Erasmus University Rotterdam
Erasmus School of Economics

Bachelor Thesis B.Sc.[2] Econometrics and Economics

A Study on the Identification Loss in GARCH-MIDAS Models

Author: Panagiotis Manolopoulos (422117) *
Supervisor: Dr. Andrea A. Naghi
Second Assessor: Dr. Martina D. Zaharieva

July 7, 2019

* The views stated in this thesis are those of the author and not necessarily those of Erasmus School of Economics or Erasmus University Rotterdam.

Abstract

The aim of this research is to study the identification properties of the GARCH-MIDAS model, introduced by Engle, Ghysels, and Sohn (2013). Since there is no formal research outcome that the specific model suffers from identification problems, its small estimated parameter values suggest this suspicion. To verify this notion, we estimate three distinct GARCH-MIDAS models with stock market and macroeconomic data to check the range of the estimated parameter values. It is found that two out these models possibly suffer from identification issues, due to their small t-statistic values. Next, to formally verify their identification issues, a Monte Carlo simulation study is performed according to the methodology of Andrews and Cheng (2012). Through this simulation, it is found that the GARCH-MIDAS model suffers from identification issues, and new critical values should be computed to make valid inferences from the model. Nevertheless, the new robust critical values, namely the Least Favorable and the Type-I critical values, do not solve the identification problems in the two models and create identification issues to the third model as well.

Contents

1 Introduction

Modeling and forecasting volatility with stock market data has been an intertemporal issue among experts in the quantitative finance area for many decades now. However, the strand of literature that incorporates macroeconomic indicators into such models started a few years ago. The idea of creating such models comes from the links between stock market volatility and business cycle fluctuations, but also with real and nominal macroeconomic volatility among other macroeconomic variables (Officer (1973), Schwert (1989)). Precise volatility forecasts are vital for practitioners in numerous fields, like asset management and risk management, as such estimates lead to important decision-making processes. Having the means to construct accurate volatility forecasts can give an edge to financial institutions, as it allows them to change their investment positions in short notice.

Volatility modeling became widespread with the seminal research on Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) models by Engle (1982) and Bollerslev (1986), respectively. These two models are known for modeling time variation in volatility while addressing the volatility clustering patterns. However, GARCH models cannot incorporate data with different frequencies, that is low-frequency macroeconomic data and high-frequency stock market data. The solution to this issue is given by the Mixed Data Sampling (MIDAS) model introduced by Ghysels, Santa-Clara, and Valkanov (2004). In this model, a lag polynomial term basically "maps" the high-frequency data with the low-frequency data using a weighting function (Ghysels, Sinko, & Valkanov, 2007). Since the introduction of the MIDAS model, there have been many extensions of it; the one considered on this report is the GARCH-MIDAS model firstly introduced by Engle et al. (2013).

The GARCH-MIDAS framework models volatility with two components, first, the short-term component which is a GARCH(1,1) process, and second, the long-term component which is the MIDAS model. Even though the GARCH-MIDAS model has been extensively used in the literature, there is a suspicion that it might suffer from identification issues. Such issues arise due to the non-linearity of the MIDAS model, and the small (in magnitude) parameter estimates that it generates. If this is the case, there is a logical reason to cast doubts regarding the results and relations presented in Engle et al. (2013). Taking into account the aforementioned introduction, in this report, we address the following research question:

To what extent do GARCH-MIDAS models suffer from identification issues, and can we get reliable results if we account for such issues?

The first step to answer the aforementioned research question is to estimate three different GARCH-MIDAS models with stock market and macroeconomic data, as in Engle et al. (2013). Once the models are estimated, the relevant MIDAS parameters are investigated whether they possibly demonstrate from identification issues. Next, a Monte Carlo simulation study is employed to formally conclude whether the models suffer from identification problems by examining the parameter distribution densities with the methodology of Andrews and Cheng (2012). Finally, the last step to answer the research question of this report is to calculate the robust critical values and review whether such critical values improve the inference of the models estimated in the first step.

The estimation of the three different GARCH-MIDAS models showed that two out of three models might suffer from identification issues. In particular, the fixed-timed span model with the realized volatility and the fixed-time span model with the industrial production growth rate have insignificant MIDAS parameters, resulting in possible identification loss of the model. Next, the Monte Carlo simulation study showed that the GARCH-MIDAS model suffers from identification loss when the MIDAS parameter θ takes values between 0 and 0.1 in absolute terms. For that reason, new robust critical values were calculated in the hopes of making better inferences from the GARCH-MIDAS model. However, these critical values did not only solve the identification issues in the two aforementioned models but also turned the previously identified model to unidentified as well.

The structure of this report is as follows: Section 2 presents an overview of the existing literature on the GARCH-MIDAS model, its ancestor, and the identification problems of the specific model. Section 3 presents the data and key descriptive statistics, whereas Section 4 outlines the methodology of this report. The results are thoroughly explained in Section 5, while Section 6 describes the relevant conclusions and suggestions for future research.

2 Literature Review

The specific section is devoted to the discussion and review of the existing literature. Section 2.1 introduces the papers that research the MIDAS model, whereas Section 2.2 presents the GARCH-MIDAS model and its extensions. Finally, the relatively new research on the identification issues of the GARCH-MIDAS model is outlined in Section 2.3.

2.1 Introduction to MIDAS Models

The GARCH-MIDAS model has a long-term volatility component that is modeled according to the MIDAS specification and filtering procedure, which was introduced by Ghysels et al. (2004). The reasoning behind such a model creation is to combine data with different frequencies or to account for data availability limitations. An example of such an application is the Value-at-Risk (VaR) modeling for getting the forecasted future losses of a stock portfolio. It is common to take 10 days for the VaR variable, but the returns series can be of a daily, or even tick-by-tick, frequency. The "mapping" between the different horizon variables is done by an (in)finite lag polynomial, which in most cases is the normalized exponential Almon or the Beta function. The MIDAS model produces parameter estimates that are more efficient in comparison to the ordinary model cases where there is an aggregation of the least frequent sampling series (Ghysels et al., 2004).

The MIDAS model has been applied and extended many times by researchers with several macroeconomic indicators. A notable case that combines both cases is the paper written by Clements and Galvão (2008). The authors extended the MIDAS model by incorporating an autoregressive term and used this extended model to forecast the US output growth more efficiently compared to more standards methods. According to their results, the extra autoregressive term added explanatory value and improved the forecasting power of the model. On the other hand, Guérin and Marcellino (2013) differentiated themselves by adjusting the MIDAS model to incorporate different regimes; in particular, they allowed parameters to change according to the regime that they are. The Markov-Switching MIDAS model that they introduced performs well in contrary to the standard MIDAS model, a conclusion drawn from empirical applications and simulations. Furthermore, Armesto, Engemann, and Owyang (2010) selected a variety of macroeconomic variables, i.e. monthly employment growth and quarterly GDP growth among others, and forecast them with either averaging the higher-frequency data or the MIDAS model. The authors showed that averaging the data yields results that are comparable to the results obtained with the MIDAS model; however, the latter ones are preferred due to the beneficial properties of forecasting.

2.2 Overview of GARCH-MIDAS Models

Engle and Rangel (2008) introduced a nonparametric approach to model equity volatility, namely the spline-GARCH model. In this framework, volatility is modeled by linking the dynamic properties of a time series together with macroeconomic information. This model combines stock market data of higher-frequency with macroeconomic indicators of lower-frequency. To be more specific, a unit GARCH model is used for the high-frequency component, whereas an exponential quadratic spline is used for the low-frequency component. An appealing property of the spline-GARCH model is that it relaxes the assumption of mean reversion, which is in contrast with most stochastic volatility models (Engle & Rangel, 2008). In the empirical part of their paper, Engle and Rangel found that short-term interest rates and gross domestic product are the primary cause for a fluctuating low-frequency market volatility. A notable observation is that the higher the inflation volatility, the higher the market volatility. However, this result might not be robust due to the sensitivity of including the case of Argentina in the data set.

A few years later, Engle et al. (2013) presented the GARCH-MIDAS model, which is an enhanced version of the spline-GARCH model. In particular, volatility is again decomposed into two distinct components. However, an adjustment in the long-term component is applied. Instead of using an exponential quadratic spline, the authors incorporated the MIDAS model to account for data with different frequencies in a more elaborated way. In a similar manner with the spline-GARCH, the short-term component is the unit GARCH model. The authors state that one advantage of the introduced model is that it avoids the two-step procedure used by Schwert (1989). Regarding the empirical part of the paper, the authors used a historical time series of 120 years and concluded to promising results regarding the longer horizon forecasting. More specifically, the macroeconomic forecasts are approximately the same as the stock market forecasts when measured over the quarter horizon, although they do prevail in the biannual horizon. With respect to the shorter horizon, the authors found that the IP growth rate and the PPI inflation rate explain 30% (on average) of the one-day-ahead volatility in almost all the considered subsamples. These results show that macroeconomic indicators are important for both short-term and long-term horizons.

Since its introduction, the GARCH-MIDAS model has often been applied by researchers, the most recent one being the one by Conrad, Custovic, and Ghysels (2018). The authors applied the GARCH-MIDAS framework to examine potential drivers of the short- and long-term components of the Bitcoin volatility. An interesting, yet irregular, result of this research is that the Bitcoin's long-term volatility is negatively affected by the S&P 500 realized volatility. Nevertheless, this finding may not hold in reality due to the relatively small sample size, which is only four years. Next, Fang, Chen, Yu, and Qian (2018) use the GARCH-MIDAS model to investigate if the Global Economic Policy Uncertainty (GEPU) Index can accurately forecast the gold futures volatility components. The authors found that the GEPU Index has a significantly positive effect on the gold future monthly volatility forecasts. Furthermore, when comparing the out-of-sample performance of the GARCH-MIDAS with the one of the simple GARCH, they found that the former one performs better when including the GEPU and realized volatility. This outcome verifies the fact that including low-frequency macroeconomic indicators enhances the model to better explain the short- and long-term volatility estimates.

Although the GARCH-MIDAS model links the lower frequency with the higher frequency data effectively, Engle et al. (2013) state that it possibly suffers from structural breaks, and for this reason, they consider various subsamples. This shortcoming inspired Pan, Wang, Wu, and Yin (2017) to develop the Regime-Switching GARCH-MIDAS (RS-GARCH-MIDAS) model that allows the short-term component of the volatility to change among two distinct regimes. The two regimes are modeled in such a way that they account for the dynamics of the oil volatility. The in-sample estimation of the model showed that the level of the macroeconomic variables has a significantly adverse effect on the volatility of oil. On the other hand, the out-of-sample model estimation showed that adding together different macroeconomic variables can significantly increase the predictive performance of the RS-GARCH-MIDAS model.

2.3 Identification Issues in GARCH-MIDAS Models

A possible shortcoming of the GARCH-MIDAS model is that it may suffer from identification issues for some specific elements of its parameter space. Such a problem may arise when the data cannot efficiently fit the model, and as a result, "noisy" parameter estimates are consistently produced. Researchers categorize the identification problem into three types: (i) the strong identification, (ii) the semi-strong identification, (iii) weakly identification, and (iv) non-identification (Andrews & Cheng, 2012). Despite the lack of literature on the identification issues of the GARCH-MIDAS model, Andrews and Cheng (2012) developed a procedure that accounts for identification problems in a broader class of models. These models are estimated using the general method of moments, maximum likelihood, and least squares among others. Their objective is to present severely distorted estimates, confidence intervals, and statistical tests under the non- or weakly identification category using a Monte Carlo simulation framework. By applying their methodology to a unit Autoregressive Moving Average (ARMA) model, the authors ascertained a distorted (bi-modal) distribution for most of the considered parameter estimates. For this reason, they incorporate robust critical values (Type-I, Type-II, and Least Favorable) to correct the statistical inference of the considered model. After the specific publication, a strand of literature emerged on this topic. Among the most important follow-up studies are Andrews and Cheng (2013, 2014) and Cheng (2015), which produce comparable results with Andrews and Cheng (2012).

The above-discussed papers investigate and present solutions to identification issues regarding the in-sample estimation. However, loss of identification in a model may have serious effects for the out-of-sample analysis. By using an analogous approach as Andrews and Cheng (2012), Naghi (2018) showed that models with non- or weak identification type produce aggrandized forecast errors. This means that the forecast errors are larger when a model suffers from identification loss compared to forecast errors under a model with the strong identification type. This result stems from using a Smooth Transition Autoregressive (STAR) model that gives uniform and bi-modal distributions for the parameter estimates. These nonstandard distributions lead to void inferences regarding the predictive ability of the considered model. The author suggested two methods to correct the out-of-sample inference, and these are the Type-I and Least Favorable robust critical values.

3 Data

The first part of this report is a case study involving stock market and macroeconomic data. For this reason, a macroeconomic indicator is retrieved from the Federal Reserve Economic Data (FRED ®) of the Federal Reserve Bank of St. Louis[1], a publicly accessible database. The variable is the Industrial Production (IP), which is transformed into the natural IP growth rate. The IP growth rate has a time range from January 1926 to December 2010 on a quarterly basis, yielding 340 observations in total. Additionally, the stock returns of the value-weighted series are obtained from the database of the Center for Research in Security Prices (CRSP)[2]. The time range for this variable is, as before, from January 1926 to December 2010, but on a daily basis. This results in a returns series with 22528 observations.

Since the two considered series include data of almost 85 years, it is natural to wonder whether there are multiple structural breaks. To account for this issue, Engle et al. (2013) divide the entire sample into five subsample periods, inspired by the seminal work of Schwert (1989). The reasoning behind the specific division of subsamples, according to the authors, is that they address the effects of several (financial) events that occurred the previous century like the World War II and the Great Moderation era among others. In this report, we divide the sample in a similar manner into three

[1]FRED website: `https://fred.stlouisfed.org`

[2]CRSP website: `http://www.crsp.com/products/research-products/crsp-historical-indexes`

subsamples. Note that the first subsample in this report starts from 1926 and not from 1920 as in Engle et al. (2013) due to data limitations.

The descriptive statistics for both variables in the relevant subsamples and the entire sample period are presented in Table 1. Regarding the daily returns, one can observe that the third subperiod ($1985-2010$) has the highest mean returns. However, this fact implies a high standard deviation among other time periods. The lowest mean returns are observed during the first subperiod ($1926-1952$), which in addition demonstrates the highest standard deviation as well. This irregular fact can be partly explained by the effects of the Great Depression, or it may be noise in the data. As for the IP growth rate, its highest average value is observed in the first subsample, whereas the lowest values is detected in the third subsample. Another notable observation is that the standard deviation of the IP growth rate ranges from 0.0128 to 0.0629, which is a rather wide interval. A final important observation is that the returns, in all subsamples and the whole sample, do not seem to follow a normal distribution as their skewness and kurtosis values are far from the normal values of 0 and 3, respectively.

Table 1: The descriptive statistics for the daily returns and the quarterly macroeconomic variable.

Sample	Variable	Mean	St. Deviation	Skewness	Kurtosis
1926-1952	Daily returns	0.00018	0.0142	0.3151	14.8755
	IP growth rate	0.01200	0.0629	-0.1243	4.2872
1953-1984	Daily returns	0.00026	0.0078	0.0692	7.3322
	IP growth rate	0.00860	0.0227	-0.5987	3.7843
1985-2010	Daily returns	0.00038	0.0117	-0.7834	23.0842
	IP growth rate	0.00520	0.0128	-2.0102	8.9921
1926-2010	Daily returns	0.00027	0.0115	-0.0244	19.8675
	IP growth rate	0.00860	0.0387	-0.0264	9.5554

4 Methodology

The first component of this report is devoted to a case study where stock market and macroeconomic data are used to estimate the GARCH-MIDAS models in accordance to Engle et al. (2013). The specifications of these models, for both the rolling window and fixed-term span cases, are presented in Section 4.1. The second component of this report is devoted to Monte Carlo simulations of the GARCH-MIDAS model to investigate the existence of identification issues and compute robust critical values. The steps of the Monte Carlo simulation are presented in detail in Section 4.2, whereas the methodology for computing the robust critical values is described in Section 4.3.

4.1 The GARCH-MIDAS Model Specifications

The three GARCH-MIDAS models that are examined in this report have been selected from the paper published by Engle et al. (2013). More specifically, the considered GARCH-MIDAS models have a:

1. Rolling window realized volatility and two years of MIDAS lags,
2. Fixed-term span realized volatility and four years of MIDAS lags,
3. Fixed-term span IP growth rate and four years of MIDAS lags.

The specification of the GARCH-MIDAS model considered in this report is equivalent to the one introduced by Engle et al. (2013). This implies that the unexpected returns can be expressed as follows:

$$r_{i,t} = \mu + \sqrt{\tau_t \times g_{i,t}}\varepsilon_{i,t} \qquad \forall i = 1, \dots N_t, \tag{1}$$

where the return of day i on quarter t is notated as $r_{i,t}$, the mean of the entire return series is μ, the short-term volatility component is $g_{i,t}$, the long-term component of volatility is τ_t, the error-disturbance term given the information set up until day $(i-1)$ from quarter t is $\varepsilon_{i,t} \mid I_{i-1,t} \sim N(0,1)$, and the last day of each quarter is denoted as N_t. It should be noted that the index t is treated only at a fixed frequency of a quarter. Engle et al. (2013) also consider a monthly and a biannually fixed period, but these cases are not investigated here. Next, the component $g_{i,t}$ is modeled with the volatility dynamics of a GARCH(1,1) model and is written as:

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t}, \tag{2}$$

where it holds that $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$ to guarantee that $g_{i,t}$ remains positive. This short-term volatility component varies on a daily basis, and its specification does not change for the three examined models. On the other hand, the component τ_t measures volatility in the long-term with the quarterly realized volatility as a proxy, but it has a different specification for each of the three considered models. For the model with the fixed-time span realized volatility, τ_t is calculated with the following MIDAS regression model:

$$\tau_t = m + \theta \sum_{k=1}^{K} \phi_k(\omega) RV_{t-k} \quad \text{with} \quad RV_t = \sum_{i=1}^{N_t} r_{i,t}^2, \tag{3}$$

where m and θ are the intercept and the slope of the MIDAS regression, respectively, and K is the number of quarterly MIDAS lags included in the model. As a last step to complete the GARCH-MIDAS model specification for the fixed-time span realized volatility version, the weighting function $\phi_k(\omega_1, \omega_2)$ is defined with a slightly changed Beta Lag polynomial as in Ghysels et al. (2007). In particular, Engle et al. (2013) state that the optimal weighting function is monotonically declining when ω_1 is equal to 1, and by setting $\omega_2 = \omega$, the weighting function becomes:

$$\phi_k(\omega_1, \omega_2) = \frac{(k/K)^{\omega_1 - 1}(1 - k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K}(j/K)^{\omega_1 - 1}(1 - j/K)^{\omega_2 - 1}} \Leftrightarrow \phi_k(\omega) = \frac{(1 - k/K)^{\omega - 1}}{\sum_{j=1}^{K}(1 - j/K)^{\omega - 1}}, \tag{4}$$

where the weights are positive, their summation is equal to 1, and the restriction $\omega > 1$ need to hold to allow recent observations to take higher weights. An advantage of using this weighting scheme is that it can fit multiple lag structures in a very manageable way (Engle et al., 2013). Equations (1) to (4) form the GARCH-MIDAS model for the fixed-time span realized volatility case with a parameter space $\Theta^{(ft)} = \{\mu, \alpha, \beta, m, \theta, \omega\}$.

However, the restriction that τ_t is fixed for quarter t can be relaxed by making both short- and long-term volatility components change on a daily frequency. This implies that the rolling window case changes slightly the model specification above and the long-term component in Equation (3) is now defined as:

$$\tau_i^{(rw)} = m^{(rw)} + \theta^{(rw)} \sum_{k=1}^{K} \phi_k(\omega) RV_{t-k}^{(rw)} \quad \text{with} \quad RV_i^{(rw)} = \sum_{j=1}^{N'} r_{i-j}^2, \tag{5}$$

where all parameters have the same explanation as in Equation 3, N' is equal to 65 (quarterly frequency), and K is the number of daily MIDAS lags. Therefore, by dropping index t from Equations (1) and (2) and together with Equations (4) and (5) the GARCH-MIDAS model for the rolling window case with parameter space $\Theta^{(rw)} = \{\mu, \alpha, \beta, m^{(rw)}, \theta^{(rw)}, \omega\}$ is formed.

Apart from the realized volatility, a macroeconomic variable is also considered in the analysis of this report. The inclusion of the IP growth rate in the GARCH-MIDAS model alters the formula of the long-term component τ_t in Equation (3). In particular, τ_t has a log version now and Equation (4) has two weights instead of one, that is:

$$\log \tau_t = m^{(mv)} + \theta^{(mv)} \sum_{k=1}^{K} \phi_k(\omega_1, \omega_2) X_{t-k}, \tag{6}$$

where X_{t-k} is the level of the macroeconomic variable, parameters m and θ have the same explanations as in Equation (3), and the Beta Lag weighting function $\phi_k(\omega_1, \omega_2)$ is now defined as:

$$\phi_k(\omega_1, \omega_2) = \frac{(k/K)^{\omega_1 - 1}(1 - k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K}(j/K)^{\omega_1 - 1}(1 - j/K)^{\omega_2 - 1}}. \tag{7}$$

Consequently, the fixed-time span GARCH-MIDAS model with the macroeconomic variable is defined with Equations (1), (2), (6), and (7) and has a parameter space $\Theta^{(mv)} = \{\mu, \alpha, \beta, m^{(mv)}, \theta^{(mv)}, \omega_1, \omega_2\}$. Finally, the estimation of the GARCH-MIDAS models above is done with the Maximum Likelihood Estimate (MLE) approach, and the relevant Log-Likelihood Function (LLF) is:

$$\text{LLF} = -\frac{1}{2}\sum_{t=1}^{T}\left[\log(2\pi) + \log g_t(\Phi)\tau_t(\Phi) + \frac{(r_t-\mu)^2}{g_t(\Phi)\tau_t(\Phi)}\right]. \tag{8}$$

4.2 Monte Carlo Simulation of the GARCH-MIDAS model

After the estimation of the three models, a close examination of the parameter estimates and the t-statistics is performed to investigate for possible identification issues of the models under examination. To establish a clear opinion whether the GARCH-MIDAS models suffer from identification issues, we perform a Monte Carlo simulation study with monthly frequencies and 2,000 iterations for each case. The built-up procedure of the simulations is described in the paragraphs below.

The first step of the Monte Carlo simulation is to initialize the long-term volatility component, τ_t, with a unit Autoregressive (AR) model for the first twelve months, which correspond to one year of MIDAS monthly lags. The initialization for τ_t is the following:

$$\log \tau_t = \phi_0 + \phi_1 \tau_{t-1} + u_t, \tag{9}$$

where $\phi_0 = 0,\ \phi_1 = 0.6$ and $u_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. The log version of τ_t is taken to ensure that it remains positive. Following the initialization of the long-term volatility, the GARCH process, $g_{i,t}$, and the returns series, $r_{i,t}$, are initialized with Equations (2) and (1), respectively. Subsequently, the long-term component τ_t is re-estimated with Equation (3), which implies that $g_{i,t}$, and $r_{i,t}$ are also re-calculated with the same formulas as beforehand. This computational procedure of the Monte Carlo simulation can be summarized with the following Data Generating Process (DGP):

$$r_{i,t} = \sqrt{\tau_t \times g_{i,t}}\varepsilon_{i,t} \qquad \forall i = 1, \dots N_t \tag{10}$$

$$g_{i,t} = 0.05 + 0.05\frac{r^2_{i-1,t}}{\tau_t} + 0.90 g_{i-1,t} \tag{11}$$

$$\tau_t = 0.5 + \theta\sum_{k=1}^{K}\phi_k(\omega)RV_{t-k} \tag{12}$$

$$RV_t = \sum_{i=1}^{N^*} r^2_{i,t} \tag{13}$$

where $\varepsilon_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,1), K = 12, N^* = 22$, and the adjusted Beta Lag polynomial $\phi_k(\omega)$ is defined as in Equation (4).

The Monte Carlo simulation framework described above is performed for different values of $h = (\mathrm{b}, \omega) \in \mathcal{H}$, where $\mathrm{b} = \sqrt{n}\theta$ and $\mathcal{H}$ is the model sample space. More specifically, the four types of identification discussed in Section 2.3 are considered in this simulation study, and for that reason, the different values of b are chosen as in Andrews and Cheng (2012). This fact implies that θ is chosen in such a way that $\mathrm{b} = (0, 0.1, 0.5, 1)$. The selection of parameter θ is $(0, 0.07, 0.19, 0.3)$ and $\omega = (1, 1.5, 2, 3)$. The particular choice of the parameter estimates stems from the decision to check the identification issue of the GARCH-MIDAS model for different values of weights ω.

In each iteration of the Monte Carlo simulation, the parameter estimates are obtained by applying the MLE method to Equation (8). Next, the simulated t-statistics for parameter θ are calculated with the usage of the simulated and true parameter estimates in the following way:

$$T_n(\theta) = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})}, \tag{14}$$

where the standard error of $\hat{\theta}$ is obtained from the Hessian matrix of the LLF. The specific t-statistics are used to construct the two robust critical values described in the following section. Additionally, the obtained simulated values of θ are plotted on histograms for the examination of identification deficiencies. According to Andrews and Cheng (2012), the histograms for the non-identified and the weakly identification cases should look bi-modal and for the semi- and strong identification category normal. The inspection of the normality of the different distribution densities of θ is done with the Jarque-Bera normality test.

4.3 Solutions for Identification Loss

The close examination of the obtained distributions densities from the Monte Carlo simulations is performed to determine whether identification issues exist in the considered GARCH-MIDAS models. If this is the case, then the methodology of Andrews and Cheng (2012) is employed for the calculation of robust critical values.

First, the Least Favorable (LF) robust critical value is the most straightforward measure considered, as it is large enough regardless of the identification type of the model. Even though the LF critical values are asymptotically correct, they have a wide range, and are not so informative, especially when the model is under the strong identification type. The LF robust critical values are calculated in the following way:

$$c^{LF}_{1-\alpha,T} = \max\left\{\sup_{h\in\mathcal{H}} c_{1-\alpha}(h), c_{1-\alpha}(\infty)\right\}, \tag{15}$$

where $c_{1-\alpha}(\infty)$ is the standard normal critical value, $c_{1-\alpha}(h)$ is the critical value calculated with Equation (14), and $\alpha = 0.05$.

The second robust critical value considered is the Type-I critical value, which is a natural extension of the LF critical value. It basically uses a data-dependent Identification Category Selection (ICS) statistic to determine the identification category of the considered model. The ICS statistic, A_T, is calculated and compared to a certain threshold κ_T and from the outcome of this comparison, the Type-I robust critical value is computed in the following way:

$$\tilde{c}_{1-\alpha,T} = \begin{cases} c^{LF}_{1-\alpha}, & \text{if } A_T \leq \kappa_T, \\ c_{1-\alpha}(\infty), & \text{if } A_T > \kappa_T, \end{cases} \tag{16}$$

where $c^{LF}_{1-\alpha}$ is the LF robust critical value, $c_{1-\alpha}(\infty)$ the standard normal critical value, and $\alpha = 0.05$. The data-dependent constant, κ_T, is equal to $(\ln T)^{1/2}$ as indicated in Engle et al. (2013) and Andrews and Cheng (2012). The constant is defined as:

$$A_T = (T\hat{\theta}'_T\hat{\Sigma}^{-1}_{\theta,T}\hat{\theta}_T/d_\theta)^{1/2}, \tag{17}$$

where d_θ measures the dimensionality of parameter θ and $\hat{\Sigma}^{-1}_{\theta,T}$ is the covariance matrix of θ.

When $A_T \leq \kappa_T$ holds, the model is recognized to have the weak identification type, and the procedure assigns the LF robust critical value as Type-1. However, when $A_T > \kappa_T$ holds, then the model is considered to have a semi-strong or strong identification type, and the ICS procedure assigns the critical value of the standard normal distribution as Type-I.

5 Results

This section is dedicated to the explanation of the obtained results. The results from the estimation of the three GARCH-MIDAS models are thoroughly presented in Section 5.1. Next, Section 5.2 is devoted to the results of the verification of the identification issues in the GARCH-MIDAS models. Finally, the computation results of the robust critical values are outlined in Section 5.3.

5.1 Estimation Results of the GARCH-MIDAS Models

The first examined model is based on the rolling window realized volatility, including two years of daily MIDAS lags. The obtained parameter estimates for each of the three subperiods and the entire sample are reported in Table 2. The numbers in parentheses correspond to robust t-statistics, that are calculated from Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors. The specific choice of standard errors is selected because the assumption of consistent sample errors is frequently violated in time series analysis. Additionally, the last two columns refer to the Log-Likelihood Function (LLF) value and the Bayesian Information Criterion (BIC) of the examined model, which provide a perspective regarding on which period is preferred over another one.

Table 2: Parameter estimates of the GARCH-MIDAS model with rolling window realized volatility.

Sample	μ	α	β	$\theta^{(rw)}$	ω	$m^{(rw)}$	LLF	BIC
1926-1952	0.00056	0.10805	0.85550	0.10664	2.1901	0.00618	22896.07	-6.21
	(6.04)	(15.51)	(87.77)	(21.17)	(23.05)	(10.84)	—	—
1953-1984	0.00041	0.09580	0.88181	0.10741	3.4470	0.00483	26686.65	-7.07
	(3.79)	(0.74)	(6.69)	(7.19)	(0.89)	(5.49)	—	—
1985-2010	0.00063	0.09960	0.85581	0.09547	9.7246	0.00663	19566.32	-6.45
	(1.77)	(4.79)	(12.56)	(4.36)	(6.48)	(10.08)	—	—
1926-2010	0.00051	0.09725	0.87202	0.10572	3.4618	0.00545	72681.91	-6.62
	(6.81)	(18.97)	(150.82)	(9.99)	(44.01)	(14.57)	—	—

A quick look in Table 2 reveals that almost all parameters are significant, except for a few cases in the second and third subperiods. The fact that $\theta^{(rw)}$ is significant and positive indicates that the model behaves well and that the information contained in the realized volatility MIDAS lags help to explain the long-term volatility component τ_i. Moreover, the positive estimates of $\theta^{(rw)}$ reveal that higher realized volatility implies a higher volatility level in the long-term horizon. In addition, the identification issue in the model seems to be absent since the values of $\theta^{(rw)}$ are approximately equal to the one of the weak identification category in Andrews and Cheng (2012). This fact implies that the inferences made above regarding the long-term volatility component do hold in reality. An extra indication of the correct model identification is the monotonically decreasing weighting function over the number of daily MIDAS lags, as shown in Figure 1.

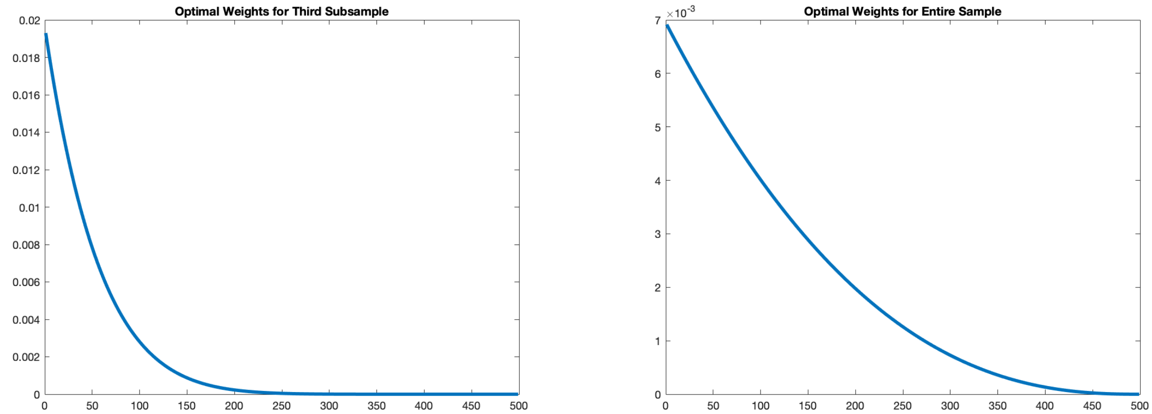


Figure 1: Optimal weights for the rolling window realized volatility model.

Another important observation is related to the values of parameters α and β. Although in a typical GARCH model, the summation of α and β is equal to 1, this summation is noticeably less than this value in this report. In particular, the summation of the specific parameter estimates ranges from 0.9554 to 0.9776, across all three subsamples and the entire sample as well. This irregularity was also observed in Engle and Rangel (2008) and Engle et al. (2013) without any formal explanation of why this phenomenon appears. Nevertheless, this finding implies a lower persistence of the short-term volatility component g_i. Furthermore, it should be mentioned that the values of the t-statistics significantly change when adjustments are made to the constraints and the starting values. As the objective of this section is to replicate the methodology of Engle et al. (2013), their constraints and starting values are also applied in the current methodology with small adjustments whenever needed.

Moving on to the model with the fixed-term span realized volatility, including four years of quarterly MIDAS lags, the obtained results are illustrated in Table 3. Once again, robust t-statistics are calculated through HAC standard errors and are presented within the corresponding parentheses. Furthermore, the Log-Likelihood Function (LLF) value and the Bayesian Information Criterion (BIC) are reported to make comparisons across the different time spans considered.

A general notice for this model is that no big differences are observed compared to the rolling window model described above. The values of θ seem to be lower than the corresponding values of the rolling window model. More specifically, θ is only significant in the first subperiod and the entire sample, whereas it is insignificant in the other two subperiods. This result implies that general conclusions from this model cannot be made for all considered time ranges. Therefore, only

Table 3: Parameter estimates of the GARCH-MIDAS model with fixed window realized volatility.

Sample	μ	α	β	θ	ω	m	LLF	BIC
1926-1952	0.00049	0.09746	0.86782	0.09696	5.7117	0.00592	20834.69	-6.20
	(4.11)	(8.07)	(45.63)	(18.38)	(47.99)	(12.24)	—	—
1953-1984	0.00046	0.08826	0.90268	0.10183	0.9596	0.00599	25013.81	-7.10
	(0.47)	(2.69)	(29.22)	(1.15)	(4.76)	(2.80)	—	—
1985-2010	0.00056	0.07410	0.89393	0.09396	13.680	0.00640	18037.07	-6.49
	(2.19)	(10.88)	(81.96)	(1.83)	(2.73)	(10.42)	—	—
1926-2010	0.00051	0.09722	0.87530	0.09768	4.9502	0.00592	77518.77	-6.64
	(10.68)	(27.46)	(314.52)	(17.32)	(123.90)	(17.21)	—	—

the significant and positive values for θ in the first subsample and the entire sample illustrate that the quarterly realized volatility MIDAS lags contain information that help model the low-frequency volatility component $\tau_{i,t}$.

Next, an irregular pattern is observed for the values of the parameter ω. In the first subsample and the whole sample, the values for parameters θ and ω are quite close. However, during the third subperiod, θ has a close value with the aforementioned periods, but the value of ω is tripled. This result might be the effect of the identification problems of the model, and this hypothesis can be verified with the two panels of Figure 2. According to Engle et al. (2013), the optimal weighting function is decreasing over the MIDAS lags, but this is not the case for the weighting function of the second subsample with an insignificant value for θ. The strong exponentially increasing weighting function in the right panel of Figure 2 poses serious identification problems in the model, which in turn results into insignificant parameter estimates.

It is worth mentioning that the t-statistics change drastically, as the restrictions and the starting values of the GARCH-MIDAS model change. This phenomenon is clear for the estimates of the whole sample, where the t-statistics of β and ω are approximately equal to 314 and 123, respectively. Regarding the α and β parameters, the fixed-term span model presents, on average, the same estimates as the rolling window model. This implies that the summation of these parameters is still lower than 1, ranging from 0.9653 to 0.9909. This observation implies a low persistence of the short-term volatility component $g_{i,t}$.

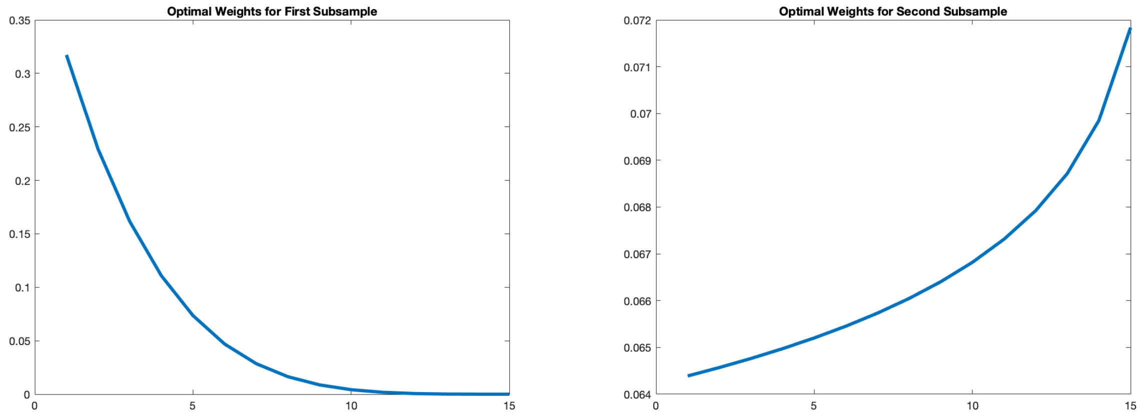


Figure 2: Optimal weights for the fixed-time span realized volatility model.

The last model examined in this section is the fixed-term span GARCH-MIDAS model with the Industrial Production (IP) growth rate instead of the realized volatility in the long-term component. For this model, $\tau_{i,t}$ is modeled with the log specification as introduced in Equation (6) and the Beta weighting function is altered compared to the previous two models. There are two weights, ω_1 and ω_2, corresponding to Equation (4). The obtained parameter estimates, together with the robust t-statistics for the four different time periods, are shown in Table 4.

Table 4: Parameter estimates of the GARCH-MIDAS model with fixed window IP growth rate.

Sample	μ	α	β	θ	ω_1	ω_2	m	LLF	BIC
1926-1952	0.00050	0.15384	0.84616	1.34396	1.0002	17.4751	0.06436	20663.60	-6.15
	(0.18)	(6.45)	(35.50)	(0.81)	(0.09)	(2.68)	(0.96)	—	—
1953-1984	0.00051	0.17482	0.82518	1.23227	1.6530	29.3820	0.04221	24853.70	-7.05
	(1.49)	(21.80)	(102.88)	(3.43)	(0.86)	(11.26)	(0.32)	—	—
1985-2010	0.00063	0.16572	0.83428	-0.00062	1.0034	36.6461	0.00032	17866.27	-6.43
	(6.18)	(18.84)	(94.82)	(-0.01)	(5.76)	(91.55)	(0.01)	—	—
1926-2010	0.00049	0.10529	0.89471	-1.78811	98.5528	80.3329	0.03648	70403.56	-6.60
	(8.96)	(13.35)	(113.48)	(-5.82)	(4.31)	(6.75)	(0.14)	—	—

The first notable observation is that all four values of the parameter θ changed drastically contrary to the two models with the realized volatility. Especially for the first two subperiods and the full period, the values of θ are clearly above one, but for the period $1985-2010$, it is even smaller than the two models presented above. The fact that the model generated such a parameter estimate poses again the identification question. Moreover, the small t-statistic values for the first and third subperiod give an extra reason to believe that the specific model is unidentified in certain time ranges. Nevertheless, inferences can be made when parameter θ is significant. The obtained estimates present results that are in contrast, that is, higher IP growth rate results into higher long-term volatility for the second subperiod and lower long-term volatility for the entire sample. However, the MIDAS lags of the IP growth rate in both time ranges contain information that helps interpret the long-term component of volatility $\tau_{i,t}$.

Additionally, the inclusion of the ω_1 term in the model, changed the values for parameter ω_2 a lot, making it even six times higher than in the model with the realized volatility. However, the ω_1 term has a value that is quite close in the three subperiods, but significantly larger for the whole sample, which seems a bit irregular. This might be the result of the identification issue of the model described a few lines above, but it can be verified with the graphs of the optimal weighting function shown in Figure 3. The introduction of the second weight in the function should not change the behavior of the function significantly. However, the right panel in Figure 3 shows the opposite. The weights do not represent a monotonic function, as in the other cases, and their values are higher for past observations.

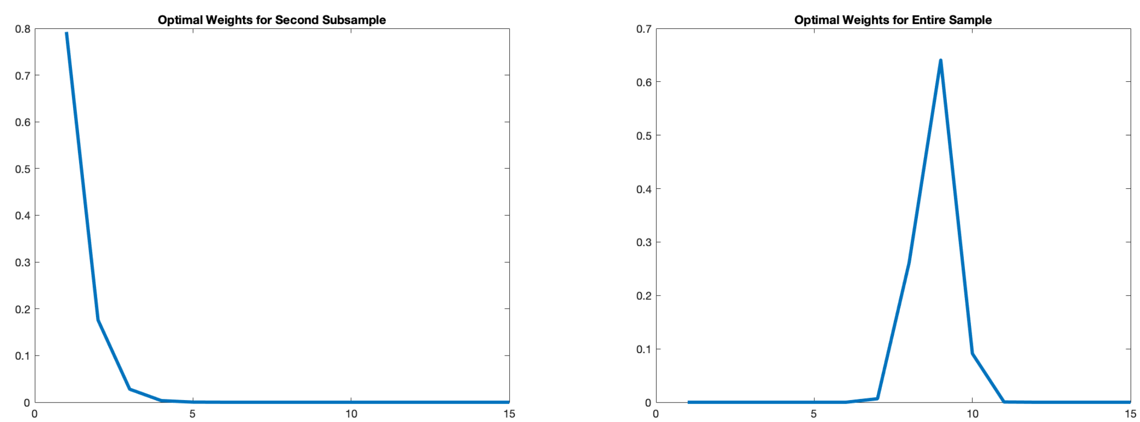
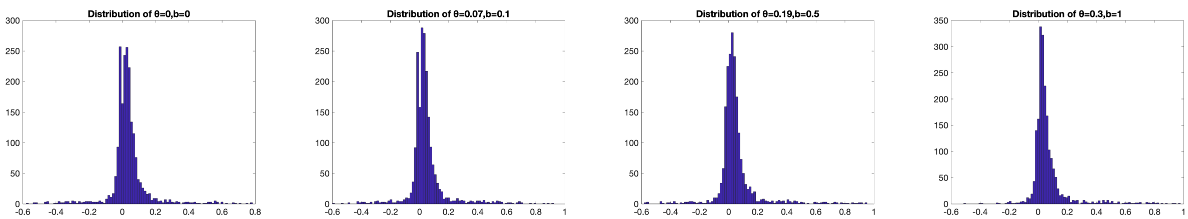


Figure 3: Optimal weights for the fixed-time span IP growth rate model.

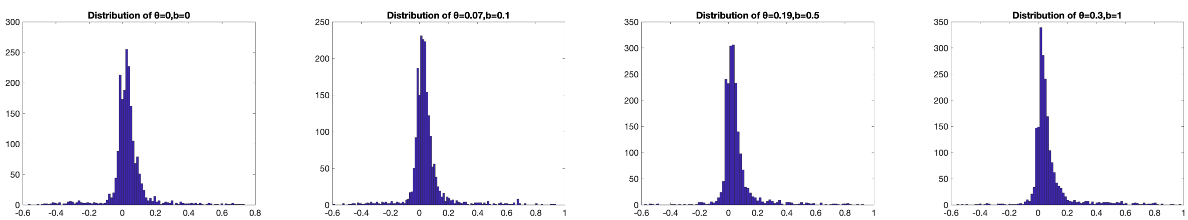
Last but not least, we observe that the summation of parameters α and β is almost equal to 1. This result is in contrast with what was observed in the previous two models. This fact leads to the outcome that constraint $\alpha + \beta < 1$ is fully satisfied in the optimization procedure and that there is a normal rate of persistence for the short-term volatility component of the GARCH-MIDAS model.

5.2 Monte Carlo Simulation Study Results

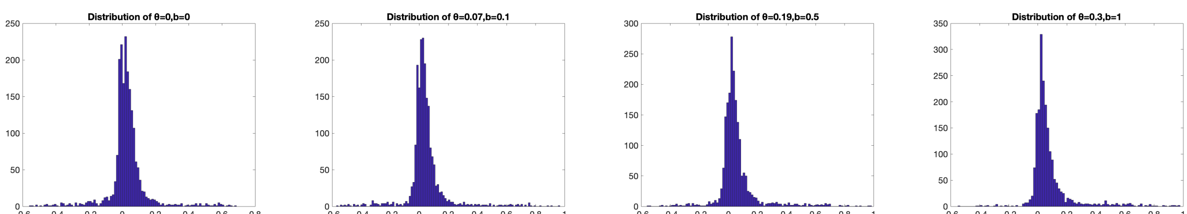
This section is devoted to the discussion of the results obtained from 16 distinct Monte Carlo simulations, each one aggregating 2000 iterations. First, the distribution densities of the different estimated parameters b and ω are presented in Figure 4. By inspecting these figures, we reach an initial conclusion that the results of Andrews and Cheng (2012) are partly verified. The authors found that the format of the distributions changes according to the real value of b, that is, smaller values of b produce bi-modal distribution densities. On the other hand, as b becomes larger, these densities approximate the normal distribution.



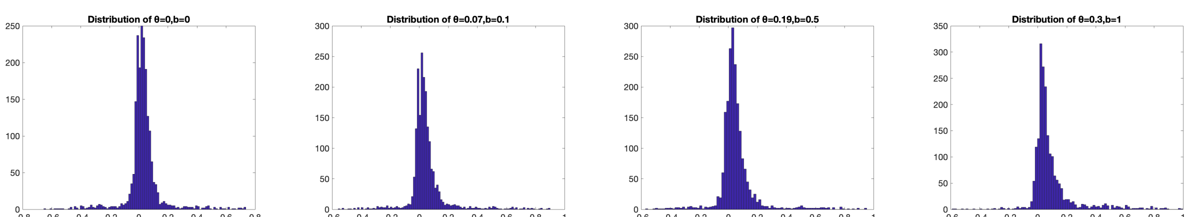
(a) $\omega = 1$



(b) $\omega = 1.5$



(c) $\omega = 2$



(d) $\omega = 3$

Figure 4: Histograms of the estimated parameter θ for different values of ω.

In this report, the distribution densities for the first two cases of b being equal to 0 and 0.1, which correspond to the non- and weakly identification categories and the underlying values of 0 and 0.07 for θ, seem to follow a bi-modal distribution. These densities do not have the normal distribution properties, a conclusion drawn by the outcome of the Jarque-Bera (JB) normality test presented in Table 5. The high skewness and kurtosis values and the low p-values for b $= (0, 0.1)$ and all cases of ω clearly indicate this finding.

One can recognize the bi-modal densities, for the non- and weakly identification categories, by the two large spikes around 0 separated by a trough in the first two columns of Figure 4, respectively. The particular type of distribution is clearly shown for the aforementioned values of b and $\omega = 1$, as a deep trough separates the two characteristic peaks of the distribution. However, increasing the value of ω decreases the bi-model effect in most of the densities presented in Figure 4. This observation implies that increasing the value of ω, while keeping a fixed parameter b, results in an improvement of the model behavior. Additionally, this effect is less severe on the densities for the weakly identification category, as the trough is not as deep as in the previous case.

These results verify the bi-modality distribution densities of the parameter estimates of the ARMA(1,1) model in Andrews and Cheng (2012). It should be noted that the explanation of this

Table 5: Outcome of the Jarque Bera normality test.

	$\omega = 1$				$\omega = 1.5$			
	b=0	b=0.1	b=0.5	b=1	b=0	b=0.1	b=0.5	b=1
Mean	0.0328	0.0440	0.0474	0.0636	0.0324	0.0422	0.0507	0.0727
St. Deviation	0.1167	0.1270	0.1402	0.1267	0.1183	0.1303	0.1224	0.1421
Skewness	1.0407	1.7934	2.3043	3.3198	0.8005	1.8327	2.6999	2.5920
Kurtosis	14.3475	14.1975	16.6486	18.6598	13.1021	15.6284	18.3805	14.1206
P-value JB test	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
	$\omega = 2$				$\omega = 3$			
	b=0	b=0.1	b=0.5	b=1	b=0	b=0.1	b=0.5	b=1
Mean	0.0286	0.0417	0.0554	0.0791	0.0291	0.0469	0.0526	0.0894
St. Deviation	0.1185	0.1384	0.1376	0.1474	0.1239	0.1328	0.1330	0.1534
Skewness	0.8524	1.5865	2.2339	2.7920	0.6563	2.1248	2.1043	2.4064
Kurtosis	12.4749	14.4931	14.3511	14.1628	13.4109	15.4377	15.8704	12.1360
P-value JB test	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010

phenomenon is not sufficiently justified in the existing literature. However, it may be associated with the value of the parameter θ being very close to zero. In such a case, the parameter ω cannot be uniquely estimated, and b takes any value close but not equal to 0.

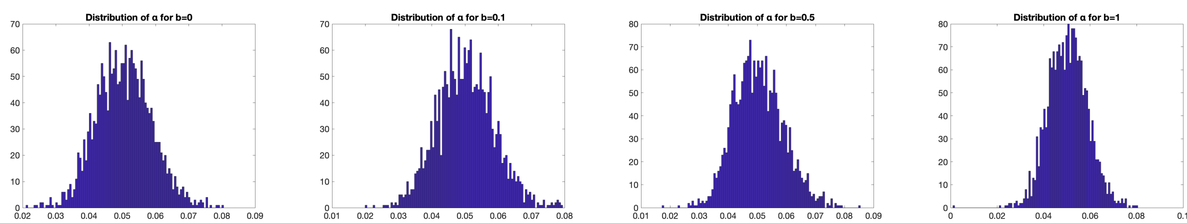
On the other hand, when b is equal to 0.5 and 1, which is the representation of the semi-strong and strong identification categories with the values for θ being equal to 0.19 and 0.3, the density distributions do not seem to follow a normal distribution. This is verified by the low p-values of the JB normality test and the values of the skewness and kurtosis presented in Table 5. The particular values differ quite a lot from the corresponding values of a normal distribution, which are 0 and 3, respectively.

These findings are in contrast with the results of Andrews and Cheng (2012), who mentioned that having at least a semi-strong identification category, results in having a normal distribution density for the estimated parameter b. More specifically, one can observe that the obtained distribution densities in the third and fourth column of Figure 4 seem to follow a Student-t or Cauchy distribution rather than a normal distribution. The large number of extreme values, higher than 0.2 in absolute terms, seems to raise the skewness value, whereas the observed high peaks between 0 and 0.1 imply a high value for the kurtosis of the distribution.

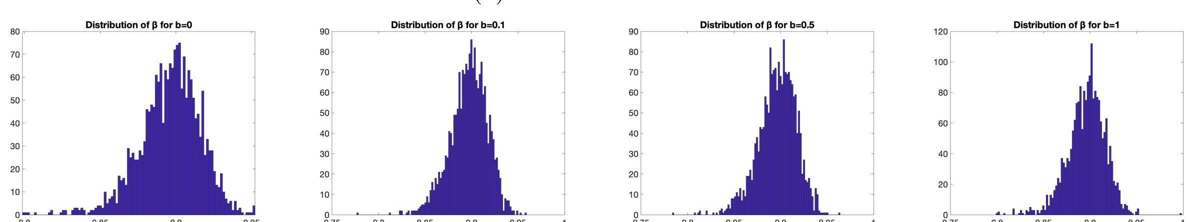
An explanation for the non-normal distribution densities for the specific identification categories might be the small number of iterations in the Monte Carlo simulations and the fact that only a small sample size was used. One can check the asymptotic properties of a model by taking several sample sizes at an increasing order and perform at least 5,000 iterations in each of the Monte Carlo simulations. Therefore, the obtained Cauchy/Student-t distribution densities are not invalid based on the selected sample size and the amount of the simulation iterations. Another explanation for the non-normal distribution densities for the specific identification categories could be the starting values of the GARCH-MIDAS model in the optimization technique. As noted in Section 5.1, changing the starting values of the model yields a completely different set of the estimated parameters. Therefore, the fact that only a few different parameter values were input into the simulation could explain the non expected type of distribution densities. Furthermore, another cause for this irregular result could be the consideration of only monthly estimates for the long-term volatility component $\tau_{i,t}$. Engle et al. (2013) mentioned that they perform the research for monthly, quarterly, and biannually estimates of the $\tau_{i,t}$ component. However, the quarterly estimates resulted in the best results. Hence, the fact that the quarterly and biannually estimation of $\tau_{i,t}$ is not explored in this report, could clarify the obtained type of distributions.

To sum up, the findings presented above indicate that the parameter θ suffers from identification issues, especially in the weak identification category. The non-normal distribution densities for the semi-strong and strong identification categories can be tolerated due to the small sample size and the number of the Monte Carlo iterations. Nevertheless, it is interesting to check whether the identification problem of the MIDAS model transfers to the parameters α and β of the unit GARCH process.

The distribution densities for parameters α and β when $\omega = 1$ are presented in Figure 5. A quick look at the histograms reveals that in almost all cases, the densities are centered around the true values of α and β, which are 0.05 and 0.90, respectively. The obtained densities for the parameter α approximate the normal distribution in all cases, as there is a limited number of extreme values and spikes that are not so high. This observation is verified with the p-values of the JB normality test, which are on average equal to 0.03. In contrast, the densities of the parameter β do look to be centered around their true value; however, they do not approximate the normal distribution. The outcome of the JB normality test showed p-values that are smaller than 0.001, which of course implies that the distribution densities are not normal.



(a) Distributions of α



(b) Distributions of β

Figure 5: Histograms of the estimated parameters α and β for $\omega = 1$.

The conclusion drawn for the GARCH parameters above verify the results of Andrews and Cheng (2012). More specifically, the authors state that Maximum Likelihood Estimation leads to small biases in the specific parameters. This means that the identification issues of the MIDAS model transfer to a small portion also to the GARCH process, but without serious consequences.

5.3 Robust Critical Values Results

Since the GARCH-MIDAS model suffers from identification issues, especially in the weakly identification category, we explore the usage of two robust critical values. These critical values are the Least Favorable and Type-I, calculated according to the methodology of Andrews and Cheng (2012). More specifically, the different robust critical values that are calculated for each of the four identification categories of θ and for each of the four distinct values of ω are illustrated in Table 6.

Table 6: 95% Confidence Interval of the Critical Values.

	b	Least Favorable Interval	Type-I Interval
$\omega = 1$	0	[-5.0008 ; 5.0008]	[-1.9600 ; 1.9600]
	0.1	[-5.3265 ; 5.3265]	[-1.9600 ; 1.9600]
	0.5	[-6.2951 ; 6.2951]	[-1.9600 ; 1.9600]
	1	[-9.2553 ; 9.2553]	[-1.9600 ; 1.9600]
$\omega = 1.5$	0	[-5.3757 ; 5.3757]	[-5.3757 ; 5.3757]
	0.1	[-5.2132 ; 5.2132]	[-1.9600 ; 1.9600]
	0.5	[-6.0382 ; 6.0382]	[-1.9600 ; 1.9600]
	1	[-8.9778 ; 8.9778]	[-1.9600 ; 1.9600]
$\omega = 2$	0	[-5.0508 ; 5.0508]	[-5.0508 ; 5.0508]
	0.1	[-6.7837 ; 6.7837]	[-1.9600 ; 1.9600]
	0.5	[-6.8108 ; 6.8108]	[-1.9600 ; 1.9600]
	1	[-8.2190 ; 8.2190]	[-1.9600 ; 1.9600]
$\omega = 3$	0	[-5.2066 ; 5.2066]	[-5.2066 ; 5.2066]
	0.1	[-4.7712 ; 4.7712]	[-1.9600 ; 1.9600]
	0.5	[-6.5219 ; 6.5219]	[-1.9600 ; 1.9600]
	1	[-8.6249 ; 8.6249]	[-1.9600 ; 1.9600]

A general conclusion is that, in almost all cases, the Least Favorable robust critical values grow while increasing values of parameter b. However, changing the value of the parameter ω does not result in any big variation for the values of the Least Favorable criterion. As for the Type-I critical value, it is equal either to the Least Favorable value for the non-identified categories, except when $\omega = 1$, or to the asymptotic normal critical value for all other cases. This outcome is irregular because in such a case, the simulated densities should follow a normal distribution. However, the histograms in Figure 4 do not show such a behavior.

The new robust critical values, especially the Least Favorable, do change most of the results presented in Section 5.1. Nonetheless, this is not the case for the GARCH-MIDAS model with the fixed-time span realized volatility. This model is found to suffer from identification issues since the t-statistics of parameter θ in Table 3 are lower than the significance threshold value. Observing the values for the parameter θ, which are between 0.094 and 0.101, one can understand that they fall under the weakly identification category. Hence, using the robust critical values of the weakly identification category, irrespective of the value of the parameter ω, does not change the significance of the parameters. This results in the persistence of the identification problems in the specific model. Consequently, the quarterly realized volatility does not include enough financial information to help explain the long-term volatility component $\tau_{i,t}$.

Next, the GARCH-MIDAS model with the fixed-time span industrial production growth rate also presented identification problems. A quick look in Table 4 reveals either very low values for parameter θ or nearly zero values for the t-statistics of the particular parameter. The new robust critical values do not alter the statistical significance of parameter θ, resulting again in the persistence of the identification issues in the model with the macroeconomic indicator. However, an interesting observation is that these particular robust critical values result in the loss of identification when before this problem did not arise. Observing the values of θ in the second subsample and the entire sample from Table 4, one can derive that it represents the strong identification category. Taking into account the Least Favorable critical values for the specific case, irrespective of ω, and the t-statistics for the aforementioned periods, we see that the significance of θ is lost. This implies that the robust critical values make the performance of the model worse, while the relations that held before, now are invalid. That is, the relation that higher industrial production growth rate results in higher long-term volatility is now baseless.

Continuing, the model that did not present any signs of identification issues is the GARCH-MIDAS model with the rolling window realized volatility. In Table 2, one observes that the parameter θ has t-statistics that are large enough to prove their significance. However, parameter θ in the third subsample considered, has a t-statistic value that with the Least Favorable critical value of the weakly identification category becomes insignificant. This observation holds for all four different values of parameter ω and implies that the model, in the specific subperiod, losses its identification properties and any relevant relations do not hold anymore. More specifically, the long-term volatility component τ_i does not contain any financial information from the rolling window estimates of the realized volatility.

Concluding, the calculation of the two robust critical values resulted in expected and unanticipated results. Regarding the expected, we refer to the persistence of the identification issues in the aforementioned GARCH-MIDAS models. Since the robust critical values cannot be lower than the normal asymptotic critical values, the already small values of the t-statistics cannot change the significance of the MIDAS parameter θ. Regarding the unanticipated results, we notice the loss of the identification properties in the rolling window realized volatility GARCH-MIDAS model. One would expect that the calculation of the robust critical values would do the opposite, since creating identification issues in a model result in an inefficient model behavior and losing the economic relations between the variables of the model.

6 Conclusions & Discussion

The relatively new GARCH-MIDAS framework is the first model that distinguishes between the short-term component and the long-term component of volatility. The distinctive feature of this model is that it links stock market data with key macroeconomic variables, even though the particular data might have different frequencies. It was introduced by Engle et al. (2013), and since then, it has been applied numerous times in various settings by the researcher community.

After a close examination of the GARCH-MIDAS estimation results in Engle et al. (2013), there are serious concerns regarding the lack of the identification in some parts of the parameter space of the model. The fact that the MIDAS parameter θ is approximately equal to zero, whereas the parameter ω has a wide range of values led to this reasonable doubt. If this is indeed correct, then the inferences and conclusions made in the specific research are potentially invalid and do not hold in reality.

To investigate this issue, three GARCH-MIDAS models are estimated with a macroeconomic indicator and the realized volatility. To this end, the value-weighted index and the industrial production growth rate were obtained from the CRSP and FRED databases, respectively. The estimation of the GARCH-MIDAS models showed that there were signs of identification issues since the parameter θ was either insignificant or had values close to zero. The results found in this report are comparable with the results that Engle et al. (2013) reached in their paper.

Next, a Monte Carlo simulation study is performed, adopting the methodology of Andrews and Cheng (2012). The authors state that when the obtained densities from the simulation are not normally distributed, then the model suffers from identification issues. Although Andrews and

Cheng used a unit ARMA model, they state that their methodology is applicable to a broader range of models, including the GARCH-MIDAS model. By simulating the model with 16 different combinations of the MIDAS parameters θ and ω, the obtained densities were investigated to check their distribution properties. For the first two categories of the parameter θ, bi-modal densities were found, indicating that the model suffers from identification issues. On the other hand, the other two categories of θ, produced Student-t densities, a result that is not in accord with the results of Andrews and Cheng (2012). However, due to the low number of iterations and the small sample size in the simulation, the Student-t densities were tolerated, and thus, the model did not suffer from severe identification issues in these categories. Despite the aforementioned results, the identification problems from the MIDAS parameters did not transmit to the parameters α and β of the GARCH process.

Since the simulation results suggest that the GARCH-MIDAS model exhibits lack of identification in some parts of the parameter space of the model, two robust critical values were computed to possibly resolve this issue. Particularly, the Least Favorable and the Type-I critical values were calculated for each of the four identification categories and for each of the four considered values of the parameter ω. Considering the specific robust critical values instead of the standard critical values did not solve the identification issues in the models. However, the consideration of these robust critical values resulted in the loss of the identification properties in a model that did not previously have such an issue.

To sum up, this research suggests that the GARCH-MIDAS model suffers from identification problems when the estimated parameter θ ranges between 0 and 0.1 in absolute terms. This conclusion is based on a Monte Carlo simulation study, where 16 different values for the MIDAS parameters θ and ω were considered. Through this simulation, two robust critical values were computed to probably resolve this problem. However, this was not the case, and the considered models either remained with or obtained identification issues.

Nevertheless, this research has a few limitations with the most important one being that it does not accommodate or test for structural breaks in the GARCH-MIDAS model. Engle et al. (2013) state that they account for the structural breaks by considering several subsamples of the original time series range. Remarkably, however, Lamoureux and Lastrapes (1990) proved that this procedure is not efficient. More specifically, the authors mention that structural breaks is a "stylized fact" of volatility modeling and if not taken into account, it can lead to spurious results. Therefore, it would be noteworthy for future research to examine whether a GARCH-MIDAS model that accounts for structural breaks suffers from identification issues.

Another limitation of this research approach is that the parameter space of the model is not explored to an exhaustive degree. This limitation might have resulted in the unexpected results of the Monte Carlo simulation study. Although, Engle et al. (2013) provide an initial set of starting values for each of the fixed-time span and rolling window models, small changes to the specific values can result in big differences in the estimated parameters of the GARCH-MIDAS model. A promising suggestion for future research would be a more thorough examination of the model parameter space, under different degrees of identification, different degrees of volatility persistence, and different sample sizes.

References

Andrews, D. W., & Cheng, X. (2012). Estimation and Inference with Weak, Semi-Strong, and Strong Identification. *Econometrica*, *80*(5), 2153–2211.

Andrews, D. W., & Cheng, X. (2013). Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure. *Journal of Econometrics*, *173*(1), 36–56.

Andrews, D. W., & Cheng, X. (2014). GMM Estimation and Uniform Subvector Inference with Possible Identification Failure. *Econometric Theory*, *30*(2), 287–333.

Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with Mixed Frequencies. *Federal Reserve Bank of St. Louis Review*, *92*(6), 521–36.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Cheng, X. (2015). Robust Inference in Nonlinear Models with Mixed Identification Strength. *Journal of Econometrics*, *189*(1), 207–228.

Clements, M. P., & Galvão, A. B. (2008). Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States. *Journal of Business & Economic Statistics*, *26*(4), 546–554.

Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and Short-Term Cryptocurrency Volatility Components: A GARCH-MIDAS Analysis. *Journal of Risk and Financial Management*, *11*(2), 23.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock Market Volatility and Macroeconomic Fundamentals. *Review of Economics and Statistics*, *95*(3), 776–797.

Engle, R. F., & Rangel, J. G. (2008). The Spline-GARCH Model for Low-Frequency Volatility and its Global Macroeconomic Causes. *The Review of Financial Studies*, *21*(3), 1187–1222.

Fang, L., Chen, B., Yu, H., & Qian, Y. (2018). The Importance of Global Economic Policy Uncertainty in Predicting Gold Futures Market Volatility: A GARCH-MIDAS Approach. *Journal of Futures Markets*, *38*(3), 413–422.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). *The MIDAS Touch: Mixed Data Sampling Regression Models.*

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, *26*(1), 53–90.

Guérin, P., & Marcellino, M. (2013). Markov-Switching MIDAS Models. *Journal of Business & Economic Statistics*, *31*(1), 45–56.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. *The Journal of Finance*, *45*(1), 221–229.

Naghi, A. A. (2018). *Identification Robust Predictive Ability Testing.* Tinbergen Institute, Erasmus University Rotterdam.

Officer, R. R. (1973). The Variability of the Market Factor of the New York Stock Exchange. *The Journal of Business*, *46*(3), 434–453.

Pan, Z., Wang, Y., Wu, C., & Yin, L. (2017). Oil Price Volatility and Macroeconomic Fundamentals: A Regime Switching GARCH-MIDAS Model. *Journal of Empirical Finance*, *43*, 130–142.

Schwert, G. W. (1989). Why Does Stock Market Volatility Change over Time? *The Journal of Finance*, *44*(5), 1115–1153.

A Programming Code

The matlab code that was used to get the results of this thesis is long enough and cannot fit in this report. The .zip file that comes with this thesis has the following outline:

1. The **"Estimation Code"** folder contains the matlab files that were run to get the results of the thesis replication part. More specifically, the included files are:
    - betapolyn_macro.m: is code for the beta polynomial function of the GARCH-MIDAS model with the fixed-time span industrial production growth rate.
    - betapolyn.m: is code for the beta polynomial function of the GARCH-MIDAS model with the rolling window and the fixed-time span realized volatility.
    - descriptive.m: is the code for obtaining the descriptive statistics presented in the data section of the thesis.
    - garchmidas_fix_macro.m: is the code for the estimation of the GARCH-MIDAS model with the fixed-time span industrial production growth rate.
    - garchmidas_fix_rv.m: is the code for the estimation of the GARCH-MIDAS model with the fixed-time span realized volatility.
    - garchmidas_roll.m: is the code for the estimation of the GARCH-MIDAS model with the rolling window realized volatility.
    - ipgrowth.mat: includes the data of the industrial production growth rate in Matlab.
    - ipgrowth.xlsx: includes the data of the industrial production growth rate in Excel.
    - nlh_garchmidas_macro.m: is the code for the negative log likelihood function of the GARCH-MIDAS model with the fixed-time span macroeconomic variable (industrial production growth rate).
    - nlh_garchmidas.m: is the code for the negative log likelihood function of the GARCH-MIDAS model with the rolling window and the fixed-time span realized volatility.
    - returndata.xlsx: includes the value-weighted return data.
2. The **"Estimation Results"** folder contains all the results obtained for the thesis replication part.
3. The **"Graphs"** folder contains all the graphs that are included in the thesis report on a .png extension.

4. The **"Simulation Code"** folder contains the matlab files that were run to get the results of the thesis extension part. More specifically, the included files are:

    - betapolyn_allw.m: is code for the beta polynomial function of the GARCH-MIDAS model with the rolling window and the fixed-time span realized volatility (with small adjustment).
    - betapolyn.m: is code for the beta polynomial function of the GARCH-MIDAS model with the rolling window and the fixed-time span realized volatility.
    - garchmidas_fix_mc.m: is the code for the simulation of the GARCH-MIDAS model with the fixed-time span realized volatility. [3]
    - nlh_garchmidas.m: is the code for the negative log likelihood function of the GARCH-MIDAS model with the rolling window and the fixed-time span realized volatility.

5. The **"Simulation Results"** folder contains all the results obtained from the 16 different simulation that were run to get the results of the thesis extension part.

[3] Please note that you need to change the values of theta and omega at lines 16 and 17, respectively, to run the code for different identification categories and different values of omega. The values of theta that are considered in this report are 0, 0.07, 0.19, 0.3, whereas the values of omega are 1,1.5,2,3.