Alternatives to the Generalized Autoregressive Conditional Heteroskedasticity Model in case of Identification Loss

by

Hester M. Kuilboer

Thesis supervisor: Dr. Andrea Naghi
Co-reader:

A thesis submitted in partial fulfilment for the degree of
Master of Science in Quantitative Marketing & Data Analytics

at the

Department of Econometrics
Erasmus School of Economics
Erasmus Universiteit Rotterdam

August 2019

Abstract

This study shows the consequences of identification loss within the generalized autoregressive conditional heteroskedasticity (GARCH) model. It shows that identification loss leads to incorrect inferences on the significance of the parameter estimates. Parameter estimates follow nonstandard distributions in case of identification loss, ranging from uniform to bimodal distributions. The t-values associated to the parameter estimates explode for the GARCH model with error specification following the normal distribution (GARCH-N). This shows the challenge in recognizing identification loss. Therefore, the robust t-values are higher than the expected 1.98. Especially for the GARCH-N model the robust t-value is excessive: the t-value associated to the identifying parameter β should be higher than 55.7 to guarantee an identified GARCH model. For the GARCH model with error specification following the t-distribution (GARCH-t) the t-values of the identifying parameter β are often imaginary in case of nonidentification and the t-values in general are not as high as for the GARCH-N model. The robust t-value should exceed 2.9 for identification purposes. However, an added issue for the GARCH-t model is the trade-off between identification loss and nonstationarity. If the GARCH-t model is better identified, the GARCH model is dangerously close to nonstationarity. Therefore, the t-distribution should not be preferred as assumption for the error specification.

In case of identification loss, the autoregressive conditional heteroskedasticity (ARCH) model outperforms the GARCH-N model. However, this is not the case for the GARCH-t model. In general, the stochastic volatility model (SVM) outperforms the GARCH model, based on the Diebold-Mariano test. This has to do with the lower volatility estimated by the SVM with respect to the GARCH model. Therefore, the value at risk does not contain the true value of the return in more cases for the SVM than for the GARCH. For predictive purposes, based on the Diebold-Mariano test, the SVM should be preferred.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem statement

Financial time series are important for index speculation and policy decision making. In time series there are moments of less uncertainty with small changes in returns and moments of more uncertainty with relatively large changes. This indicates that there is no constant volatility, a statistical feature that is often assumed in non-financial data sets. Therefore, to understand the fluctuation and better describe and predict time series, it is essential to capture the underlying volatility. A stylized fact called 'volatility clustering' tells us that periods of large volatility tend to cluster together and periods of small volatility cluster together; this makes volatility quite predictable. This is what makes volatility models successful at capturing the changes in fluctuation strength.

One main problem in estimating (financial) econometric models is the problem of identification loss (Andrews, 2001), (Andrews and Cheng, 2012). Identification loss and its consequences for interpretation and accuracy of economical models has been understudied. There are serious implications on the empirical parameter distributions: they do not follow the normal distribution. Deviations from the normal distribution in the parameter space have consequences for the interpretation of the model. For example, there are serious implications on the t-values associated to the parameter estimates. The distribution of the t-values, required for the significance level assessment of parameter estimates, do no longer conform with the t-distribution. Since t-values are directly linked to the confidence intervals of the parameter estimates, the confidence intervals of the parameter estimates are altered and smaller than they should be. Since the parameter estimates are not reliable, there are also consequences for the predictive power of the model (Naghi, 2016).

One of the first conditional volatility models developed was the autoregressive conditional heteroskedasticity (ARCH)(p)-model (Engle, 1982). This model deals with the stylized fact of volatility clustering. The ARCH(1)-model is defined as $h_t = \alpha + \beta\epsilon_{t-1}^2$, with h_t the conditional volatility at time t, and ϵ_t the shock at time t. The ARCH model assumes that the conditional volatility is dependent on the shock in past returns. After this model had been developed, research into conditional volatility exploded and over 300 different volatility models were developed within 25 years (Lunde and Hansen, 2005).

The generalized autoregressive conditional heteroskedasticity (GARCH)(p,q)-model is a generalization of the ARCH model. It is defined as $h_t = \alpha + \beta\epsilon_{t-1}^2 + \pi h_{t-1}$, with h_t the conditional volatility at time t and ϵ_t the shock at time t. This model takes into consideration the past conditional volatility on top of the aforementioned shock in return. The GARCH(p,q)-model is widely used for dealing with time series data. In 2018 alone 600 papers cited Engle (1982). Several of those papers will be scrutinized in Chapter 2.3. Some issues can arise when employing a GARCH(p,q)-model. One issue is that the identification assumption is not being met (Bollerslev et al., 1994), (Andrews, 2001), (Ma et al., 2007). As will be explained in Chapter 3.2, if the β parameter is 0, the π parameter is not identified. However, the β parameter in GARCH models shows a very significant inclination, with large t-values and small standard errors. This means that, even if $\beta = 0$, the GARCH model does not estimate β correctly as zero. Therefore, it is not straightforward to recognize the identification loss of the GARCH model. If the identification assumption is not being met, this is a clear indication that the model chosen is not the correct or best model for a particular data set. Note that the smaller the time period, the larger the possibility of identification loss (Andrews and Cheng, 2012). This is, because β is related to the identification indicator b through the sample size.

1.2 Research objectives & hypothesis

As mentioned before, the GARCH model and more specifically the GARCH(1,1)-model is used often in financial research as benchmark model or as chosen model. The GARCH(1,1)-model is a specific case of the GARCH(p,q)-model. GARCH(1,1)-model and GARCH model are used interchangeably within this thesis. The goal of this study is to investigate the consequences when a GARCH(1,1)-model suffers from an identification problem of the π parameter. For this reason both a normal (GARCH-N) and a t-distribution (GARCH-t) is used as error specification, because both specifications lead to different conclusions and results. Note that the results of this thesis can be generalized to the GARCH(p,q)-model.

The understudied problem of identification loss in financial volatility models leads to the following research question:

> *Which volatility measures are the most useful to predict returns in financial time series in case of identification loss in the generalized autoregressive conditional heteroskedasticity model?*

To achieve this goal and answer this question, I first examine the empirical distributions of the parameters and associated t-values of the GARCH(1,1)-model in a simulation study with a sample size of 500 and increasing identification strength. I show that the unidentified π parameter is uniformly distributed in case of identification loss, showing signs of a normal distribution only in case of stronger identification. Furthermore, the α parameter is negatively correlated to the π parameter and shows erratic behavior as well. The t-statistic behaves very different for both error specifications, showing high significance for normally distributed GARCH models and nonconverged models for t-distributed GARCH models. Then, I compare two different volatility models in terms of predictive power to the GARCH(1,1)-model: ARCH and SVM. I base this on the same simulation study as used for distribution investigative purposes. Since the data originates from a GARCH model, this model should have been able to describe the data better than the other volatility models. However, as I will show, for a normally distributed error specification, the GARCH model does not outperform the ARCH(1)-model, based on the Diebold-Mariano test. The Stochastic Volatility Model (SVM) employed outperforms the GARCH model in general. This is, because the SVM uses a probabilistic approach where a (G)ARCH model uses a deterministic approach. Furthermore, I show that the study of Bollerslev (1987) suffers from identification loss.

(G)ARCH models belong to the class of deterministic conditional heteroskedasticity models. This means that the conditional variance is a function of variables that are known at time *t-1*. SVMs belong to the class of stochastic conditional heteroskedasticity models. This means that the conditional variance is a function of random variables up to time *t-1*. SVMs are state-space models and can be solved through different filters. To limit the scope of this study, I compare a Kalman filter (Kastner, 2016) to the deterministic GARCH model.

For prediction purposes I compare the ARCH and SVM to the GARCH model. The ARCH model performs similar to the GARCH-N model in case of identification loss, based on the Diebold-Mariano test (Diebold and Mariano, 1991). The ARCH model performs worse than the GARCH-t model in case of identification loss, based on the Diebold-Mariano test (Diebold and Mariano, 1991). The SVM outperforms the GARCH model in general and should be preferred, especially in case of identification loss. In addition, I also compared the value at risk (VaR) evaluation from nonidentified to well identified GARCH. In 10% of cases the VaR (for a VaR of 5%) does not contain the real return value. However, for the well identified GARCH model, a large part of the cases in which this happens is when the GARCH model nears nonstationarity, when the $\alpha + \beta$ parameter together are very close to 1.

This study builds on Ma et al. (2007) and Andrews and Cheng (2012). Ma et al. (2007) show that the Zero-Information Limit Condition (ZILC) (Nelson and Startz, 2007) holds in the GARCH(1,1) model. Ma et al. (2007) investigate the effect of weak identification of β on the standard error of π, if the real value of π is 0. The size distortion of π is large, the π parameter is estimated significantly higher than 0 in almost 50% of cases although the real value is actually 0. The larger the β parameter, the smaller the size distortion becomes. From a b of 7 upwards, the size distortion seems to have diminished from 50% to 7% significance. Ma et al. (2007) show through a simulation study that the t-statistic is unreliable for all parameters in case of weak identification.

The addition of this study to the identification literature is that it studies the overestimation of the t-value. The reasoning behind this is that in empirical finance literature often a significant ARCH effect is found. If the ARCH effect is found to be insignificant, these studies are not published or it is mentioned that the ARCH effect is insignificant. A note here should be, that it then should be mentioned that the GARCH effect is unreliable as well, based on the insignificant ARCH effect. Furthermore, this study investigates more cases of identification strength (Andrews and Cheng, 2012) than the weakly identified GARCH model and discusses the difference between the different cases. I apply the same method as Andrews and Cheng (2012) and follow the identification categories they specified (Table 3.3). It also focuses on the t-distributed error of the GARCH model, showing that the t-test has different properties for the parameter estimates in this case. In the case of nonidentification, the t-value of the β parameter is imaginary in 40% of cases, showing a misspecification of the GARCH model. This important, since it shows a clear problem with the GARCH model as used and the model should therefore not be used. Ma et al. (2007) did not investigate the t-distributed error specification in the GARCH model and assumed both models acted similarly. As I will show this is not the case and it is important to therefore distinguish clearly when making an error assumption.

On top of that, I show the effects of identification loss in terms of predictive power and compare the different identification categories. If there is a reason to suspect identification loss, one should examine the t-value of β and check whether it is smaller than the robust t-statistic of 2.91 for the GARCH-t model and 55.7 for the GARCH-N model. If the error is assumed normal, one could also perform a simulation study. One should compare the predictive power of the GARCH model to the ARCH-model with a Diebold-Mariano test. If the Diebold-Mariano test does not favor the GARCH model over the ARCH-model it indicates a nonidentified GARCH model. In this case, an SVM is the preferred conditional volatility model to be used.

1.3 Thesis structure

This thesis is organized as follows; Chapter 2 comments on papers in literature that study conditional volatility models and identification problems and their impact on econometric inference. Furthermore, it scrutinizes some recent works that are based on GARCH models. Chapter 3 describes the methodology for the estimation of volatility models, the simulation set-up and the loss functions for determining the optimal model based on the Diebold-Mariano test. Chapter 4 compares the results obtained by using (G)ARCH models and stochastic volatility models, based on a simulation study and the S&P 500 data set. Finally, Chapter 5 summarizes the conclusions, and provides recommendations for future works.

Chapter 2

Literature Review

This chapter evaluates the most recent published work on financial volatility models and identification problems. Furthermore, it comments on recent published work from 2018 and 2019 that suffer from identification loss in the generalized autoregressive conditional heteroskedasticity (GARCH) model.

2.1 Volatility models

In 1982 the autoregressive conditional heteroskedasticity (ARCH) model described the modelling of conditional variance and volatility (Engle, 1982). The GARCH model soon followed (Bollerslev, 1986, 1987). In the following two decades more than 330 ARCH-related volatility models were developed (Lunde and Hansen, 2005). Lunde and Hansen (2005) showed that none of the more ingenious models outperformed the GARCH(1,1)-model for exchange rate data. However, for IBM return data other models did outperform the GARCH(1,1)-model. This shows that not always the GARCH(1,1)-model is preferred.

Blattberg and Gonedes (1974) investigated alternatives to the normal distribution commonly used when describing financial time series. Blattberg and Gonedes (1974) studied whether the observed "fat tails", that are common in the empirical distributions of stock returns, can be described better by a Student's t-distribution or a symmetric-stable model. "Fat tails" are not common for normal distributions. They showed that a Student distribution is better in describing financial time series than a symmetric-stable model. Bollerslev (1987) took these implications of "fat tails" into account. He developed a GARCH model that explicitly allows for leptokurtic distribution. A leptokurtic distribution has greater kurtosis than the normal distribution which corresponds to the observed "fat tails".

There is a distinction between GARCH models and Stochastic Volatility Models (SVMs) (Hull and White, 1987), (Taylor, 1994) in the sense that GARCH models are deterministic and SVMs are probabilistic. GARCH models have the property that the (unobservable) volatility is directly dependent on the volatility of the previous period and the error of the previous return. There is no stochasticity involved for the conditional volatility. This corresponds to assuming that the conditional volatility is known one step ahead. SVMs are different in the sense that there is a stochastic element involved in the conditional volatility. Therefore it does not assume that it knows the volatility exactly one step ahead. The volatility is modeled as an unobserved latent variable.

SVMs have been researched extensively in literature (Bos, 2012), (Broto and Ruiz, 2004). Broto and Ruiz (2004) stated that there has been difficulty implementing SVMs, because the estimation procedure was not straightforward. They reviewed multiple methods of estimating SVMs and showed that a Kalman filter (Welch and Bishop, 1995) is most time efficient and least computational intensive. Therefore, I implement the filter in this study as comparison to the GARCH model.

2.2 Identification

Economic models use the assumption of identification to justify the model. The notion of identification issues has been around since the 1950's (Goodson and Polis, 1975). Since then, more emphasis has been placed on identification issues and the implications on standard estimation techniques (Andrews and Cheng, 2012, 2013). Furthermore, tests and properties for and of these estimators have been closely investigated (Andrews, 2001), (Andrews and Soares, 2010), (Andrews and Cheng, 2012, 2013), (Cheng, 2015), (Beg et al., 2001).

Beg et al. (2001) investigated identification loss in the ARCH-in-Mean model. The specification for this model is $y = \mu + \boldsymbol{\beta}^t \mathbf{x}_t + \phi h_t + \epsilon_t$, with $h_t^2 = \alpha_0 + \psi_1 \epsilon_{t-1}^2 + ... + \psi_p \epsilon_{t-p}^2$. Beg et al. (2001) proposed tests for the null hypothesis of constant volatility with $h_t^2 = \alpha_0$. If there is a case of constant volatility h_t becomes a constant and μ, α_0, and ϕ are not identifiable separately. Beg et al. (2001) showed that the test statistic based on the requirement that the $\boldsymbol{\psi}$ parameter is nonnegative is the best performing test statistic, based on minimization of the type-I error rate. Type-I error is the probability that the null hypothesis of constant volatility is true, so $\boldsymbol{\psi} = 0$, but gets rejected.

Bollerslev et al. (1994) mentioned that for the GARCH model, in case of the null hypothesis of constant volatility, π and β in the GARCH(1,1) specification $h_t = \alpha + \beta \epsilon_t^2 + \pi h_{t-1}$ cannot be identified separately. The Lagrange-Multiplier (LM) test for the GARCH model in that case is equal to the LM test for the ARCH model. The Wald-type tests and t-tests are non-standard. The t-statistic of π does not follow a t-distribution under the null hypothesis. This is, because there is no time-varying input. The β parameter is not identified. Andrews (2001) proposed tests to check for $\beta = 0$ in a GARCH(1,1)-model. These tests make use of established asymptotic distributions of test statistics under alternatives.

Ma et al. (2007) showed that the zero information limit condition (ZILC) (Nelson and Startz, 2007) holds for the GARCH model, based on the closed form asymptotic variance covariance matrix derived by Ma (2006). Ma et al. (2007) showed that the GARCH estimate is biased upward in case of identification loss (when the α parameter is near zero) if the real value of the β parameter is zero. The size of the t-value of the β parameter, on which the significance is based is very large, around 45%. Only for larger π and large data sets of around 5000 observations the size distortion of the t-test diminishes. This means that there exists a size distortion for non and weakly identified GARCH models. The likelihood ratio (LR) and Lagrange-Multiplier (LM) tests perform better than the t-test. A reason the LM test performs better is that it is calculated under the restriction of the weakly identified parameter π.

Ma and Nelson (2010) showed that for models with the general notation $y = \gamma \cdot g(\beta, x) + \epsilon$, the Wald test systematically has the wrong size in finite sample sizes, when the identifying parameter γ is small relative to its estimation error. As an alternative they introduced a test based on the linearization of the general model with the approximation $g(\beta, x) \approx g(\beta_0, x) + (\beta - \beta_0) \cdot g_\beta(\beta_0, x)$

(Fieller, 1954), (Breusch and Pagan, 1980). Ma and Nelson (2010) provided a reduced form test, based on this linearization, as alternative to the Wald test. This test has (nearly) correct size, because its distribution does not depend on the identifying parameter γ.

Andrews and Cheng (2012) showed that among other models, the Autoregressive Moving Average (ARMA) model suffers from identification loss. The parameter estimates of the ARMA model are affected greatly by unidentified parameters. The ARMA model is depicted as: $Y_t = (\pi + \beta)Y_{t-1} + \epsilon_t - \pi\epsilon_t - 1$ with Y_t the return series, ϵ_t the innovation or shock at time t, π the moving average (MA) parameter and $\beta + \pi$ the autoregressive (AR) parameter. It is shown that the π parameter in an ARMA(1,1)-model exhibits a uniform distribution when it is not identified. It is well approximated by a normal distribution only when strongly identified. When there is no or weak identification, the β parameter shows tendencies of a bi-modal distribution. For semi-strong to strong identification it results into a more normal distribution. This shows that for identification issues, the parameter estimates do not reflect well the underlying data generating process (DGP). Therefore, standard estimation techniques such as maximum likelihood fail to capture the true GDP and estimates are not trustworthy, which has consequences for prediction purposes. Andrews and Cheng (2012) proposed robust t-statistics and tests, considering identification problems for the ARMA(1,1)-model that can also be applied to other models such as GARCH models.

Andrews and Cheng (2013) proposed tests that are robust to identification loss. They apply these tests to a nonlinear binary choice model and a smooth transition threshold (STAR) autoregressive model that suffer from identification loss. Similar to the GARCH model, if the β parameter goes to zero, the π parameter is not identified. Andrews and Cheng (2013) showed through a simulation study the consequences of identification loss for the parameter estimates distributions. The π parameter shows a clear uniform distribution in case of identification loss. Even the β parameter shows a non-normal distribution in case of identification loss. Andrews and Cheng (2013) also showed that the empirical t-distribution does not follow a normal distribution in case of identification loss.

Cheng (2015) proposed robust tests and confidence intervals for the special case of mixed identification strength within the parameter space. She developed a local limit theory modeling this special case of mixed identification strength. The results make use of both consistent estimators with different rates of convergence and inconsistent estimators. She studied more closely the effect of $\beta_p = 0$ on not directly related parameter $\pi_j, j \neq p$. The identification of this parameter π_j is dependent on the unknown value β_j.

Naghi (2016) studies the consequence of identification loss on the predictive accuracy assessment. In case of identification loss the null-distribution of the out-of-sample predictive ability tests, which is assumed to be normal, is biased to the right of the zero mean in case of identification loss. Naghi (2016) shows that if the parameter estimation error is non-negligible the West (1996) predictive ability test is standard.

2.3 Recent studies suffering from identification loss

Over the past years there were multiple studies that (possibly) suffered from identification loss without addressing the problem of incorrect test statistics. In 2018 alone 600 papers cited Engle (1982). I reviewed the citing papers from 2018 and found multiple studies (possibly) suffering from identification loss. These studies are found in Table 2.1. I assume the parameter estimates

Table 2.1: Studies reporting GARCH estimates with identification loss

Study	t-value	β (b) value	Number of observations
Otunuga et al. (2019) Table 8		0.019 (0.40)	438
Lin (2018) Table 7	8.991761	0.060276 (1.90)	996
Gormus et al. (2018) Table 7		0.060 (3.02), 0.065 (3.28)	2541
Herrera et al. (2018) Table 2		0.0887 (3.45), 0.0722 (2,81)	1512
Patton et al. (2019) Table 7		0.052(2.61), 0.053 (2.66)	2520
Shen et al. (2018) Table 2		0.045 (2.77)	3780
Zhang and Li (2018) Table 2		0.0080 (0.27) - 0.0490 (1.66)	1138
Song and Kang (2018) Table 10		0.145 (1.99)	189
Fan et al. (2018) Table 6	8.43, 1.82-3.49	0.059 , 0.158 -0.448	1510, 171-554
Shi et al. (2017) Table 3		0.0347 (1.34)	1489
Paul and Sharma (2018) Table 6		0.083-0.104	2259-2301

of β as correct estimates for this instance. I check whether the b estimate, defined as $\beta/\sqrt(T)$, is close to 0, 2, or 4. If so, I report the β parameter estimate. This corresponds to no, weak, and semi-strong identification as shown in Table 3.3. As I will show in Chapter 4.1, in these cases, these estimates are not reliable. If a certain study mentions a number of years and did not specify the number of observations, I use 252 days per year as it is the average number of trading days per year. As shown by the number of papers suffering from identification loss, it is still a big problem to date.

Otunuga et al. (2019) proposed an alternative to the GARCH model, a discrete-time dynamic model of local sample variance making use of local lagged adapted generalized method of moments. Otunuga et al. (2019) did not report the t-values or standard errors of the GARCH parameter estimates, and only compared the point prediction with the discrete-time dynamic model. The GARCH model is not identified for the Ethanol data set with an estimated b of 0.40. They showed in Figure 17d that a consequence of the near-zero β is an almost constant volatility estimate. Otunuga et al. (2019) showed that in this case the discrete-time dynamic model performs better in estimating the volatility.

Lin (2018) compared a GARCH model to variations of the GARCH model: a TARCH, and EGARCH model. The TARCH model is a threshold GARCH model. It acknowledges the stylized fact that a negative shock has a larger effect on the volatility than a positive shock. The EGARCH model is an exponential GARCH model. For the EGARCH model, the assumption of positive parameters is not necessary. Lin (2018) concluded that the EGARCH model outperformed the GARCH model. Also this study suffers from identification loss in the GARCH model. Therefore, the EGARCH model performance should still be investigated relative to a GARCH model with solid parameter identification. Gormus et al. (2018) used an LM-GARCH model based on a GARCH variance equation. The value of the b parameter is near 3. They mention that the b parameter is significant at 1%. However, they do not mention the t-value associated to the significance. Furthermore, as will be shown, high significance for the b parameter does not imply identification.

Patton et al. (2019) used FZ minimization applied on a GARCH model to expected shortfall (ES) and value at risk (VaR) forecasting. The b parameter is rather small, around weak identification. Patton et al. (2019) did not mention this issue of identification loss in their study. They do mention that FZ minimization does not allow for identifying the constant α in the GARCH model. Shen et al. (2018) used the GARCH model for the computation of VaR. They show a small ARCH effect. The assumption they make is that there is substantial volatility clustering. They showed that there is no significant volatility spillover effect based on the GARCH model. A reason for this might be the loss of identification present in the GARCH model. There is a risk spillover based on the VaR that is based on the GARCH model. Since there is a possibility of identification loss, they should re-evaluate this estimated risk spillover. Zhang and Li (2018) also used the GARCH model for computation of VaR. In this study, the GARCH model generates a volatility that reflects the market risk. However, some of the b parameters are low, indicating identification loss. Therefore the GARCH model has not generated a reliable volatility estimate. Shi et al. (2017) aimed to provide an alternative hedging method based on VaR in the shipping market. They first apply a GARCH model. The ARCH parameter estimate seemed very significant for their study in case of hedging with a 1-month FFA contract, but the parameter estimate is very small which results in a nonidentified GARCH model. This leads to an unreliable volatility estimate, on which they base conclusions, stating that the time-varying t-copula performs best. This study would benefit from the notion of nonidentification and requires more careful interpretation.

Song and Kang (2018) proposed parameter change tests for ARMA-GARCH models. Based on these tests they proposed two change points in a real data set of won-yen exchange rate. The best model experiences loss of identification, in the sense of weak identification at the smallest time period. This is an issue that should be addressed when dealing with parameter change tests. As a side note, there is a trade-off. If there is a larger time series, the GARCH model tends to a weak nonstationary integrated GARCH model (Amado and Tersvirta, 2017), as pointed out by Diebold (1986) and Lamoureux and Lastrapes (1990). This may be due to breaks, sudden shifts, in the unconditional variance process (Mikosch and Stric, 2004). This thesis also acknowledges this problem, especially if the error specification is the t-distribution. Fan et al. (2018) first identified structural breakpoints in Chinese energy power stock price. They estimated a GARCH model between each breakpoint and used the results of the estimated volatility to analyse the impact of power market reform on the stock price volatility. The same issues arose in this study as in the last study discussed. The sample sizes in between the breakpoints are too small to have a significant ARCH effect and therefore there is an unidentified π parameter. The volatility estimate is unreliable and the conclusions taken on the base of the GARCH model should be rerun with a stochastic volatility model (SVM).

Paul and Sharma (2018) compared two variations to the GARCH model, the realized GARCH and EGARCH model, to the GARCH model in a two-stage extreme-value theory (EVT) framework. Paul and Sharma (2018) compared the performance in generating quantile forecasts in ES and VaR. The GARCH estimation shows nonsignificant values of the ARCH parameter estimate. This is a clear sign of an identification problem in the GARCH model. The two other models outperform the GARCH model, which can be related to identification loss.

Chapter 3

Methodology

This chapter presents the models used for volatility prediction. First, it elaborates on the generalized autoregressive conditional heteroskedasticity (GARCH) model, the autoregressive conditional heteroskedasticity (ARCH) model and the stochastic volatility model (SVM). Second, it goes into the set-up used for simulation, based on the research of Andrews and Cheng (2012). Third, it explains how the volatility models are compared. Last, it presents a robust critical value to help determine the identification of the GARCH model.

3.1 ARCH models

ARCH models (Engle, 1982) were the first type of models considering non-constant variance used for forecasting variance. The ARCH model is estimated in Matlab.

The equation corresponding to the ARCH(1)-model is the following:

$$\begin{aligned} y_t &= \mu + \epsilon_t \\ h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \qquad\qquad \alpha_0 > 0; \alpha_1 \geq 0, \end{aligned}$$

with y_t the return at time t, μ the mean and ϵ_t the shock or innovation at time t. The conditional variance h_t at time t is dependent on the past shock squared at time $t-1$: ϵ_{t-1}^2.

If α_1 is 0, the variance for this time series becomes constant and the error is white noise. In that case there is no pattern in the disturbances of the time series. α_0 should always be larger than zero, because the equation shows the conditional variance. Variance is per definition nonnegative, since it is the square of the error term. If α_0 was zero, the data would consist of one constant for each time point, since there would be no variation. If α_1 was negative, the variance would be negatively correlated with shocks, meaning that the larger the shock at time $t-1$, the smaller the conditional variance at time t. Financial time series however show that large shocks tend to be followed by large shocks and small by small. Furthermore, this can lead to negative conditional variance, which is not allowed.

The variance function can be generalized to an ARCH(p)-model with $h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2$ with p the ARCH order and α a p+1-dimensional vector with unknown parameters and $\alpha \geq 0$.

3.2 GARCH models

The GARCH model (Bollerslev, 1986) is an extension of the ARCH model. The ARCH model assumes that the conditional volatility is only dependent on past shocks. The GARCH model adds to this model in the sense that it can also be dependent on past conditional volatility. This transforms the error variance model in a so called autoregressive moving average (ARMA) (Whittle, 1951) model. The GARCH model is estimated and simulated in Matlab.

The equation corresponding to the GARCH(1,1)-model is as follows:

$$\begin{aligned} y_t &= \mu + \epsilon_t \\ h_t &= \alpha + \beta\epsilon_{t-1}^2 + \pi h_{t-1} \qquad\qquad \alpha > 0; \beta, \pi \geq 0, \end{aligned}$$

with y_t the return at time t, μ the mean, ϵ_t the shock or innovation at time t. In this study I consider a shock that follows a t-distribution and one that follows a normal distribution. The conditional variance h_t at time t is dependent on a constant α, the past shock squared at time $t-1$ ϵ_{t-1}^2, and on the conditional variance at time $t-1$. To prevent the occurrence of negative variance, all parameters should be greater or equal to 0. The variance function can be generalized to a GARCH(p,q)-model with $h_t = \alpha + \sum_{i=1}^{p} \beta_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \pi_i h_{t-i}$ with p the ARCH order, q the GARCH order and $\alpha > 0; \beta, \pi \geq 0$.

The (ARMA)(1,1) representation (Hamilton, 1994) that goes with the GARCH (1,1)-model is as follows:

$$\begin{aligned} y_t &= \mu + \epsilon_t \\ \epsilon_t^2 &= \alpha + (\beta + \pi)\epsilon_{t-1}^2 + \nu_t - \pi\nu_{t-1} \qquad\qquad \alpha > 0; \beta, \pi \geq 0 \\ \zeta_t &= \epsilon_t^2 - h_t = h_t(u_t^2 - 1) \end{aligned}$$

with y_t the return at time t, μ the mean, ϵ_t the shock or innovation at time t, and ϵ_t^2 the unconditional variance at time t; ζ_t is the defined error term at time t, h_t the conditional variance at time t, and u_t a random error, with dependent on the distribution of the error either $u_t \sim N(0,1)$ or $u_t \sim t(\nu)$, with ν degrees of freedom.

This is achieved by adding ϵ_t^2 to both the right- and left-hand side of the equation, subtracting h_t from both the right- and left-hand side of the equation, and both adding and subtracting $\pi\epsilon_{t-1}^2$ from the right-hand side of the equation. The parameters corresponding to the autoregressive (AR) part of the equation are $(\beta + \pi)$. The parameter corresponding to the moving average (MA) part of the equation is π. This is similar to the ARMA(1,1)-model, defined as: $Y_t = (\pi + \beta)Y_{t-1} + \epsilon_t - \pi\epsilon_t - 1$.

A problem that can arise in the GARCH(1,1)-model is the identification issue as mentioned in Chapter 1.1. Identification problems occur when the β parameters is equal to zero. This corresponds to the difference between the AR and MA parameter being zero. In the case of identification loss the GARCH(1,1)-model evolves to the following:

$$\begin{aligned} y_t &= \mu + \epsilon_t \\ \epsilon_t^2 &= \alpha + \pi\epsilon_{t-1}^2 + \nu_t - \pi\nu_{t-1} \qquad\qquad \alpha > 0; \beta, \pi \geq 0 \\ \zeta_t &= \epsilon_t^2 - h_t = h_t(u_t^2 - 1) \end{aligned}$$

with y_t the return at time t, μ the mean, ϵ_t the innovation or shock at time t; ϵ_t^2 the unconditional variance at time t; ζ_t is the defined error term at time t, h_t the conditional variance at time t, and u_t a random error, with dependent on the distribution of the error either $u_t \sim N(0,1)$ or $u_t \sim t(\nu)$, with ν degrees of freedom.

This leads to the following equation:

$$\begin{aligned} y_t &= \mu + \epsilon_t \\ \epsilon_t^2 &= \frac{\alpha}{1-\pi L_1} + \zeta_t && \alpha > 0; \pi \geq 0 \\ \zeta_t &= \epsilon_t^2 - h_t = h_t(u_t^2 - 1), \end{aligned}$$

with y_t the return at time t, μ the mean, ϵ_t the innovation or shock at time t; ϵ_t^2 the unconditional variance at time t; ζ_t is the defined error term at time t, h_t the conditional variance at time t, and u_t a random error, with dependent on the distribution of the error either $u_t \sim N(0,1)$ or $u_t \sim t(\nu)$, with ν degrees of freedom.

The first step to achieve the previous equation is subtracting $\pi\epsilon_{t-1}^2$ from both the right- and left-hand side. The second step is introducing the first lag operator L_1, which stands for time $(t-1)$. The last step is dividing through (1-πL_1). h_t is equal to $\frac{\alpha}{1-\pi L_1}$, because $h_t = \epsilon_t^2 - \zeta_t$. As is shown in the equation for ϵ_t^2, the parameters for the lagged innovation and the autoregressive (AR) term, ζ_{t-1} and ϵ_{t-1}^2 respectively, are not identified. This introduces complications such as unreliable estimators and distorted testing results.

In order to estimate the parameters of the GARCH model Maximum Likelihood is used (Engle, 2001). Berkes et al. (2003) prove the consistency and asymptotic normality of the quasi maximum likelihood estimator of the parameters in case of nonstationarity. The standard errors and the t-values corresponding to the parameter estimates are computed using the inverse of the Hessian matrix. If the optimal parameter estimates cannot be computed through the numerical optimization process, the Hessian matrix is not positive semi-definite, making the inverse negative, which violates the positive variance of the parameter estimates. If this is the case, Matlab was unable to get the model to converge and therefore get stable estimates. This means that the GARCH model cannot be properly estimated with the data available. This is a clear indication of a wrong model, and if this is the case, one should choose another model.

3.3 Stochastic volatility models

Stochastic volatility models (SVMs) take a probabilistic approach to the conditional variance problem. This means that it considers not one exact value for the conditional volatility, as is the case for the GARCH model. Rather, it estimates a distribution out of which a value is a realization. (G)ARCH models consider deterministic approaches to the conditional variance problem and assume the conditional volatility fixed and known for one time ahead, allowing for error only in the observational time series. The SVM is estimated in R.

SVMs follow the idea that the logarithm of the conditional variance follow an autoregressive (AR) process of order one. They fall into the class of state-space models, assuming conditional volatility to be an unobserved state that can be estimated through assuming a relation to past states. Therefore estimation techniques associated with state-space models, such as filtering and smoothing can be used, to find the optimal parameter distributions. Reasons SVMs are not often considered in literature are the various suboptimal estimation techniques available and the lack of software packages available (Bos, 2012). I make use of the "stochvol" package Kastner (2016) provided in R.

SVMs can be expressed in the following hierarchical form:

$$\begin{aligned} y_t|h_t &\sim N(0, exp(h_t)) \\ h_t|h_{t-1}, \mu, \phi, \sigma_\eta &\sim N(\mu + \phi(h_{t-1} - \mu), \sigma_\eta^2) \\ h_0|\mu, \phi, \sigma_\eta &\sim N(\mu, \sigma_\eta^2/(1-\phi)) \qquad \phi \in (-1, 1), \end{aligned}$$

with y_t the return at time t and h_t the conditional variance at time t, μ a constant, and σ_η^2 the variance of the conditional variance. $\phi \in (1, -1)$ is necessary for a stationary autoregressive volatility process.

There are different filters that can be used to find the optimal estimates corresponding to the state-space model. The Kalman filter (Kastner, 2016) is a straight forward to use filter that yields high in-sample accuracy and I will focus on this filter for the Stochastic Volatility Model. Other filters that exist are more computational intensive and not considered in this study. The SVM is estimated through Markov chain Monte Carlo simulation (MCMC). "stochvol" specifically uses all without a loop (AWOL) joint sampling of all volatilities to reduce correlation between draws. It is used instead of the usual forward-filtering backward-sampling (FFBS), common for hidden Markov models. This is beneficial in reducing the computational burden generally associated with SVMs.

As prior values for the parameters standard (uninformative) priors are taken. For the level parameter μ the usual normal distribution is taken: $\mu \sim N(b_\mu, B_\mu)$. The persistence parameter $\phi \in (-1, 1)$ is chosen to be: $(\phi + 1)/2 \sim B(a_0, b_0)$, with B the beta distribution. This implies: $p(\phi) = \frac{1}{2B(a_0,b_0)}(\frac{1+\phi}{2})^{a_0-1}(\frac{1-\phi}{2})^{b_0-1}$. This distribution ensures the restriction of $\phi \in (-1, 1)$ being met. The prior of the variance is$\sigma^2 \sim B_\sigma \chi_1^2 = G(\frac{1}{2}, \frac{1}{2B_\sigma})$. This is different from the commonly employed conjugate Inverse gamma prior $\sigma^2 \sim G^{-1}(c_0, C_0)$. The prior specification follows the prior specification stipulated by Frühwirth-Schnatter and Wagner (2010).

The expression of the SVM shows that for each update of volatility, the model should be re-estimated. This is, because the expression of the conditional volatility is not dependent on observable data, as is the case for the GARCH model.

3.4 Simulation set-up

To investigate the consequences of identification loss I first perform a simulation study. This simulation study is based on a GARCH model with different identification strengths for an error term following the t-distribution and one following the standardnormal distribution. For the t-distribution it is necessary to vary the degrees of freedom ν, because the volatility models estimate ν. I range the degrees of freedom from 4 to 10 as shown in Table 3.1. Larger initial values of ν are associated with overestimation of the degrees of freedom.

As shown in Table 3.1 the constant α is fixed at 0.2, because of consistent estimation (Andrews and Soares, 2010). The range of b is similar to the range in Andrews and Cheng (2012). The largest value of b is varied to satisfy the stationarity assumption for π of 0.4. The stationarity assumption (Nelson, 1990) states that, in order to obtain stable conditional variance, the sum of β and π should not exceed one, on top of the other assumptions mentioned for the GARCH model. If the stationarity assumption is violated, a possible consequence is an explosion of the

Table 3.1: Initial values for the GARCH parameters as used in the simulation study

α	0.2						
b	0	2	4	7	9	12	
π	0.2	0.4	0.6	0.8			
ν	4	5	6	7	8	9	10
T	150	250	500				

Table 3.2: Relation of b to the β parameter

Number of observations	b: 0	2	4	7	12
150	0	0.16	0.33	0.57	
250	0	0.13	0.25	0.44	
500	0	0.09	0.18	0.31	0.54

Table 3.3: Identification Categories as defined by Andrews and Cheng (2012)

Category	β_T Sequence	Identification property of π
I(a)	$\beta_n = 0 \forall n \geq 1$	Unidentified
I(b)	$\beta_n \neq 0$ and $n^{1/2}\beta_n \to b \in R^{d\beta}$ (and, hence, $\|\beta_n\| = O(n^{-1/2})$)	Weakly identified
II	$\beta_n \to 0$ and $n^{1/2}\|\beta_n\| \to \infty$	Semi-strongly identified
III	$\beta_n \to \beta_0 \neq 0$	Strongly identified

conditional volatility, which I show in Chapter 4.3. Furthermore, parameter estimates explode and the β parameter is undervalued. The identification strengths corresponding to these values are also similar to Andrews and Cheng (2012) as shown in Table 3.3.

Table 3.2 shows the values of b in relation to the real value of the parameter β, which is defined as $b/\sqrt{(T)}$. I adopt the values for b that were found optimal for the ARMA model by Andrews and Cheng (2012). Ma et al. (2007) show that the t-value shows less of a size distortion from a b of 7.07, indicating that smaller values of b indicate a weakly identified GARCH model.

I vary π from 0.2 to 0.8 as shown in Table 3.1, because this parameter is not consistent (Andrews and Soares, 2010). I vary the sample size from 150 to 500, because smaller sample sizes show more clear identification problems, because the b value is higher for larger sample sizes if the β parameter is the same. The paper for which I show that the parameter estimates are not correct (Bollerslev, 1987) uses a sample size of 453.

A number of simulation runs is fixed on 10000 to prevent different results for different simulation runs. Lower number of simulation runs show differences between two runs between the parameter estimates distributions. Since the GARCH model is the true model, one would expect that this model is the best model.

As example, for a sample size of 150, a π of 0.4, and a b of 0, first a data set is simulated with these initial values for the parameters. Then, a GARCH(1,1)-model, an ARCH(1)-model, and an SVM is estimated on this simulated data set. Afterwards, the parameter estimates of the GARCH(1,1)-model are stored for each simulation, such that the empirical distribution of the parameter estimates can be examined. This is, because parameter estimates should be normally distributed and I show that this is not the case for the not and weakly identified GARCH(1,1)-model. Furthermore, I check how large b should be in order for normal distribution of the parameter estimates of the GARCH(1,1)-model. With a not identified GARCH(1,1)-model, it is meant the GARCH(1,1)-model for which the π parameter is not identified.

The prediction of the volatility h_t of the GARCH(1,1), the ARCH(1), and the SVM are stored, in order to compare the predictive power of the three models. The predicted returns of the models are computed as follows, to minimize variation between predictions due to randomness: Providing the error is normally distributed, one vector of one thousand random draws of the standard normal distribution zero (∼N(0,1)) is adopted for all models and combinations of initial parameter values. These draws are multiplied by the square root of the predicted variance and stored as predicted returns. Note that therefore the draws do not vary between models and variation in return estimates is due to variation between models.

Supposing the error terms follow a t-distribution, a vector of one thousand draws of the uniform distribution between zero and one (∼ U(0,1))are multiplied by 100 000 and stored as constant vector s. Then, for all estimated ν, that may differ between simulations, 100 000 draws from the t-distribution with degrees of freedom ν are sorted. For each of the one thousand predicted variances, one of the draws of the vector s is taken as index of the vector of the draws of the degrees of freedom ν. The vector with the resulting draws of the t-distribution with degrees of freedom ν is multiplied with the square root of the predicted variance and stored as predicted returns.

3.5 Value at risk

The value at risk (VaR) is computed in order to compare the models in terms of outliers prediction. The VaR is the boundary of the confidence interval in which the returns should lie with 95% certainty. The expectation for VaR is that in 5% of cases the return falls outside of the boundaries, which means 50 times out of 1000. The boundary of the confidence interval is the quantile of the preferred distribution multiplied by (estimated) volatility. Provided the error follows a t-distribution, the degrees of freedom for the quantile are the estimated ν and may differ between models and simulations. Assuming that the error follows the normal distribution, the associated 95% quantile is 1.96.

3.6 Loss functions

To compare the predictive power of the aforementioned models I apply the Diebold-Mariano test (Diebold and Mariano, 1991) based on two loss functions. Namely, the Mean Squared Prediction Error (MSPE), corresponding to power of 2 and the Mean Absolute Error (MAE), corresponding to power of 1.

$$MSPE = \frac{1}{P} \sum_{t=T}^{t=T+P-1} (y_{t+1} - \hat{y}_{t+1|t})^2$$
$$MAE = \frac{1}{P} \sum_{t=T}^{t=T+P-1} |y_{t+1} - \hat{y}_{t+1|t}|$$

with P the number of 1-step ahead forecasts. The MSPE penalizes larger deviations from the real value more than the MAE.

3.7 Least favourable critical value

As robust critical value taken to be appropriate for the GARCH(1,1)-model I take the least favourable critical value. For combinations of π and b with no or weak identification, e.q. $b = 0, 2, 4$, I check which t-value corresponds. For 10 000 simulations I first sort these values and then take the 9500 largest t-value as the robust t-value for this combination of initial parameter values. This is done for each combination of parameter values. As a last step the largest value over all parameter combinations is taken as the robust critical value. I expect that the least favourable critical value is highest with higher identification, because the true value of b is higher and b is not affected by the identification problem.

Chapter 4

Evaluation

This chapter discusses the performance of two GARCH models with different specifications for the error term, in various states of identification loss (Table 3.3). I split the GARCH model in one with a normal distribution as specification for the error term (Engle, 1982) and one with a t-distribution as specification (Bollerslev, 1987). Chapter 4.1 and 4.2 evaluate the parameter estimates of the GARCH model on simulated data with defined initial parameter values (Table 3.1). Chapter 4.3 compares the predictive power of the GARCH model to a state of the art stochastic volatility model (SVM) and a baseline ARCH model. For predictive purposes only the time series with significant β parameter estimates are used to limit the scope of this study. Chapter 4.3.3 compares the results, obtained on simulated data, to those obtained from real data. The real data to which it is compared is the S&P 500 index from 1947 to 1984, the same as used in Bollerslev (1987). I show that this data suffers from identification loss. The distributions of the parameter estimates should follow the normal distribution in case of strong identification (Andrews and Cheng, 2012).

However, strong identification is not obtained in this study as shown through the empirical parameter estimates distributions. The parameter estimates obtained through Quasi Maximum Likelihood Estimation, under weak stationary assumptions, of the GARCH model are proven to be asymptotic normal (Berkes et al., 2003). An approximation of the normal distribution is necessary to obtain strong identification. The initial value of b of 12 should correspond to strong identification (Andrews and Cheng, 2012). Nonetheless, the empirical parameter estimates distributions do not follow the normal distribution according to the Jaque-Bera test (Jarque and Bera, 1980). Note that when I increased the sample size from 500 to 16000 and b from 12 to 33 the β and π parameter followed a normal distribution, based on the Jaque-Bera test.

For this study I used the initial values of the parameters as specified in Table 3.1. The results for the sample size of 500 are discussed in detail in this study, because the results for the other sample sizes do not differ much from this sample size. Moreover, studies mostly use larger sample sizes (Table 2.1), which makes a sample size of 500 more relevant. Note that the identification problem is as or more apparent for larger sample sizes than for smaller sample sizes. The number of times the b parameter is significant, in case of an unidentified GARCH model, is higher for larger sample sizes than for smaller sample sizes (Table 4.1 and 4.4). Other sample sizes considered were 150 and 250. Results corresponding to those sample sizes are available for the interested reader. The GARCH model with b parameter equal to zero and therefore an unidentified π parameter is used interchangeably with an unidentified GARCH model.

Table 4.1: Number of times the t-value associated to the β parameter is larger than 1.98, based on 10 000 simulation runs

	$b = 0$	$b = 2$	$b = 4$	$b = 7, 9, 12$
$T = 150$				
$\pi_{0.2}$	48.50%	90.84%	98.77%	99.97%
$\pi_{0.4}$	48.34%	92.80%	99.45%	99.99%
$\pi_{0.6}$	48.00%	95.95%	99.76%	
$\pi_{0.8}$	48.89%	98.43%		
$T = 250$				
$\pi_{0.2}$	52.38%	93.58%	99.44%	100%
$\pi_{0.4}$	52.65%	94.97%	99.83%	100%
$\pi_{0.6}$	52.72%	97.31%	99.93%	
$\pi_{0.8}$	53.38%	99.07%		
$T = 500$				
$\pi_{0.2}$	54.32%	95.63%	99.92%	100%
$\pi_{0.4}$	57.61%	96.99%	99.91%	100%
$\pi_{0.6}$	55.18%	98.34%	99.99%	
$\pi_{0.8}$	58.05%	99.66%	100%	

4.1 Simulation results normally distributed error

This section shows the issues concerning the parameter estimates in case of identification loss for the GARCH model with a normally distributed error specification. The empirical distributions of the unidentified parameter estimate π clearly does not follow the normal distribution for b of 0, 2, and 4 (Chapter 4.1.1). The distributions become uniform and bi-modal in case of identification loss. Furthermore, the empirical distributions of the associated t-values do not follow the t-distribution. The distributions of the empirical t-values seem more geometrically distributed with large outliers, assigning more significance to lower values than they should. Therefore, the robust critical value taken as measure to determine the identification of the GARCH model should be much higher than the non-robust critical t-value of 1.98.

Table 4.1 re-emphasizes the issue of identification loss. If nonidentification occurs, the GARCH model is classified as correct or identified model in 50% of cases instead of the expected 5%. Furthermore, it shows that in general the significance of the GARCH model is highly overvalued. This is, because the model will be picked as correct and identified model in 100% of cases in case of b of 12, instead of the expected 95%.

Chapter 4.1.1, 4.1.2, and 4.1.3 show the plots containing the empirical parameter distributions show with a vertical line the initial value of the corresponding parameter. The initial values correspond to the values as specified in Chapter 3.4 Table 3.1. The plots containing the t-values show with a vertical line the value of 1.98 as critical value for a 5% significance at a sample size of 500. *A*-*D* stands for the different values of π ranging from 0.2 to 0.8 respectively. The color blue corresponds with a t-value larger than 1.98, e.q. a significant β parameter estimate. The color red corresponds with a t-value smaller or equal to 1.98, an insignificant β parameter and clear nonidentified GARCH model. Firstly, I show the results for the π parameter in Chapter 4.1.1. Secondly, I show the results for the α parameter in Chapter 4.1.2. Thirdly, I show the results for the β parameter in Chapter 4.1.3. Lastly, I show the robust critical values to determine the identification of the GARCH model in Chapter 4.1.4.

4.1.1 Results for π

This section shows the empirical parameter estimate distributions and associated t-values of the π parameter. There is clear variation between the parameter estimate distributions for the various specifications of the initial parameter values (Table 3.1) as shown in Figure 4.1 to 4.4.

Andrews and Cheng (2012) describe in Figure 1 the empirical parameter distributions for the nonidentified parameter in the ARMA model. Similar to the ARMA model, the GARCH model parameter estimates encompass the entire parameter space in case of identification loss as shown in Figure 4.1B to 4.3B. However, whereas the ARMA model has a large build-up at both boundaries of the parameter space in case of identification loss, the GARCH model only has a large build-up around zero. The fact that there is no build up around one has to do with the nonstationary restriction placed upon the GARCH model, which means that $\beta + \pi < 1$. The restriction is in place to ensure the estimation of unconditional volatility. Furthermore, the b of 4 does not seem to correspond to a semi-strongly identified GARCH model, due to the fact that a large build-up around zero remains. If the build-up around zero is ignored, the results of the ARMA model are similar to the results of the GARCH model (Figure 4.1B to 4.4B). The empirical densities of the t-statistic as shown in Figure 4.5B to 4.8B are not similar to the empirical densities of the t-statistic of the ARMA model as shown in Andrews and Cheng (2012) Figure 3. The scale of the t-statistic is much larger for the GARCH model than for the ARMA model. Furthermore, the boundaries on the ARMA model ensure a different scale for both models. Figure 4.8B shows that in case of b of 12 the distribution of the t-statistic approximates a t-distribution. In this instance, the π estimates seem to be overestimated, because the t-statistic is estimated rather largely.

Figure 4.1A to 4.4A show that for small initial values of π, the GARCH model cannot correctly estimate the π value. Only for b of 12, semi-strong identification occurs as shown in Figure 4.4A. This is in line with the results of Ma et al. (2007). They show that for small π, large b of 7 or higher should be used for correct identification. Figure 4.5A to 4.8A show the corresponding t-values of the π parameter. It is clearly visible that in all cases other than for $b = 12$, the t-values explode and do not converge to a t-distribution. This is understandable, since the parameter estimates are located on the whole parameter space of π. Larger t-values indicate very small standard errors, an issue also addressed by Ma et al. (2007). The parameter π is therefore often overestimated or deemed more significant. Similar to what Ma et al. (2007) show, even though the GARCH model often estimates π near zero, the t-value is rejected in 40% of cases. It means that the confidence interval of the parameter estimate is infinitesimal, such that zero is not in the confidence interval.

For larger values of π as shown in Figure 4.1C,D to Figure 4.3C,D the parameter estimates show larger build ups around the initial value of π than for the aforementioned π of 0.2 and 0.4. Figure 4.3D shows that the combination of large initial values of π and b ensures no build up around zero. The corresponding t-values as shown in Figure 4.5C,D to 4.7C,D show explosion in the t-statistic values. Figure 4.7D shows that the combination of large initial values of π and b ensures the empirical distribution approximating the t-distribution.

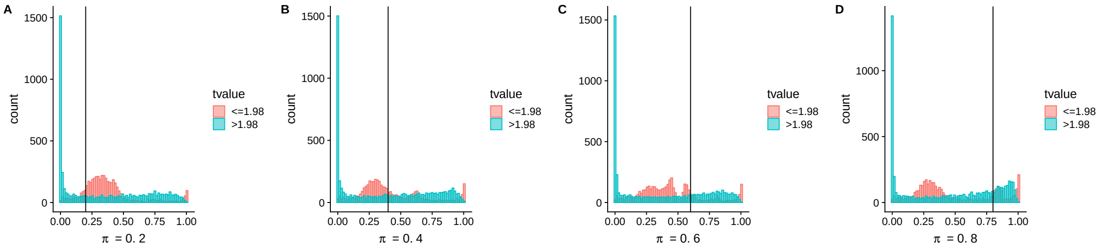

Fig. 4.1: Histogram of π for b = 0, T = 500, $\alpha = 0.2$

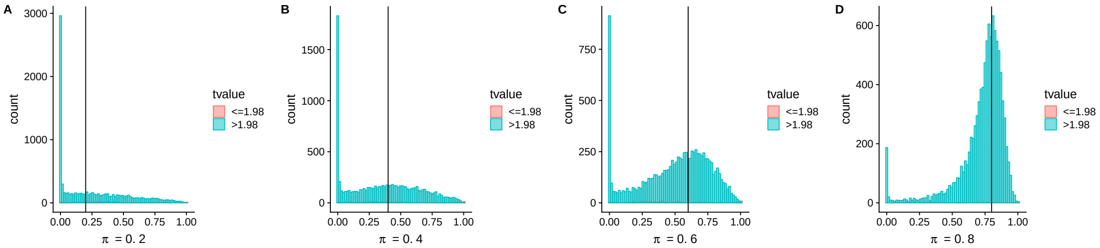

Fig. 4.2: Histogram of π for b = 2, T = 500, $\alpha = 0.2$

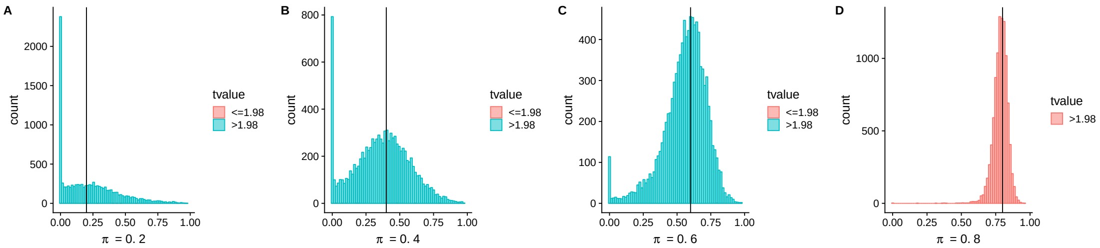

Fig. 4.3: Histogram of π for b = 4, T = 500, $\alpha = 0.2$

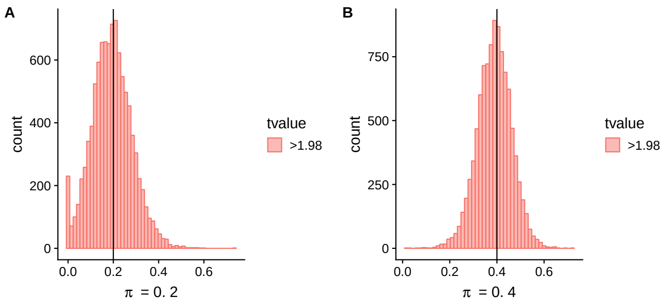

Fig. 4.4: Histogram of π for b = 12, T = 500, $\alpha = 0.2$

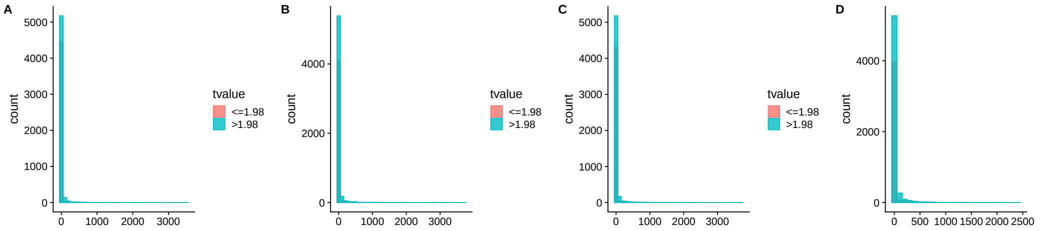

Fig. 4.5: Histogram of the t-values corresponding to the π parameter for b = 0

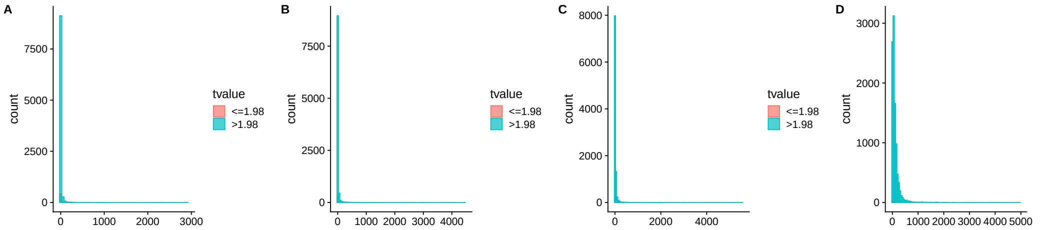

Fig. 4.6: Histogram of the t-values corresponding to the π parameter for b = 2

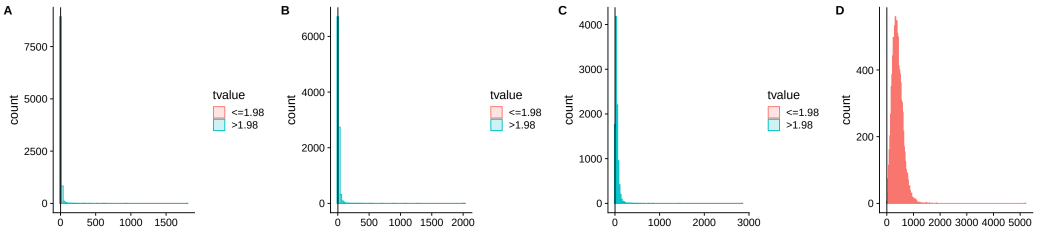

Fig. 4.7: Histogram of the t-values corresponding to the π parameter for b = 4

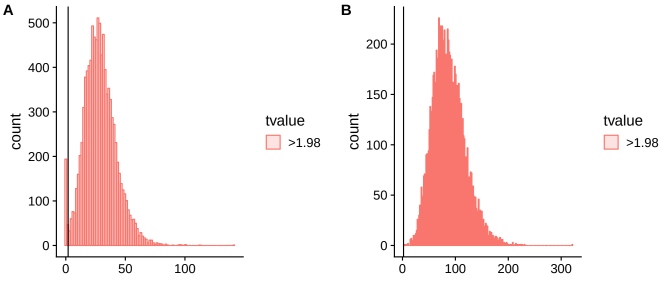

Fig. 4.8: Histogram of the t-values corresponding to the π parameter for b = 12

4.1.2 Results for α

This section shows the empirical parameter estimate distributions and associated t-values of the α parameter. Interestingly, there is less variation between the parameter estimate distributions for the various specifications of the initial parameter values (Table 3.1) as shown in Figure 4.9

to 4.12 than for the π parameter. The variation is more visible in case of more identification loss as shown in Figure 4.9 and 4.10. The reason of the large variation in case of more identification loss is the negative correlation between the α and π parameter. These findings are similar to the findings of Ma et al. (2007). Another difference to the empirical distributions of the π parameter is the lack of build-up around zero.

Figure 4.9A to 4.12A show that the distribution of α changes from bimodal in case of non-identification to nearnormal in case of b of 12 and π of 0.2. The associated t-values change from a large peak around zero, to more left centered at larger values, such as 120 in case of strong identification and approximating a t-distribution. This is shown in Figure 4.13A to 4.16A. The peak of α is only at its true initial value in case of strong identification. In other cases, the α parameter is overestimated slightly. In case of nonidentification, there is a small build up on the right and a larger build up on the left of the α parameter.

Figure 4.9B to 4.12B also show that the distribution of α changes from bimodal in case of non- and weak identification to nearnormal for b of 4 and 12 for π of 0.4. The associated t-values show similar findings as for π of 0.2 as shown in Figure 4.13B to 4.16B. However, the range of the estimated t-values decreases. There is a peak around the initial value of α of 0.2 if the GARCH model is at least weakly identified with b of 2. Results are similar for π of 0.6 as shown in Figure 4.9C to 4.12C. Furthermore, the t-values are similar as shown in Figure 4.13C to 4.16C. The scale of the α parameter increases, however, the scale of the t-values decreases. The α parameter is more significant if the π parameter is smaller. This has to do with the negative correlation of the α parameter with regards to the π parameter.

Figure 4.9D to 4.12D show the parameter estimates distributions for π of 0.4. Only in case of nonidentification there are signs of a bimodal distribution. If the model is at least weakly identified, the empirical distributions approximate the normal distribution with some outliers to higher values of α. Figure 4.13D to 4.16D show that the scale of the t-values is smaller for larger π. It is interesting to note that negative skewness occurs for small π as shown in Figure 4.9A to 4.11A. For higher values of π more positive skewness occurs.

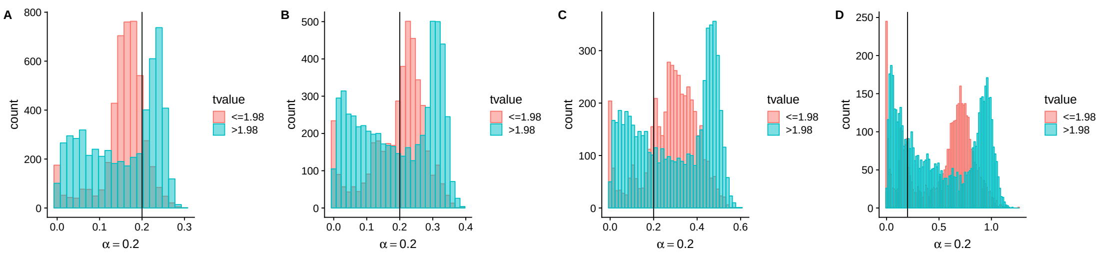

Fig. 4.9: Histogram of α for b = 0, T = 500, α = 0.2

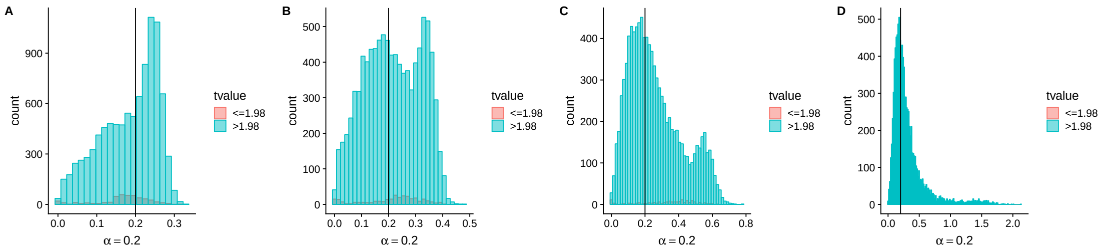

Fig. 4.10: Histogram of α for b = 2, T = 500, α = 0.2

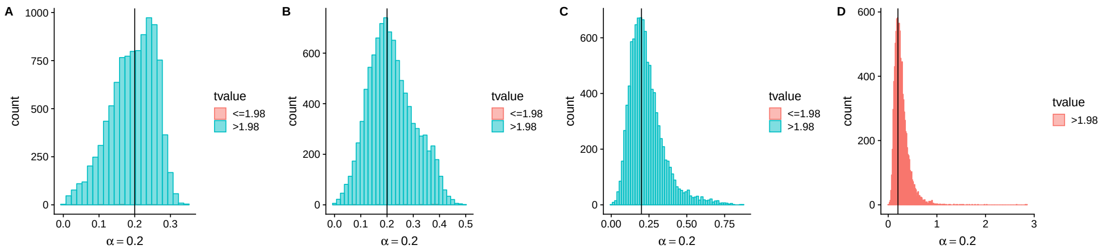

Fig. 4.11: Histogram of α for b = 4, T = 500, α = 0.2

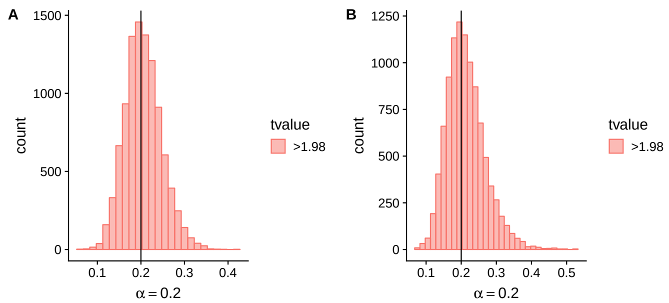

Fig. 4.12: Histogram of α for b = 12, T = 500, α = 0.2

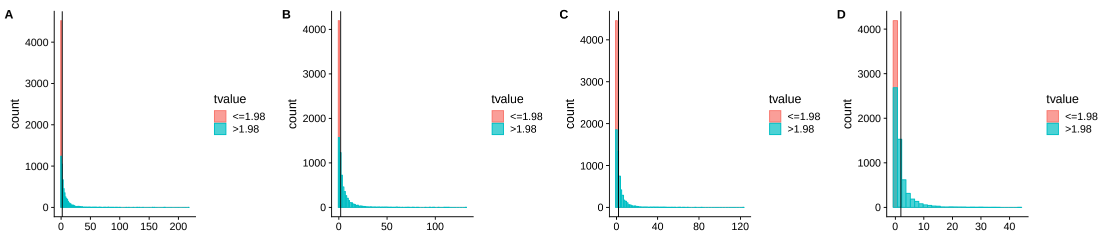

Fig. 4.13: Histogram of the t-values corresponding to the α parameter for b = 0

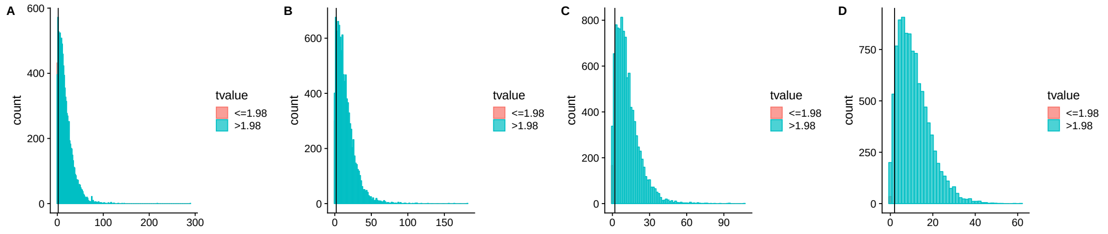

Fig. 4.14: Histogram of the t-values corresponding to the α parameter for b = 2

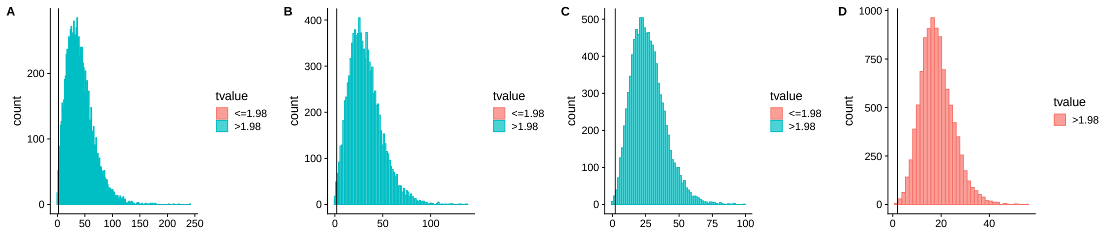

Fig. 4.15: Histogram of the t-values corresponding to the α parameter for b = 4

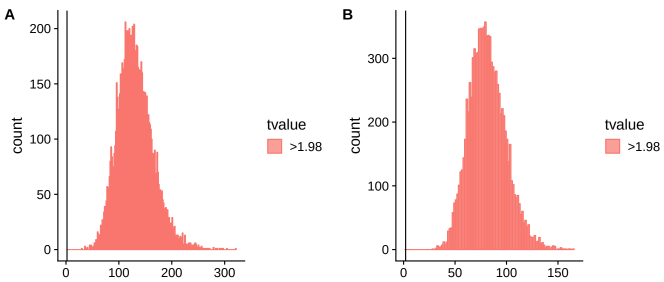

Fig. 4.16: Histogram of the t-values corresponding to the α parameter for b = 12

4.1.3 Results for β

This section shows the empirical parameter estimate distributions and associated t-values of the β parameter. For the other parameters it is more interesting to mention the complete identification space, because the values of b clearly affect the empirical distributions. I show the non- and weakly identified GARCH model for β. This is, because there is no clear difference between the results for the different b and π values as is the case for both α and π. The results for b of 4 and 12 are found in Appendix A Figure A.1 to A.4. The distributions of the t-values follow a near t-distribution and the distributions of the parameter estimates follow a near normal distribution for b of 4 and 12.

The value of β corresponding to b of 0, 2, 4, and 12 are 0, 0.09, 0.18, and 0.54 respectively as shown in Table 3.2. The β parameter is shown not to suffer the same problem, similar to the findings of Andrews and Cheng (2012), since this parameter causes the problem. Only in case of nonidentification, the β parameter is not well identified as zero. In that case, there is a large peak around zero, however, the t-values are in around 55% of cases higher than 1.98

as shown in Table 4.1. Interestingly, for all identification cases, the value of the β parameter deemed significant overlaps with the value in case of insignificance. It reiterates the problem with t-values of the GARCH model. Very small values of β have too large t-values, showing that the critical value of the t-value should be larger, to account for identification loss.

The empirical t-values are very large in case of non-identification, which causes an issue, in recognizing the identification problem. Especially, since in case of $b = 12$, the range of t-values is smaller than in case of weak identification as shown in Appendix A Figure A.4. Larger t-values correspond to smaller estimated standard errors. It shows that in case of nonidentification, the standard errors are severely underestimated.

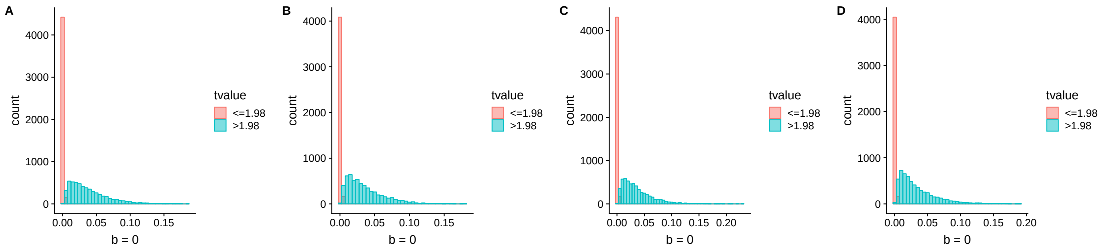

Fig. 4.17: Histogram of β for b = 0, T = 500, $\alpha = 0.2$

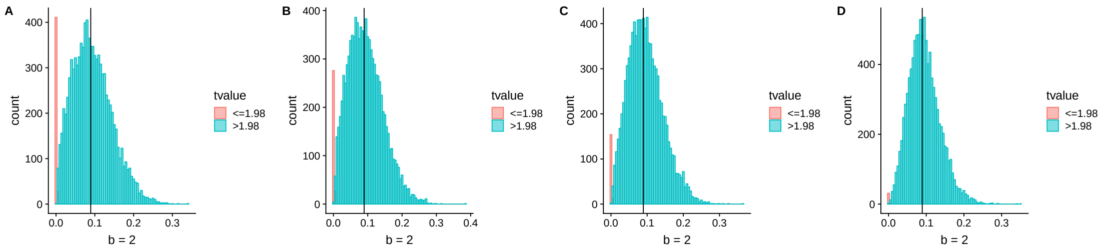

Fig. 4.18: Histogram of β for b = 2, T = 500, $\alpha = 0.2$

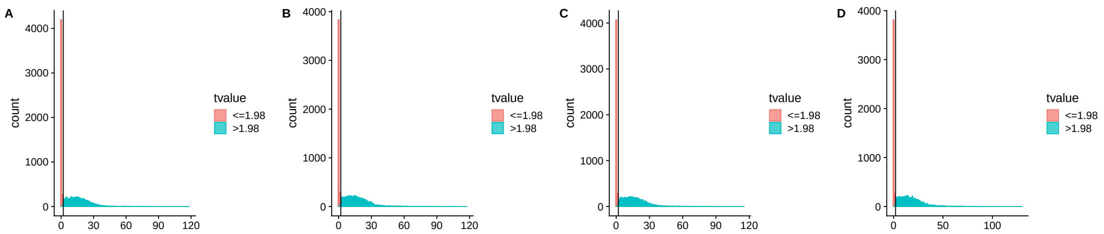

Fig. 4.19: Histogram of the t-values corresponding to the β parameter for b = 0

Table 4.2: Robust critical values for different identification categories for b

	b		
π	0	2	4
0.2	30.803	44.069	52.571
0.4	32.023	46.609	55.699
0.6	31.671	54.68	
0.8	34.876		

Table 4.3: Robust critical values for different identification categories for π

	b	
π	2	4
0.2	21.379	40.155
0.4	20.598	39.163

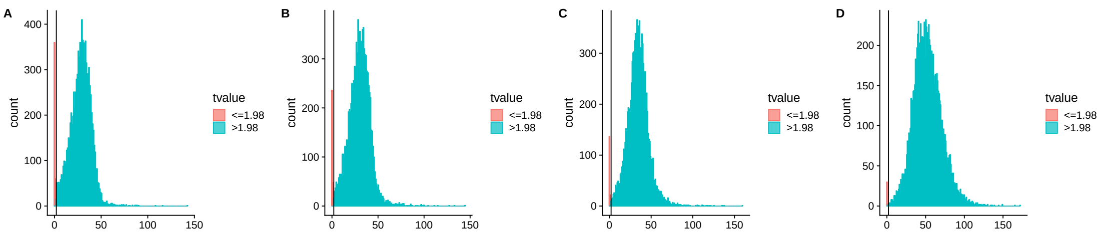

Fig. 4.20: Histogram of the t-values corresponding to the β parameter for b = 2

As an important note, Figure 4.21 shows that there is often an overestimation of the sum of $\pi + \beta$, if the initial sum is high. This means that the model is dangerously close to nonstationary, which is clearly visible, even if stationary constraints are put in place. The sum in case of identification loss is shown in Appendix A Figure A.5 to A.7. It shows that there is no pattern visible in the sum in case of identification loss. This is due to the erratic behaviour of the π parameter in case of identification loss as shown in Figure 4.1 to 4.3. It is not due to the estimated value of β, since I showed the near normality for this parameter.

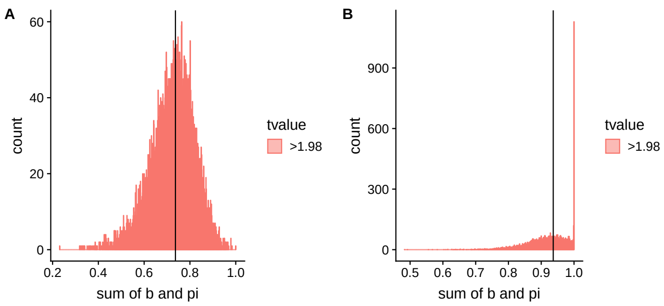

Fig. 4.21: Histogram of the sum of $\pi + \beta$ for b = 12

4.1.4 Robust critical values for the t-test

Chapter 4.1.1 and 4.1.2 showed that the parameter estimates are not reliable in case of identification loss. Therefore, it is important to have a method to determine the identification, or

Table 4.4: Number of times β parameter significant (and imaginary) based on the t-value $\nu = 4$ and 10 000 simulation runs

	$b = 0$	$b = 2$	$b = 4$	$b = 7, 9, 12$
$T = 150$				
$\pi_{0.2}$	0.40% (32.30%)	3.96% (10.89%)	16.66% (4.08%)	46.98% (1.11%)
$\pi_{0.4}$	0.55% (32.94%)	3.86% (9.27%)	19.87% (2.75%)	60.60% (0.51%)
$\pi_{0.6}$	0.76% (32.42%)	5.13% (7.91%)	32.43% (2.22%)	
$\pi_{0.8}$	0.84% (33.12%)	16.07% (6.42%)		
$T = 250$				
$\pi_{0.2}$	0.39% (30.62%)	6.52% (7.91%)	26.52% (2.23%)	82.50% (0.09%)
$\pi_{0.4}$	0.46% (30.55%)	6.45% (6.79%)	31.68% (1.44%)	89.96% (0.06%)
$\pi_{0.6}$	0.68% (30.72%)	8.85% (4.79%)	45.42% (0.86%)	
$\pi_{0.8}$	0.96% (30.93%)	23.53% (3.12%)		
$T = 500$				
$\pi_{0.2}$	0.39% (30.57%)	12.59% (5.65%)	48.24% (1.00%)	99.36% (0.03%)
$\pi_{0.4}$	0.42% (30.67%)	13.93% (4.22%)	56.03% (0.51%)	99.84%
$\pi_{0.6}$	0.63% (29.83%)	19.15% (2.77%)	71.99% (0.16%)	
$\pi_{0.8}$	0.92% (29.85%)	44.45% (1.08%)	96.38% (0.05%)	

loss thereof, of the GARCH model. One method to determine the identification of the GARCH model is using a robust critical t-value. The GARCH model is not well identified for b of 0, 2, and 4 for different values of π and Table 4.2 shows that the robust critical value of b should be 55.699 instead of 1.98. Furthermore, the GARCH model is not well identified if π is 0.2 and 0.4 for small b and Table 4.3 shows that the robust critical t-value of π should therefore be 40.155 instead of 1.98. Note that in case of nonidentification, the critical value of b is enough, because the π parameter does not show any peak at that stage. The reason for the robust critical value of the π parameter is the large overestimation of the t-values associated to the π parameter estimates.

4.2 Simulation results t-distributed error

This section shows the issues concerning the parameter estimates in case of identification loss for the GARCH model with a t-distributed error specification (GARCH-t model). For large sample sizes, as conferred by the law of large numbers, the distribution of the parameter estimates converges to a normal distribution. Similar to the normal distribution, the empirical distributions of the unidentified parameter estimate π clearly does not follow the normal distribution for b of 0, 2, and 4. The distributions become uniform in case of identification loss. However, the t-values show the main contrast between the error specification following a normal and one following a t-distribution. The scale of the empirical distribution of the t-statistic associated with the β parameter is smaller for the t-distributed error than for the normally distributed error. Furthermore, the β parameter is significant in more than 90% of cases if the GARCH model was at least weakly identified for the normally distributed error as shown in Table 4.1. This is far greater than the number of times the t-distribution attributes significance to the β parameter as shown in Table 4.4.

Interestingly, given nonidentification (b of zero), the t-statistic is imaginary in 30% of cases as shown in Table 4.4. The t-statistic is never imaginary for the error following the normal distribution. This shows that the problem of identification loss is less imminent if the error is assumed to follow the t-distribution than if it is assumed to follow the normal distribution. The t-values and standard errors of parameter estimates are based on the inverse of the Hessian matrix. Imaginary t-values are indicators of a Hessian matrix that is not positive semi-definite. 0.54% of the β parameter estimates show a t-value of larger than 1.96. If imaginary t-values occur, one could see that the model is incorrect, because of incorrect standard errors, e.q. negative ones. Therefore, I do not show the estimated t-values of the nonidentified case.

Initially, I varied the degrees of freedom ν from 4 to 10. The parameter estimate distributions do not differ much in between, except for the degrees of freedom parameter ν. It explodes for larger initial values of ν. I only show the results for a ν of 4. Results for other initial ν are available upon request. The t-values are shown for the weakly identified to the strongly identified GARCH model.

The plots containing the empirical parameter distributions show with a vertical line the initial value of the corresponding parameter. The plots containing the t-values show with a vertical line the value of 1.98 as critical value for a 5% significance at a sample size of 500. *A-D* stands for the different values of π ranging from 0.2 to 0.8 respectively. The color green corresponds with a t-value larger than 1.98, e.q. a significant β parameter estimate. The color red corresponds with a t-value smaller or equal to 1.98, an insignificant β parameter and clear nonidentified GARCH model. The color blue corresponds with an imaginary t-value, e.q. a non-converged GARCH model. Firstly, I show the results for the π parameter. Secondly, I show the results for the α parameter. Thirdly, I show the results for the β parameter. Fourthly, I show the results for the ν parameter. Lastly, I show the robust critical values to determine the identification of the GARCH model.

Note that this section will be a comparison to Section 4.1. This is, because the only difference between the two models is the error specification and therefore the parameter estimates distributions show similar distributions for both specifications. The main difference lies in the scale of the t-statistic as will be shown. I will show you that a t-distribution as error specification is not a preferred error assumption for the GARCH model, due to identification loss on one side for low b and nonstationarity on the other side for high b. For more detailed comparison between the different initial values I refer to Section 4.1.

4.2.1 Results for π

This section mentions the empirical parameter estimate distributions and associated t-values of the π parameter. The distributions are found in Appendix B Figure B.1 to B.4 and the t-values are found in Appendix B Figure B.5 to B.5. The parameter estimate distributions as shown in Appendix B Figure B.1 to B.4 are very similar to those shown in Section 4.1.1 Figure 4.1 to 4.4. Similarly, there is clear variation between the parameter estimates distributions for the various specifications of the initial parameter values (Table 3.1) as shown in Appendix B Figure B.1 to B.4. Furthermore, the t-values also explode for the π parameter in case of identification loss and approximates a t-distribution for b of 12 and π of 0.4 as shown in Appendix B Figure B.5 to B.7.

The main difference between the two specifications of the error term is that if the assumption is that the GARCH model follows a t-distribution, the β parameter is not as often significant. It is noteworthy to mention that if the β parameter is insignificant, for higher initial values of b, the π parameter estimate shows a peak around the real value as well. This is different from the normal distribution, where the insignificant β do not allow for the π to show a peak within the parameter estimate distributions. This is mainly due to the fact that there are not many insignificant β for the normally distributed error specification.

Another difference is that the scale of the parameter estimate distribution is larger for b of 12, whilst the scale of the t-statistic is smaller. This is shown in Appendix B Figure B.4 and B.7 and Chapter 4.1.1 Figure 4.4 and 4.8. This means that the significance attributed to π is larger if the error specification follows the normal distribution than if it follows the t-distribution. An explanation can be the nonstationarity issue as shown in Figure 4.36. The number of times the model is dangerously close to one is more than two times as high for the error specification following the t-distribution.

4.2.2 Results for α

This section shows the empirical parameter estimate distributions and associated t-values of the α parameter. The parameter estimate distributions as shown in Figure 4.22 to 4.25 are similar to those shown in Chapter 4.1.2 Figure 4.9 to 4.12. Similarly, the α and π parameters are negatively correlated, which explains the erratic behavior of the α parameter in case of identification loss. Furthermore, the scale of the parameter estimate distributions show similar lengths for both error specifications.

The main difference is in the size of the scale of the t-statistic as shown in Figure 4.26 to 4.28. The size of the scale is much smaller for an error following the t-distribution. This means that the confidence interval of α is larger in this case. A normally distributed error shows a smaller confidence interval for α and therefore infers more certainty in the estimated value of the parameter α. Another difference is that if the assumption of the error specification follows a t-distribution, the β parameter is not as often significant as if the error specification follows a normal distribution. It is noteworthy to mention that if the β parameter is insignificant, for higher initial values of b, the α parameter estimate shows a peak around the real value as well. This is different from the normal distribution, where the insignificant β do not allow for the α to show a peak within the parameter estimate distributions. This is mainly due to the fact that there are not many insignificant β for the normally distributed error specification.

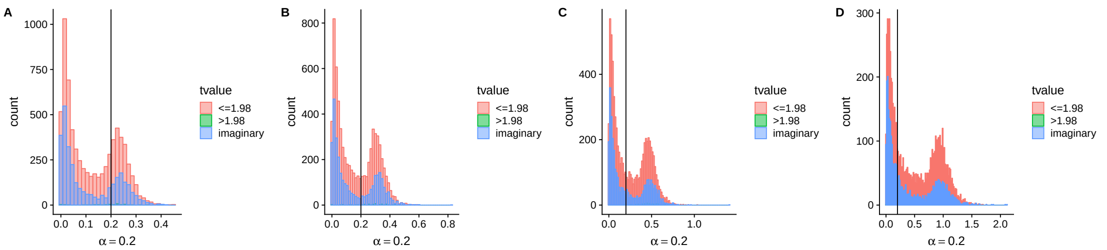

Fig. 4.22: Histogram of α for b = 0, T = 500, α = 0.2

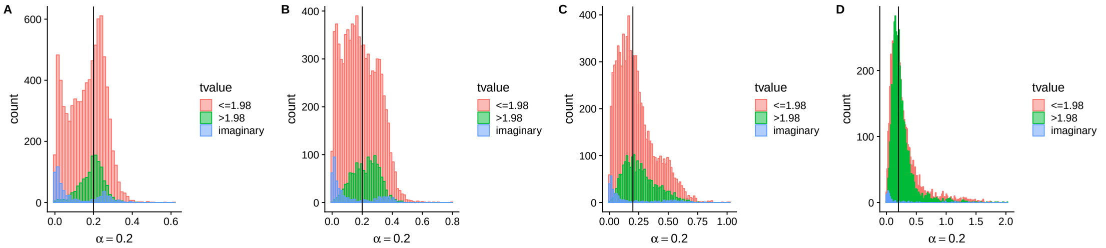

Fig. 4.23: Histogram of α for b = 2, T = 500, $\alpha = 0.2$

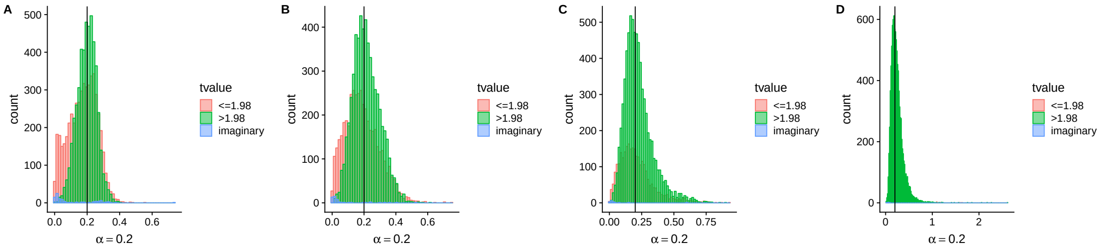

Fig. 4.24: Histogram of α for b = 4, T = 500, $\alpha = 0.2$

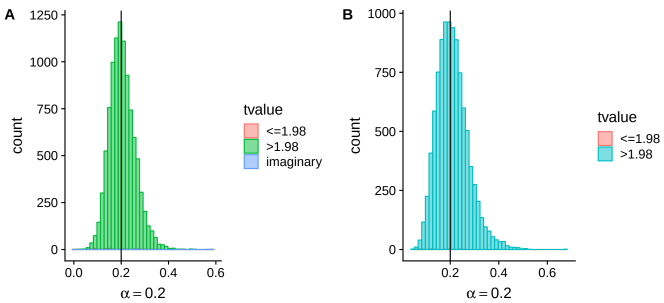

Fig. 4.25: Histogram of α for b = 12, T = 500, $\alpha = 0.2$

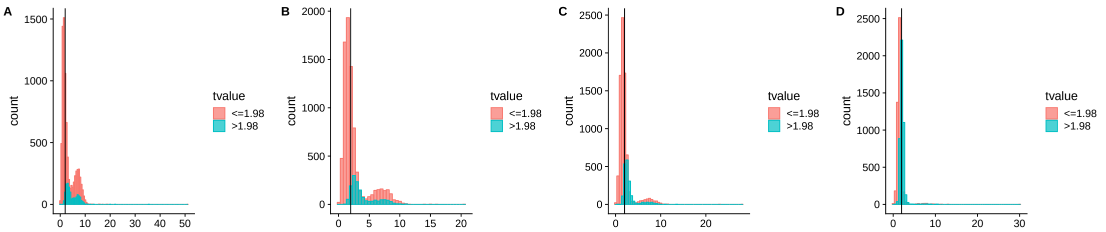

Fig. 4.26: Histogram of the t-values corresponding to the α parameter for b = 2

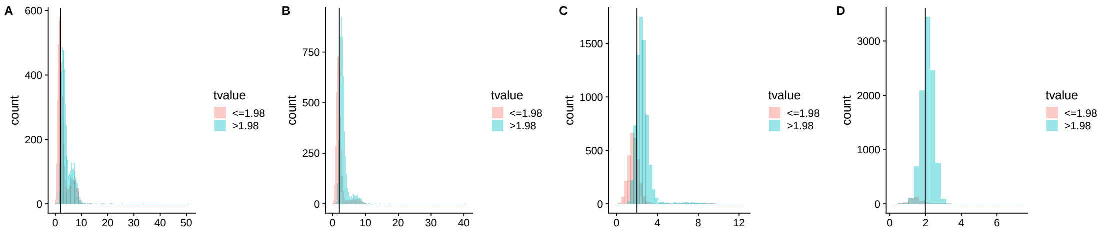

Fig. 4.27: Histogram of the t-values corresponding to the α parameter for b = 4

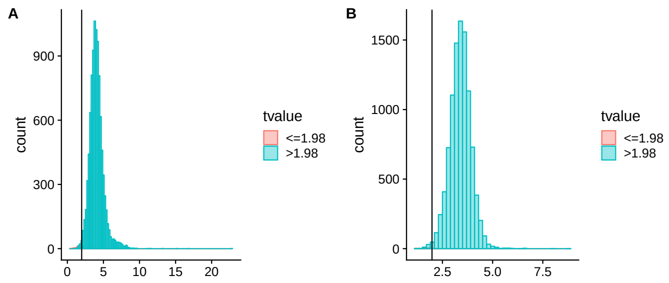

Fig. 4.28: Histogram of the t-values corresponding to the α parameter for b = 12

4.2.3 Results for β

This section shows the empirical parameter estimate distributions and associated t-values of the β parameter. The parameter estimate distributions as shown in Figure 4.29 to 4.32 are overall similar to those shown in Chapter 4.1.3 Figure 4.17 to Appendix A Figure A.2. Similarly, there is not much difference for the different values of π. However, there are some major differences between both error specifications for the β parameter. One difference is the range of the t-statistic of the β parameter. As is recurrent, the range is much smaller as shown in Figure 4.33 to 4.35 than for the normally distributed error specification as shown in Chapter 4.1.3 Figure 4.19 to Appendix A Figure A.4. Furthermore, the distribution of the t-statistic approximates the t-distribution only for b of 12 and b of 4 and large π of 0.8. However, the problem that will become apparent is the nonstationary issue in these cases.

Another major difference is that the β parameter is imaginary in part of the parameter space as shown in Figure 4.29. Furthermore, there is a clear indication that for small b of 2, the real value corresponds to an insignificant β. This is, because the peak, around the real value of b of 2, corresponds with a t-statistic lower than 1.98 as clearly shown in Figure 4.30. It becomes more clear if I consider Figure 4.31. There are two peaks, of which one corresponds with an insignificant β and one with a significant β. If you look closely, you will see that the first peak corresponds with the b value of 2.

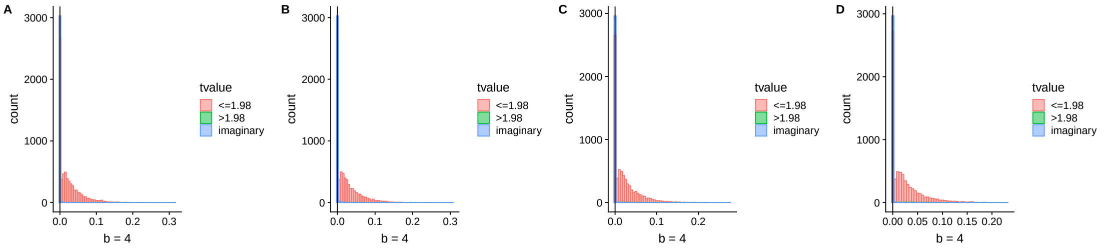

Fig. 4.29: Histogram of β for b = 0, T = 500, $\alpha = 0.2$

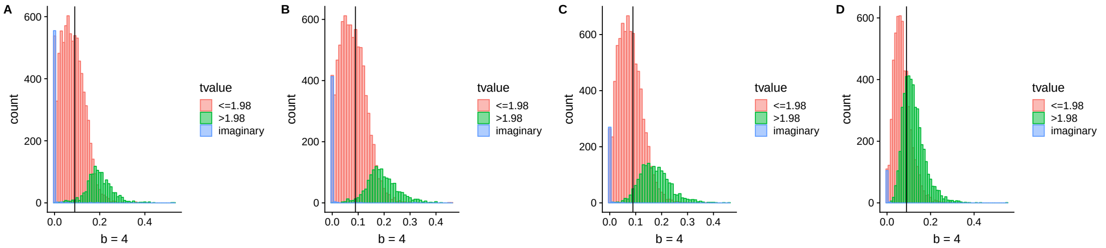

Fig. 4.30: Histogram of β for b = 2, T = 500, $\alpha = 0.2$

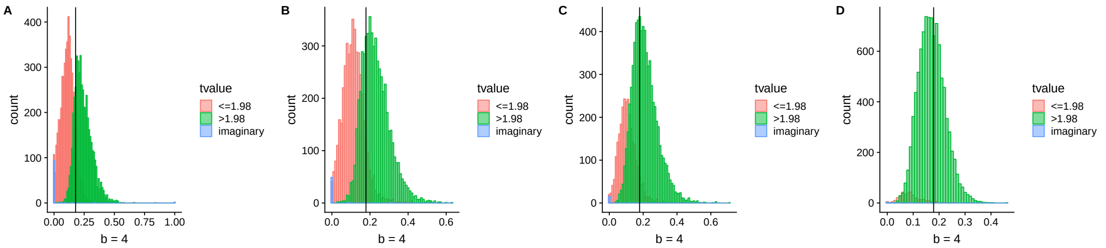

Fig. 4.31: Histogram of β for b = 4, T = 500, $\alpha = 0.2$

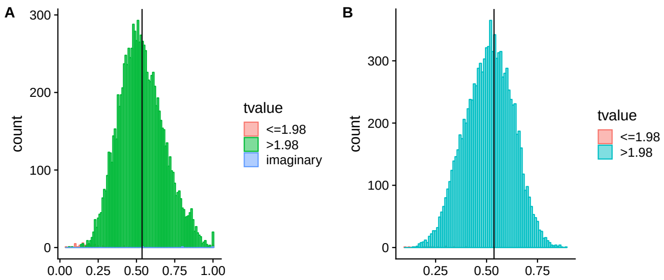

Fig. 4.32: Histogram of β for b = 12, T = 500, $\alpha = 0.2$

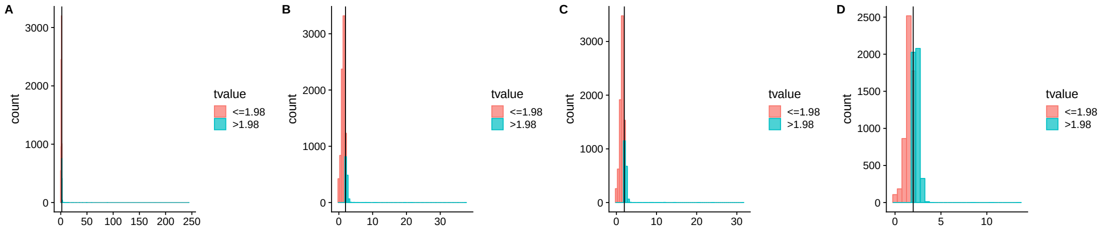

Fig. 4.33: Histogram of the t-values corresponding to the β parameter for b = 2

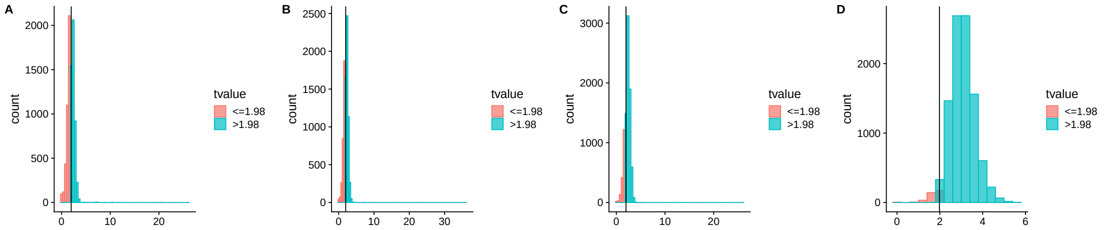

Fig. 4.34: Histogram of the t-values corresponding to the β parameter for b = 4

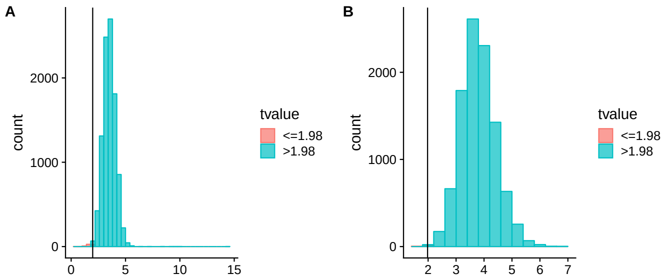

Fig. 4.35: Histogram of the t-values corresponding to the β parameter for b = 12

As an important note, the problem of nonstationarity becomes more apparent in case of an error specification following the t-distribution than for one following the normal distribution. The sum of the $\beta + \pi$ should be smaller than one to comply with the assumption of stationarity. However, as shown in Figure 4.36, the sum shows a peak at one in case of b of 12. In case of higher sum of β and π the sum itself gets overestimated as is shown in Appendix B Figure B.8 to B.10. This is similar to the results of the normally distributed error, however, the peak at 1 is larger for π of 0.2. Furthermore, there is a peak around one for larger π in general. Similarly to the normal distribution, the sum shows erratic behavior in case of non and weak identification.

Table 4.5: Robust critical values for different identification categories for b

	b	
π	2	4
0.2	2.2517	2.863
0.4	2.289	2.9097
0.6	2.3499	

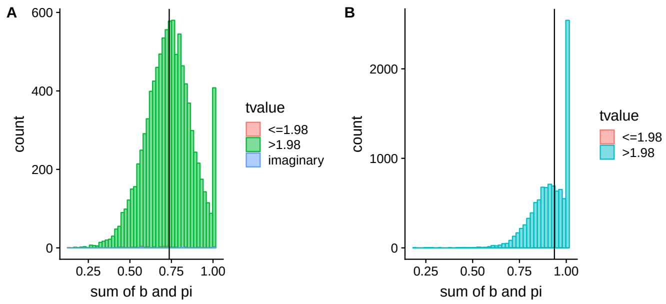

Fig. 4.36: Histogram of the sum of $\pi + \beta$ for b = 12

4.2.4 Results for ν

This section mentions the results for the parameter ν. The distributions are found in Appendix B Figure B.11 to B.14 and the t-values are found in Appendix B Figure B.15 to B.17. There does not seem to be much variation between the different identification strengths. The distributions approximate a normal distribution with the exception of some outliers and positive skewness. Note that for higher initial values of ν there are more outliers and the distribution is not clearly visible anymore. The t-values approximate a t-distribution in case of strong identification. However, one should keep in mind that the t-values for the degrees of freedom are difficult to interpret. This is, because the ν parameter cannot be lower than two. Therefore, the t-values only show the certainty of the estimate for ν. A larger t-value is associated with a smaller confidence interval around the estimated value of ν.

4.2.5 Robust critical values for the t-test

Chapter 4.2.1 and 4.2.2 show, similar to the normal distribution, that the parameter estimates are unreliable if identification loss occurs. One method to determine the identification loss of the GARCH model is a robust test statistic. This study makes use of a robust critical t-value. The GARCH model is not well identified for b of zero, two, and four and π of 0.2 to 0.6. In general, as mentioned before, the t-values are small for the t-distributed error in comparison to the normally distributed error. Table 4.5 shows that the desired critical value for a sample size of 500 should be 2.9097. This is larger than 1.98. It is much smaller than the desired critical value of 56 for a normally distributed error term for the GARCH model. It confirms the idea that the GARCH model with t-distributed error better distinguishes the nonidentified GARCH model.

4.3 Comparison of the GARCH model to the ARCH and SVM

As shown in Chapter 4.1 and 4.2, identification loss has serious impact on the reliability of the parameter estimates. Especially, the (non)identified π parameter is impacted. The parameter estimates are scattered over the parameter space and the t-values of the π parameter explode, indicating infinitesimal standard errors. The π parameter only shows signs of reliability in case of near nonstationarity and if the GARCH model is well identified (Chapter 4.1.1 and 4.2.1).

This section shows that the out-of-sample performance of the GARCH model is not significantly better than that of the baseline ARCH model in case of identification loss. Furthermore, a state-of-the-art stochastic volatility model (SVM) always outperforms the GARCH model (Table 4.7 and 4.9) and should therefore be preferred to (G)ARCH models. Note that the stars (***) behind the p-values of the Diebold-Mariano test stand for 5%, 1%, and 0.1% significance respectively. Interestingly, (G)ARCH models overestimate the volatility, leading to higher hit rate of the value at risk (VaR) than the SVM (Table 4.6 and 4.8. This means that (G)ARCH models are better at providing a confidence interval in which 95% of the returns fall.

Note that this is done for one thousand of the ten thousand simulated time series. It is only done for those time series of which the empirical t-value of the β parameter is significant. This is, because a first glance on the simulated returns (Appendix A Figure A.8 to A.11 and Appendix B Figure B.18 to B.21) did not show clear variation between the simulated returns. Only in case of stationary problems Figure 4.21 and 4.36 show some return outliers far removed from the mean indicating high volatility. Furthermore, the computational time associated with the SVM is more substantial than that associated with the GARCH model. This is mentioned as one of the pitfalls of the model.

Firstly, I comment on the prediction performance of the GARCH model with normally distributed error specification. Secondly, I compare these results to those obtained from the t-distributed error specification. Lastly, I compare the results of the simulations for the t-distributed error specification to those obtained from Bollerslev (1987).

4.3.1 Normally distributed error

This section mentions the volatility predicted by the GARCH, ARCH, and SVM for an error specification following the normal distribution. The volatility estimates are found in Appendix A Figure A.12 to A.23. Furthermore, I show that the ARCH model is better at approximating the VaR than the other volatility models (Table 4.6). Moreover, the GARCH model does not outperform the ARCH model in case of identification loss. It does outperform the ARCH model if stronger identification occurs (Table 4.7). The SVM generally outperforms the (G)ARCH models.

Appendix A Figure A.12 to A.15 show that the GARCH variance estimated is smaller in case of nonidentification and that the value of the outliers increases as the identification strength increases. This is expected, since higher values of b in combination with higher values of π should result in higher conditional volatility. It shows however problems of nonstationarity in Figure A.14D and A.15B, since the volatility estimates explode. Appendix A Figure A.16 to A.19 show that the ARCH variance acts similar to the GARCH model. This is also as expected, since the GARCH model is an extension of the ARCH model. Appendix A Figure A.20 to A.23 show

Table 4.6: Number of times return (out of 1000) not within boundaries set by value at risk, calculated as $1.96\sqrt{h_t}$

	$b=0$	$b=2$	$b=4$	$b=12$
GARCH				
$\pi_{0.2}$	58	56	53	46
$\pi_{0.4}$	42	55	45	64
$\pi_{0.6}$	65	62	58	
$\pi_{0.8}$	63	47	54	
ARCH				
$\pi_{0.2}$	55	54	56	52
$\pi_{0.4}$	42	52	50	59
$\pi_{0.6}$	67	66	56	
$\pi_{0.8}$	58	40	49	
SVM(median)				
$\pi_{0.2}$	60	57	68	66
$\pi_{0.4}$	45	53	57	94
$\pi_{0.6}$	67	68	67	
$\pi_{0.8}$	62	52	75	
SVM(mean)				
$\pi_{0.2}$	61	59	68	64
$\pi_{0.4}$	45	53	57	94
$\pi_{0.6}$	68	69	68	
$\pi_{0.8}$	63	53	73	

Table 4.7: Results of the Diebold-Mariano test for various identification stages of the GARCH model with normally distributed error specification

	Nonidentification			Weak identification $b=2$			Weak identification $b=4$			Semi-strong identification $b=12$		
	ARCH	SVM med	SVM mean	ARCH	SVM med	SVM mean	ARCH	SVM med	SVM mean	ARCH	SVM med	SVM mean
$p=1$												
$\pi_{0.2}$	0.3671	1	1	0.8639	1	1	0.0912	0.9983	1	0.3338	1	1
$\pi_{0.4}$	0.6081	1	1	0.8056	0.9823	0.9999	0.0426*	1	1	0.0030**	1	1
$\pi_{0.6}$	0.4176	0.9546	1	0.1443	1	1	0.0091**	1	1			
$\pi_{0.8}$	0.7873	1	1	0.0835	1	1	$4.05 \cdot 10^{-7}$***	1	1			
$p=2$												
$\pi_{0.2}$	0.4049	0.9999	1	0.7989	0.9999	1	0.3902	1	1	0.7182	1	1
$\pi_{0.4}$	0.6383	0.9998	1	0.8707	0.9901	1	0.1406	1	1	0.9328	0.9996	0.9995
$\pi_{0.6}$	0.1810	0.8604	1	0.1751	1	1	0.1120	1	1			
$\pi_{0.8}$	0.5779	0.9992	1	0.3075	1	1	0.1863	1	1			

that the estimated volatility is more spread out for the SVM than for the (G)ARCH models. In general, the variance of the SVM is smaller than that estimated by the (G)ARCH models.

Table 4.6 shows that there is no clear indication that stronger identification ensures a higher hit rate of the VaR. The hit rate of the VaR should be around 50. If the hit rate is smaller than 50, the potential losses as computed by the VaR exceed the actual losses too often. If it is larger, the actual losses exceed the potential losses too often. If preferring a higher hit rate, the estimated volatility is larger. This is not preferable for return prediction, based on the Diebold-Mariano test as shown in Table 4.7. If considering only the GARCH model the number of times the boundaries set by the VaR do not contain the true value as simulated by the GARCH model, increases in case of stronger identification (b of 12). Chapter 4.1.3 Figure 4.21 shows that in this case the GARCH model suffers from nonstationarity. The ARCH model often predicts larger volatility than the GARCH model and therefore contains more often more true values than the GARCH model. The VaR set by the SVM is smaller than that set by the (G)ARCH models. This is, because the number of times the return is not within the boundaries set by the VaR is larger for the SVM than for the (G)ARCH models. It is an indication that the GARCH model overestimates the conditional volatility, since the Diebold-Mariano test (Table 4.7) shows that the SVM outperforms the GARCH model in general. If these observations are removed and the

Table 4.8: Number of times return (out of 1000) not within boundaries set by value at risk, calculated as 5% quantile of the estimated degrees of freedom*$\sqrt{h_t}$

	$b=2$	$b=4$	$b=12$
$GARCH$			
$\pi_{0.2}$	27	28	16
$\pi_{0.4}$	17	18	24
$\pi_{0.6}$	14	24	
$\pi_{0.8}$	15	21	
$ARCH$			
$\pi_{0.2}$	26	27	16
$\pi_{0.4}$	17	20	20
$\pi_{0.6}$	15	20	
$\pi_{0.8}$	12	22	
$SVM(median)$			
$\pi_{0.2}$	98	113	99
$\pi_{0.4}$	84	97	84
$\pi_{0.6}$	89	98	
$\pi_{0.8}$	78	127	
$SVM(mean)$			
$\pi_{0.2}$	98	112	97
$\pi_{0.4}$	85	98	83
$\pi_{0.6}$	90	99	
$\pi_{0.8}$	81	129	

Diebold-Mariano test is then performed, the SVM still outperforms the GARCH model, because observations for which the GARCH model outperforms the SVM are removed.

Table 4.7 shows that the GARCH model does not outperform the ARCH model in case of identification loss or for lower π based on the Diebold-Mariano test. Only for π from 0.4 on and a mean absolute error loss function the GARCH model outperforms the ARCH model. It shows that in case of near nonstationarity the GARCH model should be preferred over the ARCH model. For π smaller than 0.4, the GARCH model does not outperform the ARCH model. These results are indicative of the stronger identification problem of the π parameter if the initial value of π is small as mentioned by Ma et al. (2007). It is shown in Figure 4.3 that even if the GARCH model is semi-strongly identified, there is at most weak identification for π of 0.2. However, the SVM outperforms the (G)ARCH models in general. Since the SVM predicts smaller volatility than the (G)ARCH models this is a clear indication that the (G)ARCH models overestimate volatility.

4.3.2 t-distributed error

This section mentions the volatility predicted by the (G)ARCH models and the SVM for an error specification following the t-distribution. It focuses on the difference with regards to the error specification following the normal distribution. The volatility estimates are found in Appendix B Figure B.22 to B.30. A first glance at these predicted volatilities shows the tendency of the (G)ARCH models to explode, an indicator of nonstationary (G)ARCH models. The VaR does not show any difference to the VaR for the normal distributed error specification (Table 4.8). The Diebold-Mariano test (Table 4.9) shows that the GARCH model with t-distributed error specification outperforms the ARCH model in case of identification loss. This is different to the results of the normally distributed error specification. It shows, just as the t-values in Chapter 4.2, that both models are intrinsically different and have different responses to identification loss.

Figure B.22 to B.24 show that the GARCH model with t-distributed error specification suffers more from nonstationarity than the GARCH model with normally distributed error specification.

Table 4.9: Results of the Diebold-Mariano test for various identification stages of the GARCH model with t-distributed error specification

	Weak identification $b = 2$			Weak identification $b = 4$			Semi-strong identification $b = 12$		
	ARCH	SVM med	SVM mean	ARCH	SVM med	SVM mean	ARCH	SVM med	SVM mean
$p = 1$									
$\pi_{0.2}$	$2.028{\cdot}10^{-4}$***	1	1	$3.099{\cdot}10^{-4}$***	1	1	0.0066**	1	1
$\pi_{0.4}$	$1.283{\cdot}10^{-5}$***	1	1	$3.977{\cdot}10^{-5}$***	1	1	0.0637	1	1
$\pi_{0.6}$	$9.394{\cdot}10^{-3}$**	1	1	$1.055{\cdot}10^{-8}$***	1	1			
$\pi_{0.8}$	$4.168{\cdot}10^{-14}$***	1	1	$1.977{\cdot}10^{-7}$***	1	1			
$p = 2$									
$\pi_{0.2}$	$2.158{\cdot}10^{-3}$**	1	1	0.0189*	1	1	0.1845	0.9996	0.9995
$\pi_{0.4}$	$7.213{\cdot}10^{-5}$***	1	1	0.00416**	1	1	0.6886	0.9999	0.9998
$\pi_{0.6}$	0.0928	1	1	0.02893*	1	1			
$\pi_{0.8}$	$3.924\text{e}{\cdot}10^{-6}$**	1	1	0.7339	1	1			

The range of volatility estimates is higher for the GARCH model with t-distributed error specification. The nonstationary problem is shown through the log scale used in Figure B.23D and B.24. These figures clearly show that the GARCH model with error following the t-distribution should not be preferred, because if there is no identification loss, there is a problem of nonstationarity. Figure B.28 to B.30 show that the variance estimated by the SVM is smaller than the variance estimated by the GARCH model. The exploded volatility estimates of the SVM model as shown in Figure B.29D for $\pi = 0.8$ and B.30B for $\pi = 0.4$ show that the explosion is smaller for the SVM than for the GARCH model. This is similar to the results of the GARCH model with normally distributed error specification. The SVM overall predicts smaller volatility than the (G)ARCH models. This is also shown in Table 4.8 through the fact that the confidence interval as provided by the SVM is smaller than that provided by the (G)ARCH models.

The Diebold-Mariano test shows similar to the GARCH model with normally distributed error specification that the SVM always outperforms the GARCH model. However, a large difference between the two error specifications is that the GARCH model is in general significantly better than the ARCH model. If considering the mean squared prediction error (MSPE)($p = 2$) and the model becomes nonstationary, the GARCH model does not significantly outperform the ARCH model. A reason why the GARCH model outperforms the ARCH model can be that the ARCH model overestimates the volatility in comparison to the GARCH model. This can also be the reason that the SVM outperforms the GARCH model, because the GARCH model overestimates the volatility if compared to the SVM. Interestingly, the SVM could not estimate the degrees of freedom ν. This means that the simulation of the GARCH model with given ν does not provide clear t-distributed time series.

4.3.3 Results for S&P 500 data

This section compares the results obtained through a simulation study as described in Chapter 3.4 to the S&P 500 index data from Bollerslev (1987). First, I simulated data on the assumption that the parameter estimates are the true parameter estimates as provided by Bollerslev (1987). 6.8% of the simulated data did not result in a converged GARCH model and conveyed imaginary t-values. In 26.5% the t-value is larger than 1.98. If I compare these results to the number of times β is significant as shown in Chapter 4.2 Table 4.4, it seems to correspond with a b of two. The GARCH and ARCH model have large volatility estimates as shown in Figure 4.37 and the true return as simulated is never larger than the boundaries provided by the value at risk. This means that the GARCH and ARCH models overestimate the variance in general. The variances as estimated by the (G)ARCH models explode, as shown by the scale, which is transformed to a log-scale. It became a stationarity issue, because of the assumption that the estimates were the correct estimates as provided by Bollerslev (1987). This is similar to other results for the GARCH model with an error specification following the t-distribution as shown in Chapter 4.3.2. The prediction error also exploded for the (G)ARCH models. The prediction error of the SVM is between -0.1 and 0.1. The real value as simulated lies between -0.1 and 0.1. These results show that the SVM should be preferred.

The Diebold-Mariano test does not show that the GARCH model is significantly better than any of the other models, with a p-value of 0.84. As shown before in Table 4.9 an insignificant GARCH-t model does not indicate a weakly identified GARCH model. It actually shows that the model is well identified, however, nonstationary. The SVM estimates smaller volatility and the estimated VaR does not contain the true value of the return in 8.5% of cases.

I estimated the various volatility models on 25% of the S&P 500 data. Figure 4.40 shows the parameter estimates as estimated for the last 25% of data. Especially in the later period, the β parameter estimates are insignificant. Through close investigation it is visible that in general the GARCH model is weakly identified, because the parameter estimates of β that are significant are within the boundaries of those that are insignificant. Furthermore, the conditional variance estimates do not explode, which happens in case of stronger identification. The VaR does not contain the real value of the return in 3,5,9, and 8 cases. This is, because the estimated variance of the GARCH model is larger than the estimated variance of the SVM. The estimated variance as shown in Figure 4.41 shows that the outliers of the variances are much larger for the GARCH model with respect to the SVMs.

Figure 4.42 shows the estimated return by the various volatility models. The GARCH and the ARCH model show the highest (and lowest) returns and the median SVM shows the smallest returns in most of the data. The Diebold-Mariano test is significant (4.9%) for the ARCH model with the MSPE as loss function. It confirms the idea of weak identification as shown by Table 4.9. The t-value is always under 2.16 and the robust critical t-value is 2.90, therefore the β parameter is always insignificant, confirming nonidentification.

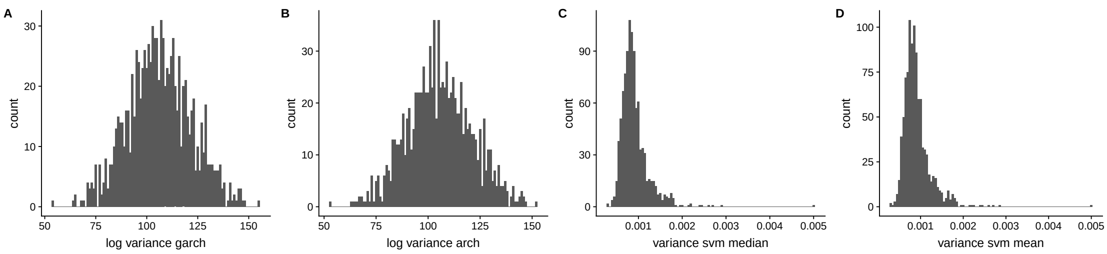

Fig. 4.37: Estimated variances on simulated data S&P 500

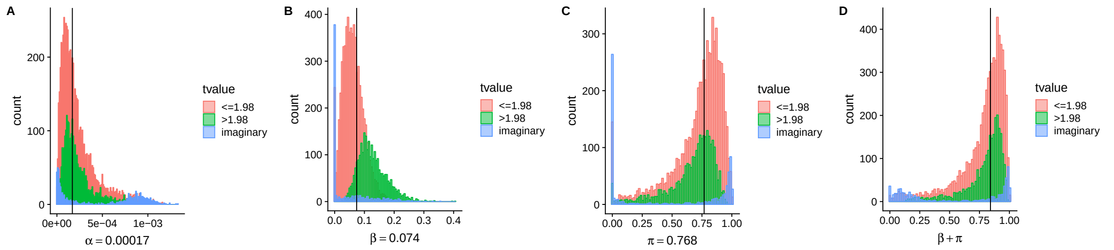

Fig. 4.38: empirical parameter distributions S&P 500

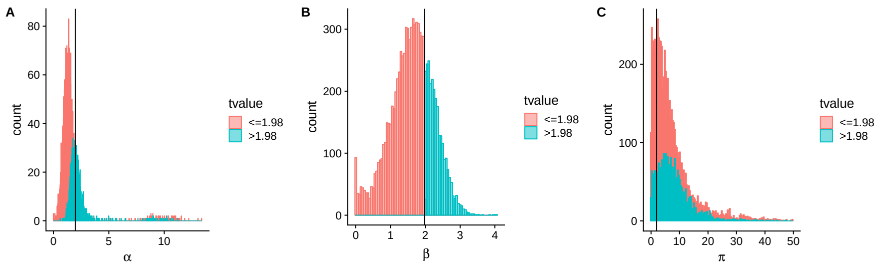

Fig. 4.39: empirical distributions t-values S&P 500

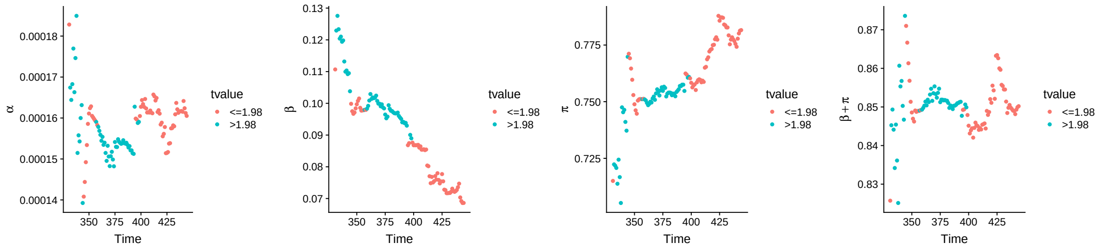

Fig. 4.40: Parameter estimates S&P 500

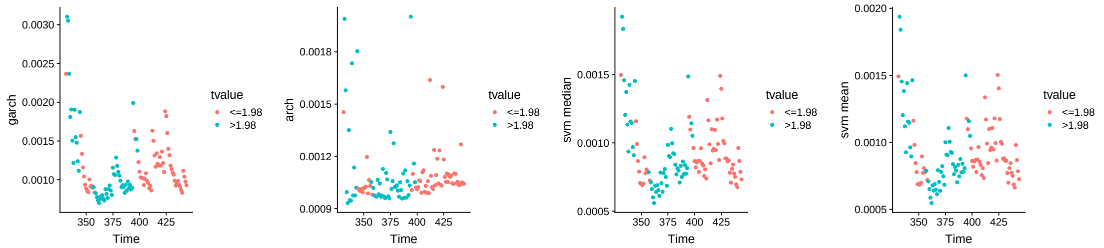

Fig. 4.41: Estimated variances S&P 500

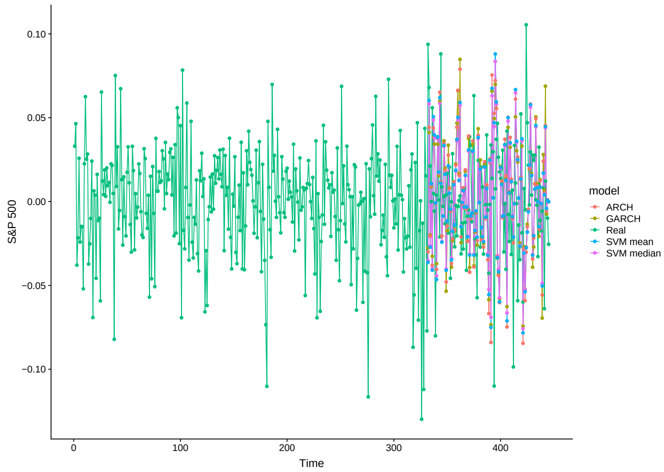

Fig. 4.42: Prediction S&P 500

Chapter 5

Concluding Remarks

5.1 Conclusions

The goal of this study was to present the problems encountered when facing identification loss in the generalized autoregressive conditional heteroskedasticity (GARCH) model. Furthermore, the question I aimed to answer in this thesis was which volatility model performed best in terms of predictive power in case of identification loss in the GARCH model. This, to provide an alternative to the GARCH model, since there are severe consequences if the GARCH model suffers from identification loss. Consequences are such as unreliable parameter estimates and exploded t-values for the identifying parameter β and the parameter suffering from identification loss π. The stochastic volatility model (SVM) is superior to the GARCH model, because it outperforms the GARCH model in general, based on the Diebold-Mariano test.

The parameter estimates are unreliable, showing uniform distributions for the parameter suffering from identification loss π. There are only signs of a normal distribution in case of strong identification from b of twelve. Even for the largest true values for b as used in the simulation study, the parameter estimates do not show a normal distribution. This means that in this study I do not obtain a strong identification for the GARCH model. This has implications for tests that are based on the normality assumption of the parameter estimates, such as t-tests. These implications include the (in)significance of the parameters, since this is based on t-tests. An issue that arises for larger b is nonstationarity. This is visible for the GARCH model with error specification following the normal distribution (GARCH-N) and it is most prominent if the error specification follows the t-distribution (GARCH-t). Therefore, sample sizes should be larger for the parameter estimates to be reliable and to prevent nonstationarity.

The identification issue is most severe in the GARCH-N model. β is significant in 50% of cases, based on the t-value, leading to very small confidence intervals. This is the case if the true value is zero for the GARCH-N model, leading to high robust t-values. Even though the GARCH model should outperform the ARCH model, this does not happen in case of nonidentification. Only for b of four the GARCH model outperforms the ARCH model. The GARCH-t model also suffers from the problem of nonidentification. However, the t-distributed GARCH model classifies the model correctly as incorrect if nonidentification occurs. Different from the GARCH-N model, the GARCH-t model does outperform the ARCH model in case of weak and semi-strong identification. This is, because the ARCH model estimates stronger volatility in this case than the GARCH model. Since the empirical distributions of the parameters do not show a normal tendency in this case, errors should not be assumed t-distributed. Provided stronger identification, the GARCH-t model explodes and a state of nonstationarity is entered. Therefore, in general a t-distribution should not be preferred as error assumption for the GARCH model.

In general, the SVMs outperform the GARCH model. The reason that the SVMs outperform the GARCH model, is because the SVM estimates smaller variances compared to the GARCH model. This is also shown through the value at risk (VaR), which is rejected more often for the SVM than for the (G)ARCH models. Therefore, the SVM should be preferred to the GARCH models. In general, the VaR is rejected in more than 5% of cases, which needs to be considered if using these models for computing VaR.

The robust t-test as applied in this study shows that the β parameter is not significantly higher than zero for the S&P 500 data set. This means that the π parameter is not identified and the GARCH model does not show correct estimates for S&P 500 data. I used the estimates as provided by Bollerslev (1987) as true values to simulate a time series and showed that the variances as estimated by the GARCH model exploded. This means that, if those were the correct estimates, the GARCH model was nonstationary. Since the variance as predicted by the GARCH model on the S&P 500 data did not explode, the estimates are not correct. I examined the predictive power on the S&P 500 data and showed that the GARCH model is weakly identified. The last prediction as generated by the GARCH model approximated the data the best, however, the other time periods showed that other models approximated the data better.

5.2 Future work

In this study I checked the assumptions that the simulations of the GARCH model are based on. An important one is the assumption that the parameter estimates show a normal distribution when the parameters are strongly identified, in an ideal world. I did not find a normal distribution for the parameter estimates of α and the degrees of freedom estimates for data sets of 16000 observations and 10 000 simulation runs. Due to time limitations I could not increase the number of observations further. It is important for future research to check whether with a higher number of observations the assumption of a normal distribution is fulfilled.

Furthermore, this study encountered a problem of nonstationarity. Therefore, a follow-up study should consider higher b, upwards from seven, as proposed by Ma et al. (2007). Moreover, it should consider higher number of observations than five hundred, since strong identification is not reached.

Appendix A

Normal distribution

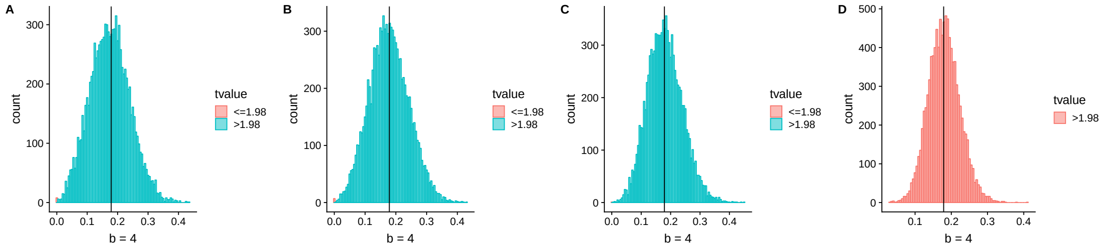

Fig. A.1: Histogram of β for b = 4, T = 500, $\alpha = 0.2$

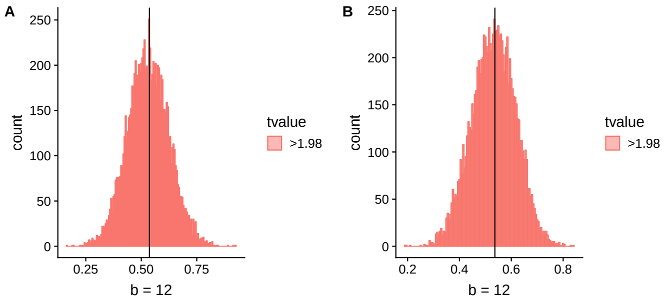

Fig. A.2: Histogram of β for b = 12, T = 500, $\alpha = 0.2$

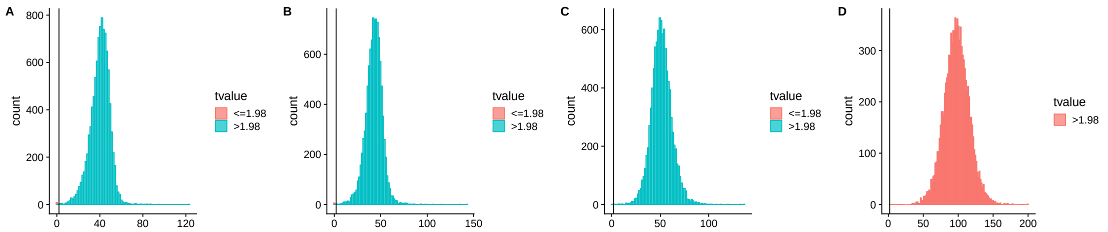

Fig. A.3: Histogram of the t-values corresponding to the β parameter for b = 4

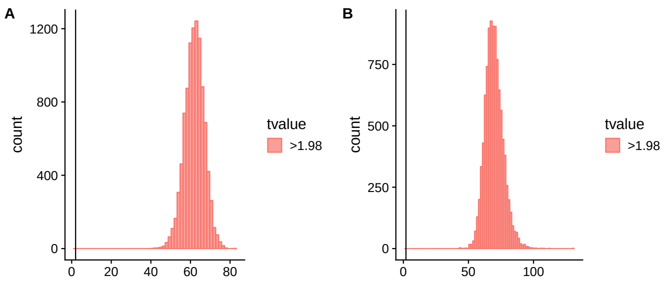

Fig. A.4: Histogram of the t-values corresponding to the β parameter for b = 12

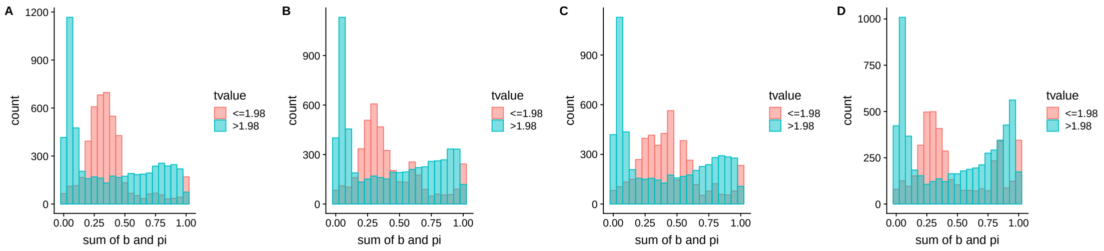

Fig. A.5: Histogram of the sum of $\pi + \beta$ for b = 0 ; normal distribution

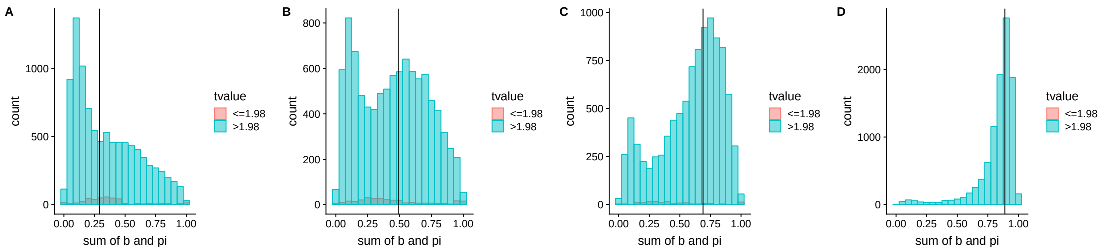

Fig. A.6: Histogram of the sum of $\pi + \beta$ for b = 2 ; normal distribution

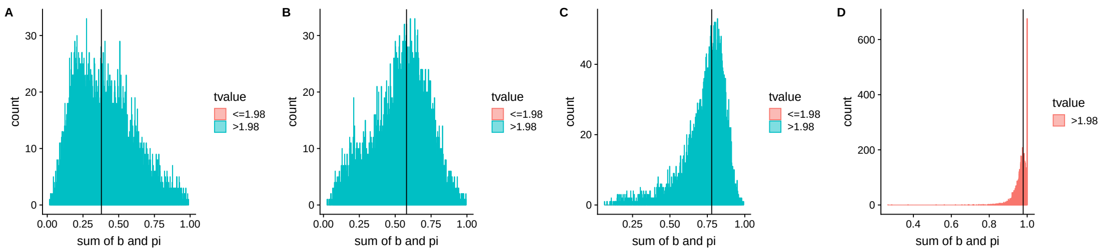

Fig. A.7: Histogram of the sum of $\pi + \beta$ for b = 4 ; normal distribution

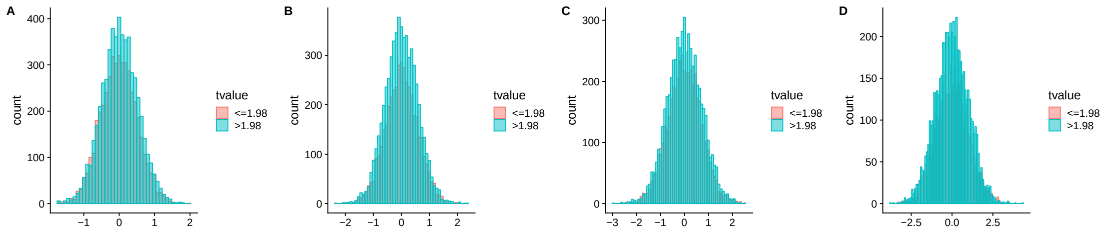

Fig. A.8: Normally distributed error - simulated returns when b = 0

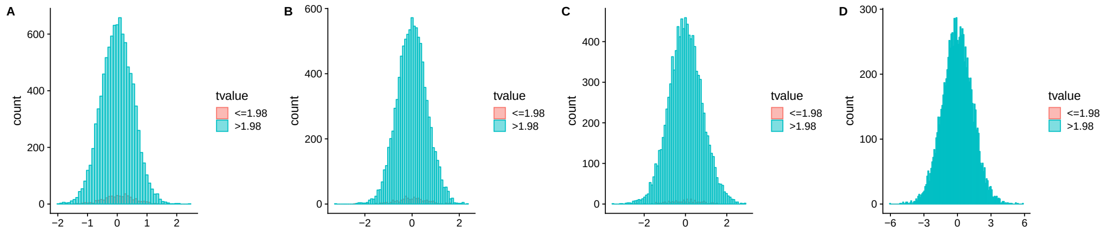

Fig. A.9: Normally distributed error - simulated returns when b = 2

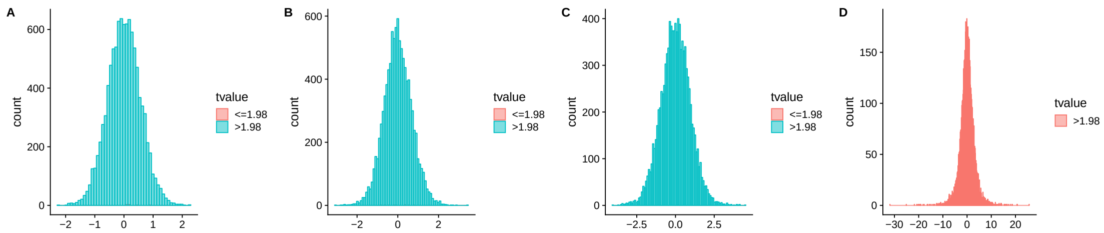

Fig. A.10: Normally distributed error - simulated returns when b = 4

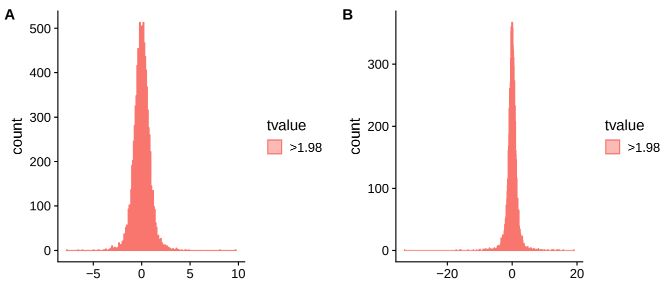

Fig. A.11: Normally distributed error - simulated returns when b = 12

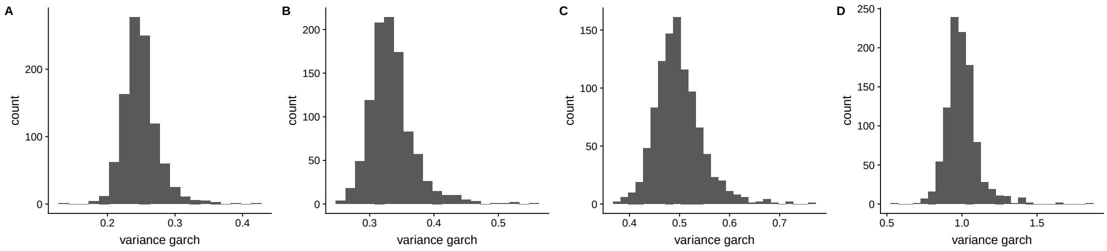

Fig. A.12: Normally distributed error - estimated garch variance when b = 0

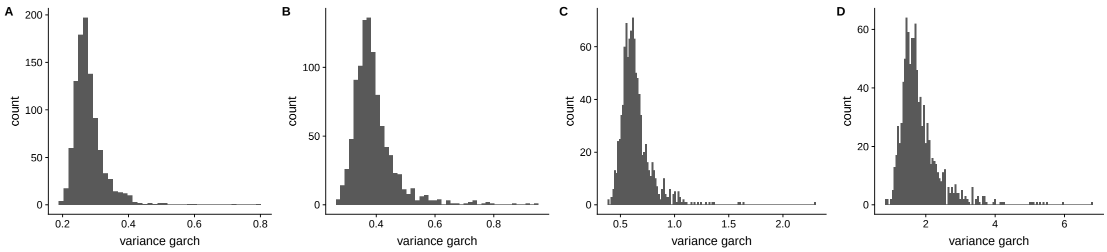

Fig. A.13: Normally distributed error - estimated garch variance when b = 2

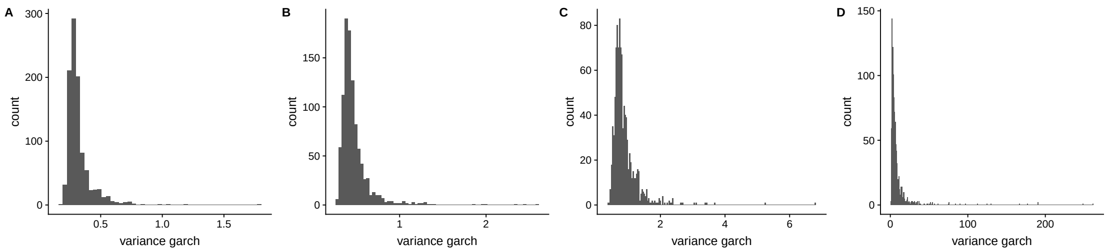

Fig. A.14: Normally distributed error - estimated garch variance when b = 4

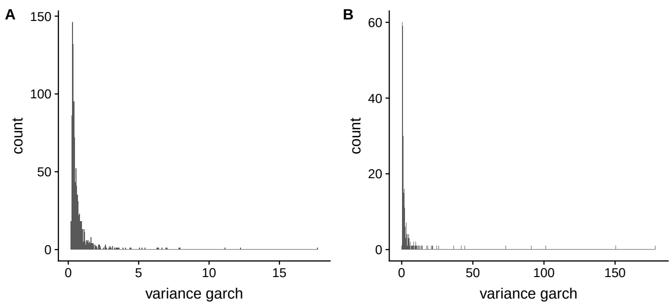

Fig. A.15: Normally distributed error - estimated garch variance when b = 12

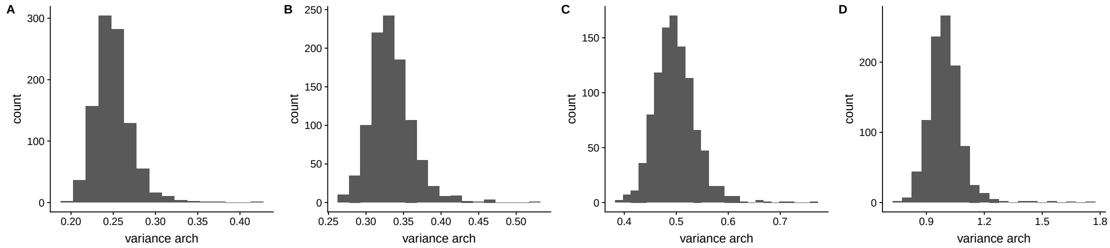

Fig. A.16: Normally distributed error - estimated arch variance when b = 0

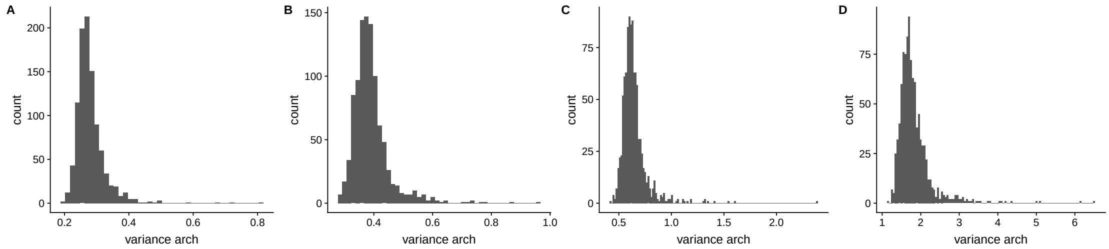

Fig. A.17: Normally distributed error - estimated arch variance when b = 2

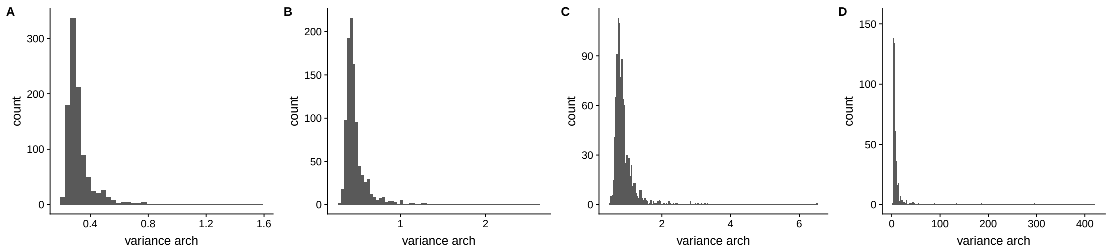

Fig. A.18: Normally distributed error - estimated arch variance when b = 4

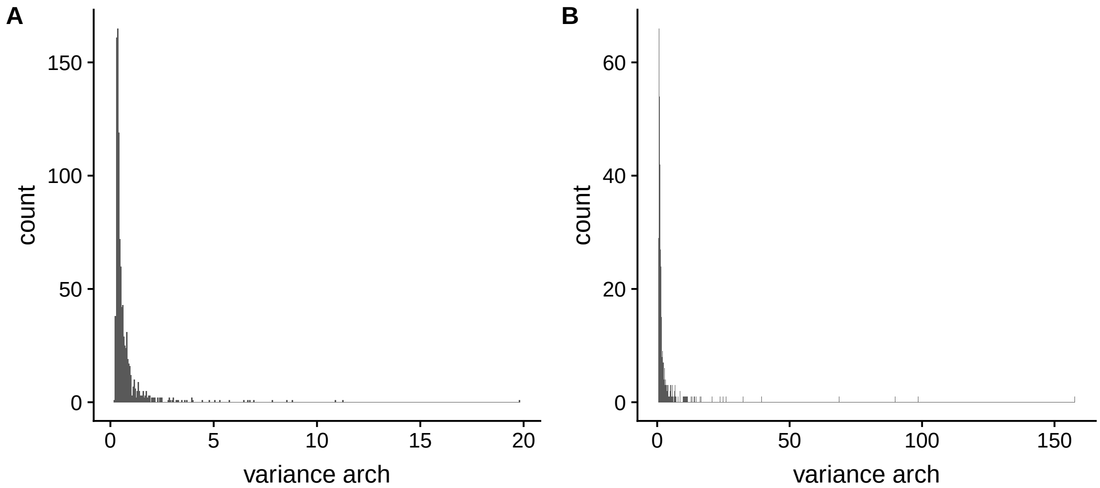

Fig. A.19: Normally distributed error - estimated arch variance when b = 12

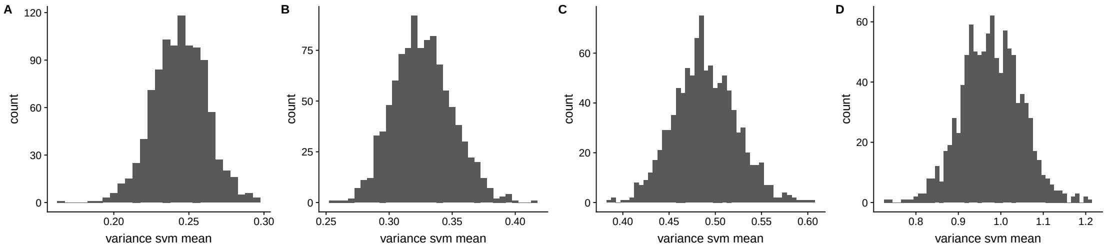

Fig. A.20: Normally distributed error - estimated svm variance (mean) when b = 0

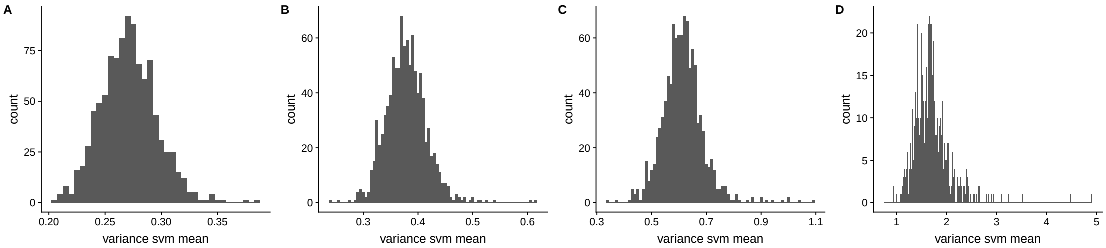

Fig. A.21: Normally distributed error - estimated svm variance (mean) when b = 2

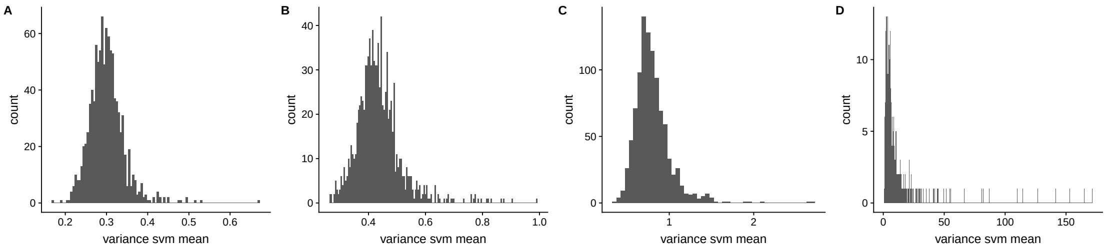

Fig. A.22: Normally distributed error - estimated svm variance (mean) when b = 4

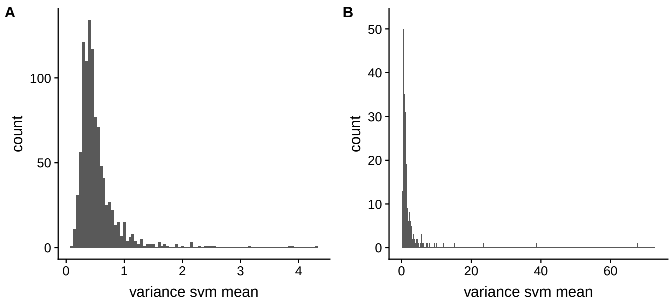

Fig. A.23: Normally distributed error - estimated svm variance (mean) when b = 12

Appendix B

t-distribution

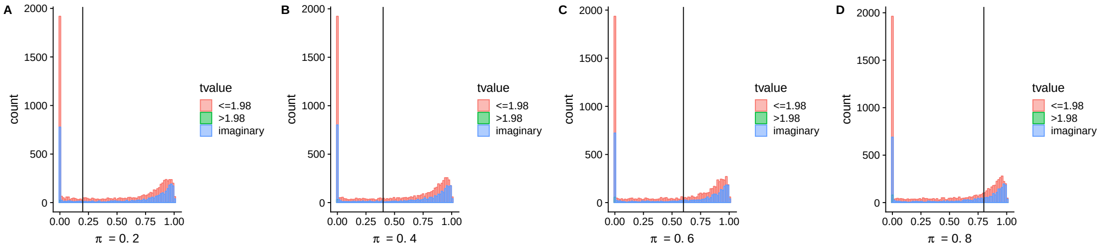

Fig. B.1: Histogram of π for b = 0, T = 500, $\alpha = 0.2$

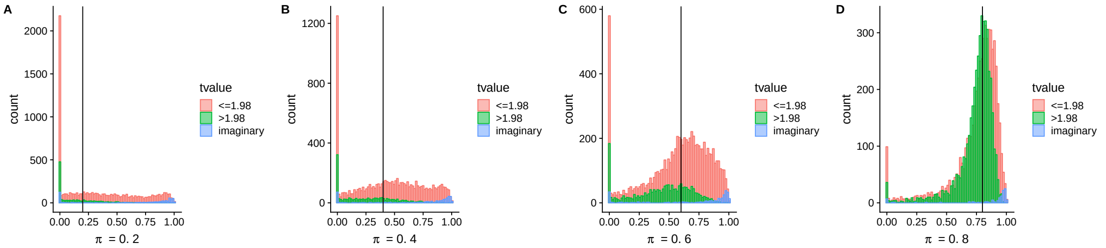

Fig. B.2: Histogram of π for b = 2, T = 500, $\alpha = 0.2$

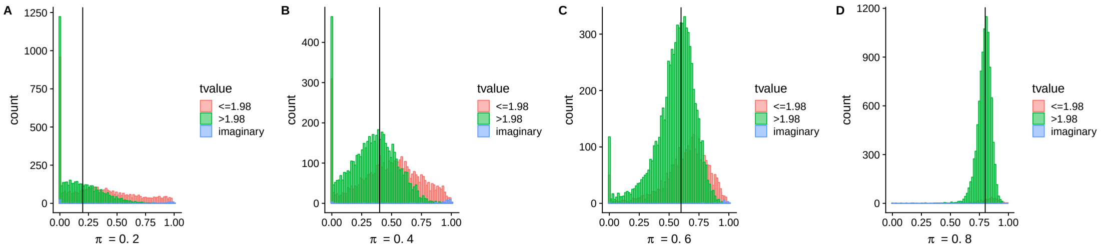

Fig. B.3: Histogram of π for b = 4, T = 500, $\alpha = 0.2$

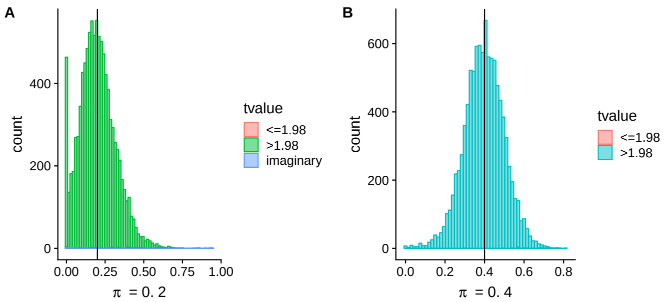

Fig. B.4: Histogram of π for b = 12, T = 500, $\alpha = 0.2$

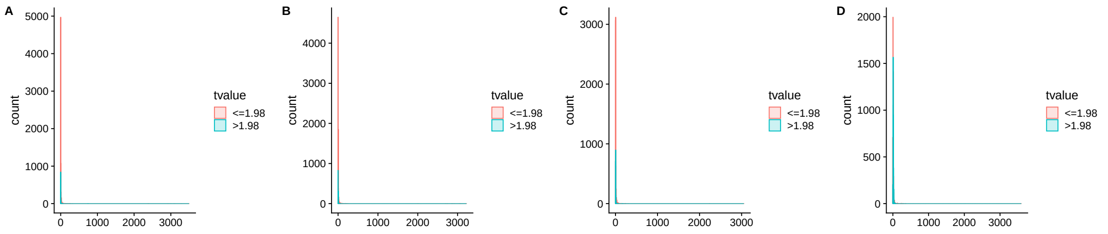

Fig. B.5: Histogram of the t-values corresponding to the π parameter for b = 2

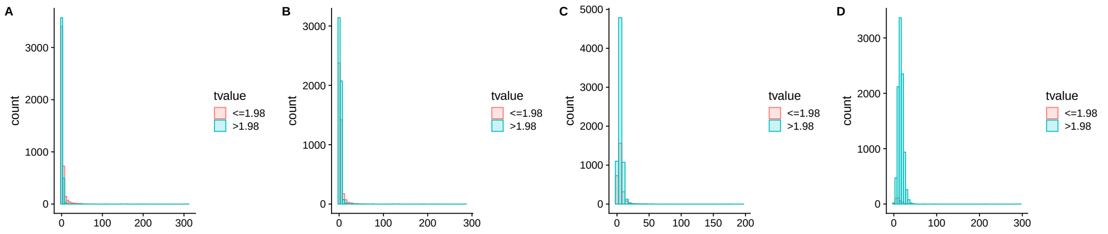

Fig. B.6: Histogram of the t-values corresponding to the π parameter for b = 4

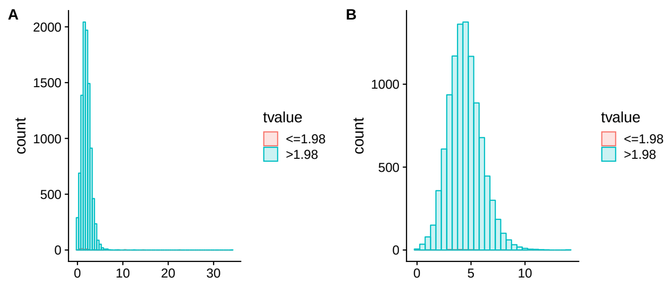

Fig. B.7: Histogram of the t-values corresponding to the π parameter for b = 12

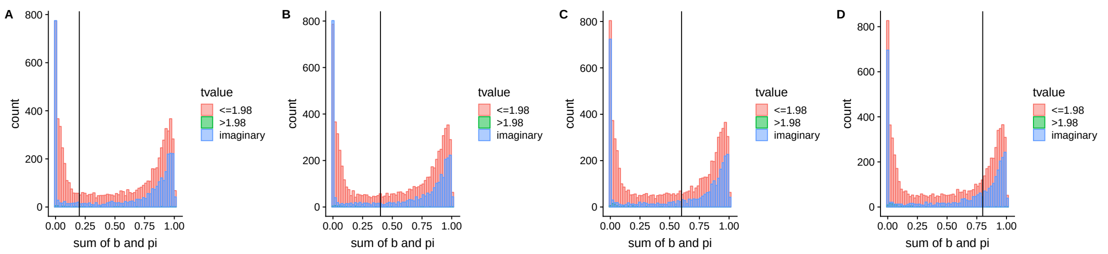

Fig. B.8: Histogram of the sum of $\pi + \beta$ for b = 0 ; t-distribution

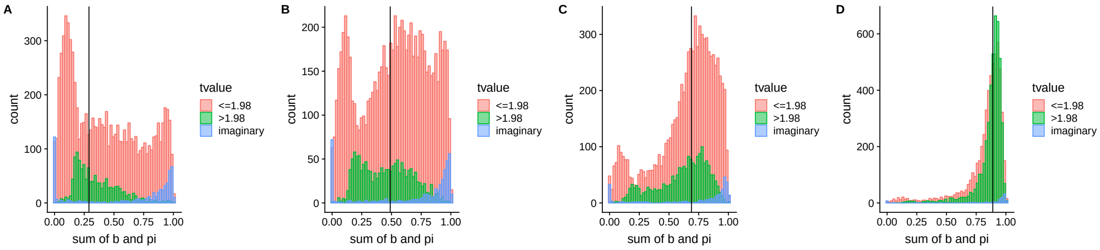

Fig. B.9: Histogram of the sum of $\pi + \beta$ for b = 2 ; t-distribution

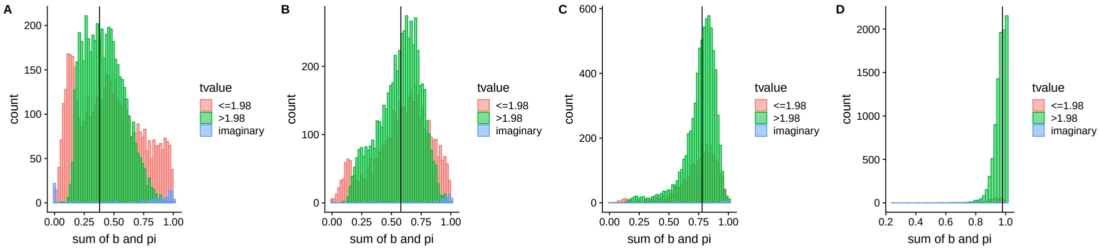

Fig. B.10: Histogram of the sum of $\pi + \beta$ for b = 4 ; t-distribution

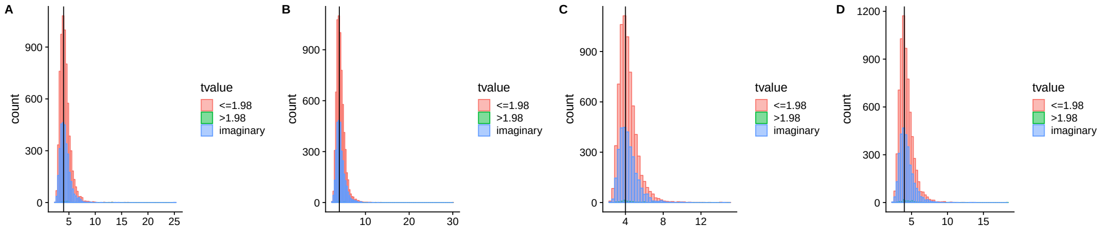

Fig. B.11: Histogram of ν for b = 0, T = 500, $\alpha = 0.2$

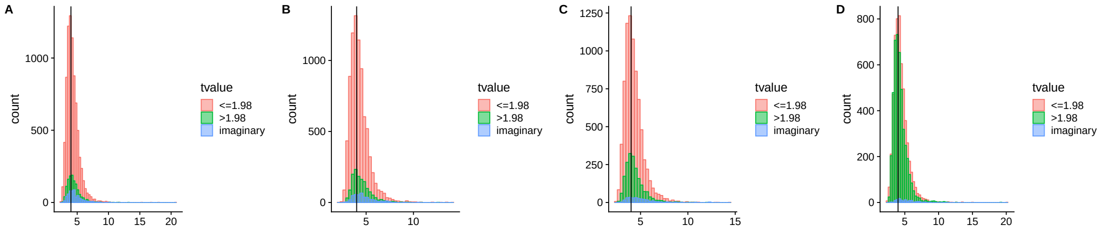

Fig. B.12: Histogram of ν for b = 2, T = 500, $\alpha = 0.2$

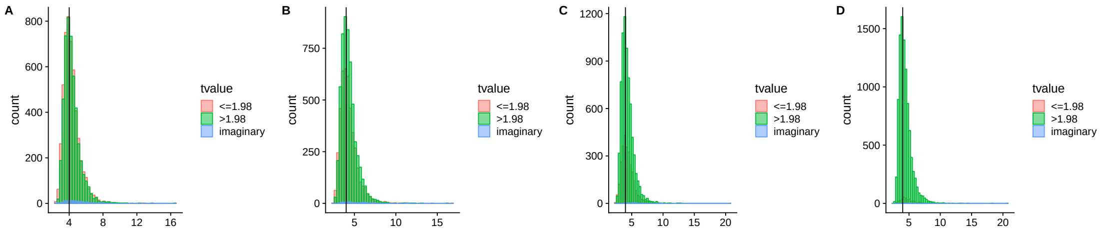

Fig. B.13: Histogram of ν for b = 4, T = 500, $\alpha = 0.2$

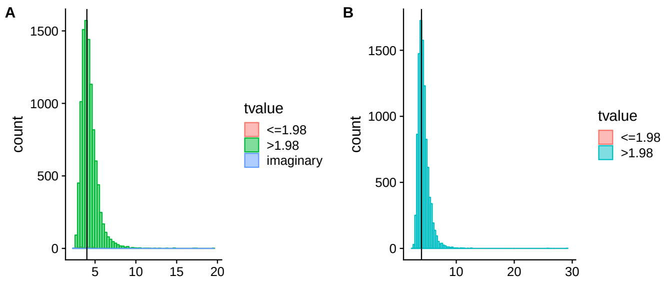

Fig. B.14: Histogram of ν for b = 12, T = 500, $\alpha = 0.2$

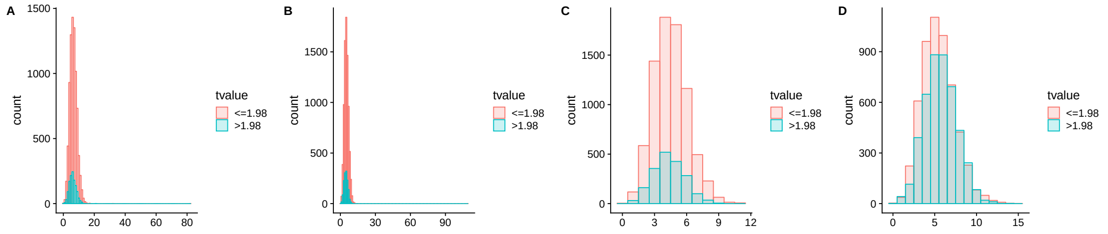

Fig. B.15: Histogram of the t-values corresponding to the ν parameter for b = 2

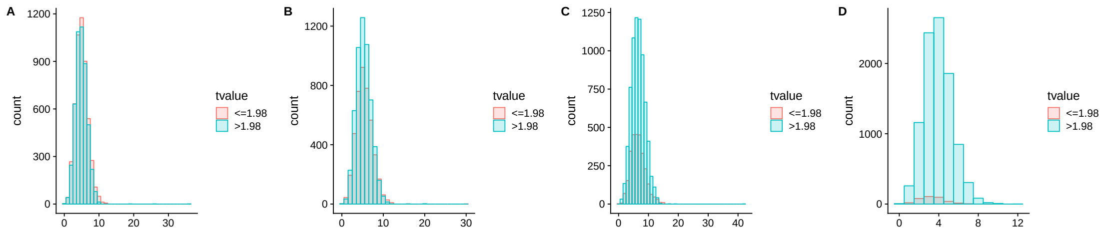

Fig. B.16: Histogram of the t-values corresponding to the ν parameter for b = 4

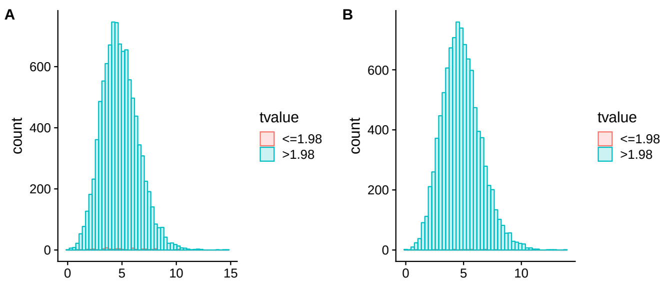

Fig. B.17: Histogram of the t-values corresponding to the ν parameter for b = 12

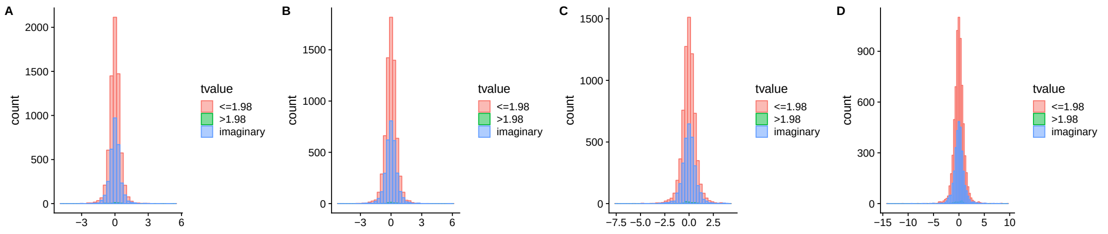

Fig. B.18: t-distributed error - simulated returns when b = 0

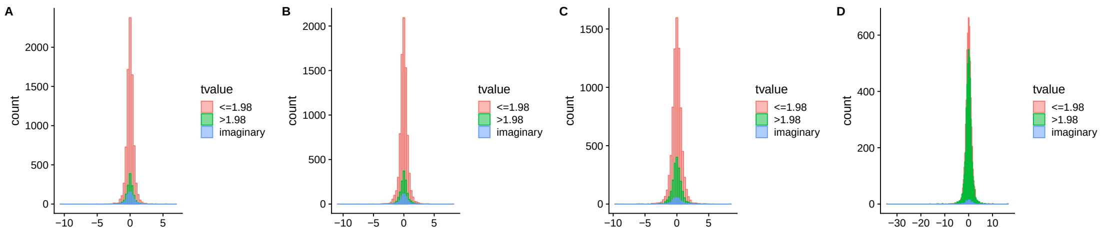

Fig. B.19: t-distributed error - simulated returns when b = 2

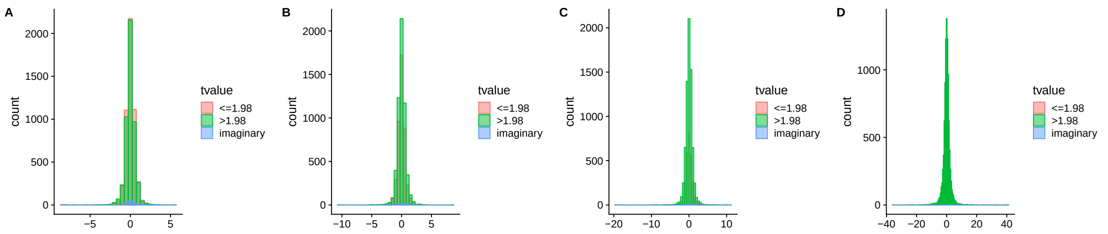

Fig. B.20: t-distributed error - simulated returns when b = 4

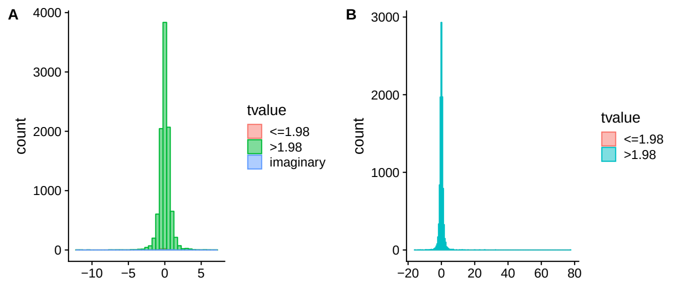

Fig. B.21: t-distributed error - simulated returns when b = 12

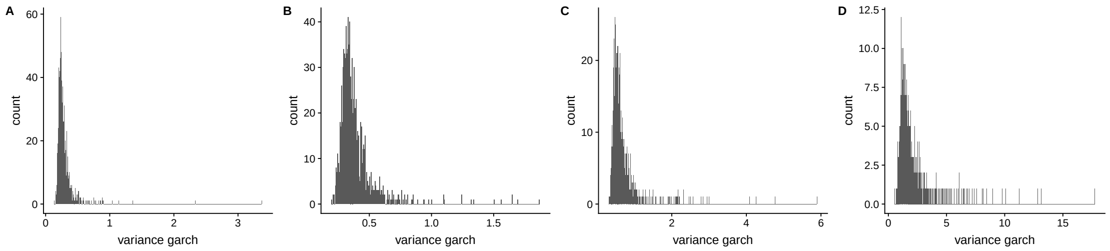

Fig. B.22: t-distributed error - estimated garch variance when b = 2

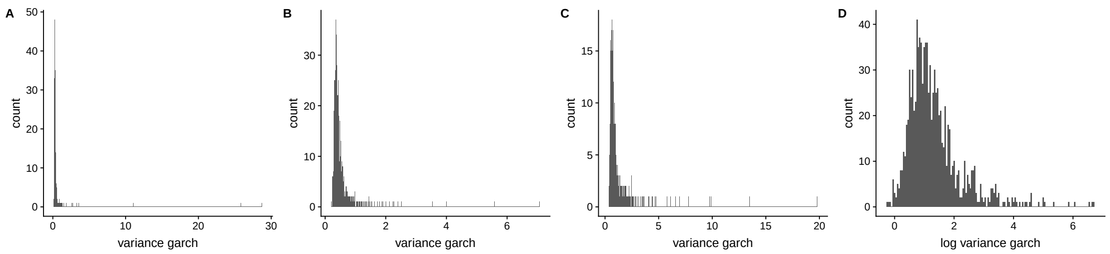

Fig. B.23: t-distributed error - estimated garch variance when b = 4

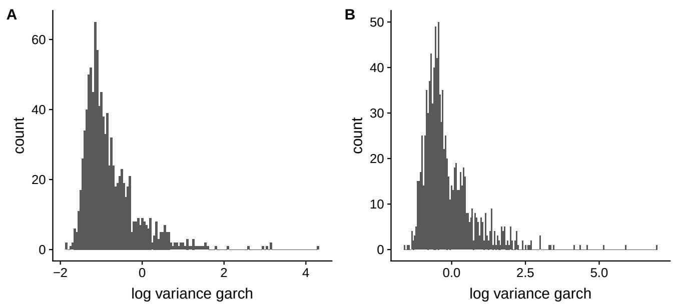

Fig. B.24: t-distributed error - estimated garch variance when b = 12

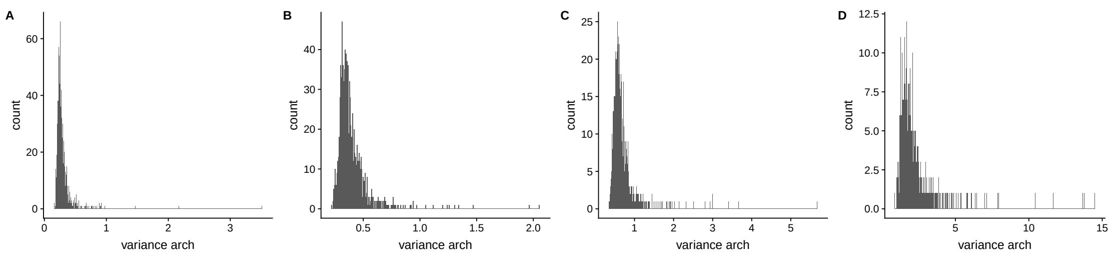

Fig. B.25: t-distributed error - estimated arch variance when b = 2

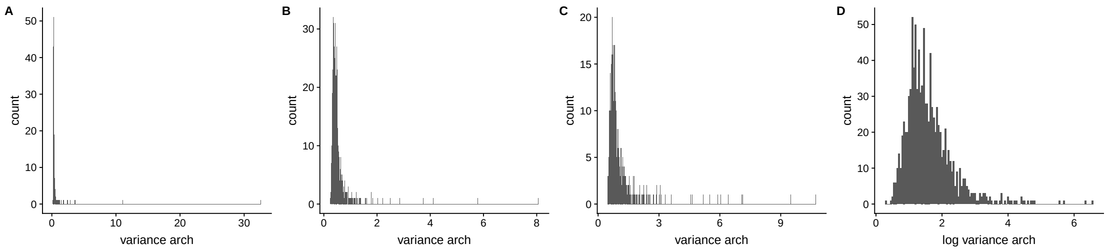

Fig. B.26: t-distributed error - estimated arch variance when b = 4

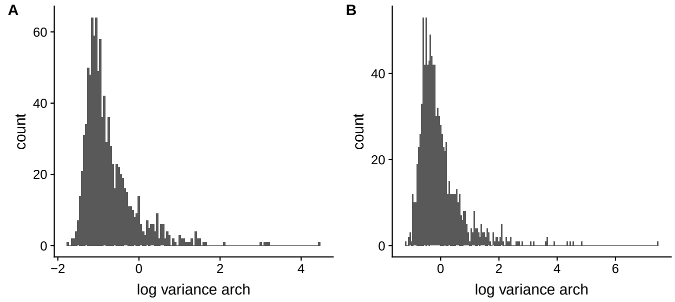

Fig. B.27: t-distributed error - estimated arch variance when b = 12

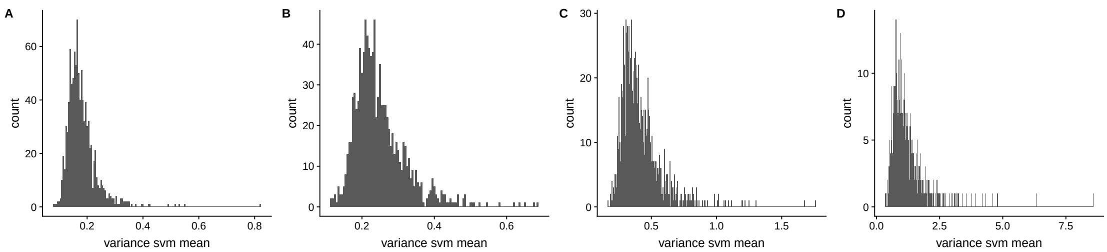

Fig. B.28: t-distributed error - estimated svm variance (mean) when b = 2

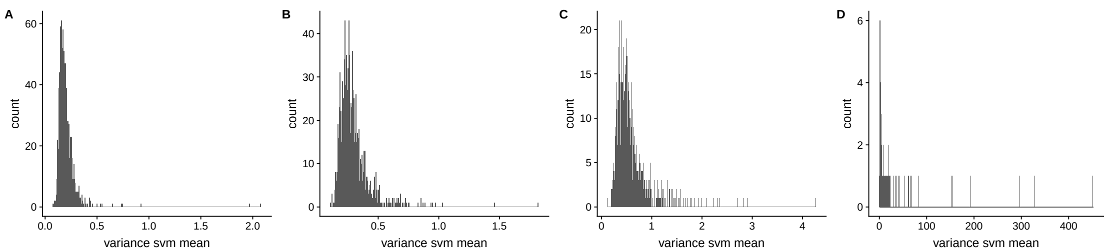

Fig. B.29: t-distributed error - estimated svm variance (mean) when b = 4

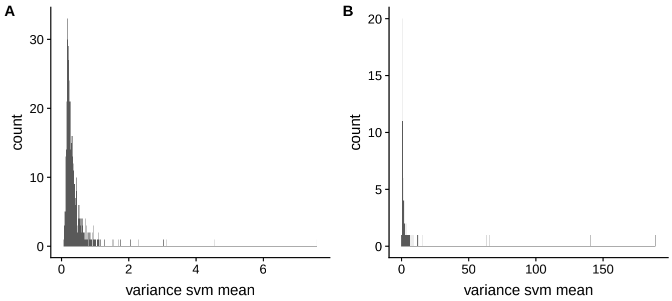

Fig. B.30: t-distributed error - estimated svm variance (mean) when b = 12

Bibliography

C. Amado and T. Tersvirta. Specification and testing of multiplicative time-varying garch models with applications. *Econometric Reviews*, 36(4):421–446, 2017. cited By 6.

Donald W. K. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734, 2001. ISSN 00129682, 14680262.

Donald W. K. Andrews and Xu Cheng. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211, 2012. ISSN 00129682, 14680262.

Donald W. K. Andrews and Gustavo Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.

Donald W.K. Andrews and Xu Cheng. Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics*, 173(1):36 – 56, 2013. ISSN 0304-4076.

A.B.M.R.A. Beg, M.J. Silvapulle, and P. Silvapulle. Tests against inequality constraints when some nuisance parameters are present only under the alternative: Test of arch in arch-m models. *Journal of Business and Economic Statistics*, 19(2):245–253, 2001. cited By 11.

I. Berkes, L. Horvth, and P. Kokoszka. Garch processes: Structure and estimation. *Bernoulli*, 9 (2):201–227, 2003. cited By 229.

Robert C Blattberg and Nicholas J Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *The Journal of Business*, 47(2):244–80, 1974.

T. Bollerslev, R.F. Engle, and D.B. Nelson. Chapter 49 arch models. *Handbook of Econometrics*, 4:2959–3038, 1994. cited By 482.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. 1986.

Tim Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547, 1987. ISSN 00346535, 15309142.

Charles S. Bos. *Relating Stochastic Volatility Estimation Methods*, chapter Six, pages 147–174. John Wiley Sons, Ltd, 2012. ISBN 9781118272039.

T. S. Breusch and A. R. Pagan. The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253, 1980. ISSN 00346527, 1467937X.

Carmen Broto and Esther Ruiz. Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, 18(5):613–649, 2004.

Xu Cheng. Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics*, 189(1):207 – 228, 2015. ISSN 0304-4076.

Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy I: an asymptotic test. Technical report, 1991.

F.X. Diebold. Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, 5(1):51–56, 1986. cited By 213.

R. Engle. Garch 101: The use of arch/garch models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168, 2001. cited By 350.

Robert Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

Y. Fan, Z. Zhang, X. Zhao, and H. Yin. Interaction between industrial policy and stock price volatility: Evidence from china's power market reform. *Sustainability (Switzerland)*, 10(6), 2018. cited By 1.

E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(2):175–185, 1954. ISSN 00359246.

Sylvia Frühwirth-Schnatter and Helga Wagner. Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85–100, 2010.

Raymond E. Goodson and Michael P. Polis. A survey of parameter identification in distributed systems. *IFAC Proceedings Volumes*, 8(1, Part 1):50 – 61, 1975. ISSN 1474-6670. 6th IFAC World Congress (IFAC 1975) - Part 1: Theory, Boston/Cambridge, MA, USA, August 24-30, 1975.

A. Gormus, S. Nazlioglu, and U. Soytas. High-yield bond and energy markets. *Energy Economics*, 69:101–110, 2018. cited By 2.

James Douglas Hamilton. *Time series analysis*. Princeton Univ. Press, Princeton, NJ, 1994. ISBN 0691042896.

A.M. Herrera, L. Hu, and D. Pastor. Forecasting crude oil price volatility. *International Journal of Forecasting*, 34(4):622–635, 2018. cited By 5.

John Hull and Alan White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

Carlos M. Jarque and Anil K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255 – 259, 1980. ISSN 0165-1765.

Gregor Kastner. Dealing with stochastic volatility in time series using the r package stochvol. *Journal of Statistical Software, Articles*, 69(5):1–30, 2016. ISSN 1548-7660.

C.G. Lamoureux and W.D. Lastrapes. Persistence in variance, structural change, and the garch model. *Journal of Business and Economic Statistics*, 8(2):225–234, 1990. cited By 505.

Z. Lin. Modelling and forecasting the stock market volatility of sse composite index using garch models. *Future Generation Computer Systems*, 79:960–972, 2018. cited By 2.

Asger Lunde and Peter Hansen. A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005.

J. Ma, C.R. Nelson, and R. Startz. Spurious inference in the garch (1,1) model when it is weakly identified. *Studies in Nonlinear Dynamics and Econometrics*, 11(1), 2007. cited By 23.

Jun Ma. A Closed-Form Asymptotic Variance-Covariance Matrix for the Maximum Likelihood Estimator of the GARCH(1,1) Model. Working Papers UWEC-2006-11-R, University of Washington, Department of Economics, October 2006.

Jun Ma and Charles R. Nelson. Valid Inference for a Class of Models Where Standard Inference Performs Poorly: Including Nonlinear Regression, ARMA, GARCH, and Unobserved Components. Economics Series 256, Institute for Advanced Studies, September 2010.

T. Mikosch and C. Stric. Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *Review of Economics and Statistics*, 86(1):378–390, 2004. cited By 242.

Andrea A. Naghi. Identification robust predictive ability testing (job market paper 2) most recent version : here. 2016.

C.R. Nelson and R. Startz. The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, 138(1):47–62, 2007. cited By 28.

Daniel B. Nelson. Stationarity and persistence in the garch(1,1) model. *Econometric Theory*, 6 (03):318–334, 1990.

O.M. Otunuga, G.S. Ladde, and N.G. Ladde. Local lagged adapted generalized method of moments: An innovative estimation and forecasting approach and its applications. *Journal of Time Series Econometrics*, 11(1), 2019. cited By 1.

A.J. Patton, J.F. Ziegel, and R. Chen. Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 2019. cited By 0.

S. Paul and P. Sharma. Quantile forecasts using the realized garch-evt approach. *Studies in Economics and Finance*, 35(4):481–504, 2018. cited By 0.

Y. Shen, X. Shi, and H.M.P. Variam. Risk transmission mechanism between energy markets: A var for var approach. *Energy Economics*, 75:377–388, 2018. cited By 4.

W. Shi, K.X. Li, Z. Yang, and G. Wang. Time-varying copula models in the shipping derivatives market. *Empirical Economics*, 53(3):1039–1058, 2017. cited By 5.

J. Song and J. Kang. Parameter change tests for armagarch models. *Computational Statistics and Data Analysis*, 121:41–56, 2018. cited By 4.

Stephen J. Taylor. Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4(2):183–204, 1994.

Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA, 1995.

Kenneth D. West. Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084, 1996. ISSN 00129682, 14680262.

Peter Whittle. Hypothesis testing in times series analysis. 1951.

X. Zhang and J. Li. Credit and market risks measurement in carbon financing for chinese banks. *Energy Economics*, 76:549–557, 2018. cited By 0.