observe for the freight series. We can also see

that the maximum values of the simulated series are larger than the bootstrap ones in every series, whereas the minimum is about the same (i.e. both methods reach the “floor“). In the case with the crisis-regime, the main difference is the shape of the regime as we can see in Figure 10.9. It is now obvious that the bootstrap variant returns the same shape of the crisis, and the simulation variant does not. Here the differences in the variances do not influence the model in the same way it did in the case without the regime change because the regime itself influences so much. We can also see that the value of χ_0 for the LIBOR rates are too low.

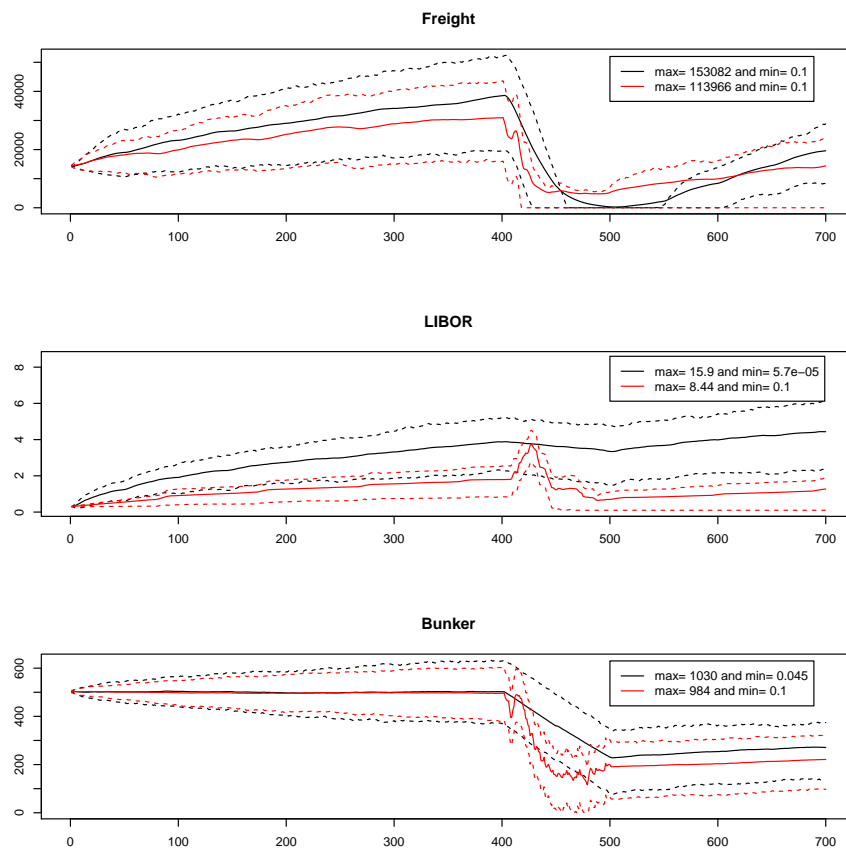


Figure 10.9: The average trend series over 1000 runs, where black is simulated from the model and red is the bootstrap equivalent, with a crisis-regime from $t = 400$ until $t = 500$ and 75% envelopes (dotted lines).

We should choose a value that's higher, to get results that better fit the observed fall in rates during the financial crisis in 2008. We might also increase the value of χ_0 for the oil rates.

10.5 Combining the models

We can now combine what we have found above to simulate realisations with both stochastic trend and stochastic volatility. First we simulate the trend as described in the section above. We choose the same trend for BSI and BPI, and we use a scaled version of the bunker trend as the crude trend. We scale the bunker χ_t by the $\frac{\max(\chi_{\text{crude}})}{\max(\chi_{\text{bunker}})}$ -ratio from the filtered series, and then use this to create μ_t for the crude series. We now find the compounded returns of the trend in the usual way (this one of the reasons we need the trend to be positive, since the logarithm of a negative number doesn't exist). When we denote the compounded return of the trend series at time t as m_t we can find a realisation of r_t by drawing it from a normal distribution with m_t as its expected value; $r_t \sim N(m_t, \Sigma_t)$, when Σ_t is the simulated correlation matrix found in the way described in a section above. As before we can either do this by simulating from the model, or by bootstrapping from filtered state-spaces. A single realisation of this can be seen in Figure 10.10 for the case without a crisis-regime. As we have seen in the previous section, it is the trend-model that affect the realisations the most, since they are the part with the largest variance. It can therefore be interesting to see how the different volatility sampling behaves when we have given a common trend. It turns out that there is basically no difference between bootstrapping the volatilities and simulating from the model when we have given a trend (with the exception of the LIBOR rate which still can give extreme values on occasions when the volatility is simulated). An example of this is given in Figure 10.11, where we have given the mean of 1000 realisations of the BPI rates with a common trend and volatilities both simulated and bootstrapped together with the 75% largest and 25% smallest value at each time-step for both series. The figure shows quite clearly why stochastic volatility model with a trend can outperform models without trend. Since the volatility is modelled on a log-scale, large values will give large variance on the simulated prices. When a simulated trend tends downward, it will force the drawn returns to lie close to the trend. Therefore when a trend tends upwards it can lead to large differences between the prices within the same segment, but when or if it tend downwards the prices will come together again. This actually reflects what we see for the freight rates in the historical data (in Figure 2.1), although not so much for the scaled oil prices. In Figure 10.12 we have plotted *one* realisation of bootstrapped volatility without trend, with an upwards trend and with a downwards trend. Since this figure have the same volatility-structure about different trends, it shows the differences very clearly. It can also be noted that since the initial value was quite low, the range of the model without a trend is much smaller than those with a trend.

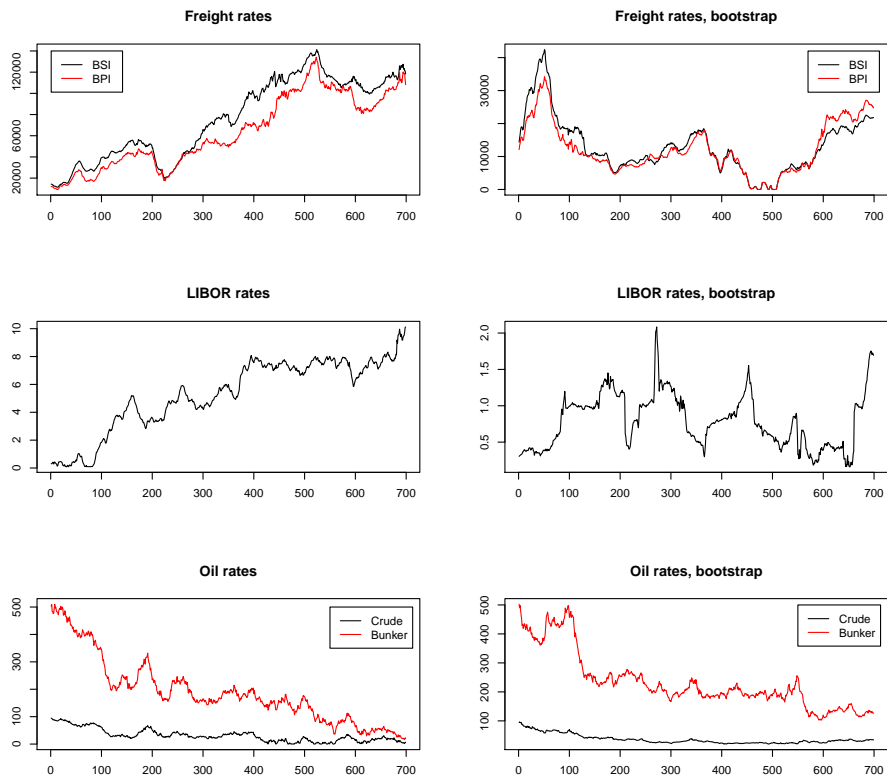


Figure 10.10: A single realisation of the full model with both stochastic trend and volatility.

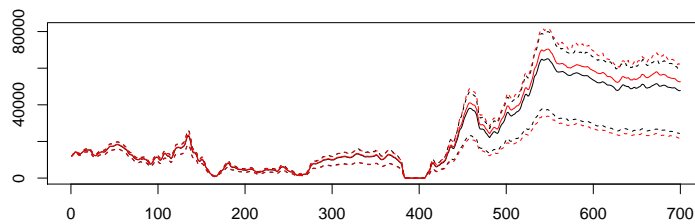


Figure 10.11: The mean of 1000 realisations of BSI with simulated (black) and bootstrapped (red) volatilities around a common trend, with 75% envelopes.

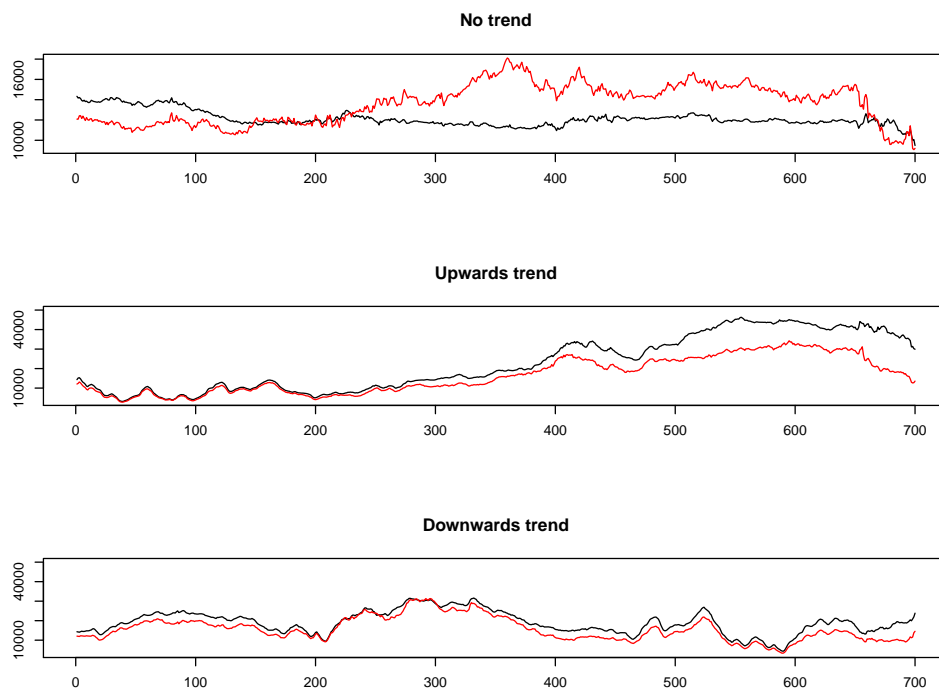


Figure 10.12: One realisation of a fixed volatility-structure without a trend, with an upwards trend and with a downwards trend. As usual the black series is the BSI and the red is the BPI

11 Reviewing the model and the results

First we will review our model and the difference between bootstrapping and simulating the underlying state-space, then we will recommend improvements and further work.

11.1 About the model

The main problem in this thesis was to describe and implement a model to generate reasonable realisations of five given price series. The realisations should exhibit the same internal correlation-structure as the historical prices, and it should be possible to enforce an outside *crisis-regime* where all the prices would plummet. To achieve this we have used the state-space approach to time series, and divided the model in two parts; a stochastic trend and a stochastic volatility part. When implementing the model we have also considered two different ways of doing this; by simulating directly from the model we have chosen for the underlying state-space, and by bootstrapping those state-space time series.

Our main hope was that the stochastic volatility model was enough to describe the prices' behaviour, but it was sadly not the case. The filtered correlation-structure was not strong enough to mirror what we see in the historical case, so we needed a common trend-structure to achieve this. It then became most natural to use the trend-structure to model the fall in prices during a crisis-regime, even though it would have been possible to do so without modelling an explicit stochastic trend (for example by defining $r_t \sim N(S_t, \Sigma_t)$, where $S_t = 0$ outside of the crisis-regime and $S_t = a$ for some $a < 0$ inside). That our stochastic volatility model did not return a strong enough correlation-structure might be because the data do not exhibit it, but it might also be because our model underestimate it. When we view how our model were able to filter out the correlations for the series of the generated data it did not perform as well as it did for the volatilities, and this could lead us to wonder if especially $\delta_{1,t}$ have been underestimated by our algorithm. It should also be noted that both oil rates were well within the normal distribution assumption on their returns.

By implementing the stochastic trend model we are able to get realisations more in line with historical observations. A problem with the trend model is how to ensure that it stays strictly positive, but by using a stochastic trend-model we also get some extra benefits. One problem with the stochastic volatility model was that when the prices started at sufficiently small values, it had problems with drawing larger numbers. We can consider the LIBOR rates which has its last observed price at about 0.3. To be able to get a value of, for example, 1 it needs an increase of 1.20 on the log scale ($r = \log(\frac{1}{0.3}) = 1.2$). Of course it doesn't need to jump from 0.3 to 1 in one time step, but it still need the accumulated increase of 1.20 (i.e.

$\sum_{t=1}^T r_t = 1.2$). This can be difficult for a model that is basically a series of normally distributed variables with an average variance of $\exp(-12.4) = 4.12 \cdot 10^{-6}$. And since the model is on a log scale, it is more difficult to go from 0.3 to 1 than from 1 to 1.7 ($\log(\frac{1.7}{1}) = 0.53$), even though the increase in the rate is the same (0.7). The stochastic trend model, however, does not draw log-scaled variables from the normal distribution, and will not suffer from this.

When it comes to the difference between bootstrapping and simulating from the underlying state-space, there are hardly any differences on average for the stochastic volatility model. As the Figure 10.2 shows the average and the envelopes are almost the same for both models. The only difference being that the maximum value are somewhat larger for the simulation case. In other words, the models perform similar on the average case, but since the simulated versions are not constrained by the filtered series it can sometimes give more extreme values. The exception being the LIBOR rate which has extreme values so large they distort the averages.

The trend model has larger differences between its simulated and bootstrapped versions. The bootstrapped versions give more stable realisations, and since the volatility model will affect the series as well, it might be better to use this when simulating without any crisis-regime. When adding a crisis-regime all series are drawn downwards, and this will lessen the upwards trend the simulated trend exhibit over a long time-span. The bootstrap version will also have a trend-structure that is identical during the crisis-regime, no matter the value of the series when the crisis start. In the case where we simulate the trend, it is possible to increase χ_0 if one thinks that the value of a series at the beginning of a crisis in a given realisation will be larger than it was in 2008, or just if one thinks the estimated values of χ_0 are too conservative (as it indeed seems it is for the LIBOR and oil rates when we watch Figure 10.9).

11.2 Improvements and further work

For parsimonious reasons we would ideally like to keep the model as simple as possible, and therefore add as few stochastic variables as we can. It could therefore be fruitful to try different multivariate volatility models to see if they might get better results for the correlation, especially for the freight rates. It could also be smart to try to add a mean-reverting property to the LIBOR rate. Most interest-rate models have this, and it would stop the rate from getting “stuck” at small values and from getting the extremely large values simulation sometimes leads to. Instead of focusing on modelling the stochastic volatility, we could just as well focus on the stochastic trend model, and use a simple normal or student-t distribution for the volatilities. We would then of course use a multivariate model for all the series, which might better capture the correlation between the freight series and between

the oil series. The problem would still be to find a smart way to keep the trends from going negative. In this thesis we have also assumed that the trend between freight and oil is uncorrelated. When talking to the risk group at Western Bulk it seems there have historically been a slow moving negative correlation between those series, but that this might have changed lately. Underlying macro-trends that affect both the freight rates (of dry bulk) and the oil rates might force both prices up in periods of industrial expansion, but this might not be observable on daily or even weekly prices and will also be influenced by the fleet utilisation. So oil and freight will presumably be more correlated in a market where the oil rates are high and few ships are available, but have little or no correlation when the market is different from this. It can be helpful to implement these factors within the model.

When looking at the freight return series we observed that there might be some memory property present. So we used a built-in function in R to fit both series to an AR(1) model, and used the fitted residuals rather than the return-series directly in the stochastic volatility model. We then simulated the residuals of the freight series with the returns of the LIBOR and oil rates. We could now simulate realisations of the residuals, and together with the fitted ϕ value for the return series, we could create the realisations for the returns for the freight rates. The time did not permit us to do extensive research, but it would seem that the residuals had somewhat less internal correlation as well as being less volatile than the return series directly. However, combined with the trend model we would then get series that seemed more correlated, since the volatility about the trend decreased, even though there seemed to be less correlation between the two series of residuals. The simulated values also looked more like the historical observations, since they now had the same smooth characteristic. When now applying the Jarque-Bera test on the observed r_t , the freight residuals still did not come from a normal distribution; they were now more asymmetric and the BSI-series had quite heavy tails. It could be interesting to try different models for the freight returns (e.g AR(k) for $k > 1$ or even a moving average model), and see how they are correlated and what kind of returns simulating their residuals would give.

References

- Casella, G. and Berger, R. L. (2002). *Statistical Inference*, second edn, Duxbury Press.
- Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*, Oxford Univeristy Press.
- E. Carlstein et al. (1998). Matched-block bootstrap for dependent data, *Bernoulli* **7**: 305–328.
- Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*, Chapman and Hall.
- Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo*, second edn, Chapman and Hall.
- Gilbert, P. (2011). R package available at <http://cran.r-project.org/web/packages/dse/index.html>.
- Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime, *Journal of Economic Dynamics and Control* **12**: 385–423.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**: 357–384.
- Hoffman, D., Raffenetti, R. and Ruedenberg, K. (1972). Generalization of euler angles to n dimensional orthogonal matrices, *Journal of Mathematical Physics* **13**: 528–533.
- Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals, *International Statistical Review* **55**: 163–172.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility:inference and likelihood comparison with arch models, *Review of Economic Studies* **65**: 361–393.
- Lahiri, S. (1999). Theoretical comparisons of block bootstrap methods, *Annals of Statistics* **27**: 386–404.
- McAleer, M., Asai, M. and Yu, J. (2004). Multivariate stochastic volatility: A review, *Econometric Reviews* **25**: 145–175.
- Plataniotis, A. and Dellaportas, P. (2010). A multivariate stochastic volatility model.

- Rencher, A. and Schaalje, G. (2008). *Linear Models in Statistics*, second edn, Wiley-Interscience.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Chapman and Hall.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations, discussion paper, *Journal of Royal Statistical Society series B* **71**: 319–392.
- Ruiz, E. and Pascual, L. (2002). Bootstrapping financial time series, *Journal of Economic Surveys* **16**: 271–300.
- Smeekes, S. (2009). *Bootstrapping Nonstationary Time Series*, PhD thesis, Maastricht University.
- Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes - a study of the daily sugar prices 1961-79, *Time Series Analyses: Theory and Practice* **1**: 203–226.
- Wei, W. (2006). *Time Series Analysis*, second edn, Pearson.
- Wilmott, P., Howison, S. and Dewynne, J. (1995). *The Mathematics of Financial Derivatives*, Cambridge University Press.

A Probability Distributions

There are different parametrisations for some of the probability distributions, so here is a list of the definitions for the densities of the distributions we have used.

Distribution 1 (The normal). *A continuous random variable, X , has a normal distribution, denoted $X \sim N(\mu, \sigma^2)$, if its density follows*

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

when $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$.

Distribution 2 (The beta). *A continuous random variable, X , has a beta distribution, denoted $X \sim \beta(\omega, \gamma)$, if its density follows*

$$f(x) = \frac{\Gamma(\omega + \gamma)}{\Gamma(\omega)\Gamma(\gamma)} x^{\omega-1}(1-x)^{\gamma-1}, \quad 0 \leq x \leq 1,$$

when $E(X) = \frac{\omega}{\omega+\gamma}$, $\text{Var}(X) = \frac{\omega\gamma}{(\omega+\gamma)^2(\omega+\gamma+1)}$ and $\Gamma(\cdot)$ is the gamma function (not to be confused with the gamma distribution).

Distribution 3 (The gamma). *A continuous random variable, X , has a gamma distribution, denoted by $X \sim \Gamma(\alpha, \beta)$, if its density follows*

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}, \quad x \geq 0,$$

when $E(X) = \frac{\alpha}{\beta}$, $\text{Var}(X) = \frac{\alpha}{\beta^2}$ and $\Gamma(\cdot)$ is the gamma function.

B Stochastic volatility plots

B.1 Generated data

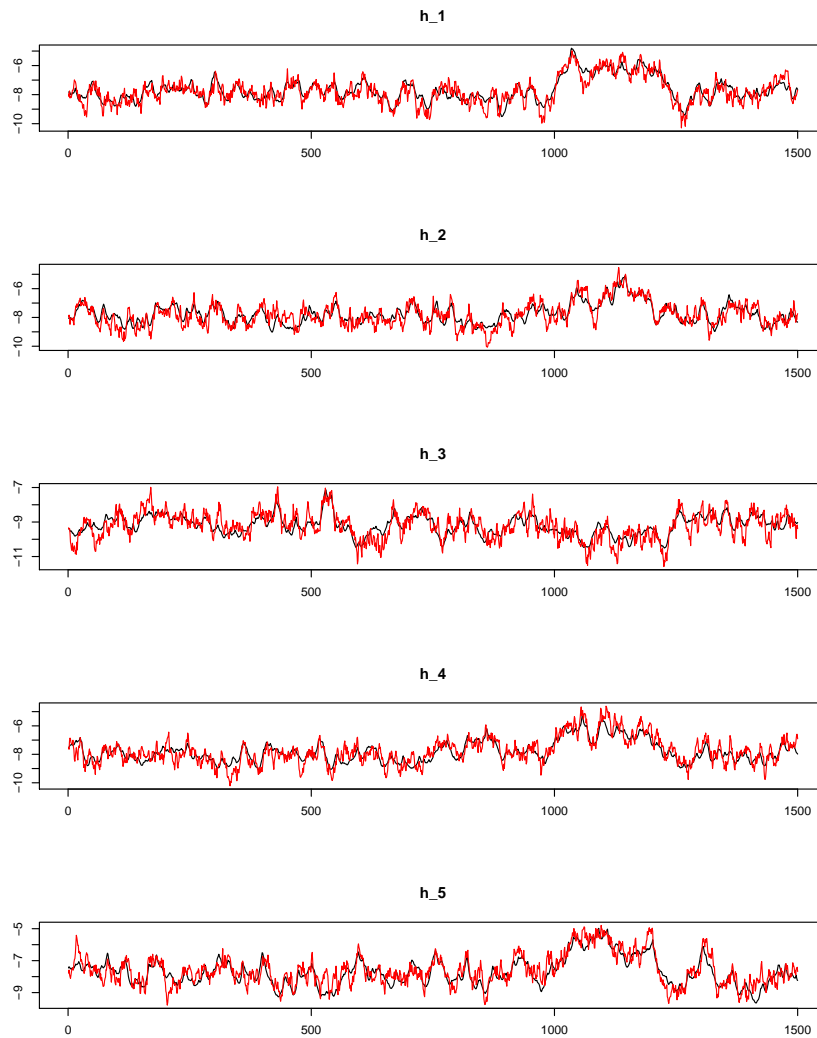


Figure B.1: The filtered series for h_t with the correct series in red.

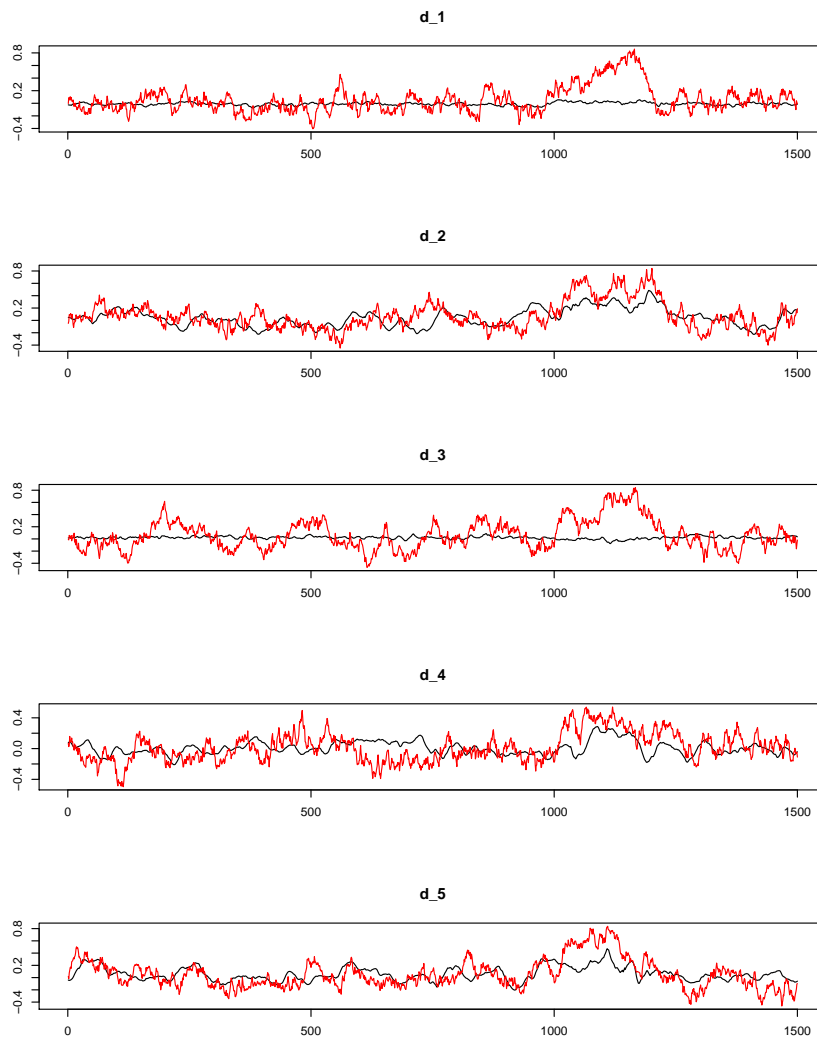


Figure B.2: The filtered series for $\delta_{t,1:5}$ with the correct series in red.

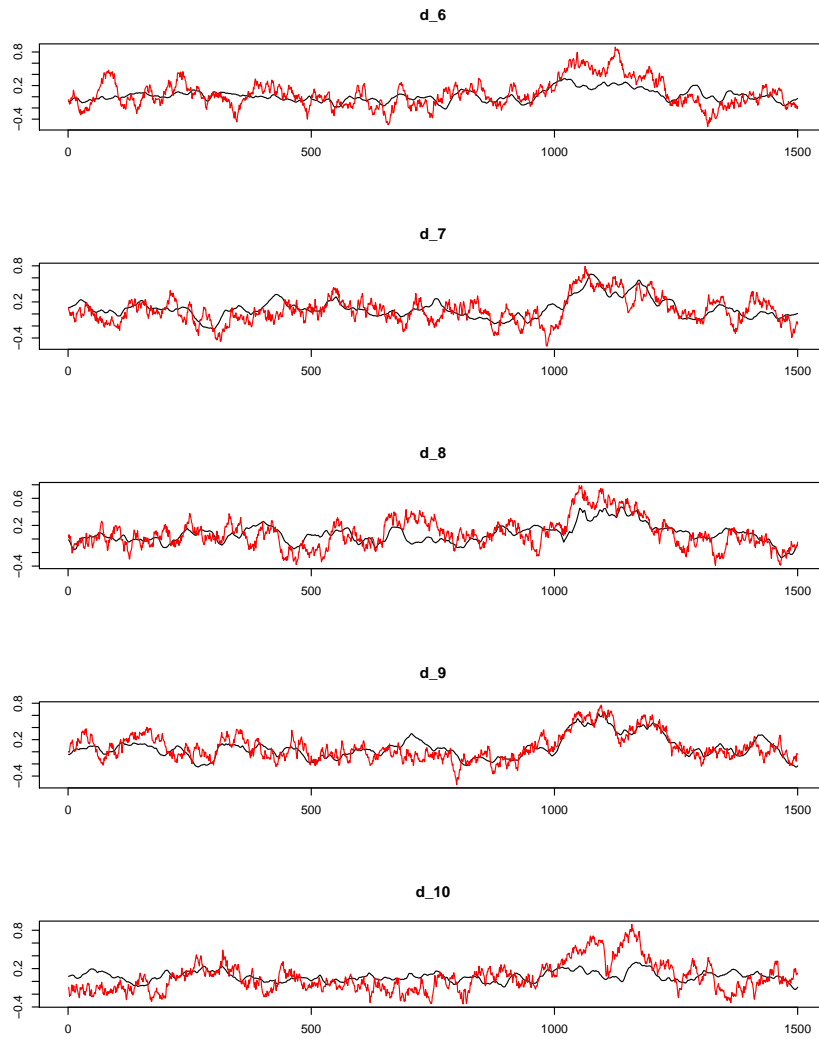


Figure B.3: The filtered series for $\delta_{t,6:10}$ with the correct series in red.

B.2 Real data

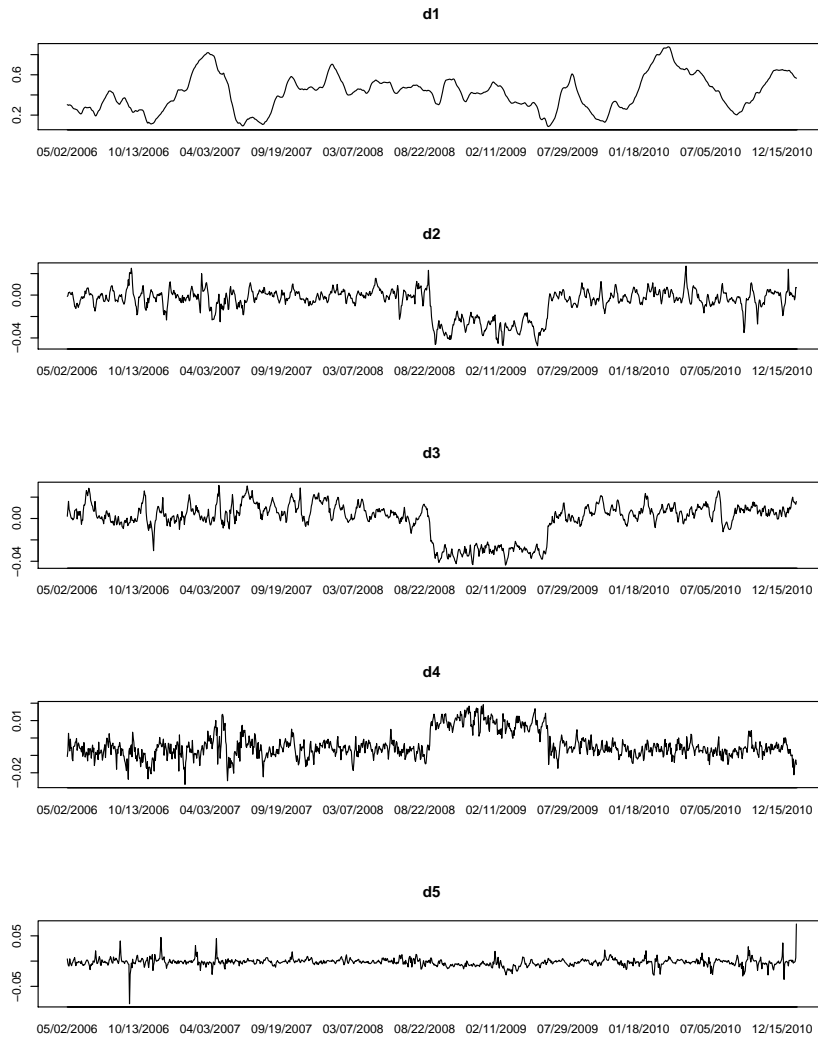


Figure B.4: The filtered series for $\delta_{t,1:5}$.

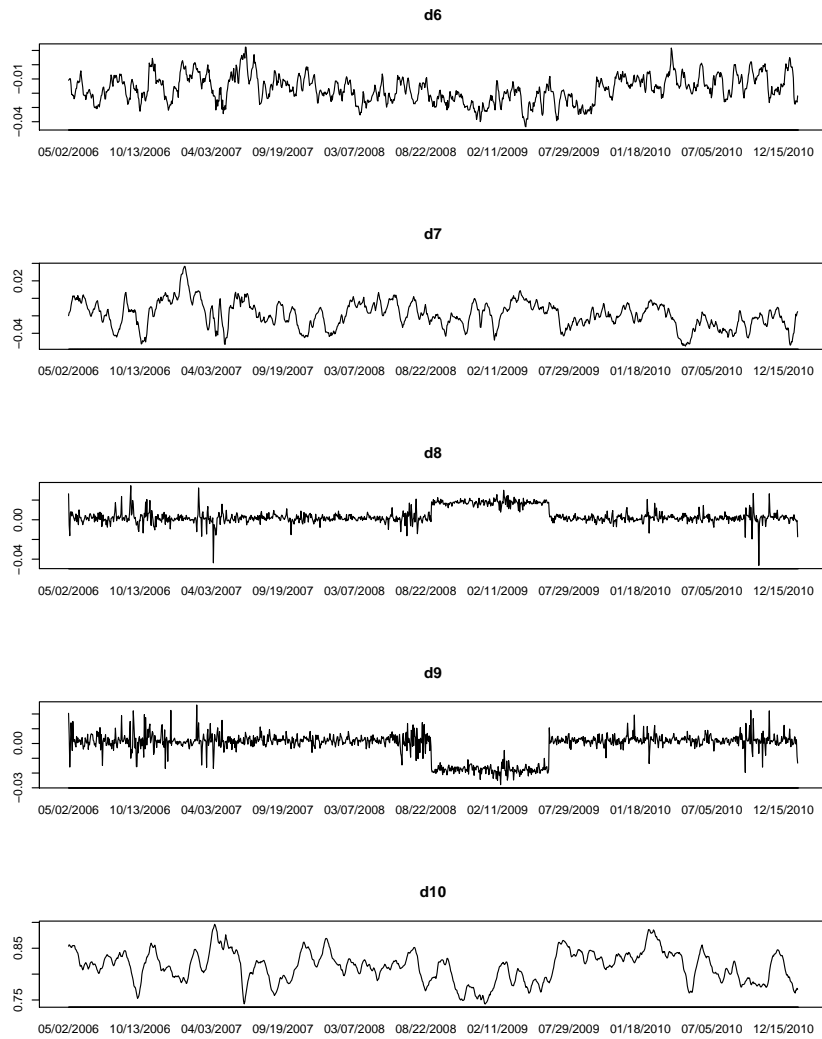


Figure B.5: The filtered series for $\delta_{t,6:10}$.