IDENTIFICATION PROBLEMS IN WORKHORSE ECONOMETRIC MODELS: THE CASE OF THE GARCH MODEL

Alexander Kiel
386879

Casper Witlox
426233

Jan Jakob Heijenga
432951

Sander van Veen
434780

April 23, 2018

ABSTRACT

The invalidity of estimation results obtained under unidentified or weakly identified econometric model parameters has concrete implications. The following paper outlines those issues specifically for the GARCH(1,1) model. Under this context, the reliability of Bollerslev's (1987) results are examined. Finally, we propose methods to make the aforementioned empirical results of Bollerslev more robust to identification problems.

1 INTRODUCTION

The GARCH(p, q) model is an approach to estimate volatility in financial markets. In general, these models are able to capture potential volatility clusters, which is a frequently seen phenomenon in the financial sector. The intuitive idea behind the GARCH(p, q) model is that volatility will change only gradually over time, such that σ_t^2 is closely related to σ_{t-1}^2 up to σ_{t-p}^2. Moreover, the squared previous excess returns ε_{t-1}^2 until ε_{t-q}^2 are an 'ex post' measure of volatility, accordingly, these may also provide useful information on how volatility is changing. The temporal dependence of speculative prices in which time series are serially uncorrelated, however, not independent, has been given various attention in previous literature such as Mandelbrot (1963). Furthermore, subsequent research has focused on variance mixture models due to their greater descriptive validity. We refer to Boothe and Glassman (1987) for a more extensive discussion on the class of autoregressive conditional heteroskedastic (ARCH) models. Deviations from the general GARCH(p, q) model have gained tremendous popularity in Financal Econometris. A prime example is Bollerslev's (1987), GARCH(p, q)-t model which shall be later scrutinized in detail.

Perhaps less known to the interested reader is that the introduced GARCH model can suffer from important identification issues. Andrews and Cheng (2012) stresses that models that are unidentified retrieve unreliable parameter estimates. Specifically, if the parameters of a volatility capturing model are not well identified, standard estimation techniques, such as ordinary least squares, maximum likelihood or generalized method of moments, fail to capture the exact salient features of the underlying data generating process. In this scenario, Andrews and Cheng (2012) illustrate for the ARMA model that the parameter estimates of the model do not reflect the 'true' values of the underlying DGP. Consequentially, the estimators exhibit non-standard distributions and remain inconsistent. This results in incorrect constructed parameter confidence intervals. Furthermore, identification problems generate not only invalid estimation results but also affect test results one would like to perform a posteriori such as a Student's t-test, a Wald test or a Likelihood Ratio test. Likewise, Andrews and Cheng (2013) disclose that these statistical hypothesis tests are affected by unidentified parameters, leading to falsified hypothesis testing results. This can lead to incorrect conclusions regarding the significance of parameters.

Similar to the findings of Andrews and Cheng (2012, 2013) for the ARMA model, this paper aims to analogously analyze the GARCH model. We suspect that a considerable amount of research has overlooked the problem of identification. This unawareness has direct consequences. In a first step, we illustrate that for both different sample sizes and 'true' values of the underlying data generating process estimates, the distribution of the parameter estimates is not normal and thus they are unreliable. Disregarding potential identification problems in models which are for example employed by policy institutions in order to obtain forecasts, could ergo lead to erroneous policy implications. Current research on the effect of identification problems, specifically with respect to generalized autoregressive conditional heteroskedasticity (GARCH) models is rather limited. Prior to 2012, there were hardly any theoretical identification loss results obtained for this statistical model.

The following paragraphs canvass prior research conducted on the robustness of identification issues. Most research established around this field has been composed between 2012 and 2015, undeniably indicating the contemporary potential of this promising young field. Ultimately, additional research is required in order to gain further awareness of the vast consequences identification problems can have for empirical modeling results. Andrew and Cheng (2012, 2013, 2014), widely regarded as pioneers in the field of potential identification losses, firstly introduced methods to make tests and confidence sets less prone towards identification problems. In doing so, they refer to a class of models which is estimated by using maximum likelihood, least squares, quantiles, generalized method of moments, generalized empirical likelihoods, minimum distances, and semi-parametric estimators. Andrews and Cheng (2014) focus on the robustness of standard generalized method of moments (GMM) estimators. Instead, we solely focus on maximum likelihood estimation techniques in our paper. Furthermore, Andrews and Cheng (2014) analyze asymptotic properties of extremum estimators, t-tests, quasi-likelihood ratio (QLR) tests and confidence sets (CS). Thus, their research introduces statistical tests and confidence sets that are robust to identification issues by establishing estimator criterion functions. Consequently, the paper provides asymptotic results that are uniform over distributions that generate the observations. Based on their research, we will give a similar analysis for Bollerslev's results. Furthermore, Andrew and Cheng (2012) define the so-called *semi-strong* identification case, which bridges the gap between weak and strong identification.

The aforementioned asymptotic testing results are empirically verified on various model classes. Andrews and Cheng (2012) apply their analysis on an ARMA(1,1) model as a running example, which serves as the groundwork for our paper. Similarly, Andrews and Cheng (2013) employ their extended results to two models: a nonlinear binary choice model and the smooth transition threshold autoregressive model (STAR). Andrews and Cheng (2014) robust results' are applied to both a non-linear regression model with endogeneity and a probit model exhibiting weak instrumental variables. Finally, Cheng (2015) explicitly analyzes nonlinear regression models. In doing so, she studies the inference in regression models composed of nonlinear functions which encompass transformation parameters as well as loading coefficients that measure the importance of each component. Non-identification and weak identification are present in multiple parts of the parameter space, resulting in mixed identification strength for multiple unknown parameters. Cheng proposes robust tests, confidence intervals for sub-vectors and linear functions of the unknown parameters. In order to construct these robust inference procedures, she develops a local limit theory that models mixed identification strength. The asymptotic results involve both inconsistent estimators that depend on a localization parameter and consistent estimators with different rates of convergence. A sequential argument is used to peel the criterion function based on the identification strength of the parameters.
A popular extension of the general GARCH model by Blattberg and Gonedes (1974) goes along with the stylized facts of financial time series and introduces the idea that price changes and rates of returns might be better described by an alternative, unimodal symmetric distribution, for example, a t-distribution instead of a normal distribution. Bollerslev's GARCH(p, q)-t model, takes the aforementioned aspects into account and allows for both conditional heteroskedasticity and a conditional leptokurtic distribution (t-distributed errors). In contrast to Bollerslev's predecessors, his GARCH model considers explicitly the conditional dependence in the second moment, allowing for a temporal dependence within the time series. The main discrepancy between this model and the 'standard' GARCH$(1, 1)$ is that Bollerslev's GARCH$(1, 1)$-t model exhibits t-distributed errors instead of normal distributed errors. Our paper aims to enrich the extant but rather limited literature about unidentified volatility clustering models by analyzing the well-entrenched modeling results of Bollerslev (1987). We expect his results may be affected by weak or non identification. The results presented in his paper confirm that speculative price changes and rates of return series are approximately uncorrelated over time but characterized by tranquil and volatile periods. According to Bollerslev estimation results, the relatively simple GARCH(1,1)-t model fits his data series considered here quite well. Eager to extend the range of analyzed model classes, our paper adds value to the existing literature by explicitly focusing on the uninvestigated identification robustness of the aforementioned volatility capturing GARCH(1,1)-t model.
Both the parameter estimates and the statistical significance tests of the GARCH(1,1)-t model validate in Bollerslev's work the hypothesis that speculative price changes and rates of return series are approximately uncorrelated over time but characterized by tranquil and volatile periods. Bollerslev, however, does not take potential identification problems into account. Our paper empirically verifies Bollerslev's results by explicitly acknowledging possible parameter identification complications. Our main contribution is a critique on his empirical findings. The structure of this paper is as follows. In the *Methodology* we will familiarize the reader with the problem by mathematically annotating the unidentified case in the general GARCH(1,1) model. Furthermore, we depict an algorithm for our iterative estimation procedure. In the *Data* section, we swiftly describe the construction of our aforementioned data generation process and our use of Bollerslev's results. In the *Results* section, we firstly present distributions for both parameter estimates and corresponding t-test statistics (critical values). Moreover, in this section, we analyze to what extent these distributions are well-modeled by a normal distribution. Finally, we discuss potential extensions which go in line with the discussed literature and conclude our main findings (*Discussion/Conclusion*). The associated Matlab code is kindly available online. [1].

[1] Replication files (*MATLAB*) available at
`https://github.com/C-o-r/Econometrics_Seminar`

2 METHODOLOGY

2.1 THE UNIDENTIFIED GARCH(1,1) MODEL

A general GARCH(1,1) model typically take the following form:

$$\varepsilon_t = \sqrt{h_t} u_t, \qquad u_t \sim i.i.d. N(0,1) \tag{1}$$

$$\text{where } h_t = \alpha + \beta \varepsilon_{t-1}^2 + \pi h_{t-1} \tag{2}$$

It can be shown that this GARCH model has the following ARMA(1,1) representation: [2]

$$\varepsilon_t^2 = \alpha + (\pi + \beta)\varepsilon_{t-1}^2 + v_t - \pi v_{t-1} \tag{3}$$

with innovation

$$v_t = \varepsilon_t^2 - h_t = h_t(u_t^2 - 1) \tag{4}$$

The parameters of the GARCH(1,1) model are not identified when β is zero or close to zero. Notice here that when the difference between the AR and MA parameter of this model is zero (or very close to zero), i.e. $\beta = 0$, the model evolves into:

$$\varepsilon_t^2 = \alpha + \pi \varepsilon_{t-1}^2 + v_t - \pi v_{t-1} \tag{5}$$

where the loss of identification can be illustrated in the following manner:[3]

$$\varepsilon_t^2 - \pi \varepsilon_{t-1}^2 = \alpha + v_t - \pi v_{t-1} \tag{6}$$

$$\varepsilon_t^2 (1 - \pi L_1) = \alpha + v_t (1 - \pi L_1) \tag{7}$$

$$\varepsilon_t^2 = \frac{\alpha}{(1 - \pi L_1)} + v_t \tag{8}$$

In this case, $h_t = \frac{\alpha}{(1-\pi L_1)}$ [4]. In other words, the parameters for the AR term ε_{t-1}^2 and the lagged innovation v_{t-1} are not identified when they are equal or close to equal. This leads to various problems regarding the model's estimators, such as being unreliable but also testing results are distorted. In order to investigate the unreliability of the model's parameters, we simulate an underlying data generating process and estimate the GARCH(1,1) model parameters. Subsequently, we visualize the unreliability of the obtained estimators by graphically, depicting the estimation results for both the parameters, as well as the corresponding Student's t-tests via density functions.

2.2 IDENTIFICATION CATEGORIES

The aforementioned derivation outlines the unidentified case. Additionally, there are more stages of model identification: a model can be unidentified but also weakly, semi-strongly and strongly identified. In general, the more identified a model is, the more reliable its estimations are. Table 1 below outlines the different stages of identification for the GARCH-coefficient π as described by Andrews and Cheng (2012).

[2]By adding ε_t^2 to the left- and right-hand sided of (2) and subtracting h_t

[3]Where L_1 denotes the first lag operator.

[4]Because $h_t = \varepsilon_t^2 - v_t$

Category	$\{\beta_T\}$ Sequence	Identification Property of π
I(a)	$\beta_T = 0 \; \forall T \geq 1$	Unidentified
I(b)	$\beta_T \neq 0$ and $T^{\frac{1}{2}}\beta_T \rightarrow b \in R^{d\beta}$ (and, hence, $\|\beta_T\| = O(T^{-\frac{1}{2}})$)	Weakly identified
II	$\beta_T \rightarrow 0$ and $T^{\frac{1}{2}}\|\beta_T\| \rightarrow \infty$	Semi-strongly identified
III	$\beta_T \rightarrow \beta_0 \neq 0$	Strongly identified

Table 1: Identification Categories ranked by Andrews and Cheng (2012)

2.3 Stationarity

According to Nelson (1990) the concept of stationarity contemplates the intuitive idea that the underlying DGP exhibits invariant properties such as a stable conditional variance. In a covariance stationary GARCH(1,1) process, $E_t[h_{t+k}]$ converges to $\bar{h}$ as $k \rightarrow \infty$. In order to obtain a stationary GARCH(1,1) model various restrictions on the parameter space are applied:

$$\beta + \pi < 1 \tag{9}$$

$$\alpha > 0 \tag{10}$$

$$\beta \geq 0 \tag{11}$$

$$\pi \geq 0 \tag{12}$$

A GARCH(1,1) model which exhibits the above restrictions is deemed to be stable. When these constraints are not met, the process exhibits a potentially unstable and explosive structure. Whereas we initially attached relatively little importance to this restricted parameter space, it later during our research became a vital prerequisite for our robustness analysis (see *Results* section).

2.4 Sample sizes and Simulation runs

We select sample sizes of $T = 150, 250, 500$. In economic applications, these sample sizes are often used, thus working with quite small sample sizes seems appropriate. Moreover, based on Andrews & Cheng (2014) findings identification issues seem to disappear for large sample sizes of around 1000 or bigger. Thus, it is most interesting to focus on this range of sample sizes. In order to estimate the densities of the coefficients, we conduct a vast amount of simulations. The simulations were performed on a HP Elitedesk 705 computer equipped with an AMD64 Family 21 Model 48 Stepping 1 AuthenticAMD processor and 16 GB of RAM running a 64-bit version of Windows 10. We combine all obtained coefficients in order to construct the density functions. As a consequence, the trade-off we encounter is time efficiency against accuracy. If we increase the amount of simulation runs, the densities of the coefficients attain not only a much more continuous and smoother shape but also become more accurate. Nonetheless, we are exposed to a much longer completion time. In the figures underneath, we graphically depict the densities for α, β and π from left to right. Keeping the amount of simulation runs constant in each figure, we compare the estimation results of two independent estimation runs which are separately depicted in each row. We compare the density distributions of the corresponding parameter estimates for the two independent estimation runs and examine whether those runs yield similar or equal results. Based on these comparisons, we try to obtain the best suited value for n .

We start with 500 simulations runs ($n = 500$). In Figure 1, we graphically depict the densities from left to right that we find for α, β and π respectively. Both in the first and second row, we find densities for the exact same variables, thus the regressions we run as shown in the first row we ran again in the second row. The histograms don't look completely similar, also the densities are not really smooth yet. Thus, we conclude that it might be more appropriate to increase the amount of simulation runs in order to retrieve both smoother and more accurate densities.

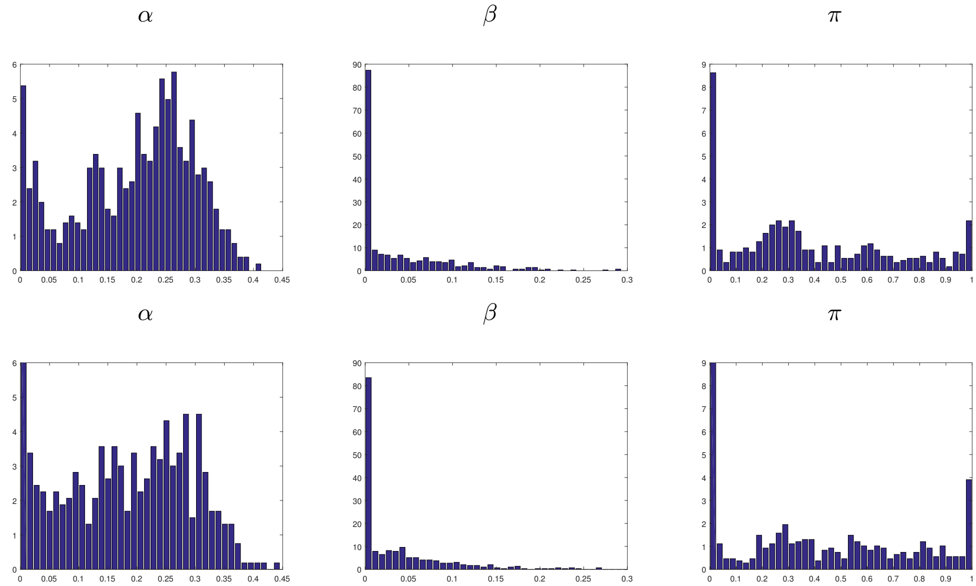

Figure 1: Finite-sample ($T = 150$) densities of the estimators of α, β and π in the GARCH(1,1) model with 500 simulations when $\alpha_0 = 0.2$, $\beta_0 = 0$ and $\pi_0 = 0.4$.

By using 1000 simulation runs ($n = 1000$), we obtain smoother and more accurate densities. However, especially the density for α clearly differs between the two different runs. Accordingly, we further increase the amount of simulation runs.

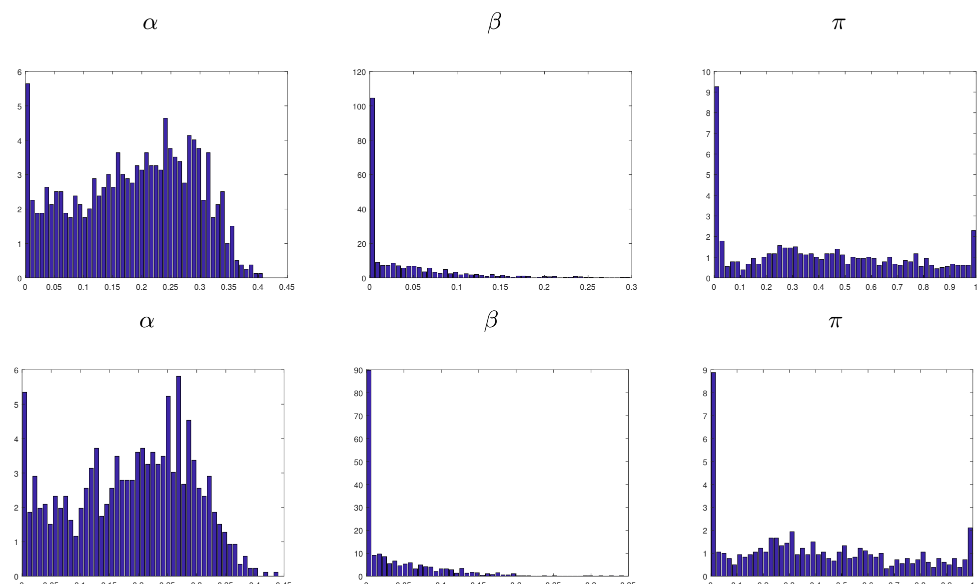

Figure 2: Finite-sample ($T = 150$) densities of the estimators of α, β and π in the GARCH(1,1) model with 1000 simulations when $\alpha_0 = 0.2$, $\beta_0 = 0$ and $\pi_0 = 0.4$.

Lastly, we perform 10000 simulations ($n = 10000$). In doing so, we obtain consistent and smooth densities as shown in Figure 3. In conclusion, we annotate as an intermediate result that it is reasonable to use 10000 simulations for the densities.

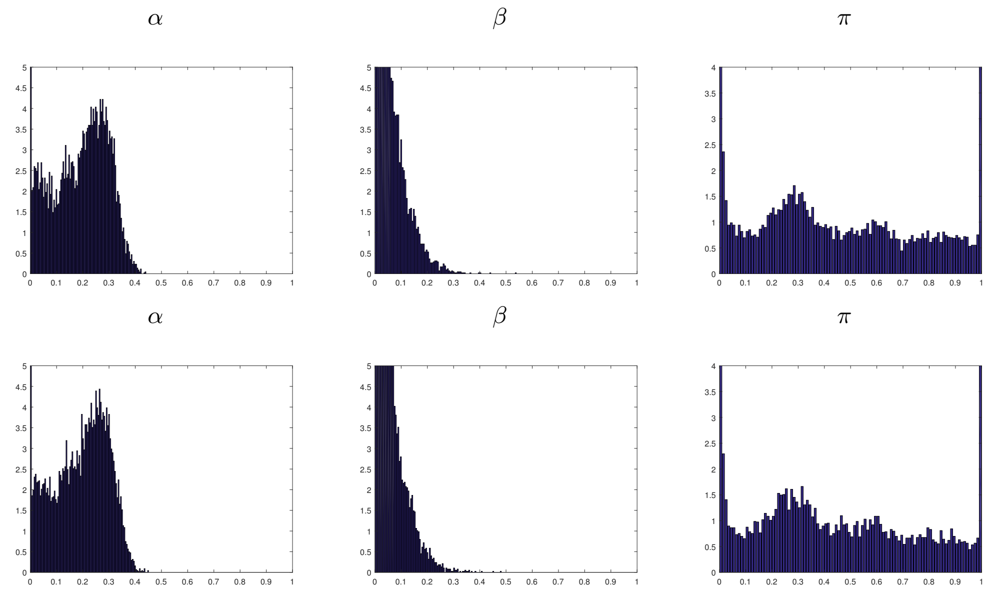

Figure 3: Finite-sample ($T = 150$) densities of the estimators of α, β and π in the GARCH(1,1) model with 10000 simulations when $\alpha_0 = 0.2$, $\beta_0 = 0$ and $\pi_0 = 0.4$.

2.5 True Values for β and the iterative Parameter estimation

To get a sense of the problems caused by different identification categories, see section 2.2, we check the densities of the coefficients for different true values of β. We decide to use the same values for b as done by Andrews and Cheng (2012), namely $b = 0, 2, 4, 12$. Only for the small sample sizes, (150 and 250) we cannot use b equal to 12 due to stationarity problems as explained in section 2.3. In those cases we use the maximum possible integer value for b that ensures that the model remains stable, which is 7 and 9 respectively. The iterative procedure of simulating an underlying data generating process and subsequently estimating both the GARCH(1,1) model parameters as well as the corresponding t-statistic is carried out via the numerical computing environment MATLAB. Note that in the pseudocode below we disdain Bollerslev's t-distributed error terms on purpose for now and depict the algorithm for obtaining parameter estimations in a first step. The algorithm can be depicted as follows:

Algorithm 1 GARCH(1,1)-Estimation(T,n,α_0,π_0 ,b)

create double[][] $Estimations \leftarrow$ new double[3][n]
create double[] $Errors \leftarrow$ new double[n]
create double[][] $Parameters \leftarrow$ new double[3][1]

//control for asymptotically normally distributed estimators
$\beta_0 \leftarrow \frac{b}{\sqrt{T}}$

$Parameters \leftarrow$[α_0;β_0;π_0]

if ($\beta_0 + \pi_0 \geq 1$ or $\alpha_0 \leq 0$ or $\pi_0 < 0$ or $\beta_0 < 0$)
IllegalArgumentException('GARCH(1,1) stationarity assumptions violated')

for j:=1 to n **do**

// @Author: Kevin Sheppard (kevin.sheppard@economics.ox.ac.uk)
$Errors$←GARCH(1,1)-simulate[5](T, $Parameters$);

$Estimations$[:,j]← GARCH(1,1)-Estimation[6]($Errors$);
end

//depict α_{est}
print histogram($Estimations$[1, :])
//depict β_{est}
print histogram($Estimations$[2, :])
//depict π_{est}
print histogram($Estimations$[3, :])

Similarly to obtaining parameter estimates for the GARCH(1,1) model in section 2.5, we estimate the parameters and t-values from the GARCH(1,1)-t model.

2.6 Least favorable critical value analysis

Andrews & Soares (2010) argue that the parameter estimate for the constant in this model, α, is consistent. Moreover, β is consistent, but cannot be plugged into the GDP, as the model's identification strength is unknown. Meanwhile, their finding argue that π is not consistent. Accordingly, we need to vary both β and π when constructing the least favorable critical values for the GARCH coefficient π.

3 DATA

3.1 SIMULATED DATA

For the examination of identification problems, we opt to simulate an underlying data generating process in order to determine the reliability of our model estimation results. We analyze the estimates of the GARCH model parameters α, β and π at different 'true' underlying values of β for different sample sizes (T). Subsequently, we artificially construct an underlying DGP which generates data following a GARCH-process. Using our DGP we could easily create this data (see Methodology). Thus, we evaluate the estimation results for different underlying generated data. For every sample size and b we generate data of this sample size 10000 times in order to get 10000 different coefficient estimates. This gives us useful insights about the distribution of the different coefficients and t-test statistics. Because large sample sizes retrieve asymptotically normal parameter results, we set β of the DGP equal to ${}^{b}/_{\sqrt{T}}$, where b varies from 0 to x, where x is the largest possible integer for which the model is still stationary. In that manner, we re-parameterize β as a drifting sequence that depends explicitly on the sample size. Accordingly, we retain weak identification problems at every sample size.

In a second step, we replicate data from Bollerslev (1987). Bollerslev's first data set consists of daily spot prices from the New York foreign exchange market on the U.S. dollar versus the British pound and the Deutschmark from March 1, 1980 until January 28, 1985 for a total of $1,245$ observations excluding weekends and holidays. The spot prices are converted to continuously compounded rates of return. The second data set includes five different monthly stock price indices for the U.S. economy, namely Standard and Poor's 500 Composite, Industrial, Capital Goods, Consumer Goods and Public Utilities price indices. These indices are monthly averages of daily prices from 1947.1 to 1984.9, i.e., 453 observations. Bollerslev's coefficient results for these two datasets are subsequently used in our research in order to create an underlying GARCH-t-process. Based on this underlying data generating process we are able to obtain (un)identified GARCH$(1,1)-t$ estimates.

4 RESULTS

4.1 ESTIMATORS

In the figures below, the obtained density results are depicted. All estimators closely approximate a normal distribution as b is increasing. From left to right, we illustrate densities for $b = 0, 2, 4, x$, where x is the maximum possible integer value, respectively 7, 9 and 12, of b for which the model's stationary conditions are satisfied, see 2.5. For the different estimators b, we find densities for sample sizes $T = 150, 250, 500$. We also observe that an increase in sample size results in a smoother normal distribution.

We have apply the Jarque-Bera test for normality on every density below, with a significance level of 5%. Even though a few parameter densities seem to take on the normal distribution, we reject normality in every case.

Table 2 shows how b is converted into β_T, by using the following formula: $\beta_T = b/\sqrt{T}$

T/b	0	2	4	7	9	12
150	0	0.1633	0.3266	0.5715	-	-
250	0	0.1265	0.2530	-	0.5692	-
500	0	0.0894	0.1789	-	-	0.5367

Table 2: Conversion from b to β_T

In what follows, we have plotted the densities of the estimator of α, for different sample sizes.

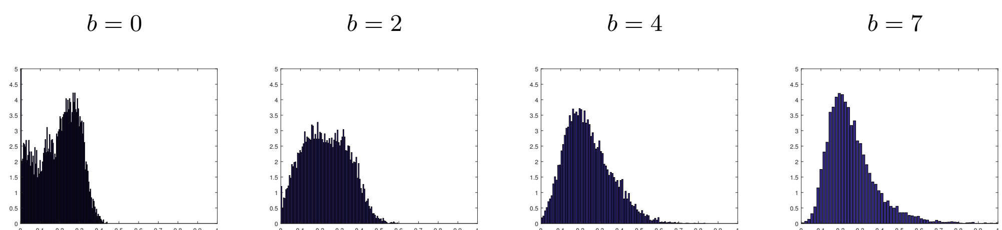

Figure 4: Finite-sample ($T = 150$) densities of the estimator of α in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{150}}$ and $\pi_0 = 0.4$.

As you can see in figure 4, an increment in b results in a smoother density function. None of these figures, however, appear to act like a normally distributed function. In case of $b = 0$, the model is not identified, which explains the not-normal distribution in the first density. In case of $b = 2$ and $b = 4$ the model is weakly identified. For $b = 7$, the model is still weakly identified, but is approaching a normal distribution steadily and getting closer to semi-strong identification. As mentioned before, all Jarque-Bera tests reject the null hypothesis of normality thus, with this sample size, the estimation results of α are unreliable.

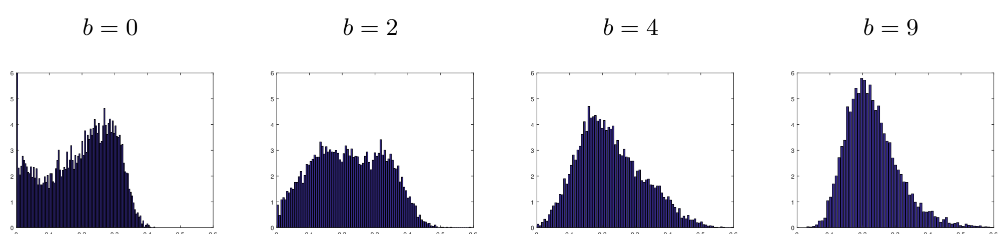

Figure 5: Finite-sample ($T = 250$) densities of the estimator of α in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{250}}$ and $\pi_0 = 0.4$.

Both for a sample size of $T = 150$ as for $T = 250$, the same conclusion can be drawn. In case of $b = 0$, the model is not identified, which explains absence of a normal distribution in the first histogram in Figure 5. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 9$, which looks more like a normal distribution, the model is closer to being semi-strongly identified. All Jarque-Bera tests reject the null hypothesis of normality with probability $p < 0.00$. Even though, for $b = 9$, the function looks quite normal, there still remains a small positive skewness, thus the estimation results are not totally reliable,

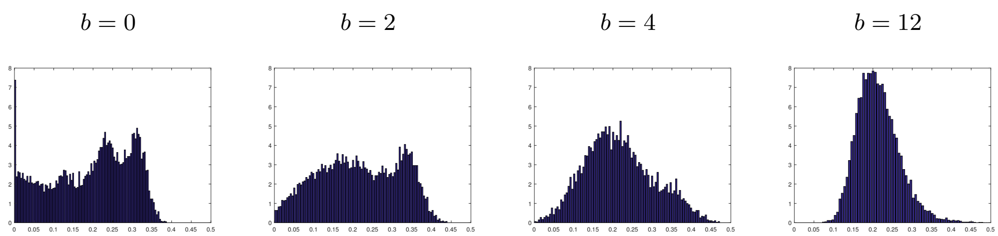

Figure 6: Finite-sample ($T = 500$) densities of the estimator of α in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{500}}$ and $\pi_0 = 0.4$.

As one can clearly see, the first three densities in Figure 6 are far from being normally distributed. The last figure, however, seems to be normally distributed, but also for this density, just like the other ones in Figure 6, its Jarque-Bera test rejects the null hypothesis of normality. In case of $b = 0$, the model is not identified, which explains the absence of a normal distribution in the first density in 6. In case of $b = 2$ and $b = 4$, -the model is weakly identified and for $b = 12$, which looks quite like a normal distribution, the model is close to being semi-strong identified.Thus, we can conclude that the estimation results are not fully reliable.

In what follows, we have plotted the densities of the estimator of β.

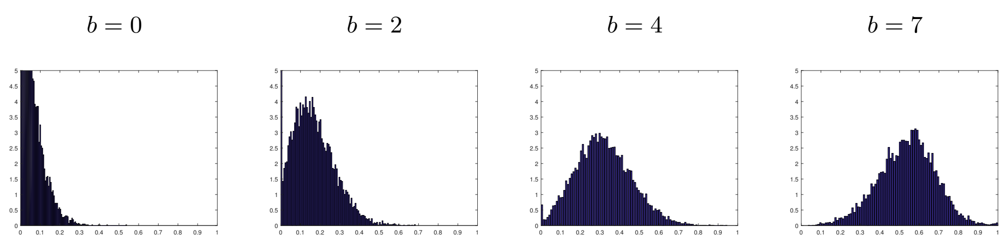

Figure 7: Finite-sample ($T = 150$) densities of the +estimator of β in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{150}}$ and $\pi_0 = 0.4$.

Clearly, the first density in Figure 7 is not normally distributed. The same observation applies to the other densities, because their skewness's are different from zero. For the second and third densities the skewness is positive, whereas negative skewness occurs in the last histogram. This is also shown by applying Jarque-Bera tests on this densities, which all reject the null hypothesis of normality with probability $p < 0.00$. In case of $b = 0$, the model is not identified, which explains the not-normal distribution in the first density in 7. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 7$, which looks more like a normal distribution, the model is getting closer to being semi-strong identified.

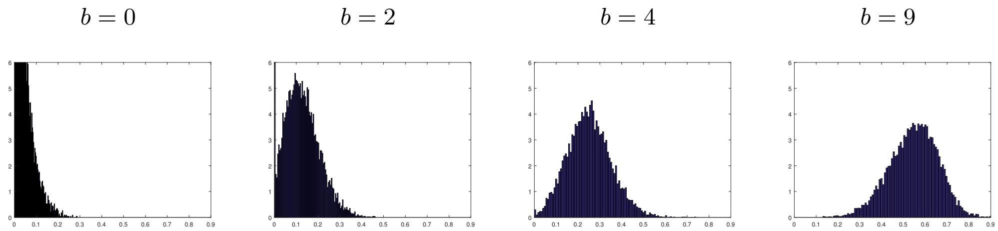

Figure 8: Finite-sample ($T = 250$) densities of the estimator of β in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{250}}$ and $\pi_0 = 0.4$.

As one can clearly see, the first density in Figure 8 is far from being normally distributed. Accordingly, the other figures are not normally distributed due to a skewed distribution. On one hand a positive skewness in the second and third density, on the other a negative skewness in the last density. This observation is supported by applying a Jarque-Bera test on these densities, where all tests reject the null hypothesis of normality. This notion can be explained by considering identification issues. In case of $b = 0$, the model is not identified, which explains the not normal distribution in the first density in Figure 8. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 9$, which looks more like a normal distribution, the model is closer to semi-strong identified.

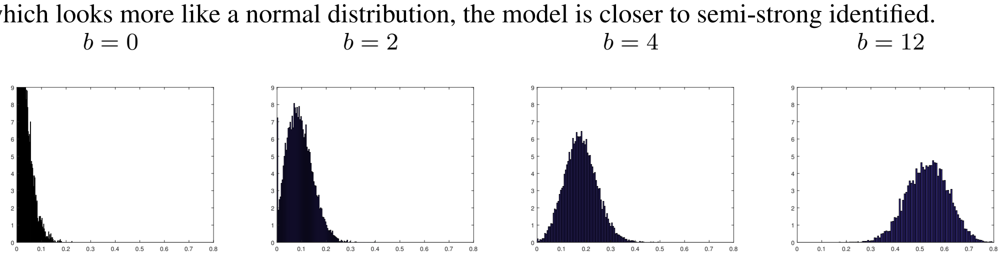

Figure 9: Finite-sample ($T = 500$) densities of the estimator of β in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{500}}$ and $\pi_0 = 0.4$.

Moreover, for a sample size of $T = 500$, the first density in Figure 9 is clearly not normally distributed. The second and third are also not normal due to their distribution's skewness. By applying Jarque-Bera tests on these three densities, normality is convincingly rejected with probability $p < 0.00$. The last density also is not normally distributed, with a Jarque-Bera test rejecting normality with a test statistic of 18.48863 and probability $p = 0.000097$. In case of $b = 0$, the model is not identified, which explains the not-normal distribution in the first density in Figure 3. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 12$, which looks more like a normal distribution, the model is close to semi-strong identified. Conclusively, our observations are in accordance with possible identification issues.

In what follows, we have plotted the densities of the estimator of π.

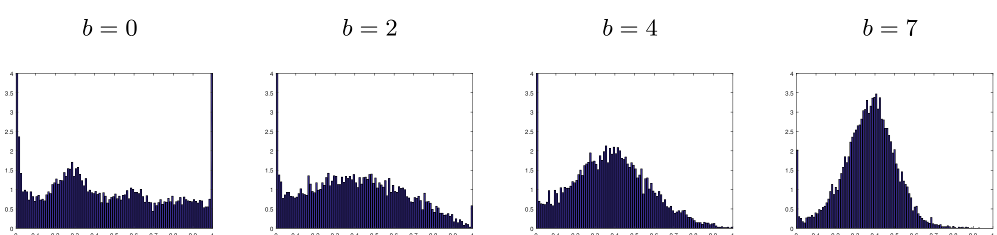

Figure 10: Finite-sample ($T = 150$) densities of the estimator of π in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{150}}$ and $\pi_0 = 0.4$.

As the densities in Figure 10 show, the first three densities are clearly not normally distributed. The last density, however, looks to be normally distributed, but due to a steep peak for smaller values, normality is also rejected in this case. Accordingly, all Jarque-Bera tests reject the null hypothesis of normality with probability $p < 0.00$. These observations are in accordance with identification issues. In case of $b = 0$, the model is not identified, which explains the not-normal distribution in the first density in Figure 10. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 7$, which looks more like a normal distribution, the model is getting closer to semi-strong identified.

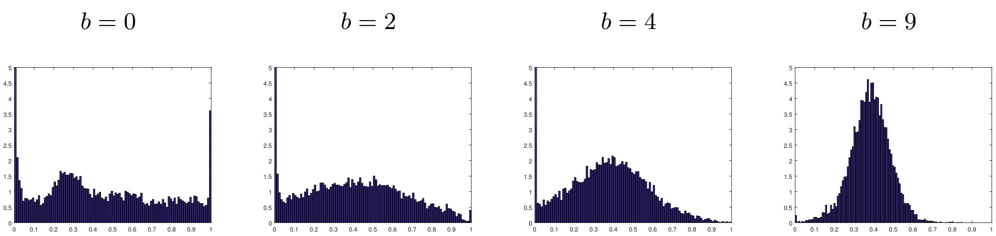

Figure 11: Finite-sample ($T = 250$) densities of the estimator of π in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{250}}$ and $\pi_0 = 0.4$.

Clearly, Figure 11 indicates that the first three densities are not normally distributed. When looking at the shape of the last density, the density might be normally distributed. However, by applying Jarque-Bera tests, normality is convincingly rejected for all densities. These observations are in accordance with identification issues. In case of $b = 0$, the model is not identified, which explains the not-normal distribution in the first density in Figure 11. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 9$, which looks more like a normal distribution, the model is closer to semi-strong identified.

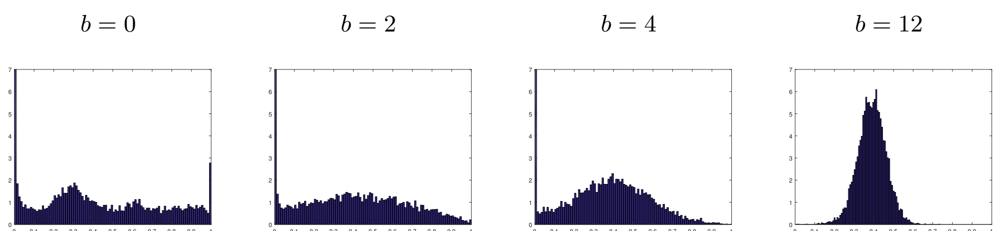

Figure 12: Finite-sample ($T = 500$) densities of the estimator of π in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{500}}$ and $\pi_0 = 0.4$.

Also in case of a sample size of 500, the first three densities, as shown in Figure 12, are very far from normal. The last density appears to be closer to a normal distribution, but by applying a Jarque-Bera test, also in this case, just like in the first three densities, normality is rejected , even tough it appears to be already close to normal. In case of $b = 0$, the model is not identified, which explains the not-normal distribution in the first density in Figure 12. In case of $b = 2$ and $b = 4$ the model is weakly identified and for $b = 12$, which looks more like a normal distribution, the model is close to semi-strong identified. Thus these observations are in accordance with identification issues.

4.1.1 Normally distributed estimators

In short, we don't find any normal distributed densities for any of the coefficients, there are multiple reasons for that. First of all due to time issues it is impossible to increase sample size and number of simulations any further (to infinity). The computation time for the densities is already very high and increasing sample size and simulation both to some very big number will make computing densities on a normal computer not doable. We do know from previous literature that doing this would actually result in a density much closer to normal.

This problem also may be caused due to stationarity issues. For our π of 0.4 is relatively large and $\pi + \beta$ should be smaller than 1 (in case of stationarity). When trying to solve this problem, our distributions get skewed to the left if we use a big β (bigger than 0.4). If we use a β smaller than 0.4, however, we also don't obtain normal densities due to identification issues. Thus besides identification issues, the GARCH estimator distributions can also be affected by stationarity restrictions.

If we reduce the true value of π these problems are solved for β. Unfortunately, these problems remain for α and π due to their small values. This leads to their distributions being truncated on the left side, because α and π should be bigger than zero each.

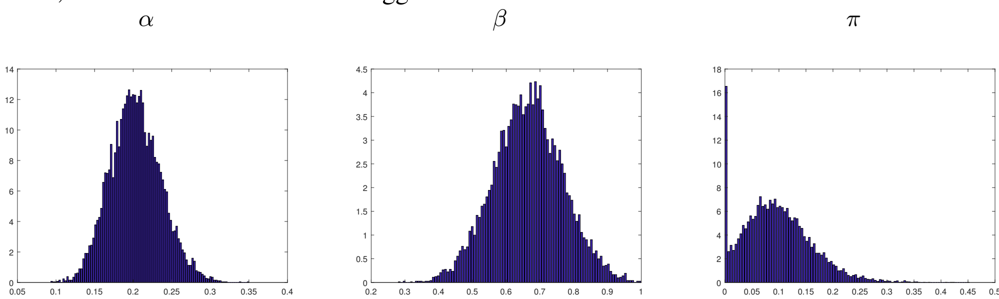

Figure 13: Finite-sample ($T = 500$) densities of the estimators of α, β and π in the GARCH(1,1) model with 10000 simulations when $\alpha_0 = 0.2$, $\beta_0 = \frac{15}{\sqrt{500}}$ and $\pi_0 = 0.1$.

By computing densities for $\pi_0 = 0.1$, we obtain histograms as shown in Figure 13. Now, by performing a Jarque-Bera test on the coefficients α, β, π, we find that β is normally distributed (see figure 14) with a Jarque-Bera test statistic of 2.84. The other two coefficients are, as to be expected, still not normally distributed.

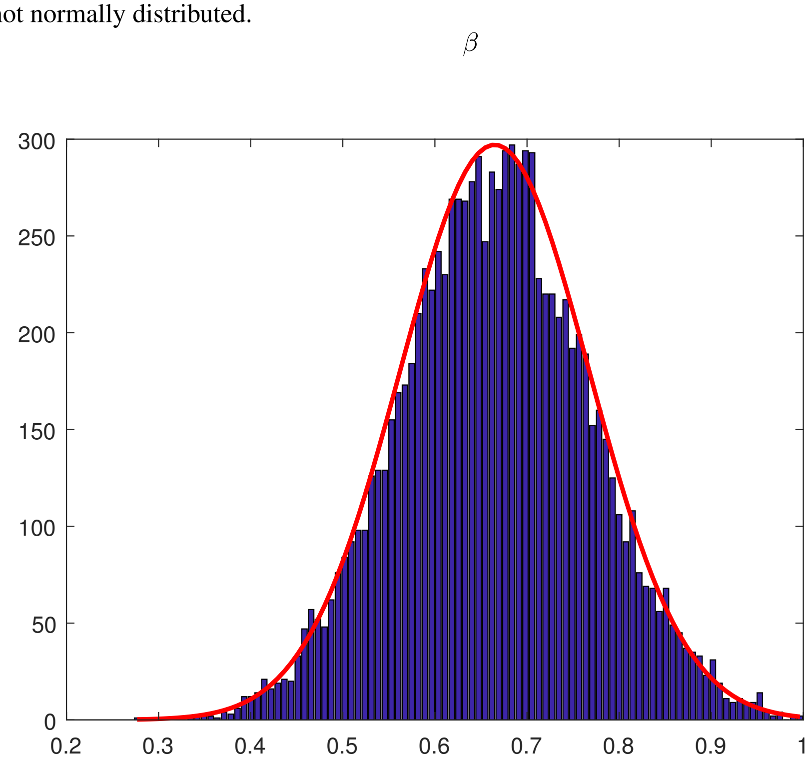

Figure 14: Beta histogram (not a density) with a normal N(0.66, 0.01) distribution line to illustrate normality of the β coefficient

4.1.2 Confidence Intervals

These results cause major problems when constructing confidence intervals. For a normal-distributed estimator, its 95% confidence interval can be constructed as follows:

$$\text{CI} = \text{estimated value} \pm z_{0.975} * \frac{\sigma}{\sqrt{T}} \tag{13}$$

where, $z_{0.975}$ is the 97.5 percent quantile of the normal distribution, $\frac{\sigma}{\sqrt{T}}$ is the distribution's standard error, where σ is the standard deviation and T is the sample size. Therefore, confidence intervals for normal-distributed estimators are dependent on its quantiles. When an estimator is not normally distributed, the values for its quantiles differ from those of a normal distribution. Consequently, we cannot use the standard $z_{0.975}$, the 97.5 percent quantile of the normal estimate distributions, to construct a 95 percent confidence interval, thus the constructed confidence intervals are unreliable when the model's coefficients are not identified.

4.2 DENSITIES OF THE t-TEST STATISTICS

In Andrews and Cheng (2012) research on the ARMA model they found that also distributions of the t-test statistic are affected by weak or non identification. Which is quite easy to see when looking at the formula for the t-test statistic. For example the t-test statistic for coefficient estimate β;

$$t = \frac{\hat{\beta} - \beta_0}{SE_\beta} \tag{14}$$

As seen in formula 14 the distribution of t is directly affected by the distribution of $\hat{\beta}$. As the distributions of the coefficients are not normal when there is non or weak identification we can expect the t-test statistics to also be not.

This is an interesting issue to address regarding the GARCH(1,1) model. A non-normal distribution for the t-test statistic means incorrect conclusions when testing for significance.

Thus t-test statistic distributions for the different parameters for different values of β_0, a null hypothesis of $\hat{\beta} = 0$ and different sample sizes are shown below.

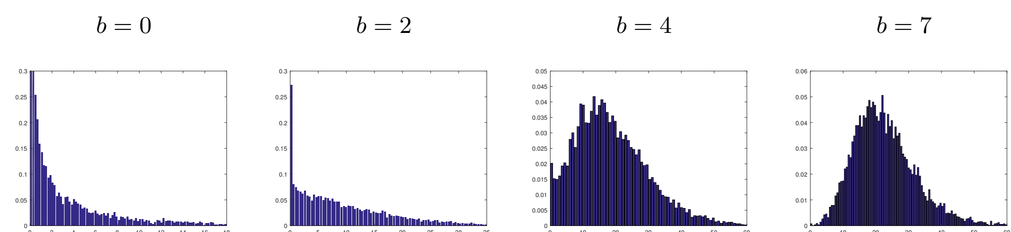

Figure 15: Finite-sample ($T = 150$) densities of the t-test statistic for the constant parameter α in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{150}}$ and $\pi_0 = 0.4$.

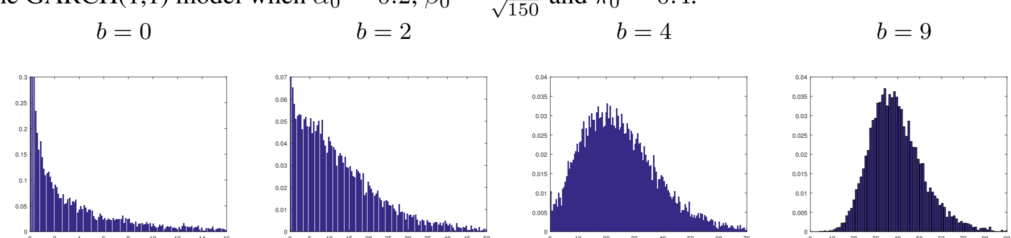

Figure 16: Finite-sample ($T = 250$) densities of the t-test statistic for the constant parameter α in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{250}}$ and $\pi_0 = 0.4$.

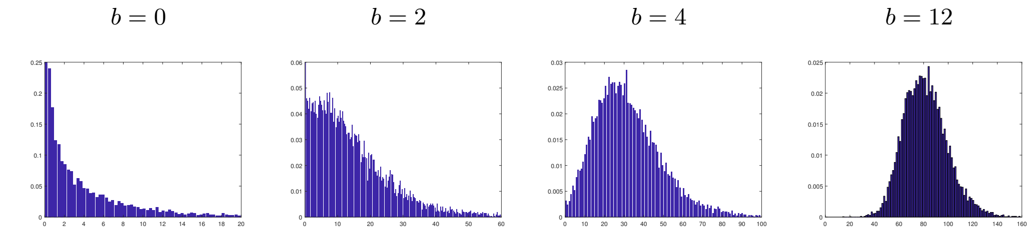

Figure 17: Finite-sample ($T = 500$) densities of the t-test statistic for the constant parameter α in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{500}}$ and $\pi_0 = 0.4$.

We see the same phenomenon for the t-test statistics of α as for the coefficient estimates of α. That is, as true value β_0 is increasing, we see the distribution of the t-test statistics approaching normality. Even for the highest possible values for β_0 Jarque-Bera tests for normality are rejected. Also we see that a higher sample size results in more normal distributed t-test statistics, with the closest to normal distribution attained when $b = 12$ and $T = 500$.

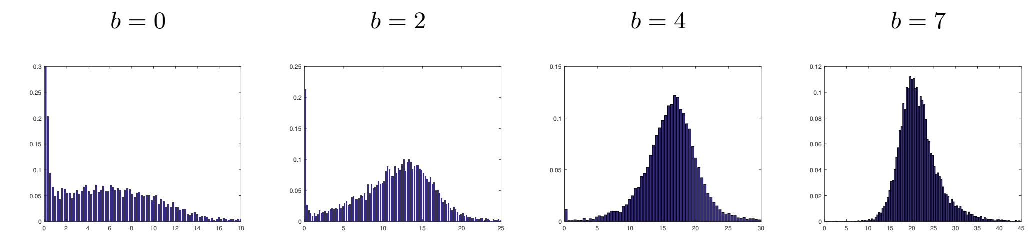

Figure 18: Finite-sample ($T = 150$) densities of the t-test statistic for the ARCH parameter β in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{150}}$ and $\pi_0 = 0.4$.

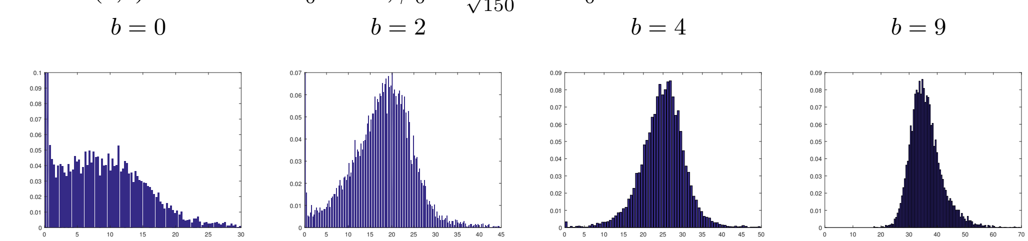

Figure 19: Finite-sample ($T = 250$) densities of the t-test statistic for the ARCH parameter β in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{250}}$ and $\pi_0 = 0.4$.

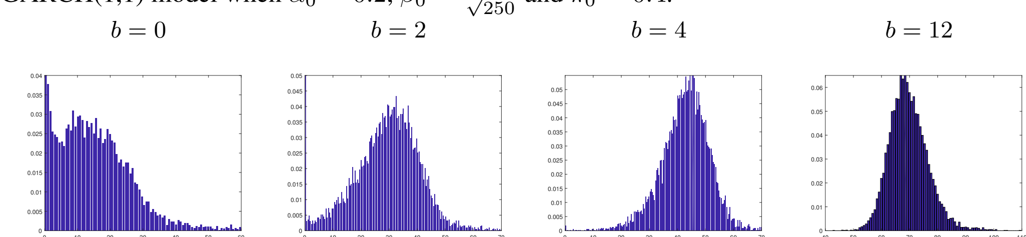

Figure 20: Finite-sample ($T = 500$) densities of the t-test statistic for the ARCH parameter β in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{500}}$ and $\pi_0 = 0.4$.

We see the same phenomenon happening for the t-test statistics of β as for the coefficient estimates of β. That is, with true value β_0 increasing we see the distribution of the t-test statistics getting closer to normal. Where we start with a chi-squared looking distribution for the α coefficient, the β coefficient looks more uniformly distributed when there is weak or non identification. As with the α estimates even with the highest possible values for β_0 we find a high Jarque-Bera test statistic for normality, thus we reject normality. Also we see that a higher sample size results in more normal distributed t-test statistics, with the closest to normal distribution attained when $b = 12$ and $T = 500$. For the smaller sample sizes 150 and 250 we find a fat right tail.

$b = 0$ $b = 2$ $b = 4$ $b = 7$

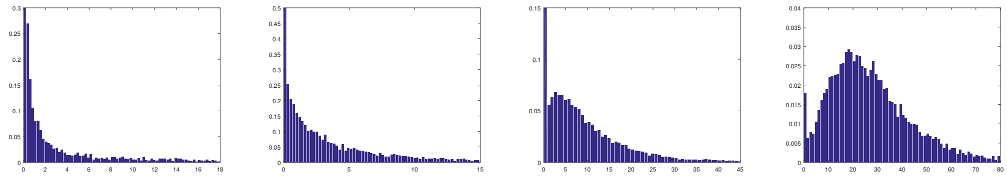

Figure 21: Finite-sample ($T = 150$) densities of the t-test statistic for the GARCH parameter π in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{150}}$ and $\pi_0 = 0.4$.

$b = 0$ $b = 2$ $b = 4$ $b = 9$

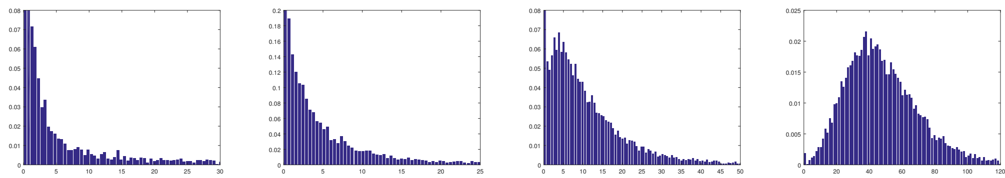

Figure 22: Finite-sample ($T = 250$) densities of the t-test statistic for the GARCH parameter π in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{250}}$ and $\pi_0 = 0.4$.

$b = 0$ $b = 2$ $b = 4$ $b = 12$

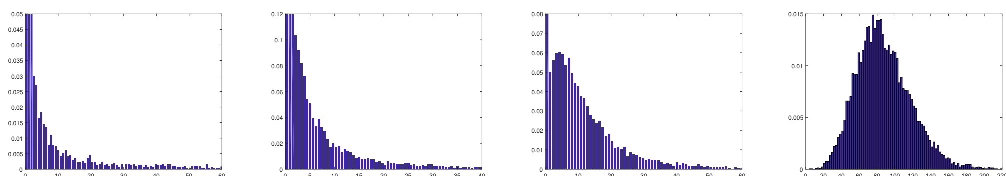

Figure 23: Finite-sample ($T = 500$) densities of the t-test statistic for the GARCH parameter π in the GARCH(1,1) model when $\alpha_0 = 0.2$, $\beta_0 = \frac{b}{\sqrt{500}}$ and $\pi_0 = 0.4$.

We see the same phenomenon happening for the t-test statistics of π as for the coefficient estimates of π. That is, with true value β_0 increasing we see that the distribution of the t-test statistics is getting closer to normal. The t-test statistic densities of π look most like the t-test statistic densities of α. Nonetheless, we see a notably high peek at zero for $b = 4$. Also as for the other coefficients, even with the highest possible values for β_0, we reject normality. Also we see that a higher sample size results in more normal distributed t-test statistics, with the closest to normal distribution attained when $b = 12$ and $T = 500$. Again we see fat right tails for smaller sample sizes.

4.3 Published research paper example: Bollerslev (1987)

4.3.1 Bollerslev's estimation results

If you look at the coefficient estimates for β in Bollerslev (1987), you see that these values are very small. Also the sample sizes of the data Bollerslev (1987) used were quite small, 453 and 1245. Therefore there seemed to be a high change of a non-identified or weak identified model, it is interesting to look at his results.

There are two ways to evaluate the results, table 3. One way assumes that the estimation results in the paper are wrong due to identification problems and thus we can say his results are unreliable. Option 2: the estimation results are right, we can use them as true values and create densities of the coefficients. Let's assume option 2 is the case for the data in the Bollerslev paper. Using the estimation results, we have simulated density functions of the estimators of α, β and π. We changed our Matlab code to be able to create GARCH data with student-t errors as in Bollerslev (1987). We also used student-t errors in the GARCH estimation as in the Bollerslev paper.

We used the 'true' values as in the Bollerslev results to create the GARCH-data in our DGP. Thus, for example for the Standard and Poor's 500 Composite data for α_0 we used 0.00017, for β_0 we worked with 0.074 and for π_0 and the degrees of freedom we used 0.768 and 7.194 respectively. Besides, we used the same sample size as in the paper ($T = 453$) for all stock indices and a sample size of 1245 for the exchange rates. These values are shown in table 3

	μ	α	β	π	$1/dof$
Exchange Rates					
Britain	-4.56×10^{-4}	0.96×10^{-6}	0.057	0.921	0.123
	(1.67×10^{-4})	(0.46×10^{-6})	(0.017)	(0.023)	(0.024)
Germany	-5.86×10^{-4}	0.13×10^{-6}	0.095	0.881	0.072
	(1.78×10^{-4})	(0.59×10^{-6})	(0.021)	(0.027)	(0.022)
Stock Price Indices					
500 Composite	5.63×10^{-3}	0.17×10^{-3}	0.074	0.768	0.139
	(1.46×10^{-3})	(0.13×10^{-3})	(0.045)	(0.148)	(0.055)
Industrial	5.73×10^{-3}	0.19×10^{-3}	0.078	0.756	0.129
	(1.51×10^{-3})	(0.14×10^{-3})	(0.046)	(0.150)	(0.054)
Capital Goods	5.78×10^{-3}	0.27×10^{-3}	0.065	0.751	0.140
	(1.73×10^{-3})	(0.23×10^{-3})	(0.048)	(0.183)	(0.052)
Consumer Goods	5.53×10^{-3}	0.14×10^{-3}	0.103	0.782	0.161
	(1.48×10^{-3})	(0.07×10^{-3})	(0.041)	(0.081)	(0.052)
Public Utilities	4.32×10^{-3}	0.05×10^{-3}	0.123	0.820	0.131
	(1.27×10^{-3})	(0.03×10^{-3})	(0.047)	(0.066)	(0.049)

Table 3: Bollerslev (1987)

We find the following densities for the estimators.

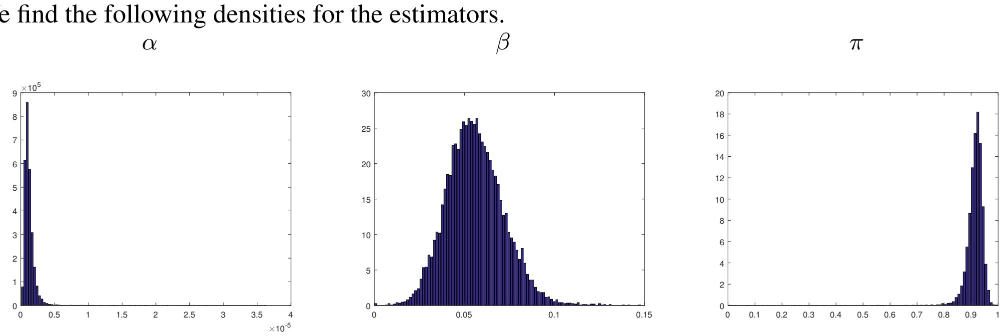

Figure 24: Finite-sample ($T = 1245$) densities of the estimators of the Britain Exchange Rate with α, β, π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.96 \times 10^{-6}$, $\beta_0 = 0.057$, $\pi_0 = 0.921$ and $dof = 8.13$.

For the α coefficient of Britain exchange rate we see a large build up of mass between 0 and 0.5×10^{-5}. Besides that we see a uniformly looking distribution between 0.5×10^{-5} and 3.5×10^{-5}, thus nothing close to normal. The β coefficient looks closer to normal distributed but is slightly positively skewed. We reject normality. For the π coefficient of Britain exchange rate we see a large build up of mass between 0.85 and 1. Besides that we see a uniformly looking distribution between 0 and 0.85, thus nothing close to normal.

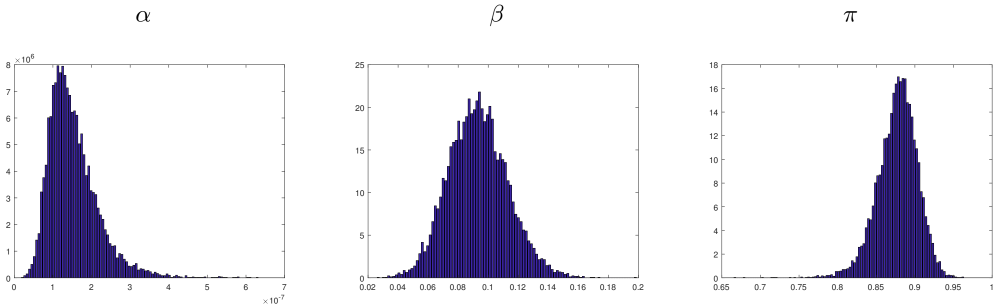

Figure 25: Finite-sample ($T = 1245$) densities of the estimators of the Germany Exchange Rate with α, β, π in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.13 \times 10^{-6}$, $\beta_0 = 0.095$, $\pi_0 = 0.881$ and $dof = 13.89$.

The coefficients of Germany exchange rate look closer to normal distributed than those of the Britain exchange rates. This is in line with our expectations as β_0 is quite bigger than for the German exchange rate. We see that α is quite substantially positively skewed and π is quite substantially negatively skewed. However β looks quite normal distributed. We still reject normality due to a quite significant skewness of 0.19, we find a Jarque-Bera test statistic of 76.47.

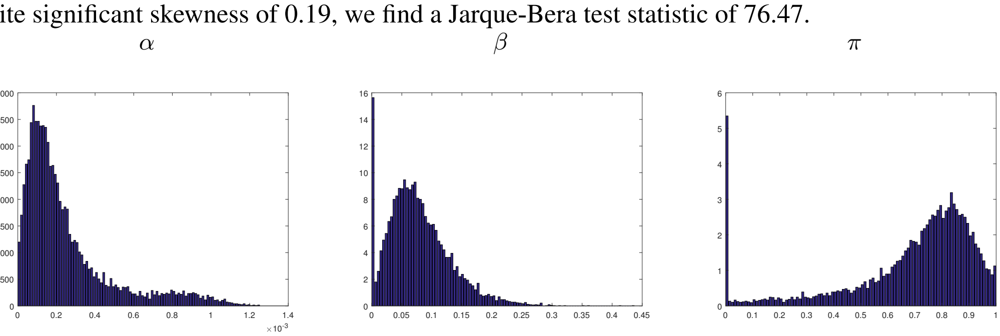

Figure 26: Finite-sample ($T = 453$) densities of the estimators of Standard and Poor's 500 Composite Stock Price Index with α, β, π in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.17 \times 10^{-3}$, $\beta_0 = 0.074$, $\pi_0 = 0.768$ and $dof = 7.19$.

For the Standard and Poor's 500 Composite Stock Price Index we see that the coefficient distributions are far from normal distributed. We see other characterstics in the densities than we saw in the exchange rates coefficients densities. This is probably due to the smaller sample sizes for the stock price index. For α we see a far from normal double peak density. With a second peak at around 0.8×10^{-3}. For β and π we find a large build up of mass at zero followed with a respectively positively and negatively not normal distribution.

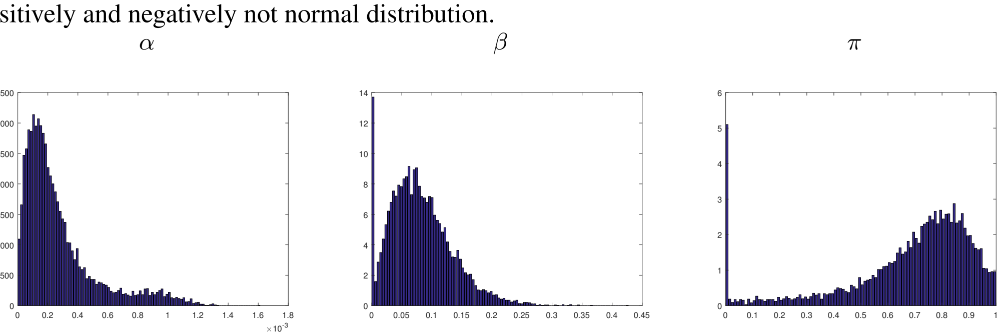

Figure 27: Finite-sample ($T = 453$) densities of the estimators of Industrial Stock Price Index with α, β, π in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.19 \times 10^{-3}$, $\beta_0 = 0.078$, $\pi_0 = 0.756$ and $dof = 7.75$.

For the Industrial Stock price we find similar results as for the Standard and Poor's 500 Composite Stock Price Index, this is as expected as the coefficients values and sample size are similar.

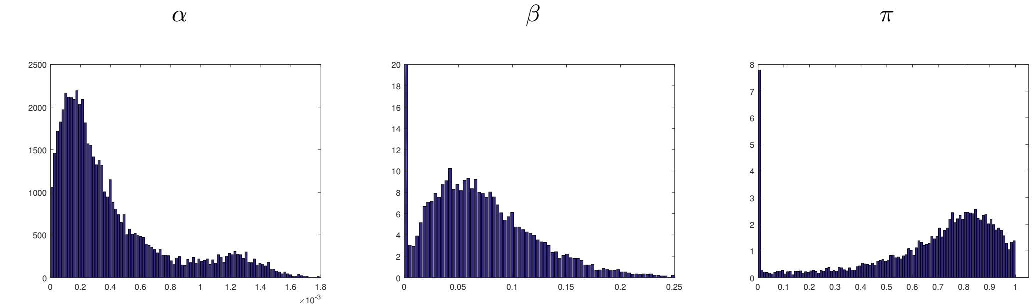

Figure 28: Finite-sample ($T = 453$) densities of the estimates of the Capital Goods Stock Price Index with α, β, π in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.27 \times 10^{-3}$, $\beta_0 = 0.065$, $\pi_0 = 0.751$ and $dof = 7.14$.

For the Capital Goods Stock we find similar results as for the Standard and Poor's 500 Composite Stock Price Index, this is as expected as the coefficients values and sample size are similar.

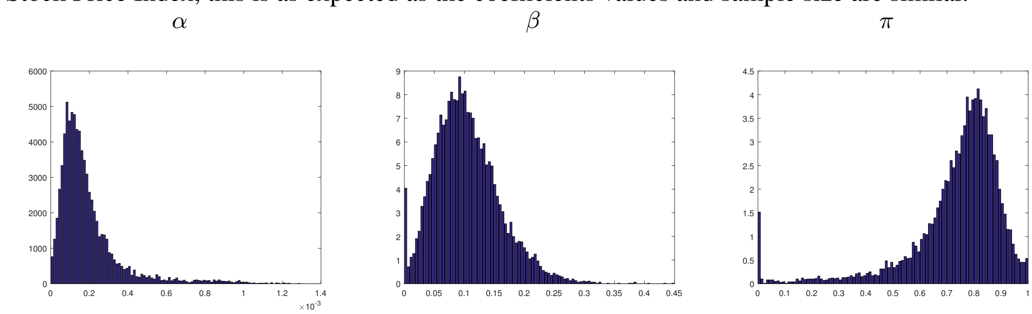

Figure 29: Finite-sample ($T = 453$) densities of the estimates of the Consumer Goods Stock Price Index with α, β, π in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.14 \times 10^{-3}$, $\beta_0 = 0.103$, $\pi_0 = 0.782$ and $dof = 6.21$.

For the Consumer Goods stock we find closer to normal densities, this is as expected with the bigger $\beta_0 = 0.103$. For α we see a positively skewed distribution and for π a negatively skewed distribution. For β we still see the build-up of mass at zero and the positively skewed distribution.

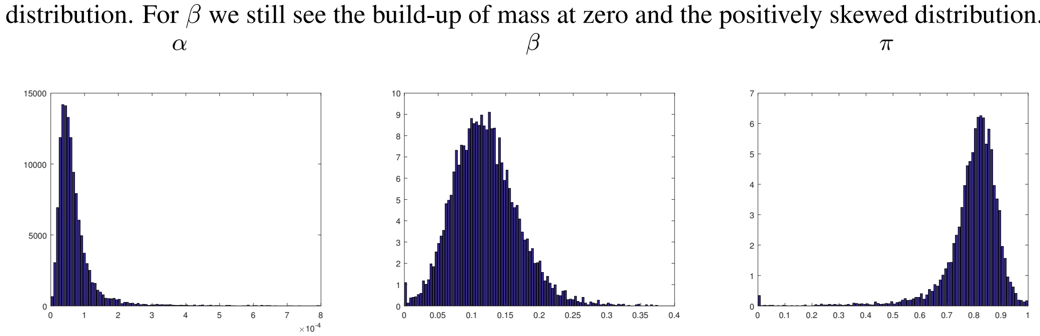

Figure 30: Finite-sample ($T = 453$) densities of the estimates of the Public Utilities Stock Price Index with α, β, π in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.05 \times 10^{-3}$, $\beta_0 = 0.123$, $\pi_0 = 0.820$ and $dof = 7.63$.

For the Public Utilities Stock Price index we finally lose the build-up of mass at zero for the β coefficient. This is explainable as β is quite big, which leads to a better identified model, causing the distribution to be closer to normal. Besides this, we see slightly closer to normal distributions for α and π compared to the Consumer Goods stock.

In short, the densities show that the coefficient estimates are not close to being normally distributed, with exception of the Germany exchange rate. You can also clearly see that a higher sample size results in a closer-to-normal distribution of the coefficients. This was also expected from known literature. As we can see, the coefficients for the exchange rate (Britain and Germany) are more normally distributed than the other stock coefficients. Although the exchange rates have similar or lower values of β than the stocks, the estimation results are more reliable. We can also see that when β increases for Public Utilities and Consumer Goods, the densities become closer to normal. However, for all coefficient densities we reject normality. Thus we can conclude that all of the found coefficients in the Bollerslev paper are not completely reliable. Especially for the Standard and Poor's, Industrial and Capital Goods stocks, the densities are very far from normal and thus it's very likely that the estimates Bollerslev found are not equal or even far from the true values. This means that a lot of the conclusions he made, for example about a significant GARCH effect for the Standard and Poor's data, can not be made.

4.3.2 Bollerslev's t-test statistics

After finding the Bollerslev's coefficient densities we continued our research in order to find t-test statistic distributions. These are t-test statistics in which we test the null hypothesis of a not-significant coefficient against the alternative hypothesis of a coefficient significant different from zero. We tested this for all coefficients in all exchange rates and stock price indices Bollerslev used.

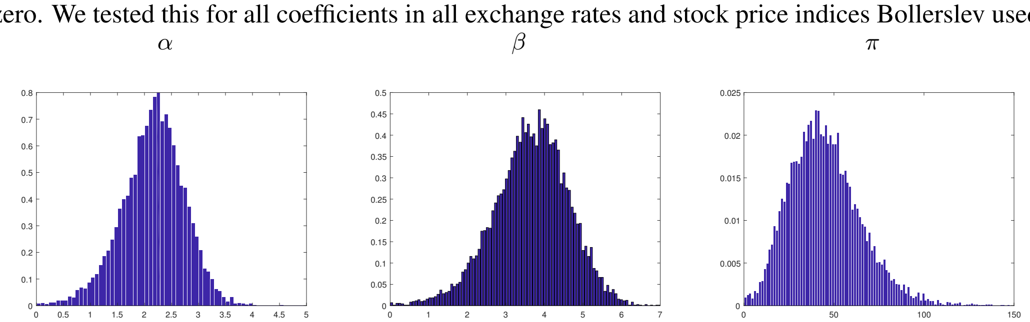

Figure 31: Finite-sample ($T = 1245$) densities of the t-test statistics of the Britain Exchange Rate with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.96 \times 10^{-6}$, $\beta_0 = 0.057$, $\pi_0 = 0.921$ and $dof = 8.13$.

The t-test statistic densities look close to normal for all coefficients of the Britain Exchange rate, we do observe some negative skewness for α and β and some positive skewness for π.

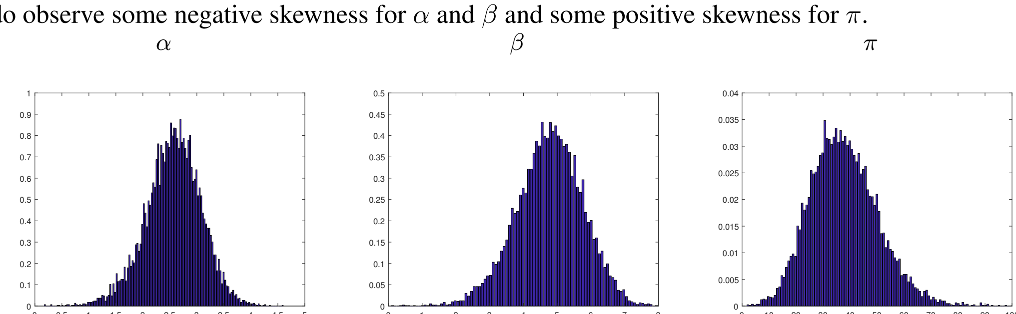

Figure 32: Finite-sample ($T = 1245$) densities of the t-test statistics of the Germany Exchange Rate with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.13 \times 10^{-6}$, $\beta_0 = 0.095$, $\pi_0 = 0.881$ and $dof = 13.89$.

As parameter coefficients and sample size are quite similar for Britain and Germany exchange rate, we observe similar densities for the t-test statistic for the Germany as for the Britain exchange rate.

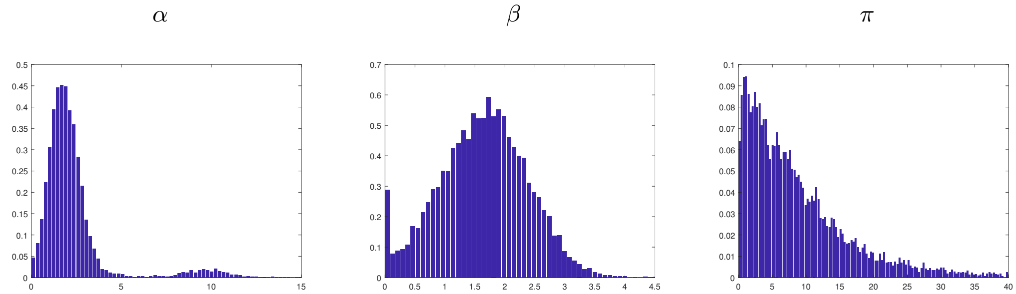

Figure 33: Finite-sample ($T = 453$) densities of the t-test statistics of Standard and Poor's 500 Composite Stock Price Index with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.17 \times 10^{-3}$, $\beta_0 = 0.074$, $\pi_0 = 0.768$ and $dof = 7.19$.

We see a first estimate peak around 2 for the α t-test statistic. Normality is clearly rejected, due to big positive skewness. Also, we see a second peak at around 10. Also for the β coefficient, normality is rejected. This time due to a large peak around zero. The π t-test statistic distribution does not look anywhere close to a normal distribution, with a large peak at around 1 and decreasing mass for larger values.

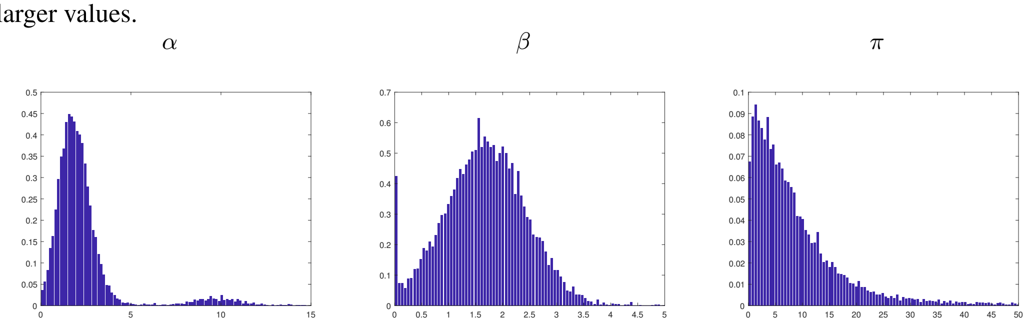

Figure 34: Finite-sample ($T = 453$) densities of t-test statistics of the Industrial Stock Price Index with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.19 \times 10^{-3}$, $\beta_0 = 0.078$, $\pi_0 = 0.756$ and $dof = 7.75$.

Figure 34 shows that α's t-test statistic distribution is positively skewed and its peak is centered around 2, again we find the second peak around 10. The β t-test statistic distribution is centered around 1.7. Also in this case, normality is rejected. Again a large peak around zero stands out. For the π t-test statistic we again find a from 0 on decreasing distribution.

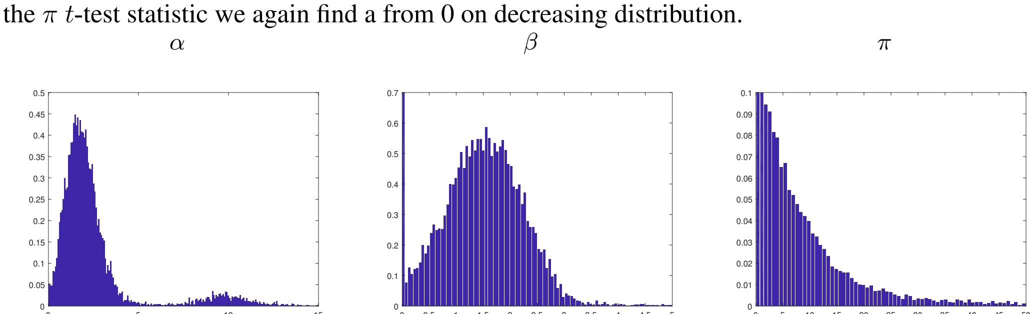

Figure 35: Finite-sample ($T = 453$) densities of the t-test statistics of the Capital Goods Stock Price Index with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.27 \times 10^{-3}$, $\beta_0 = 0.065$, $\pi_0 = 0.751$ and $dof = 7.14$.

For the Capital Goods stock we find quite similar observations as in the Industrial Stock Price, due to similar parameter coefficients and sample size. What stands out for the Capital Goods stock is the extremely high peak at around 0 β.

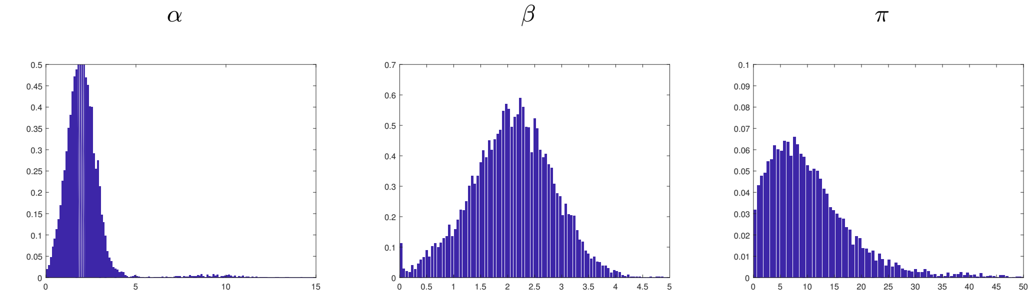

Figure 36: Finite-sample ($T = 453$) densities of the t-test statistic of the Consumer Goods Stock Price Index with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.14 \times 10^{-3}$, $\beta_0 = 0.103$, $\pi_0 = 0.782$ and $dof = 6.21$

The t-test statistic distributions for Consumer Goods stock looks closer to normal, this is explained by the relatively high β_0 coefficients. We see that the second peak for the t-test statistic of α at around 10 has disappeared. We do see some uniformly looking distribution between 5 and 10. For the t-test statistic of the β coefficient we see that the peak at zero has reduced drastically. For the t-test statistic distribution of π we now see a peak at around 5. With a distribution looking closer to normal, while still being very far from normal due to extremely positive skewness.

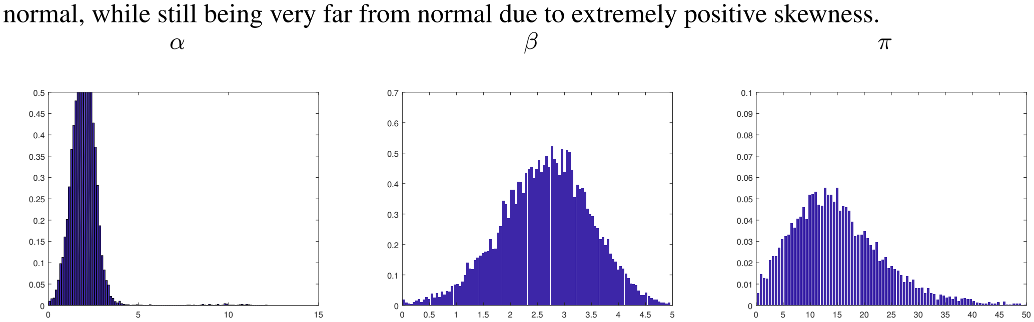

Figure 37: Finite-sample ($T = 453$) densities of the t-test statistics of the Public Utilities Stock Price Index with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.05 \times 10^{-3}$, $\beta_0 = 0.123$, $\pi_0 = 0.820$ and $dof = 7.63$.

For the Public Utilities stock we see quite similar densities as in the Consumer Goods stock. While being again slightly improved due to the again larger β_0. We now see the peak for the t-test statistic of β at around zero completely disappeared and the t-test statistic distribution of π being less positively skewed.

Summary: as for the coefficient distributions we see that our distributions start to look better when increasing sample size or β_0. We again find no normal distributions in any case. We can conclude that the critical values used in the case of a normal distribution are unreliable.

4.4 Critical Values for the t-test statistic

Due to the found non-normal densities of the t-test statistic, we have to construct new critical values. The normal critical values can not anymore be used when checking for significance of the parameters. As the GARCH-parameter is the most vital in the GARCH(p, q) model, we should construct robust critical values for the t-test statistic of this coefficient. As when the GARCH-coefficient would be equal to zero it does not make sense to use a GARCH-model.

4.4.1 Calculating the Critical Value of the t-test statistic of the GARCH-coefficient

To test for significance of GARCH-coefficient π we can calculate the t-test statistic. We test here the null hypothesis $\pi = 0$ against the alternative hypothesis $\pi \neq 0$, we find the t-test statistic using formula 15, here we use for $\pi_0 = 0$ to test for significance.

$$t = \frac{\hat{\pi} - \pi_0}{SE_\pi} \tag{15}$$

Because of the non-normal distribution of the t-test statistics we cannot use the standard normal value of 1.96 as a critical value for the significance of the GARCH-coefficient. Thus it is necessary to find our own critical value in order to determine the significance of the GARCH-coefficient. Due to time shortage, we construct the least favorable critical values, as these are very robust and easy to compute. Because the histograms for sample size $T = 500$ are closest to being normally distributed, we decide to create critical values for this sample size. Only for 3 values $b = (0, 2, 4)$, as in these cases we find no to weak identification. We will find the one sided 5 percent CV of the t-test statistic of π. Because our coefficients cannot be negative our t-test statistics cannot be negative as well, thus there is no need to do a two sided test. We do this by finding the CV for different values of $\pi_0 = (0.2, 0.4, 0.6)$ and choosing the largest. We need to check for different values of π_0 as π_0 is not consistent. We can just plug in one value for α_0 in the DGP, namely 0.2, as α_0 is consistent. We found the following results:

π/b	0	2	4
0.2	35.18	21.94	18.99
0.4	52.54	37.34	40.38
0.6	49.92	80.17	110.77

Table 4: Critical values, $T = 500$, $\alpha = 0.2$

These results follow from the following densities.

$\pi = 0.2$ $\pi = 0.4$ $\pi = 0.6$

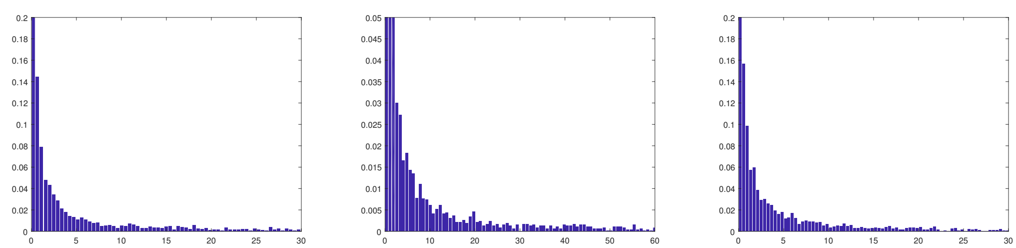

Figure 38: Finite-sample ($T = 500$) densities of the π GARCH coefficient. In the GARCH(1,1) model with 10000 simulations when $\alpha_0 = 0.2$ and $\beta_0 = 0$

When we look at the densities of $\beta_0 = 0$ we clearly see that $\pi = 0.4$ gives us the highest critical value for GARCH coefficient π. For finding the least favorable critical values we take the 95 percent critical value from this data and conclude that for the sample size of 500 and no identification it is wise to use a CV of 52.54 for the t-test statistic when checking for significance of your parameter.

$\pi = 0.2$ $\pi = 0.4$ $\pi = 0.6$

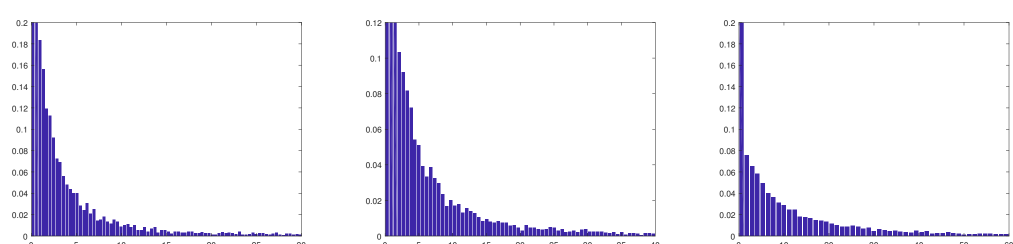

Figure 39: Finite-sample ($T = 500$) densities of the π GARCH coefficient. In the GARCH(1,1) model with 10000 simulations when $\alpha_0 = 0.2$ and $\beta_0 = \frac{2}{\sqrt{500}}$

When we look at the densities of $\beta_0 = \frac{2}{\sqrt{500}}$ we clearly see that $\pi = 0.6$ gives us the highest critical value for GARCH coefficient π. For finding the least favorable critical values we take the 95 percent critical value from this data and conclude that for sample size = 500 and no to weak identification it is wise to use a CV of 80.17 for the t-test statistic when checking for significance of your parameter.

$\pi = 0.2$ $\pi = 0.4$ $\pi = 0.6$

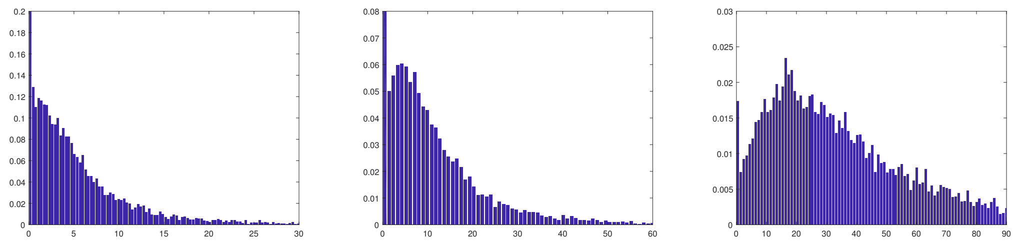

Figure 40: Finite-sample ($T = 500$) densities of the π GARCH coefficient. In the GARCH(1,1) model with 10000 simulations when $\alpha_0 = 0.2$ and $\beta_0 = \frac{4}{\sqrt{500}}$

When we look at the densities of $\beta_0 = \frac{4}{\sqrt{500}}$ we clearly see that $\pi = 0.6$ gives us the highest critical value for GARCH coefficient π. For finding the least favorable critical values we take the 95 percent critical value from this data and conclude that for sample size = 500 and weak identification it is wise to use a CV of 110.77 for the t-test statistic when checking for significance of your parameter.

4.4.2 Improving the robustness of Bollerslev's estimation results

In his paper, Bollerslev assumed that the t-test statistics are normally distributed. As seen in the previous section, this is not always the case. If this is not the case, the 95 percent critical value for determining the significance of coefficients can be drastically affected. Therefore the conclusions Bollerslev made may be unreliable.

We obtained the least favorable critical value for the Britain exchange rate and for the S and P stock index. We did this by generating data in our DGP with the same sample size and constant as respectively the Britain exchange rate and the S and P stock index. Following a GARCH$(1, 1) - t$ process. We took these two as an example as they are the two least identified indexes for the two different sample sizes. This way we will be able to showcase the critical value problem in the best way.

We can't precisely determine in what identification state these financial instruments are, we do know that they are either non- are weak identified. Therefore we used multiple different β_0. Because π is not consistent we used multiple different π_0. We continued determining the 95 percent t-test statistic critical value in all these cases. We concluded taking the biggest 95 percent critical value. We took this value as our least favorable critical value.

See table 5 for the conversion from b to β we used in this example.

T/b	0	1.58	2.01	4
453	0	0.074	-	0.1879
1245	0	-	0.057	0.1134

Table 5: Conversion from b to β_T

Britain Exchange Rate

In table 6 you can find the critical values for the Britain Exchange rate for different values of π and β. As you can clearly see, weaker identification results in higher critical values for the t-test statistic. We find the highest critical value for $b = 0$ and $\pi = 0.921$ for the Britain Exchange rate. For $b = 4$ we used $\pi = 0.88$ instead of $\pi = 0.921$ as the model wouldn't be stationary otherwise.

π/b	0	2.01	4
0.2	110.52	24.09	8.59
0.5	115.67	23.36	9.46
0.88	-	-	65.49
0.921	**149.45**	80.00	-

Table 6: Critical values for the Britain Exchange rate, with $T = 1245$, $\alpha_0 = 0.96 \times 10^{-6}$ and $dof = 8.13$

In figures below you see all the densities of the GARCH-coefficient for the different values of b and π. You can clearly see that the distributions of the t-test statistic of the GARCH-coefficient π become more normal when b increases, resulting in lower critical values.

$\pi_0 = 0.2$ $\pi_0 = 0.5$ $\pi_0 = 0.921$

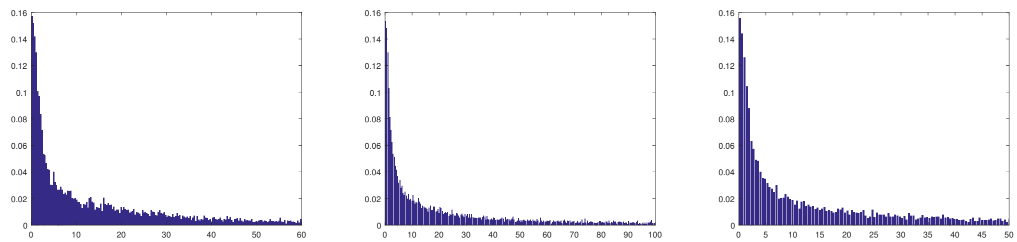

Figure 41: Finite-sample ($T = 1245$) densities of the t-test statistic of the Britain Exchange Rate with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.96 \times 10^{-6}$, $\beta_0 = 0$ and $dof = 8.13$

$\pi_0 = 0.2$ $\pi_0 = 0.5$ $\pi_0 = 0.921$

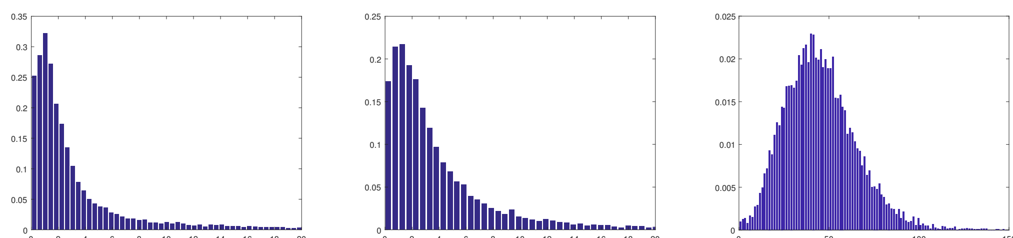

Figure 42: Finite-sample ($T = 1245$) densities of the t-test statistic of the Britain Exchange Rate with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.96 \times 10^{-6}$, $\beta_0 = 0.057$ and $dof = 8.13$

$\pi_0 = 0.2$ $\pi_0 = 0.5$ $\pi_0 = 0.88$

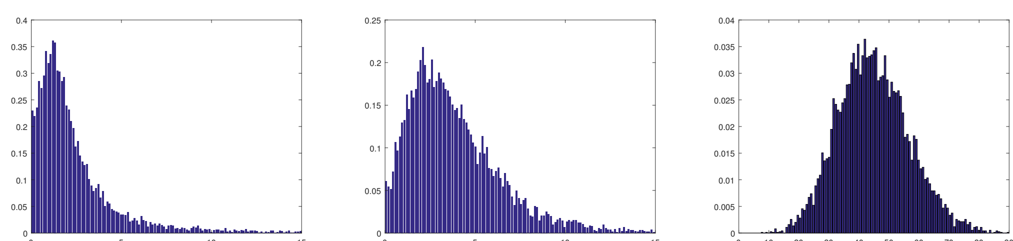

Figure 43: Finite-sample ($T = 1245$) densities of the t-test statistic of the Britain Exchange Rate with α, β,π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.96 \times 10^{-6}$, $\beta_0 = \frac{4}{\sqrt{1245}}$ and $dof = 8.13$

For the Britain exchange rate Bollerslev found a t-test statistic of $0.921/0.023 = 40.04$. This is much bigger than the 1.96 critical value Bollerslev used and thus he concluded that the GARCH effect is significant for the Britain exchange rate. However, when we look at table 6 we see that the more appropriate critical value is 149.45. Because $40.40 < 149.55$ we cannot reject the null hypothesis $\pi_0 = 0$ and thus conclude that the GARCH coefficient is not significant.

Standard and Poor's 500 Composite Price Index

In table 7 you can find the critical values for the Standard and Poor's 500 Composite Price Index for different values of β and π. As you can clearly see, weaker identification results in higher critical values for the t-test statistic. We find the highest critical value for $b = 0$ and $\pi = 0.5$.

π/b	0	1.58	4
0.2	91.94	40.28	10.10
0.5	**93.20**	38.59	10.11
0.768	90.50	35.76	25.31

Table 7: Critical values of the 500 Composite Stock Price Index, with $T = 453$, $\alpha_0 = 0.17 \times 10^{-3}$ and $dof = 7.19$

In the figures below, all densities of the GARCH-coefficient for the different values of b and π are shown. Clearly, the t-test statistic distributions of the GARCH-coefficient π become closer to normal as b increases, resulting in lower critical values.

$\pi_0 = 0.2$ $\pi_0 = 0.5$ $\pi_0 = 0.768$

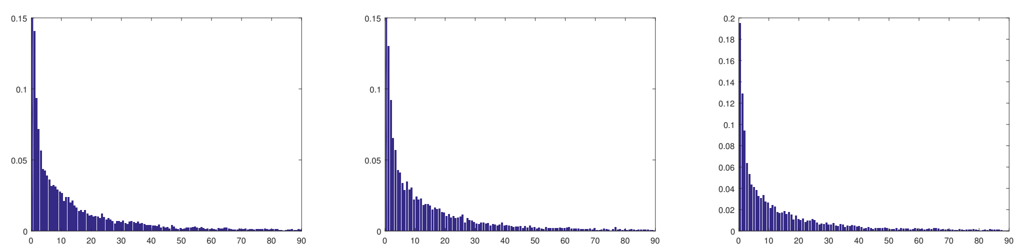

Figure 44: Finite-sample ($T = 453$) densities of the t-test statistic of the 500 Composite Stock Price Index with α, β, π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.17 \times 10^{-3}$, $\beta_0 = 0$ and $dof = 7.19$

$\pi_0 = 0.2$ $\pi_0 = 0.5$ $\pi_0 = 0.768$

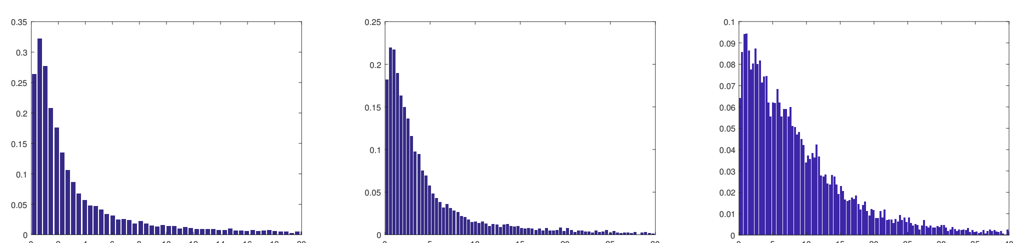

Figure 45: Finite-sample ($T = 453$) densities of the t-test statistic of the 500 Composite Stock Price Index with α, β, π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.17 \times 10^{-3}$, $\beta_0 = 0.074$ and $dof = 7.19$

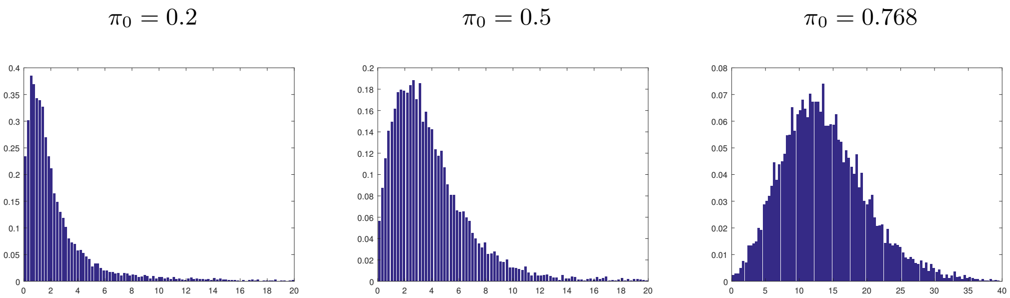

Figure 46: Finite-sample ($T = 453$) densities of the t-test statistic of the 500 Composite Stock Price Index with α, β, π and dof in the GARCH(1,1)-t model with 10000 simulations when $\alpha_0 = 0.17 \times 10^{-3}$, $\beta_0 = \frac{4}{\sqrt{453}}$ and $dof = 7.19$

For the Standard and Poor's 500 Composite Price Index, Bollerslev found a t-test statistic of $0.768/0.148 = 5.19$. This is bigger than the 1.96 critical value Bollerslev used and thus he concluded that the GARCH effect is significant for the Standard and Poor's 500 Composite Index. However, when we look at table 7 we see that the more appropriate critical value is 93.20. Because $5.19 < 93.20$ we cannot reject the null hypothesis $\pi_0 = 0$ and thus conclude that the GARCH coefficient is not significant.

5 Discussion

For we have made certain assumptions when conducting our research about identification issues in a GARCH$(1,1)$ model, there is room for variations and extensions to our research. Firstly, we have conducted our research by choosing a true value for π equal to 0.4. By choosing this value, we noticed that we had to deal with not only expected identification issues but additionally also stationarity conditions. In addition to our research, one could choose a smaller π, for example of 0.2. In this case, β's distribution is not (much) affected by the stationarity condition $\beta + \pi < 1$ because π takes on a much smaller value. This could lead to normality in β's distribution, if identification issues are not interfering. We expect, however, that, in case of $\pi = 0.4$, π's distribution cannot be normal because of the condition $\pi > 0$ and thus its distribution would be truncated from the left-hand side. On the other hand it can also be interesting to take a larger π. In doing so, the GARCH coefficient π will be less affected by stationarity issues and it would be interesting to see how π will be distributed for different categories of identification when stationarity issues don't play a role. The same that holds for π also applies to the intercept coefficient α. Taking different values for α may also change conclusions. Furthermore, it may be interesting to find correct and robust confidence intervals for all coefficients for semi and weak identified GARCH$(1,1)$ models. Also it may be useful to use critical values with less drawbacks for π. As mentioned before, constructing critical values can be really complicated and due to time issues we were not able to find more favorable critical values. By using the least favorable critical values we showed that the critical values used in case of strong identification are incorrect. Moreover, we illustrated that Bollerslev should have used other critical values.

6 Conclusion

The aforementioned results of this paper imply substantial identification problems for both the GARCH$(1,1)$ and the GARCH$(1,1)$-t model. Even for relatively high values of the 'true' underlying parameter β_0, we solely detect semi-strong identification. We are not able to obtain normally distributed parameter coefficients for even the highest possible values of β_0, while equaling π_0 to 0.4. Furthermore, we reject normality for the t-test statistic ubiquitously, however, we notice that both for higher values of β_0 and a bigger sample size, the t-test statistics tend to be much closer to a normal distribution. As the model becomes semi-strong identified, the parameter estimates become sufficiently reliable in these cases.

Compared to Andrews & Cheng (2012) findings, identification issues in the GARCH(1,1) model are to an even greater extend omnipresent than in the ARMA model, especially for small sample sizes. As shown in previous literature, identification issues lessen when sample sizes increases. On the other hand, for small sample sizes identification issues are truly severe. In the GARCH$(1,1)$ model weak identification is predominantly present in data with a small sample size as the result of stationarity restrictions on the coefficients. For instance, the estimated β coefficient is constrained from above by 1 and $\pi_0 + \beta_0$ always should be smaller than 1 in case of stationarity.
In conclusion, our results unfold various problems. First of all, non- or weak identification results in unreliable coefficient estimates. Consequently, the construction of confidence intervals needs to be carefully reconsidered. Secondly, the t-test statistic distribution is also affected because the their statistic densities are not normally distributed. Standard critical values in order to determine the significance of the parameters become unreliable. This is a crucial issue to note because interpretation of regression results could fundamentally differ. For example, one may conclude that a financial model has a significant GARCH-effect where there would be none.
Influential previous papers have insufficiently stressed the non- or weak identification of coefficients in volatility models. Bollerslev's paper serves as a prime example, as his parameter coefficients, and thus confidence intervals, for different stocks and exchange rates are unreliable. We also conclude that various of his significant GARCH-coefficients might not be significant after all. Bollerslev's findings regarding the construction of volatility capturing models have had a vital impact on future research in this field. Concluding that a lot of his results are unreliable is therefore very much shocking and unexpected. In the last couple of decades there was no researcher that has taken the time to check for identification in Bollerslev's model. Here from we expect that a lot of researchers and people working in the financial sector just assume that their model is well-identified without checking. Therefore we emphasize the importance of identification issues when estimating volatility in financial markets by using GARCH models. We hope that this paper raises awareness regarding identification issues and causes researchers to stop taking wrong conclusions.

REFERENCES

D. W. K. Andrews and X. Cheng. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211, 2012.

D. W. K. Andrews and X. Cheng. Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics*, 173(1):36–56, 2013.

D. W. K. Andrews and X. Cheng. Gmm estimation and uniform subvector inference with possible identification failure. *Econometric Theory*, 30(2):287–333, 2014.

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

R. C. Blattberg and N. J. Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business 47*, pp. 244–280, 1974.

T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, pp. 542–547, 1987.

P. Boothe and D. Glassman. The statistical distribution of exchange rates: empirical evidence and economic implications. *Journal of international economics*, 22(3-4):297–319, 1987.

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

B. B. Mandelbrot. The variation of certain speculative prices. *Journal of Business 36*, pp. 394–419, 1963.

Daniel B Nelson. Stationarity and persistence in the garch (1, 1) model. *Econometric theory*, 6(3): 318–334, 1990.