ERASMUS UNIVERSITY ROTTERDAM

BACHELOR THESIS DOUBLE DEGREE ECONOMICS/ECONOMETRICS

Breaking Down Stock Market Volatility
A Component Model Analysis

Submitted by:	**Wesley van Driel (425342)**
Supervised by:	**B. van Os**
Second assessor:	**M.D. Zaharieva**

Abstract

This paper analyzes the link between long-term stock market volatility and macroeconomic variables through the GARCH-MIDAS model. By using daily stock returns and monthly macroeconomic variables, the results show that the long-term volatility is strongly influenced by the realized variance, producer price index and industrial production. Specifically, an increase in inflation leads to an increase in the long-term volatility, whereas changes in industrial production have the opposite effect on this component. These results are confirmed and validated by the new GAS-MIDAS and GARCH-AMIDAS models in terms of an out-of-sample forecasting exercise. The GARCH-AMIDAS model outperforms the other component models in the short-run, while the GAS-MIDAS model works best in the long-run.

Date: July 6, 2019

Contents

1. Introduction

A primary goal of investors is to hedge the risk on outstanding portfolios. One way of doing so is by assessing the portfolio's volatility over time. Traditionally, volatility is estimated by different ARCH-type models that capture the size of stock market volatility based on its time-variation and day-to-day clustering (Engle, 1982). However, recent studies have shown that the dynamics of volatility are best captured by so-called component models. Lee and Engle (1993) introduced one of these models by separating the conditional volatility into an additive short- and long-term component. A vast literature, including Schwert (1989b), Glosten, Jagannathan, and Runkle (1993) and Timmermann (2000), has shown that this long-term component can be identified using monthly macroeconomic variables and the state of the economy.

However, macroeconomic variables are often measured over periods spanning several months, whereas time-varying volatility is based on high-frequency data. That is why similar research that uses traditional volatility models is bounded on a limited set of macroeconomic variables. Ghysels, Sinko, and Valkanov (2007) helped to overcome this issue by introducing the mixed data sampling (hereafter MIDAS) model. The key feature of this model is its ability to use data of different frequencies in a single regression. It is therefore the perfect instrument to regress time-varying volatility on macroeconomic variables. That is why Engle, Ghysels, and Sohn (2013) use this MIDAS specification in combination with the GARCH model of Bollerslev (1986) to explain long-term volatility in a new class of component models: the GARCH-MIDAS models. An advantage of this model over other component models is that it allows low-frequency macroeconomic variables to directly affect the conditional variance via the long-term volatility component.

This paper analyzes the connection between long-term volatility and economic variables through the GARCH-MIDAS model. In particular, it attempts to answer the research question how the growth rates of production price index inflation and industrial production influence the total volatility of daily U.S. stock returns. The foundation of the research is a replication of the results by Engle et al. (2013). The main contribution of this paper to the existing literature is an extension of the component models by introducing the GAS-MIDAS (Generalized Autoregressive Score model) and GARCH-AMIDAS (asymmetric MIDAS) models. These models are based on their unique characteristics when dealing with stock returns. For example, the GAS model is known for a better fit of unexpected returns than GARCH (see for example Creal, Koopman, and Lucas (2013)), whereas the asymmetric MIDAS model takes different effects of returns on volatility into account.

The empirical analysis covers data for the period 1885-2018, which includes daily stock returns and the monthly macroeconomic variables producer price index and industrial production. The latter variables have been transformed to quarterly volatilities by taking the geometric series over three consecutive months for the period 1920-2018. The data set is first used to replicate the results of Engle et al. (2013) and afterwards to compare the GARCH-MIDAS model with the new component models by testing forecast capacity.

The main findings can be summarized as follows. First, the GARCH-MIDAS estimates with realized variance indicate a positive and significant impact on the long-term volatility component. It should be noted that this impact is small overall, but is able in some cases to explain around 50% of the total variance. It is also found that the macroeconomic variables can explain a significant part of the long-term component. Specifically, including the producer price index corresponds to an overall increase of the variance, whereas the inclusion of industrial production is double-sided. The impact of leveled industrial production to the persistent volatility corresponds to counter-cyclical effects, while industrial production expressed as variances lead to business effects. Finally, the forecast evaluations show that the GARCH-AMIDAS specification outperforms the traditional GARCH-MIDAS model with macroeconomic variables for 1- and 2-month ahead forecasts. The latter model fits the data well when using realized variances for the same time period. The GAS-MIDAS model outperforms the other models with any economic variable for longer horizons.

The remainder of the paper is organized as follows. The literature and its relevance to this research are discussed in Section 2. Next, the methods of Engle et al. (2013) are thoroughly explained in Section 3, whereas the new component models are explained in Section 4. The data that is used to derive empirical results is examined in Section 5. Sequentially, the obtained results are presented in Section 6, after which the overall conclusions and limitations are discussed in Section 7.

2. Literature Review

During the past century, researchers attempted to find models that can capture stock market volatility. Even though progress was made, an inevitable obstacle was the result that stock returns have asymmetric effects on their volatility (Singleton & Wingender, 1986). The present value model of Campbell (1991) is one of the first to encapsulate the thought that publicly accessible news may have different effects on unexpected returns. It is in fact this information that influences the horizon of the expectations on future cash flows and by doing so determines the magnitude of these effects (Campbell, 1991). That is why previous research, including Engle and Lee (1999) and Chernov, Gallant, Ghysels, and Tauchen (2003), considered volatility component models that separate volatility based on the effects on the short- and long-horizon. In fact, Chernov et al. (2003) found that at least two components are needed to capture the total volatility of stock prices based on different stock price models. The first component is influenced by day-to-day fluctuations, whereas the latter is determined by general conditions of the economy (Chernov et al., 2003). These two components combined in a single model should capture the total dynamics of volatility based on the impact of news on the macro economy and day-to-day fluctuations.

However, it has been unclear which variables influence this long-term component. This changed when academics revisited the results of Schwert (1989b), who tries to characterize the relation between stock market volatility and macroeconomic variables. These variables include economic activity, macroeconomic volatility, financial leverage, and various other measurements. The motivation behind this research is the observation that stock market volatility changes significantly over time. As a matter of fact, stock market volatility reached up to twenty percent over the period 1857-1987 (Schwert, 1989b). This result is reached by using data from the same period to determine the realized variance by aggregating daily to monthly returns. Using a Vector Autoregressive model, Schwert (1989b) found counter-cyclical patterns in industrial production and volatility. The intuition behind this observation is that high confidence in the economy corresponds to higher demand for products, while there is less uncertainty on the market at the same time (Schwert, 1989b).

Continuing the investigation in this field, Engle et al. (2013) used a two-component volatility model to capture different effects on stock market volatility. More precisely, the authors specified a unit GARCH process for the short-term component and a MIDAS model for the long-term component. An advantage of this model is that the macroeconomic variables can be directly examined by scaling them to the same frequency as the daily returns using a Beta weighting scheme (Engle et al., 2013).

Overall, Engle et al. (2013) found that the explanatory power of the total volatility stays approximately the same for volatility models with a quarterly time horizon, but performs better for the biannual horizon. Specifically, by directly including the macroeconomic variables producer price index and industrial production directly into the model, the GARCH-MIDAS model seems to be more appealing when considering a longer time horizon. The macroeconomic variables seem to explain between 10% and 35% of the daily stock return's volatility, which implies that the model can also be useful for the short-run (Engle et al., 2013). One should note that these results are obtained by identifying break-points in the data because, similar to the conclusions of Schwert (1989b), it is hard to capture economic changes such as crises.

The results of Engle et al. (2013) have not gone unnoticed by other academics. After its publication in 2013, other researchers looked extensively at the implications of the GARCH-MIDAS model in various fields. For instance, Conrad and Loch (2015) report additional macroeconomic variables - on top of the ones used in the original paper - that can explain the behaviour of U.S. stock returns. They support their findings by changing the GARCH(1,1) process to a threshold GARCH(1,1) process to take into account different effects of the shocks on the short-term component. This gives rise to the idea that changing specifications of the GARCH-MIDAS model might improve its explanatory power. The GARCH-MIDAS model can be further augmented, as shown by Asgharian, Hou, and Javed (2013). They used principal component analysis to see if the predictive power of the GARCH-MIDAS model can be improved. And indeed, by including the first principal component the GARCH-MIDAS model outperforms the original specification and can explain business cycles (Asgharian et al., 2013).

In a similar way, the MIDAS model can be applied in other fields. For instance, Clements and Galvão (2009) used this model to show that the latest monthly data can directly predict U.S. output growth. Their results are innovative in the field of macroeconomic forecasting, because they showed that applying the MIDAS model to determine the direction of output growth works better than using previously calculated forecasts (Clements & Galvão, 2009). This gave rise to the strength of the MIDAS model when forecasting several periods ahead, instead of compounding them by individual forecasts. Bai, Ghysels, and Wright (2013) actually investigated this predictive power by comparing the MIDAS model with the Kalman filter. The authors looked at the balance between forecasting performances and computational power. Bai et al. (2013) found that the forecasting performances of the two models are close, while the computational cost of the MIDAS model is significantly less. This is attractive to the users of the latter model as it will provide results faster, which is for instance preferred in the field of financial analysis. However, this model does not need to be static and can be altered to take different regimes into account. Pan, Wang, Wu, and Yin (2017) applied these so-called regime-switching MIDAS models on macroeconomic variables to predict future oil price movements. In fact, a model with two regimes has a better forecasting performance than the original model, because it takes structural breaks in the data into account (Pan et al., 2017).

3. The GARCH-MIDAS Model

Originally, Lee and Engle (1993) express unexpected returns as $\sqrt{\tau_t + g_{it}}\varepsilon_{it}$, with ε_{it} being the shock. In this decomposition g_{it} represents the short-run component that captures daily fluctuations that are implicitly connected to stock movements. On the other hand, τ_t represents a trend in the return's volatility that is directly influenced by the state of the economy.

The equation of unexpected returns has been revised in Engle and Rangel (2008) to fit their spline-GARCH model. More precisely, assume that there are N_t days in month t, hereby allowing the number of days to differ between months. If r_{it} is the return on day i in month t, then it holds that:

$$r_{it} = \mu + \sqrt{\tau_t g_{it}}\varepsilon_{it}, \qquad \forall i = 1, \ldots, N_t, \tag{1}$$

where $\varepsilon_{it} \mid \mathcal{I}_{i-1,t} \sim N(0,1)$ with $\mathcal{I}_{i-1,t}$ being the information set available at day $i-1$ in month t. It is also assumed that the expected daily return μ is constant for all i and t. In line with the research of Engle and Rangel (2008), the short-term component g_{it} in equation (1) follows a mean-reverting GARCH(1,1) process of the form:

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

with $\alpha, \beta \geq 0$ and $\alpha + \beta \leq 1$ to ensure that the GARCH(1,1) process is covariance stationary. On the other hand, τ_t can be expressed using either a fixed time span or a rolling window.

3.1. Fixed Time Span

As mentioned in Section 2, Schwert (1989b) measured the long-run volatility using realized volatilities by aggregating daily returns over a certain spanning horizon. The realized variance over a month t is denoted by RV_t. However, extensive research has shown that determining the realized volatility in this way is noisy, because outliers can rapidly change the measure's magnitude and precision. That is why Engle et al. (2013) regard the GARCH-MIDAS model as a filter of the realized volatility by considering the following MIDAS regression:

$$\tau_t = \xi + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) RV_{t-k}, \tag{3}$$

$$RV_t = \sum_{i=1}^{N_t} r_{it}^2, \tag{4}$$

where φ_k is the so-called smoothing function and is given by the Beta weighting scheme:

$$\varphi_k(\omega_1, \omega_2) = \frac{(k/K)^{\omega_1 - 1}(1 - k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K} (j/K)^{\omega_1 - 1}(1 - j/K)^{\omega_2 - 1}}, \tag{5}$$

with $\varphi_k(\omega_1, \omega_2) \geq 0$ for $k = 1, ..., K$. Traditionally, the exponential weighting scheme is used due to its computational easiness. This paper, on the other hand, uses the Beta lag function because it can incorporate different lag structures and weighting schemes (Ghysels et al., 2007). For example, setting $\omega_1 = 1$ and $\omega_2 > 1$ ensures a that the weights decay over their domain, where the slope is determined by the magnitude of ω_2. The maximum number of lags K is determined by minimizing the Bayesian Information Criterion (hereafter BIC).

The GARCH-MIDAS model with fixed time span RV is given by equations (1)-(5), where the parameters of interest are given by the set $\Theta = \{\alpha, \beta, \mu, \xi, \theta, \omega_1, \omega_2\}$.

3.2. Rolling Window

Another way to express the realized variance is by using a rolling window, which is obtained by removing the constraint that τ_t is fixed during a month. Instead, it can fluctuate throughout a month such that the revised MIDAS regression is formulated as:

$$\tau_i^{(rw)} = \xi^{(rw)} + \theta^{(rw)} \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) RV_{i-k}^{(rw)}, \tag{6}$$

$$RV_i^{(rw)} = \sum_{j=1}^{N'} r_{i-j}^2, \tag{7}$$

where N' is expressed in days and determines the size of the rolling window. Similar to Engle et al. (2013) the GARCH-MIDAS model with rolling windows specification is obtained by also dropping the monthly aspect in equations (1) and (2), and combining these with equations (5)-(7).

3.3. GARCH-MIDAS with Macroeconomic Variables

It is also possible in the GARCH-MIDAS framework to replace the realized variance by other variables. As mentioned in Section 1, this paper considers the macroeconomic variables industrial production growth rate and producer price index inflation rate. The one-sided MIDAS filter that is used in Engle et al. (2013) with these variables is given by:

$$\log \tau_t = \xi_n + \theta_n \sum_{k=1}^{K_n} \varphi_k(\omega_{1,n}, \omega_{2,n}) X^m_{n,t-k}, \tag{8}$$

where $X^m_{n,t-k}$ represents measure type n (either l for level or v for volatility) of macroeconomic variable m. The exact estimation procedure of the macroeconomic volatility is explained thoroughly in Section 5.

An interesting feature of this model is that the levels and volatilities of a single macroeconomic variable m can be combined in one model that becomes:

$$\log \tau_t = \xi_{lv} + \theta_l \sum_{k=1}^{K_l} \varphi_k(\omega_{1,l}, \omega_{2,l}) X^m_{l,t-k} + \theta_v \sum_{k=1}^{K_v} \varphi_k(\omega_{1,v}, \omega_{2,v}) X^m_{v,t-k}, \tag{9}$$

where K_l and K_v can be of different size. The reason behind this is that both parameters represent the lasting effects of either levels or volatility on the long-term volatility. These effects may have a different impact and should be scaled accordingly, in which case K_l and K_v might not be equal (Engle et al., 2013).

3.4. Estimation Procedure

The GARCH-MIDAS models are estimated using quasi-maximum likelihood (hereafter QML). Wang and Ghysels (2015) showed that the estimators are asymptotically normal when using realized variance, but did not show whether this also holds for macroeconomic variables. Conrad and Loch (2015) filled this gap this by using Monte Carlo simulation to show that the asymptotic distribution of the estimators stays the same.

To estimate the parameters using QML the probability density function of the returns needs to be established. A generic distribution of the returns in equation (1) is given by:

$$f(r_{it}) = \frac{1}{\sigma_{it}} f(\varepsilon_{it}), \qquad \forall i = 1, ..., N_t, \tag{10}$$

where $\sigma_{it} = \sqrt{\tau_t g_{it}}$. A traditional approach in basic ARCH-type models is to assume that ε_{it} is standard normally distributed. That is, if the monthly short-term volatility is given by $g_t = \sum_{i=1}^{N_t} g_{it}$

and Θ is the set of unknown parameters, the log-likelihood function becomes:

$$\ell(\Theta) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log\ \tau_t g_t + \log(2\pi) + \frac{(r_t-\mu)^2}{\tau_t g_t}\right]. \tag{11}$$

However, a common result in financial research is that (daily) stock returns are often not normally distributed; see for instance Schwert (1989b), Chernov et al. (2003), and Campbell and Hentschel (1992). In fact, returns are more diverse around their mean such that large returns occur more often than assumed by the normal distribution. The underlying distribution has in general a higher peak and fatter tails than the normal distribution to capture these characteristics. It might therefore be of interest to also investigate the Student's t-distribution, which is known for its high peak and fat tails. In that case, the log-likelihood function changes to:

$$\begin{aligned}\ell(\Theta) = \sum_{t=1}^{T}\Bigg[&-\frac{1}{2}\log\ \tau_t g_t + \log\Gamma\left(\frac{\nu+1}{2}\right) - \frac{1}{2}\log(\nu\pi)\\ &-\log\Gamma\left(\frac{\nu}{2}\right) - \frac{\nu+1}{2}\log\left(1+\frac{(r_t-\mu)^2}{\nu\tau_t g_t}\right)\Bigg].\end{aligned} \tag{12}$$

A full derivation of the two log-likelihood functions can be found in Appendices A.1 and A.2, respectively.

It should be noted that the equation above needs a predetermined number of lags in the MIDAS model to capture the dynamics of τ_t. Because this paper uses the same data set as Engle et al. (2013), it abstains itself from investigating the optimal number of lag years and applies the same number as used by the authors. Therefore, τ_t is determined using four MIDAS lag years.[1]

The code that is used to estimate the parameters of the GARCH-MIDAS models is inspired by the MATLAB files of Engle et al. (2013).[2] These files only contain functions for the GARCH-MIDAS model with fixed time span RV, which needed to be modified to suit the data and extensions better. The optimization for all other models has been programmed manually and their description can be found in Appendix B.

[1]In case of using the fixed window RV, four MIDAS lag years corresponds to $K = 16$. This number changes to $K = 1000$ when using the rolling window RV. For the macroeconomic variables, the number of lag years is similar to fixed window for the RV.

[2]Link to original code of Engle et al. (2013): `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/27513`

4. Alternative Component Models

4.1. GAS-MIDAS model

For years ARCH-type models have been put under a magnifying glass by researchers to see if any improvements were possible. Conversely, only a modest stream of academic literature focuses on an alteration within the GARCH-MIDAS framework. An example of such research is conducted by Conrad and Loch (2015), where the authors used an asymmetric mean-reverting GARCH(1,1) model to confirm that the long-term component behaves counter-cyclically. However, there are circumstances in which other model specifications are more suitable than ARCH-type models.

Consider for instance the Generalized Autoregressive Score (hereafter GAS) model by Creal et al. (2013). When the observation density is normal this model is similar to a GARCH(1,1) process, whereas the contrary is true when the this distribution is Student's t. The underlying idea is that observations that are large in magnitude may occur in the fat tails of the data's distribution. Because these values fall within the thickness of the distribution they should not cause the variance to increase to their full extent (Creal et al., 2013). The Student's t-distribution of the unit GAS model will follow this intuition by not letting the score of the underlying distribution change the volatility too drastically.[3] This might be more suitable for stock returns, because large and small returns occur more often than expected under a normal distribution (Glosten et al., 1993).

That is why this paper considers the GAS model within the component model framework. Particularly, let the time-varying parameter of interest be the short-term component g_{it} such that the GAS(1,1) model is given by:

$$g_{i+1,t} = c + A_1 s_{it} + B_1 g_{it}\,, \tag{13}$$

where s_{it} is a function of past data. A more traditional approach to express this function is through:

$$s_{it} = S_{it} \nabla_{it}\,, \tag{14}$$

$$\nabla_{it} = \frac{\partial \log p\,(\varepsilon_{it}|g_{it};\Phi)}{\partial g_{it}}\,, \tag{15}$$

$$S_{it} = \mathbb{E}_{i-1,t}\left[\nabla_{it}\nabla_{it}'\right]^{-1}\,, \tag{16}$$

where $\varepsilon_{it} \,|\, \mathcal{I}_{i-1,t} \sim p\,(\varepsilon_{it}|g_{it};\Phi)$, ∇_{it} represents the score and S_{it} is calculated using the inverse Fisher information matrix. In fact, this way of determining S_{it} allows for straightforward deriva-

[3] A mathematical derivation is given by Creal et al. (2013)

tions of different ARCH-type models (Creal et al., 2013).

The main aim of using the Student's t-distributed GAS(1,1) model (hereafter t-GAS(1,1)) in this paper is to compare its fit and predictive power with the t-GARCH(1,1) model in the GARCH-MIDAS framework. The t-GAS(1,1) model is created by assuming that ε_t are Student's t-distributed with ν degrees of freedom, such that the unit GARCH specification can be replaced by:

$$g_{i+1,t} = 1 + \alpha\left(1 + 3\nu^{-1}\right)\left[\frac{\nu^{-1} + 1}{\nu^{-1} + (1 - 2\nu^{-1})/\varepsilon_{it}^2} - 1\right] g_{it} + (\alpha + \beta)(g_{it} - 1) \ , \quad \nu > 2 \, , \tag{17}$$

with $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$. See Appendix A.3 for a full derivation of this model specification. The GAS-MIDAS model can be obtained by replacing equation (2) in the GARCH-MIDAS specification by equation (17). Note that the GAS model is a GARCH(1,1) process if ν goes to infinity, because in that case the Student's t-distribution converges to the normal distribution (Creal et al., 2013).

4.2. Asymmetric-MIDAS model

Another potential improvement of the GARCH-MIDAS model can be found in the MIDAS specification. There has been tremendous research on the asymmetric effects of positive and negative returns on stock market volatility; see Singleton and Wingender (1986) and Campbell and Hentschel (1992). Models that do not take these asymmetries into account often fail to capture the full market effects. For example, Engle et al. (2013) conclude that the GARCH-MIDAS model fails to take economic shifts into account. Therefore, it is the question if a symmetric MIDAS model is the best option to capture asymmetric effects of stock returns on volatility.

To capture asymmetric effects, Ghysels, Santa-Clara, and Valkanov (2005) introduced the asymmetric MIDAS (hereafter AMIDAS), which in this framework is given by:

$$\tau_t = \xi + \theta\left(\phi \sum_{k=1}^{K} \mathbb{1}_{t-k}^{+} \varphi_k(\omega_1, \omega_2) X_{t-k} + (1 - \phi) \sum_{k=1}^{K} \mathbb{1}_{t-k}^{-} \varphi_k(\omega_3, \omega_4) X_{t-k}\right), \tag{18}$$

where $\mathbb{1}_{t-k}^{+}$ denotes the indicator function for strictly positive underlying returns or levels, $\mathbb{1}_{t-k}^{-}$ for nonpositive values, and $\phi \in (0, 2)$ such that the weights sum up to one.

Note that the indicator function looks at the underlying values rather than those in X_t directly. The reason for this is that realized variance for daily returns and quarterly macroeconomic variance are strictly nonnegative, which implies that they cannot be separated based on different signs. This issue is solved by separating the values in X_t based on those used in their calculation. This means that the volatilities that are used to calculate the realized variance are grouped by positive and negative returns. This approach comes from Campbell and Hentschel (1992), who found that the volatility feedback-effect makes negative returns stronger in magnitude. Note that a similar approach is used for quarterly macroeconomic volatilities by using levels.

An advantage of the above specification is that it can be used in combination with GARCH and GAS models, irrespective of their underlying probability distribution. That is why these new specifications are called the GARCH-AMIDAS and GAS-AMIDAS models.

4.3. Estimation Procedure

The estimation procedure for the GAS-(A)MIDAS and GARCH-AMIDAS is similar to the original model of Engle et al. (2013). The only difference compared to the log-likelihoods of equations (11) and (12) is that the set of unknown parameters Θ is extended. More specifically, ϕ is added to determine the asymmetric effect of positive and negative returns on long-term volatilities, and two extra shape parameters of the additional Beta weighting function are added.

4.4. Forecast Evaluation

Finally, different forecast scenarios are considered to evaluate the predictive power of the models. As mentioned by Engle et al. (2013), the long-term component is predetermined at time t and the conditional expectation of the short-term component $\mathbb{E}_{t-1}(g_{it}) = 1 + (\alpha+\beta)^{i-1}(g_{1,t}-1)$ converges to one. This means that the volatility forecast for month t of the GARCH-MIDAS model is given by:

$$\begin{aligned} \mathbb{E}_{t-1}\left(\sum_{i=1}^{N_t} g_{it}\tau_t\varepsilon_{it}^2\right) &= \tau_t \sum_{i=1}^{N_t} \mathbb{E}_{t-1}(g_{it}) \\ &= \tau_t\left(N_t + \frac{1-(\alpha+\beta)^{N_t}}{1-\alpha-\beta}(g_{1,t}-1)\right), \end{aligned} \tag{19}$$

where $\varepsilon_{it} \mid \mathcal{I}_{i-1,t} \sim N(0,1)$. It should be noted that this equation does not hold for models that have a Student's t-distribution as underlying distribution due to a difference in variance. More specifically, if ε_{it} follows a Student's t-distribution with ν degrees of freedom, then the equation for volatility forecasts changes to:

$$\mathbb{E}_{t-1}\left(\sum_{i=1}^{N_t} g_{it}\tau_t\varepsilon_{it}^2\right) = \tau_t\left(\frac{\nu}{\nu-2}\right)\left(N_t + \frac{1-(\alpha+\beta)^{N_t}}{1-\alpha-\beta}(g_{1,t}-1)\right), \qquad \nu > 2. \tag{20}$$

Similar to Conrad and Loch (2015), it is assumed that the long-term volatility stays equal to its one-step ahead prediction; that is, $\hat{\tau}_{t+h|t-1} = \hat{\tau}_{t|t-1}$ for $h > 0$. For every volatility forecast a Mincer-Zarnowitz (hereafter MZ) regression is performed to evaluate the fit of the forecast to the actual volatility. The individual fits are also evaluated by means of the Root Mean Squared Error and Mean Absolute Error (hereafter RMSE and MAE, respectively). Both measures are included, because the RMSE puts more emphasis on large deviations, whereas the MAE takes the absolute distance of errors into account. This implies that they provide similar results in stable economic periods, while the RMSE is stricter in crises as it punishes large errors.

5. Data

This chapter starts by discussing the retrieval of the data in Section 5.1. Next, the descriptive statistics for the daily stock returns and macroeconomic variables are discussed in Sections 5.2 and 5.3, respectively.

5.1. Data Composition

The empirical quality of the models is assessed using U.S. daily stock returns and macroeconomic variables of the period 1885-2018. More specifically, the daily stock returns were used before by Schwert in his paper "Indexes of United States stock prices from 1802 to 1987" and are obtained from his online data library.[4] However, his data set merely covers the period February 16, 1885 to July 2, 1962 and does not provide a prevailing interpretation when used in the models (Engle et al., 2013). That is why this set is augmented by CRSP value-weighted returns to December 2018, such that forecasting evaluations can be conducted. The second set of variables includes monthly producer price index (PPI) inflation rate and industrial production (IP) growth rate. Both variables are obtained from the Federal Reserve Bank of St. Louis and cover the period January 1, 1920 to December 31, 2018. The period 1885-1919 has been omitted for the macroeconomic data, because it is not publicly accessible.

5.2. Daily Stock Returns

The descriptive statistics for daily stock returns can be found in Table 1. The data set of all variables is split into different sub-samples to account for potential structural breaks. According to Engle et

[4]Link to the online library: `http://schwert.ssb.rochester.edu/dstock.htm`

Table 1: Descriptive statistics for daily stock returns.

Sample	Mean	St. Dev.	Skewness	Kurtosis	Jarque-Bera	Observations
Daily Stock Returns (1890 - 2010)						
1884-1919	0.00027	0.00846	-0.324	9.993	0.000	10,355
1920-1952	0.00038	0.01315	0.287	16.242	0.000	9,727
1953-1984	0.00042	0.00763	0.016	7.230	0.000	8,043
1985-2010	0.00048	0.01096	-0.929	25.280	0.000	6,558
1953-2010	0.00045	0.00927	-0.680	23.964	0.000	14,601
1890-2010	0.00037	0.01041	-0.094	20.048	0.000	33,211
Daily Stock Returns (1890 - 2018)						
Sample	Mean	St. Dev.	Skewness	Kurtosis	Jarque-Bera	Observations
1985-2008	0.00044	0.01053	-1.128	30.146	0.000	6,054
2009-2018	0.00048	0.01060	-0.253	8.199	0.000	2,516
1890-2018	0.00037	0.01034	-0.109	19.712	0.000	35,223

Notes. Daily stock returns are obtained from Schwert's online library and CRSP. The columns represent the mean, standard deviation, skewness, kurtosis, Jarque-Bera p-value and the number of observations.

al. (2013), three sub-samples need to be formed to represent the effects before World War I (1884-1919), the Great Depression (1920-1952) and after World War II (1953-2010). Extra sub-samples have been created in the lower part of the same table to represent the sample forecasts. Table 1 shows that the descriptive statistics for the daily stock returns are very close to the values of Engle et al. (2013). They also reflect one of the key features in financial research, namely that stock returns are not normally distributed. In fact, they have excess kurtosis and skewness, which indicates that a probability distribution with fatter tails might be more appropriate to fit the underlying data. This observation is supported by the Jarque-Bera p-values which reject the normality hypothesis. A final remark on Table 1 is that most returns are negatively skewed; indicating that there were frequent small gains and a few large losses.

5.3. Macroeconomic Variables

While the daily stock returns do not need any transformations, the opposite holds for the macroeconomic variables. The reason for this is that they need to be scaled to the frequency that is used in the MIDAS filter. Engle et al. (2013) have shown that a quarterly frequency works well, because it ensures a good balance between covariance stability and fit. That is why the monthly leveled PPI and IP are transformed to quarterly rates by means of a geometric series using three consecutive months. These time series are shown in the first row of Figure 1 for PPI and IP, respectively. The corresponding descriptive statistics can be found in Table 2, where the separation of the sub-samples is similar to Table 1.

Table 2: Descriptive statistics for quarterly leveled macroeconomic data.

Sample	Variable	Mean	St. Dev.	Skewness	Kurtosis	Jarque-Bera	Observations
Macroeconomic Variables (1920 - 2010)							
1920-1952	PPI	0.00027	0.01373	-1.634	12.930	0.000	396
	IP	0.00339	0.02436	0.221	9.023	0.000	396
1953-1984	PPI	0.00332	0.00478	1.763	7.259	0.000	384
	IP	0.00277	0.00787	-0.916	4.638	0.000	384
1985-2010	PPI	0.00197	0.00731	-2.338	19.943	0.000	312
	IP	0.00173	0.00477	-1.769	8.610	0.000	312
1953-2010	PPI	0.00272	0.00608	-1.483	21.317	0.000	696
	IP	0.00230	0.00668	-1.013	5.832	0.000	696
1920-2010	PPI	0.00183	0.00965	-2.224	22.414	0.000	1,092
	IP	0.00270	0.01560	0.363	19.558	0.000	1,092
Macroeconomic Variables (1920 - 2018)							
Sample		Mean	St. Dev.	Skewness	Kurtosis	Jarque-Bera	Observations
1985-2008	PPI	0.00177	0.00740	-2.395	20.221	0.000	288
	IP	0.00176	0.00440	-1.858	10.148	0.000	288
2009-2018	PPI	0.00138	0.00679	-0.248	2.726	0.372	120
	IP	0.00143	0.00433	-1.516	7.202	0.000	120
1920-2018	PPI	0.00173	0.00946	-2.161	22.298	0.000	1,188
	IP	0.00260	0.01498	0.395	21.150	0.000	1,188

Notes. Macroeconomic variables are obtained from the FRED. The columns represent the variable name, mean, standard deviation, skewness, kurtosis, Jarque-Bera p-value and the number of observations. PPI stands for producer price index and IP for industrial production. The series used to calculate these measures are the quarterly growth rates, which are determined using geometric series.

In addition to these growth rates, volatilities are constructed to be used as input to the different models. The approach to create quarterly volatilties is taken from Schwert (1989b) and Engle et al. (2013). Specifically, the quarterly growth rates are regressed on four quarterly dummies and on four lagged quarterly growth rates, which produce the following regression:

$$X_t = \sum_{i=1}^{4} \alpha_j D_{jt} + \sum_{j=1}^{4} \beta_j X_{t-i} + u_t \,, \tag{21}$$

where the squared residuals $\hat{u}_t^2$ are used as an estimate for quarterly volatilities of macroeconomic variable X. The resulting volatilities are depicted in the bottom row of Figure 1 for PPI and IP, respectively.

The quarterly levels in Figure 1 depict the crises in the United States. For example, the growth rates during the 1920s are mostly negative, which corresponds to the aftermath of the Great Depression. Similarly, industrial production almost halved after the Wall Street Crash in 1929 compared to the years before. While most peaks in the time series correspond to economic crises, they are

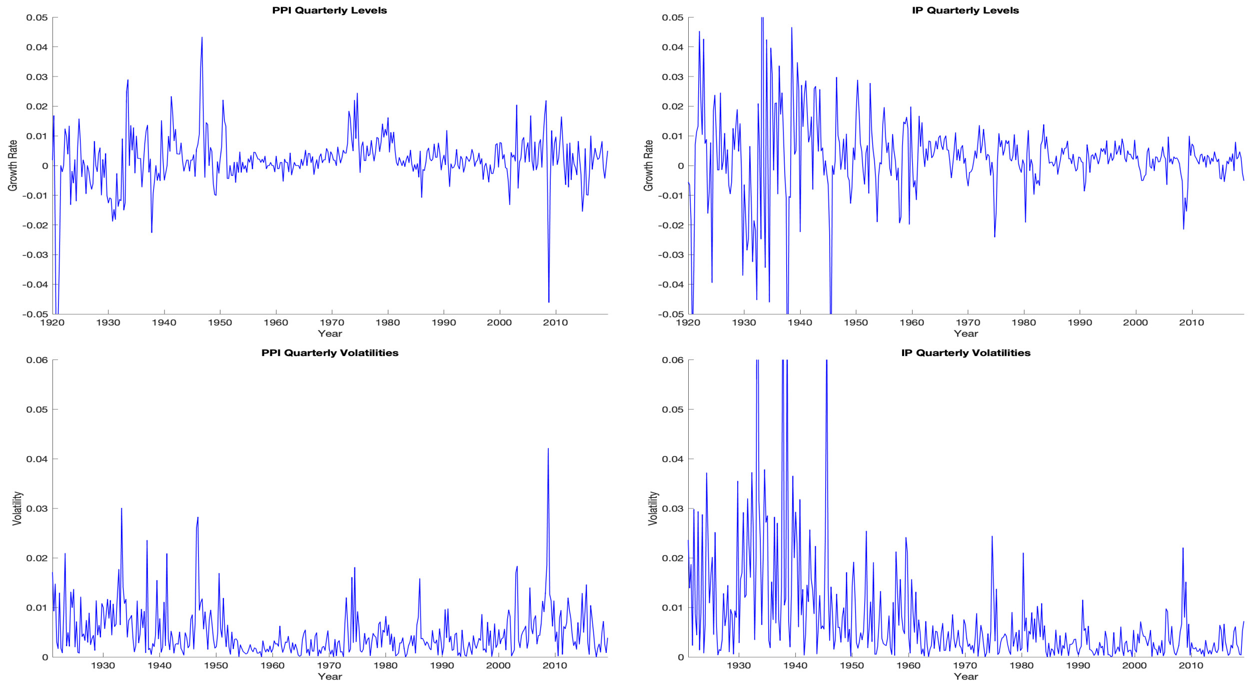

Figure 1: The quarterly levels and volatilties for PPI and IP for the period 1920-2018

not of the same magnitude for both time series. For instance, until the 1980s industrial production decreased hardest by crises. The reason behind this is not only economic crises, but also the arms race during the Cold War between the United States and Soviet Union. After this period the producer price index took most of the negative effects with its minimum during the latest economic crisis of 2008.

6. Empirical Results

Section 6.1 discusses the estimates of the GARCH-MIDAS model with realized variance and then turns to the macroeconomic variables in Section 6.2. Next, the estimates of the GARCH-MIDAS model with both PPI and IP are analyzed in Section 6.3. The economic contribution of the aforementioned models are compared in Section 6.4. Finally, Section 6.5 discusses the results of the new component models and analyzes the forecasts of all component models.

6.1. GARCH-MIDAS with Realized Variance

The estimates of the GARCH-MIDAS model with realized variance can be found in Table 3. The estimates are created by optimizing the log-likelihood function in equation (11), where ω_1 has been set to one to ensure that the weights are decaying over their domain. The parameter ω_2 in

Table 3: Parameter estimates for GARCH-MIDAS with realized variance.

Fixed Realized Variance								
Sample	μ	α	β	ϑ	ω	ξ	LLF	BIC
1890-2010	0.00063	0.10985	0.85782	0.00894	4.48675	0.00003	113,551.4	-6.7802
	(16.97)	(28.92)	(168.11)	(25.89)	(154.54)	(25.75)		
1890-1919	0.00053	0.15438	0.78515	0.00449	34.48643	0.00005	31,788.2	-6.9362
	(7.77)	(16.71)	(59.15)	(11.56)	(2.13)	(23.86)		
1920-1952	0.00075	0.10472	0.85551	0.00929	5.41706	0.00003	31,267.1	-6.4233
	(9.02)	(5.66)	(40.34)	(15.56)	(1.20)	(3.37)		
1953-1984	0.00060	0.09322	0.88826	0.00832	5.16332	0.00003	28,697.7	-7.1294
	(2.33)	(12.71)	(84.35)	(1.95)	(6.21)	(7.03)		
1985-2010	0.00074	0.09743	0.87218	0.00702	9.64940	0.00005	21,857.0	-6.6577
	(7.85)	(12.24)	(81.61)	(11.67)	(35.00)	(12.77)		
1953-2010	0.00064	0.09363	0.88562	0.00872	4.36340	0.00004	50,543.2	-6.9193
	(11.66)	(19.38)	(190.43)	(12.66)	(4.90)	(11.22)		
Rolling Realized Variance								
Sample	μ	α	β	ϑ	ω	ξ	LLF	BIC
1890-2010	0.00063	0.11397	0.84306	0.01043	9.23857	0.00003	114,023.3	-6.7830
	(6.18)	(4.25)	(20.63)	(8.94)	(29.53)	(8.43)		
1890-1919	0.00051	0.15707	0.76180	0.00765	31.67387	0.00004	32,257.8	-6.9440
	(4.42)	(28.16)	(52.05)	(12.97)	(37.93)	(12.10)		
1920-1952	0.00075	0.10504	0.85342	0.01109	5.26680	0.00003	31,639.6	-6.4226
	(8.87)	(20.35)	(120.34)	(28.26)	(39.58)	(10.18)		
1953-1984	0.00063	0.10036	0.86943	0.01130	9.85999	0.00002	28,880.4	-7.1349
	(8.48)	(1.99)	(15.90)	(2.07)	(10.61)	(7.73)		
1985-2010	0.00075	0.10432	0.84576	0.00846	21.20215	0.00004	21,640.5	-6.6526
	(7.17)	(14.10)	(37.78)	(8.62)	(1.39)	(5.79)		
1953-2010	0.00067	0.10016	0.86505	0.00981	13.0180	0.00003	50,720.7	-6.9223
	(6.38)	(8.37)	(41.17)	(21.06)	(12.84)	(13.94)		

Notes. The GARCH-MIDAS estimates are created using four MIDAS lag years. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

the same equation is for this purpose renamed to ω in the current subsection. Almost all estimates are significant, which is in line with the results of Engle et al. (2013). The parameter of interest from Table 3 is ϑ, which represents the influence of the realized variance on the long-term volatility. The parameter is statistically different from zero in almost all sub-periods, but remains close to zero. This allows ω to take on any value on its domain, which means that the parameter is unidentified. The idea that this parameter is unidentified is intensified by noticing that almost all values of ω are different from Engle et al. (2013).

Another way of recognizing this identification issue is through the optimization procedure. The estimates in the log-likelihood function are extremely sensitive to their starting values. For instance,

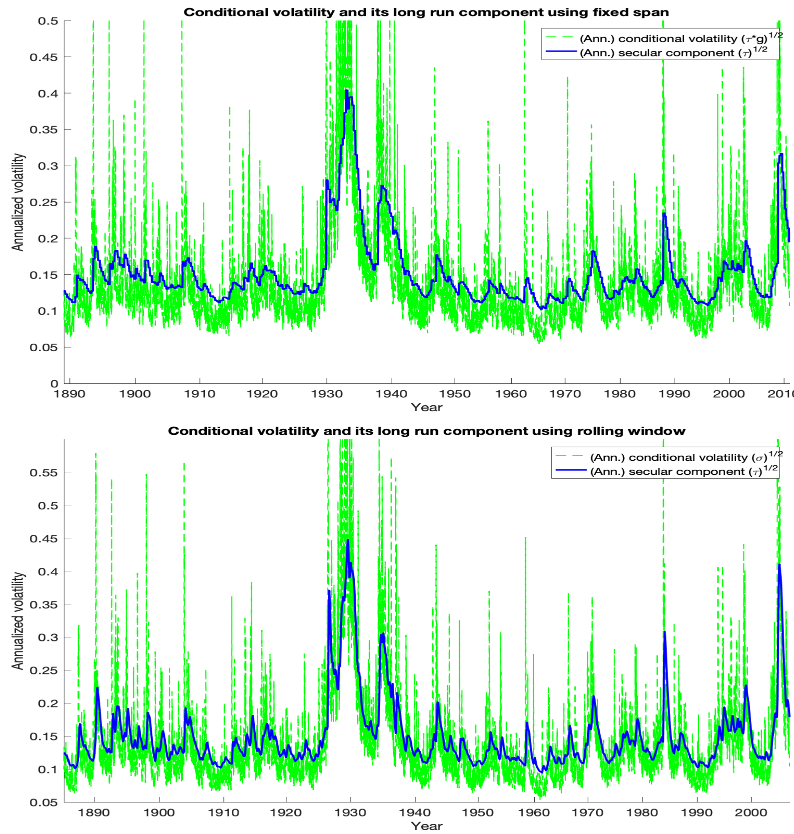

Figure 2: Conditional volatility for the GARCH-MIDAS model using realized volatility.

changing the starting value of ϑ can result in a completely different set of parameters, whereas the same does not hold for μ. This issue can be solved by adapting the procedure of Andrews and Cheng (2012) which constructs robust critical values using simulations. This paper does not attempt to solve these problems and merely mention its potential existence.

Next, a brief look is taken at the estimates of the GARCH(1,1) process. By construction, the sum of α and β converges to one. Yet, as noted by Engle et al. (2013), none of the sums exceed the maximum of 0.9815 in the period 1953-1984. When using the rolling window specification to determine the realized variance, it seems that most values lay around 0.9685. The reason why these sums are not close to one has to do with the long-term volatility component of the model. The unconditional variance of the GARCH-MIDAS model is equal to τ_t due to the fact that the unconditional expectation of g_{it} is equal to one. That is, the long-term component is predetermined and dominates the persisting variance of the model as a whole. This implies that the GARCH(1,1) process values persistence less, which results in lower values of the coefficients that control for this effect (Bollerslev, 1986).

Alternatively, the long-term component can be expressed as the log of τ_t. The estimates for this GARCH-MIDAS model can be found in Table 10 in Appendix C. Again, the values are similar to those of Engle et al. (2013) and follow similar a trend in the t-statistics. It is also here that changing the starting values for the optimization results in drastical shifts in the t-statistics, which indicates that there are possible identification issues in ϑ and ω.

Overall, the results for the GARCH-MIDAS model with realized variance seem to be aligned with those of Engle et al. (2013). The parameter estimates significantly differ across periods, which indicates that there are economic shifts throughout the sample. This thought is confirmed by the time series in Figure 2, where the conditional volatility shifts upwards after its peak in the late 1920s due to the Great Depression. Another shift occurs after the financial crisis of 2008, which corresponds to the idea that the conditional volatility is influenced by economic conditions. For instance, the trust of consumers in financial products decreases when an economic crisis arises. This may lead to less confidence in the market and more variance in their pricing. Also, several institutions invest in those markets that are hit hardest by economic crises. An example is the housing market, which crashed during the last crisis and let to enormous losses in the financial sector.

6.2. GARCH-MIDAS with Single Macroeconomic Series

The parameter estimates of ϑ_l in the GARCH-MIDAS model with PPI levels are shown in Table 4. The values range between 0.2537 and 1.1026, and are similar to those of Engle et al. (2013). In fact, they are all positive and significantly different from zero, which implies that more inflation leads - ceteris paribus - to more stock market volatility. During the period of the Great Depression, the effect of a change in inflation is captured by $\vartheta_l = 0.2537$. Simultaneously, the Beta weighting function in equation (5) assigns a weight of 0.3588 on the first lag (corresponding to $\omega_1 = 10.68$). This means that a one percent increase in inflation during this quarter results in a 3.1% increase in stock market volatility next quarter (that is: $e^{0.2537*0.3588/3} - 1 \approx 0.0308$). It can be shown in a similar way that for other periods a one percent increase in inflation corresponds to an increase of 2.3% during 1953-1984, 0.5% during 1953-2010 and has no statistical impact during 1985-2010.

Next, the parameter estimates of the leveled IP are investigated. The lower part of Table 4 shows that the values of ϑ_l range between -0.2414 and -0.9521, such that an increase in industrial production - ceteris paribus - decreases the stock market volatility. This counter-cyclical business effect is also observed by Schwert (1989b) and Engle et al. (2013). Specifically, a one percent increase in industrial production has the biggest impact during 1984-2010 when there was a 0.1%

Table 4: Parameter estimates for GARCH-MIDAS with level PPI and IP.

Sample	μ	α	β	ϑ_l	ω_1	ω_2	ξ	LLF	BIC
Producer Price Index									
1924-2010	0.00066	0.08795	0.90613	0.28910	10.68211	14.12968	-8.95922	77,649.0	-6.7108
	(2.48)	(2.10)	(21.21)	(2.65)	(6.92)	(6.65)	(-16.51)		
1924-1952	0.00077	0.09685	0.89465	0.25370	18.18765	2.65045	-8.73977	27,134.4	-6.3547
	(3.68)	(16.24)	(141.33)	(2.48)	(3.80)	(3.52)	(-46.34)		
1953-1984	0.00059	0.08851	0.89243	1.10259	7.26627	2.78565	-10.05580	28,709.4	-7.1311
	(2.50)	(12.09)	(93.62)	(7.36)	(2.40)	(2.71)	(-83.49)		
1985-2010	0.00072	0.08957	0.89807	0.87191	7.55557	13.70510	-9.2728	21,848.8	-6.6639
	(9.82)	(14.44)	(128.11)	(2.34)	(0.88)	(0.93)	(-51.03)		
1953-2010	0.00063	0.08506	0.90485	0.76829	18.10860	5.17096	-9.51986	50,536.4	-6.9177
	(3.10)	(18.07)	(168.56)	(3.45)	(1.97)	(2.78)	(-66.54)		
Industrial Production									
Sample	μ	α	β	ϑ_l	ω_1	ω_2	ξ	LLF	BIC
1924-2010	0.00067	0.08790	0.90533	-0.24139	11.93550	7.87982	-8.82542	77,649.9	-6.7109
	(9.19)	(1.63)	(16.86)	(-2.99)	(2.44)	(2.70)	(-14.69)		
1924-1952	0.00077	0.09808	0.89171	-0.36158	3.22549	2.85058	-8.54069	27,135.5	-6.3549
	(11.03)	(15.33)	(130.06)	(-3.67)	(2.46)	(2.99)	(-48.92)		
1953-1984	0.00060	0.08700	0.90032	-0.95210	5.19501	3.74770	-9.37186	28,706.9	-7.1305
	(2.11)	(10.92)	(115.70)	(-4.42)	(1.21)	(0.91)	(-57.18)		
1985-2010	0.00073	0.09121	0.89400	-0.84688	8.18203	2.42293	-8.98245	21,849.0	-6.6539
	(2.21)	(11.39)	(105.52	(-3.30)	(2.09)	(2.06)	(-47.09)		
1953-2010	0.00064	0.08640	0.90257	-0.94268	4.82288	2.98514	-9.12252	50,539.6	-6.9182
	(2.47)	(17.53)	(166.17)	(-5.15)	(4.94)	(4.73)	(-73.19)		

Notes. The GARCH-MIDAS estimates are created using four MIDAS lag years. That is why the first sample starts in 1924 instead of 1920. The parameter ϑ_v is multiplied by 10^{-2} to represent percentages. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

decrease in stock market volatility. The lowest quarterly decrease can be found during 1953-1984, which is mostly due to the end of World War II. Once the weapons were down, the global economy started to expand by having more trade agreements and unions that were not possible during the war. This might explain why an increase in industrial production has the least impact on stock market volatility during that period; given all the other macroeconomic expansions.

However, leveled PPI and IP are not the only variables that influence stock market volatility. The uncertainty of investors about these values can have a significant impact on its own. That is why this uncertainty is represented by macroeconomic variances, which are put in the GARCH-MIDAS model to determine the long-term volatility. Table 5 shows the estimates of this model with PPI and IP variances. A notable difference between these estimates with PPI and those in Table 4 is that they are smaller in magnitude. Specifically, the minimum value it can take on is 0.0413 whereas the maximum is 0.1498. It should be noted that both values are not significantly different from

Table 5: Parameter estimates for GARCH-MIDAS with variance PPI and IP.

				Producer Price Index					
Sample	μ	α	β	ϑ_v	ω_1	ω_2	ξ	LLF	BIC
1924-2010	0.00066	0.09191	0.89820	0.11642	0.68187	0.36456	-9.60176	76,554.4	-6.7037
	(2.10)	(24.96)	(222.26)	(6.86)	(1.42)	(1.09)	(-71.73)		
1924-1952	0.00074	0.09875	0.89322	0.05907	1.66980	0.50339	-9.02307	26,019.0	-6.3168
	(7.47)	(13.37)	(117.62)	(15.22)	(18.05)	(39.82)	(-43.14)		
1953-1984	0.00059	0.08482	0.90251	0.14981	8.42631	2.02360	-10.04765	28,991.0	-7.1286
	(8.74)	(1.49)	(41.20)	(0.67)	(0.58)	(2.95)	(-9.95)		
1985-2010	0.00073	0.09011	0.89880	0.04132	24.36358	215.81986	-9.22181	21,849.5	-6.6541
	(7.83)	(13.46)	(122.77)	(1.11)	(0.24)	(0.21)	(-49.84)		
1953-2010	0.00064	0.08699	0.90046	0.09703	5.61085	1.46918	-9.74143	50,535.8	-6.9176
	(2.63)	(19.73)	(173.63)	(9.05)	(2.92)	(2.23)	(-80.60)		
				Industrial Production					
Sample	μ	α	β	ϑ_v	ω_1	ω_2	ξ	LLF	BIC
1924-2010	0.00066	0.09115	0.89886	0.48676	5.23299	1.30026	-9.46134	76,557.1	-6.7039
	(2.08)	(26.51)	(125.90)	(4.31)	(0.29)	(0.34)	(-39.84)		
1924-1952	0.00075	0.10265	0.88377	0.05804	1.05914	0.48863	-9.65860	26,035.6	-6.3209
	(9.73)	(13.10)	(101.46)	(10.19)	(8.23)	(7.61)	(-67.20)		
1953-1984	0.00060	0.08423	0.90500	0.06962	2.72020	0.74399	-10.00460	28,701.1	-7.1291
	(2.74)	(15.48)	(88.91)	(12.13)	(2.77)	(3.28)	(-40.65)		
1985-2010	0.00072	0.09099	0.89579	0.04800	221.16644	14.85785	-9.24479	21,850.5	-6.6544
	(0.97)	(0.63)	(6.86)	(3.35)	(1.21)	(2.56)	(-5.23)		
1953-2010	0.00064	0.08378	0.90786	0.02767	277.42739	165.32762	-9.39829	50,535.1	-6.9175
	(13.45)	(19.46)	(179.13)	(3.68)	(6.81)	(5.99)	(-61.09)		

Notes. The GARCH-MIDAS estimates are created using four MIDAS lag years. That is why the first sample starts in 1924 instead of 1920. The parameter ϑ_v is multiplied by 10^{-4} to represent percentages. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

zero, such that the range becomes slightly smaller. Engle et al. (2013) only found a significant result during World War I and after World War II, which are clearly different from these results.

This difference can be explained through the optimization procedure. As explained in Section 3.4, Engle et al. (2013) used a standard minimization procedure in MATLAB to find the MLE estimators. However, changing the starting values slightly results in completely different MLE estimators. It is plausible that the results of Engle et al. (2013) are found in local minima. That is why this paper uses Simulated Annealing to find proper starting values, before plugging those into the minimization algorithm.[5] The overall sign of the parameter estimates stay the same, such that the conclusion are similar to those drawn from the estimates in Table 4.

Most interesting are the estimates for ϑ_v in the lower part of Table 5. Originally, an increase in industrial production during the current quarter caused the stock market volatility to decrease next

[5] The Simulated Annealing algorithm that is used comes from the Global Optimization Toolbox in MATLAB

Table 6: Parameter estimates for GARCH-MIDAS with level and variance PPI and IP.

Sample	μ	α	β	ϑ_l	$\omega_{1,l}$	$\omega_{2,l}$	ϑ_v	$\omega_{1,v}$	$\omega_{2,v}$	ξ	LLF	BIC
Producer Price Index												
1924-2010	0.00067	0.09222	0.89755	0.07312	17.52337	109.18135	0.12534	0.64614	0.34566	-9.66550	76,560.0	-6.7029
	(2.36)	(25.66)	(229.15)	(2.90)	(1.79)	(1.73)	(9.95)	(3.09)	(2.05)	(-75.15)		
1924-1952	0.00074	0.09872	0.89294	0.08002	32.42805	209.18051	0.08193	1.30747	0.51090	-9.11106	26,021.8	-6.3142
	(8.86)	(18.23)	(163.60)	(10.10)	(5.02)	(5.59)	(3.65)	(4.48)	(3.86)	(-38.79)		
1953-1984	0.00059	0.08858	0.89249	0.98833	11.84099	3.51226	0.04223	61.02317	44.86674	-10.14655	28,710.6	-7.1281
	(9.09)	(12.22)	(96.84)	(6.35)	(2.11)	(3.41)	(1.91)	(1.19)	(1.16)	(-80.76)		
1985-2010	0.00073	0.09138	0.89362	0.56866	13.29352	24.20311	0.04141	9.99200	1.48536	-9.46899	21,851.0	-6.6505
	(2.51)	(14.54)	(133.98)	(0.98)	(0.41)	(0.40)	(1.62)	(2.52)	(2.10)	(-45.26)		
1953-2010	0.00064	0.08672	0.90074	0.45978	43.85341	10.17263	0.12310	0.19082	0.09843	-9.95402	50,548.3	-6.9174
	(4.34)	(1.14)	(12.38)	(2.61)	(3.10)	(3.33)	(8.26)	(0.89)	(0.76)	(-12.36)		
Industrial Production												
Sample	μ	α	β	ϑ_l	$\omega_{1,l}$	$\omega_{2,l}$	ϑ_v	$\omega_{1,v}$	$\omega_{2,v}$	ξ	LLF	BIC
1924-2010	0.00066	0.08902	0.90339	-0.00097	39.27673	161.54955	0.02580	126.17347	70.54283	-9.12611	76,549.4	-6.7019
	(2.08)	(24.63)	(245.44)	(-0.30)	(2.06)	(24.86)	(4.02)	(1.67)	(1.75)	(-66.36)		
1924-1952	0.00077	0.10416	0.87738	-0.35690	2.49065	1.02531	0.05771	0.53272	0.34680	-9.60657	26,045.7	-6.3200
	(7.27)	(4.72)	(62.82)	(-4.40)	(2.56)	(1.95)	(3.87)	(1.95)	(0.35)	(-8.00)		
1953-1984	0.00062	0.08893	0.89640	-0.11307	4.26028	3.22851	0.04301	82.89529	206.26825	-9.61767	28,716.4	-7.1295
	(9.26)	(15.28)	(133.59)	(-13.06)	(35.44)	(19.22)	(11.45)	(12.09)	(13.50)	(-88.81)		
1985-2010	0.00073	0.09094	0.89435	-0.77522	8.66631	2.52981	0.00747	224.49162	53.50549	-9.01932	21,849.0	-6.6499
	(7.94)	(13.48)	(113.59)	(-1.69)	(0.72)	(1.47)	(1.14)	(4.30)	(0.98)	(-40.10)		
1953-2010	0.00064	0.08639	0.90259	-0.09432	4.82363	2.98611	0.00277	5.19154	12.18917	-9.12360	50,539.6	-6.9162
	(2.05)	(5.6831)	(40.02)	(-0.92)	(3.08)	(0.67)	(0.02)	(7.15)	(1.61)	(-44.28)		

Notes. The GARCH-MIDAS estimates are created using four MIDAS lag years. That is why the first sample starts in 1924 instead of 1920. The parameters ϑ_l and ϑ_v are multiplied by 10^{-2} and 10^{-4} respectively to represent percentages. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

quarter. This conclusion changes when using IP volatilities, because then an increase in industrial production has a positive impact on stock market volatility. Similar observations can be found in related research, where it is concluded that the ”business cycle uncertainty matters” (Engle et al., 2013, p. 787). All the estimates are statistically different from zero, such that the impact of industrial production is not trivial.

6.3. GARCH-MIDAS with Both Macroeconomic Series

The parameter estimates for the model with both levels and volatilities appear in Table 6. It seems that nearly all estimates are similar to their equivalents in the single series. The biggest difference when using PPI is that ϑ_l is not bigger than one, such that the overall impact of inflation is slightly less than before. The estimates for IP have also decreased in magnitude. A reason for this might be that due to the inclusion of three additional parameters in the model, the overall standard errors have increased. This can occur when the dimension of the models increases, which causes the available data to be spares for the parameters to get a good fit. One way of expanding the amount of data is by using higher-frequency returns; that is, intra-day stock returns. In fact, the parameters that are estimated for the period 1924-2010 support this thought, because they are overall more significant than for other periods.

6.4. Analyzing the Economic Sources

One way of understanding the impact of economic variables on stock market volatility is by analyzing their variance ratios. These ratios represent how much a variable contributed to the total variation in that particular component model (Engle et al., 2013). This measure is calculated by dividing the variance of the log long-term component by the log of the total variance; that is, $\text{Var}(\log \tau_t)/\text{Var}(\log \tau_t g_t)$. Table 7 contains the variance ratios for different GARCH-MIDAS models.

Overall, the model with rolling window RV has the highest contribution with more than 47% during the Great Depression. This number is closely followed by the fixed span RV specification, which explains roughly 40% of the variation in quarterly volatility. The most redundant contribution of merely 2% during the same era is of the GARCH-MIDAS model with PPI level. This implies that the long-term component almost does not fluctuate over time, such that the model merely becomes the GARCH(1,1) model. Particularly, the ratios for the whole sample show that the capacity of economic contribution is 17% when including macroeconomic variables.

When both level and variance of the macroeconomic series are included, the economic contribu-

Table 7: Variance Ratios for different GARCH-MIDAS models.

			Daily Stock Returns			
Model	1890-2010	1890-1919	1920-1952	1953-1984	1985-2010	1953-2010
Fixed RV	37.26	12.88	51.12	12.22	27.09	24.70
Fixed RV (log)	27.76	11.03	45.40	9.54	12.46	11.65
Rolling window RV	45.21	24.70	53.69	30.28	47.08	39.05
Rolling window RV (log)	30.80	18.07	45.23	19.53	13.88	11.30
			Macroeconomic Variables			
Model	1924-2010	1890-1919	1924-1952	1953-1984	1985-2010	1953-2010
PPI level	2.71	-	4.14	35.83	7.32	14.08
PPI variance	11.17	-	2.19	13.20	2.39	12.14
PPI level + variance	12.33	-	3.28	36.71	10.39	17.55
IP level	4.36	-	9.39	19.23	10.03	12.47
IP variance	16.66	-	15.31	6.69	3.71	2.73
IP level + variance	5.14	-	24.72	29.02	9.66	12.48

Notes. The variance ratios are calculated by the formula: 100*Var(log τ_t)/Var(log $\tau_t g_t$). The column 1890-1919 for the macroeconomic variables is empty, because the data for this period is unavailable (see Section 5.3).

tion of the long-run component increases and reaches its height. Specifically, around 37% and 29% of the total quarterly variation is explained for PPI and IP, respectively. It is possible that similar results are obtained during 1890-1919, but this cannot be investigated due to the unavailable data.

6.5. Comparison with Alternative Models

This section compares the parameter estimates of the GARCH-MIDAS model for the full sample to those of the new component models. The GARCH-MIDAS model with Student's t-distribution is addressed to as the GARCH-MIDAS-t model. The names of the GARCH-AMIDAS and GAS-MIDAS model do not change due to their uniqueness in this framework. Table 8 contains the estimates for the different models with leveled macroeconomic series. The other tables are put in Appendix C.

It seems that the GARCH-AMIDAS model performs best for both PPI and IP when comparing the log-likelihoods and BICs across models. The reason behind this is that the threshold parameter ϕ captures the effects of positive and negative values on the long-term volatility. According to the values of this parameter, the long-term component is mostly influenced by positive values of PPI, whereas the opposite is true for IP. This result is in line with the results of Section 6.2, where it was stated that an increase in industrial production decreases the stock market volatility. Another interesting result is that value of ϑ_l is higher for all new component models compared to the traditional GARCH-MIDAS model. For example, the latter claims that a one percent increase in inflation has no

Table 8: Parameter estimates for different component models with levels PPI and IP for the period 1920-2010.

Producer Price Index												
Model	μ	α	β	ϑ_l	$\omega_{1,+}$	$\omega_{2,+}$	$\omega_{1,-}$	$\omega_{2,-}$	φ	ξ	ν	LLF/BIC
GARCH	0.00066	0.08795	0.90613	0.28910	10.68211	14.12968				-8.95922		77,649.0
	(2.48)	(2.10)	(21.21)	(2.65)	(6.92)	(6.65)				(-16.51)		/ -6.7108
GARCH-t	0.00079	0.05404	0.91312	0.39591	18.46374	3.93741				-11.17661	5.89523	78,376.0
	(2.10)	(19.10)	(212.38)	(1.95)	(3.32)	(7.14)				(-102.63)	(24.30)	/ -6.7732
GAS-t	0.00079	0.05597	0.91692	0.36378	18.57484	4.11736				-11.22148	6.42435	78,337.8
	(1.57)	(20.34)	(229.92)	(3.13)	(9.24)	(9.98)				(-125.77)	(27.72)	/ -6.7699
AMIDAS	0.00067	0.09276	0.90198	0.70305	1.47380	1.09947	98.41831	18.91826	1.54249	-9.04403		77,656.4
	(14.52)	(27.35)	(246.16)	(26.15)	(3.81)	(4.58)	(35.38)	(13.97)	(13.64)	(-118.12)		/ -6.7101
Industrial Production												
Model	μ	α	β	ϑ_l	$\omega_{1,+}$	$\omega_{2,+}$	$\omega_{1,-}$	$\omega_{2,-}$	φ	ξ	ν	LLF/BIC
GARCH	0.00067	0.08790	0.90533	-0.24139	11.93550	7.87982				-8.82542		77,649.9
	(9.19)	(1.63)	(16.86)	(-2.99)	(2.44)	(2.70)				(-14.69)		/ -6.7109
GARCH-t	0.00079	0.05371	0.91293	-0.57120	3.70301	2.22828				-10.87523	5.83800	78,376.8
	(2.34)	(14.94)	(234.20)	(-16.64)	(3.89)	(3.60)				(-116.59)	(19.52)	/ -6.7733
GAS-t	0.00079	0.05587	0.91735	-0.13214	9.16303	6.55967				-11.09712	6.41489	78,333.6
	(2.13)	(21.31)	(161.97)	(-1.42)	(2.04)	(1.78)				(-58.40)	(26.77)	/ -6.7696
AMIDAS	0.00068	0.09193	0.89512	-0.77697	16.00774	9.94227	2.02315	1.49167	0.26282	-9.46603		77,678.5
	(13.57)	(26.07)	(239.44)	(-10.05)	(7.61)	(7.61)	(5.74)	(7.25)	(3.75)	(-91.17)		/ -6.7121

Notes. The first column contains the model specifcation GARCH-MIDAS with normal distribution (GARCH) and with Student's t-distribution (GARCH-t), GAS-MIDAS with Student's t-distribution (GAS-t) and the GARCH-AMIDAS model with normal distribution (AMIDAS). The estimates are created using a four MIDAS lag years. The parameter ϑ_v is multiplied by 10^{-2} to represent percentages. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

effect on the stock market volatility during the next quarter. The estimates of the GARCH-MIDAS-t model, on the other hand, suggest that this causes a 1.8% increase in volatility. The GAS-MIDAS and GARCH-AMIDAS follow this direction by a 1.4% and 1.9% increase, respectively. The difference in effects might be due to different underlying distributions of the models and the extra shape parameters. Specifically, the shape parameter ν is for both the GAS(1,1) and GARCH(1,1) significantly different from zero. This confirms the idea of Section 5.2 that the returns have the Student's t-distribution as underlying distribution due to their high peak and fat tails.

Another way to compare the component models is to assess their predictive power using an out-of-sample forecasting exercise. The data has been split into two subsets: (i) a training set over the period 1985-2008, and (ii) a testing set over 2009-2018. The parameters are estimated using the training set and then fixed, whereafter they are used during the testing sample. There are three forecast horizons of interest: 1-, 2- and 4-months ahead. These horizons ensure that models are not only compared based on relatively small errors from the short-horizon, but also on long-run errors.

The estimates of the MZ-regression, RMSE and MAE are shown in Table 9. The values in bold represent the lowest errors for a certain forecast horizon given the model specification. Similar to the results in Section 6.2, it appears that the errors of the models with PPI and IP are smaller than those with realized variance. They also show that the GARCH-MIDAS and GARCH-AMIDAS specifications perform the best for 1-month ahead forecasts. More specifically, the former model works well with realized variance, whereas the latter outperforms the other models with the macroeconomic variables. When the forecast horizon increases, the GARCH-MIDAS-t model seems to perform better when using realized variance. An evident reason for this result is that the Student's t-distribution takes observations in its fatter tail into account. This means that it will not let the long-term component increase drastically for large changes in returns. However, the GAS-MIDAS model performs best for the 4-months ahead forecasts in almost all cases. This is due to the mathematical construction of the formula, which ensures that the score of the model is not influenced heavily by large changes in returns.

These results are validated by the estimates of the MZ-regression. None of the models seem to create accurate volatilities for the 1- and 2-month ahead forecasts. This is different for the 4-month ahead forecasts where most models seem to fit the MZ-regression well. The AMIDAS specification again beats the other models based on the joint hypothesis that $c = 0$ and $\psi = 1$, which implies that the forecasts represent the actual values well.

Table 9: Forecast estimates for different component models for the period 2009-2018.

Forecast horizon		1-month ahead					2-month ahead					4-month ahead				
		c	ψ	p	RMSE	MAE	c	ψ	p	RMSE	MAE	c	ψ	p	RMSE	MAE
GARCH	(fixed RV)	-0.0015	1.0405	0.0000	**0.0022**	**0.0017**	-0.0019	1.1140	0.0000	0.0023	0.0018	-0.0021	1.1607	0.0097	0.0029	0.0022
GARCH-t	(fixed RV)	-0.0013	1.0482	0.0000	0.0024	0.0018	-0.0020	1.1502	0.0000	**0.0021**	**0.0016**	-0.0022	1.1983	0.0055	0.0028	0.0021
GAS-t	(fixed RV)	-0.0008	0.9829	0.0000	0.0026	0.0020	-0.0015	1.0812	0.0000	0.0023	0.0017	-0.0018	1.1388	0.0165	**0.0027**	**0.0021**
AMIDAS	(fixed RV)	-0.0022	1.1087	0.0000	0.0023	0.0018	-0.0027	1.1930	0.0000	0.0023	0.0018	-0.0027	1.2237	0.0014	0.0030	0.0022
GARCH	(roll. RV)	-0.0008	1.0039	0.0000	**0.0020**	**0.0015**	-0.0013	1.0765	0.0000	0.0019	0.0015	-0.0011	1.0584	0.1182	0.0034	0.0024
GARCH-t	(roll. RV)	0.0000	0.9259	0.0000	0.0024	0.0017	-0.0007	1.0048	0.0000	**0.0019**	**0.0014**	-0.0005	0.9938	0.1276	0.0034	0.0024
GAS-t	(roll. RV)	0.0010	0.8175	0.0000	0.0034	0.0021	0.0001	0.9166	0.0000	0.0022	0.0016	0.0000	0.9311	0.0088	**0.0033**	**0.0022**
AMIDAS	(roll. RV)	-0.0011	1.0203	0.0000	0.0025	0.0018	-0.0017	1.0990	0.0000	0.0022	0.0017	-0.0015	1.0911	0.0133	0.0033	0.0024
GARCH	(level PPI)	-0.0015	1.0744	0.0000	0.0021	0.0016	-0.0021	1.1667	0.0000	0.0021	0.0016	-0.0023	0.0056	0.0000	0.0029	0.0021
GARCH-t	(level PPI)	-0.0014	1.0704	0.0005	0.0023	0.0018	-0.0021	1.1665	0.0000	0.0022	0.0011	-0.0023	1.2167	0.0071	0.0029	0.0021
GAS-t	(level PPI)	-0.0009	1.0252	0.0203	0.0028	0.0020	-0.0018	1.1469	0.0018	0.0024	0.0017	-0.0022	1.2104	0.0071	**0.0028**	**0.0021**
AMIDAS	(level PPI)	-0.0012	1.0601	0.0002	**0.0019**	**0.0015**	-0.0016	1.1362	0.0016	**0.0021**	**0.0015**	-0.0017	1.1771	0.0504	0.0030	0.0022
GARCH	(level IP)	-0.0013	1.0264	0.0000	0.0022	0.0017	-0.0019	1.1101	0.0000	0.0022	0.0017	-0.0019	1.1459	0.0169	0.0029	0.0022
GARCH-t	(level IP)	-0.0013	1.0397	0.0001	0.0024	0.0019	-0.0020	1.1408	0.0000	0.0021	0.0017	-0.0022	1.1873	0.0064	0.0028	0.0021
GAS-t	(level IP)	-0.0008	0.9868	0.0037	0.0030	0.0021	-0.0019	1.1264	0.0004	0.0024	0.0017	-0.0024	1.1995	0.0000	**0.0027**	**0.0020**
AMIDAS	(level IP)	-0.0011	1.0043	0.0000	**0.0021**	**0.0017**	-0.0017	1.0927	0.0000	**0.0021**	**0.0016**	-0.0019	1.1488	0.0213	0.0029	0.0022
GARCH	(var. PPI)	-0.0013	1.0266	0.0000	0.0023	0.0018	-0.0019	1.1159	0.0000	0.0021	0.0017	-0.0067	1.1513	0.0144	0.0029	0.0022
GARCH-t	(var. PPI)	-0.0005	0.9506	0.0000	0.0027	0.0020	-0.0014	1.0576	0.0001	0.0022	0.0017	-0.0016	1.1042	0.0241	0.0028	0.0021
GAS-t	(var. PPI)	-0.0003	0.9286	0.0004	0.0030	0.0021	-0.0013	1.0559	0.0011	0.0024	0.0017	-0.0017	1.1190	0.0000	**0.0027**	**0.0021**
AMIDAS	(var. PPI)	0.0008	0.5975	0.0000	**0.0021**	**0.0017**	0.0003	0.6564	0.0000	**0.0018**	**0.0015**	0.0004	0.6563	0.0000	0.0057	0.0044
GARCH	(var. IP)	-0.0016	1.0815	0.0000	0.0020	0.0016	-0.0022	1.1709	0.0000	0.0021	0.0016	-0.0024	1.2298	0.0045	0.0029	0.0021
GARCH-t	(var. IP)	-0.0015	1.0727	0.0001	0.0023	0.0018	-0.0022	1.1715	0.0000	0.0021	0.0016	-0.0024	0.0029	0.0028	**0.0021**	**0.0017**
GAS-t	(var. IP)	-0.0009	1.0185	0.0086	0.0027	0.0020	-0.0018	1.1392	0.0007	0.0023	0.0017	-0.0022	1.2036	0.0040	0.0027	0.0020
AMIDAS	(var. IP)	-0.0012	1.0452	0.0000	**0.0020**	**0.0016**	-0.0017	1.1308	0.0003	**0.0021**	**0.0016**	-0.0019	1.1821	0.0247	0.0029	0.0021

Notes. The first column contains the GARCH-MIDAS with normal distribution (GARCH) and with Student's t-distribution (GARCH-t), GAS-MIDAS with Student's t-distribution (GAS-t) and the GARCH-AMIDAS model with normal distribution (AMIDAS). The underlying parameters are estimated during 1895-2008 and then fixed. The parameters c and y represent the coefficients in the regression: $\sigma_{t+h} = c + \psi\hat{\sigma}_{t+h|t} + \eta_t$, for $h > 0$. The p-values of the joint hypothesis that $c = 0$ and $y = 1$ are given by p, whereas the columns RMSE and MAE represent the Root Mean Squared Error and Mean Absolute Error, respectively. The values in bold represent the lowest errors for a certain forecast horizon given the model specification.

7. Conclusion

This paper analyzes to which extend economic variables can explain movements in long-term stock market volatility using the GARCH-MIDAS model. The main aim is to answer the research question how the growth rates of production price index inflation and industrial production influence the total volatility of daily U.S. stock returns. By validating the results of Engle et al. (2013), it is confirmed that the realized variance, producer price index and industrial production strongly influence the long-term volatility. The overall effect of the realized variance on this component is positive, taking into account that some parameters are possibly not identified. The macroeconomic variables show two different effects. First, an increase in inflation during the current quarter leads to an increase in stock market volatility in the next quarter. The other effect is a decrease in this long-term component when the industrial production increases. This result shows the counter-cyclical pattern of this variable, which is also observed in the research of Engle et al. (2013).

The economic interpretation of the GARCH-MIDAS model is determined by the variance ratios, which show that the realized variance can overall explain most of the market's volatility. For the full sample, this contribution can reach around 45% when applying a rolling window. On the other hand, only 38% can be attributed to the fixed span RV during the same period. The variance ratios of the macroeconomic variables are relatively low for the full sample, but reveal a significant contribution during 1953-1984. In fact, the models that use both levels and variances of producer price index and industrial production perform better than models that use these series separately.

Sequentially, the quality of the new component models is determined based on their forecasting abilities. More specifically, these models are estimated along with the traditional GARCH-MIDAS model on data from the period 1985-2010, after which their fit is determined on data from 2009-2018. The GARCH-AMIDAS model outperforms the GARCH-MIDAS model with macroeconomic variables for 1- and 2-months ahead forecasts. The GAS-MIDAS model with Student's t-distribution consistently achieves the best results for 4-months ahead forecasts.

However, some limitations were revealed while performing this research. First of all, the solutions from the optimization procedure may be found in local minima. As explained in Section 6.3, the estimates in this paper differ significantly from those by Engle et al. (2013). The starting values strongly determine the optimal outcomes, which influence the results in this research. It should be noted that the interpretation of the variables stays similar, such that the conclusions remain valid. A second limitation can be found in the data sets that are used to conduct this research. As mentioned in Section 5.1, the macroeconomic variables of this paper are not synchronized with those of Engle et al. (2013). The reason behind this is that the series were publicly unavailable for the period 1885-1919, such that not all results could be validated.

Due to the existence of these restrictions, further research is needed to get better insights in the latest volatility component models. For example, the optimization procedure of the models can be investigated more thoroughly to avoid solutions in local optima. Moreover, the GAS-MIDAS and GARCH-AMIDAS models can be altered to find better specifications. This could be done by changing the Student's t-distribution to another distribution to fit the GAS-MIDAS model better. Another extension of the new component models is the inclusion of principal components of the macroeconomic series as explanatory variables. These implementations might enhance these models and help investors to hedge their risk better than they do at the moment.

References

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

Asgharian, H., Hou, A. J., & Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A garch-midas approach. *Journal of Forecasting*, *32*(7), 600–612.

Bai, J., Ghysels, E., & Wright, J. H. (2013). State space models and midas regressions. *Econometric Reviews*, *32*(7), 779–813.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Campbell, J. Y. (1991). A variance decomposition for stock returns. *The Economic Journal*, *101*(405), 157–179.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, *31*(3), 281–318.

Chernov, M., Gallant, A. R., Ghysels, E., & Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, *116*(1-2), 225–257.

Clements, M. P., & Galvão, A. B. (2009). Forecasting us output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics*, *24*(7), 1187–1206.

Conrad, C., & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, *30*(7), 1090–1114.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

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., & Lee, G. (1999). A long-run and short-run component model of stock return volatility. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, 475–497.

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.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, *76*(3), 509–548.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, *26*(1), 53–90.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Lee, G. G., & Engle, R. F. (1993). A permanent and transitory component model of stock return volatility. *Available at SSRN 5848*.

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. (1989a). *Indexes of united states stock prices from 1802 to 1987* (Tech. Rep.). National Bureau of Economic Research.

Schwert, G. W. (1989b). Why does stock market volatility change over time? *The Journal of Finance*, *44*(5), 1115–1153.

Singleton, J. C., & Wingender, J. (1986). Skewness persistence in common stock returns. *Journal of Financial and Quantitative Analysis*, *21*(3), 335–341.

Timmermann, A. (2000). Density forecasting in economics and finance. *Journal of Forecasting*, *19*(4), 231.

Wang, F., & Ghysels, E. (2015). Econometric analysis of volatility component models. *Econometric Theory*, *31*(2), 362–393.

A. Derivations

A.1. Log-Likelihood of Normal Distribution

The distribution of the returns is given by equation (10). If we assume that $\varepsilon_{it} \mid \mathcal{I}_{i-1,t} \sim N(0,1)$, then the log-likelihood function can be written as:

$$
\begin{aligned}
\ell &= \sum_{t=1}^{T} \log(f(r_t)) \\
&= \sum_{t=1}^{T} \log\left(\frac{1}{\sigma_t} f(\varepsilon_t)\right) \\
&= \sum_{t=1}^{T} \log\left(\frac{1}{\sigma_t}\frac{1}{\sqrt{2\pi}} e^{-\varepsilon_t^2}\right) \\
&= \sum_{t=1}^{T} \log\left(\frac{1}{\sigma_t}\frac{1}{\sqrt{2\pi}} e^{-\frac{(r_t-\mu)^2}{2\sigma_t^2}}\right) \\
&= -\sum_{t=1}^{T}\left[\log(\sigma_t) + \log\left(\sqrt{2\pi}\right) + \frac{(r_t-\mu)^2}{2\sigma_t^2}\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\left[\log(\tau_t g_t) + \log(2\pi) + \frac{(r_t-\mu)^2}{\tau_t g_t}\right]
\end{aligned}
\tag{22}
$$

A.2. Log-Likelihood of Student's t-Distribution

The distribution of the returns is given by equation (10). If we assume that $\varepsilon_{it} \mid \mathcal{I}_{i-1,t} \sim t(\nu)$, where ν represents the degrees of freedom, then the log-likelihood function can be written as:

$$
\begin{aligned}
\ell &= \sum_{t=1}^{T} \log(f(r_t)) \\
&= \sum_{t=1}^{T} \log\left(\frac{1}{\sigma_t} f(\varepsilon_t)\right) \\
&= \sum_{t=1}^{T} \log\left(\frac{1}{\sigma_t}\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{\varepsilon_t^2}{\nu}\right)^{\frac{\nu+1}{2}}\right) \\
&= \sum_{t=1}^{T}\left[-\frac{1}{2}\log(\tau_t g_t) + \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \frac{1}{2}\log(\nu\pi) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{\nu+1}{2}\log\left(1+\frac{\varepsilon_t^2}{\nu}\right)\right] \\
&= \sum_{t=1}^{T}\left[-\frac{1}{2}\log(\tau_t g_t] + \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \frac{1}{2}\log(\nu\pi) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{\nu+1}{2}\log\left(1+\frac{(r_t-\mu)^2}{\nu\tau_t g_t}\right)\right]
\end{aligned}
\tag{23}
$$

A.3. GAS: Student's t-Distribution

Suppose that ε_{it} follows a Student's t-distribution with ν degrees of freedom. The observation log-density of ε_{it} conditional on g_{it} is then given by:

$$\begin{aligned}\log p(\varepsilon_{it}|g_{it};\Phi) = \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log(\nu) - \frac{1}{2}\log(\pi) \\ -\frac{1}{2}\log(g_{it}) - \frac{\nu+1}{2}\log\left(1+\frac{\varepsilon_{it}^2}{(\nu-2)g_{it}}\right)\end{aligned} \tag{24}$$

where $\Gamma(.)$ represents the gamma function. Following the notation of Creal et al. (2013), let ∇_{it} be the score which is given by:

$$\begin{aligned}\nabla_{it} &= \frac{\partial \log p(\varepsilon_{it}|g_{it};\Theta)}{\partial g_{it}} \\ &= \frac{\varepsilon_{it}^2(\nu+1)}{2g_{it}^2(\nu-2)}\left(1+\frac{\varepsilon_{it}^2}{(\nu-2)g_{it}}\right)^{-1} - \frac{1}{2g_{it}}\end{aligned} \tag{25}$$

If the scaling matrix S_{it} is assumed to be the inverse of the Fisher information matrix, it must hold that:

$$\begin{aligned}S_{it} &= \mathbb{E}\left[\frac{\partial^2 \log p(\varepsilon_{it}|g_{it};\Theta)}{\partial g_{it}\partial g_{it}}\right]^{-1} \\ &= -\frac{2g_{it}^2(\nu+3)}{\nu}\end{aligned} \tag{26}$$

Hence, the short-term component is updated recursively using:

$$g_{i+1,t} = 1 + \alpha\left(1+3\nu^{-1}\right)\left[\frac{\nu^{-1}+1}{\nu^{-1}+(1-2\nu^{-1})/\varepsilon_{it}^2} - 1\right]g_{it} + (\alpha+\beta)(g_{it}-1)\ , \quad \nu > 2 \tag{27}$$

where the unconditional variance of g_{it} is again unity.

B. Code

Run the pipeline

1. **Main**: starts the process by accepting user input for the time period and model specifications, after which it calls each component model separately.

GARCH-MIDAS model

1. **garchmidas_frv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the GARCH-MIDAS model with fixed span RV by optimizing the log-likelihood function.

2. **garchmidas_rrv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the GARCH-MIDAS model with rolling window RV by optimizing the log-likelihood function.

3. **log_garchmidas_frv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the log-specification of the GARCH-MIDAS model with fixed span RV by optimizing the log-likelihood function.

4. **log_garchmidas_rrv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the log-specification of the GARCH-MIDAS model with rolling window RV by optimizing the log-likelihood function.

5. **garchmidas_level.m**: this function collects the leveled macroeconomic variable of interest, after which it returns the optimal parameters for the GARCH-MIDAS model by optimizing the log-likelihood function.

6. **garchmidas_var.m**: this function collects the macroeconomic variable of interest which is expressed by its variance, after which it returns the optimal parameters for the GARCH-MIDAS model by optimizing the log-likelihood function.

7. **garchmidas_combo.m**: this function combines both levels and variances of a macroeconomic variable of interest in a single equation, after which it returns the optimal parameters for the GARCH-MIDAS model by optimizing the log-likelihood function.

GAS-MIDAS model

1. **gasmidas_frv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the GAS-MIDAS model with fixed span RV by optimizing the log-likelihood function.

2. **gasmidas_rrv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the GAS-MIDAS model with rolling window RV by optimizing the log-likelihood function.

3. **gasmidas_level.m**: this function collects the leveled macroeconomic variable of interest, after which it returns the optimal parameters for the GAS-MIDAS model by optimizing the log-likelihood function.

4. **gasmidas_var.m**: this function collects the macroeconomic variable of interest which is expressed by its variance, after which it returns the optimal parameters for the GAS-MIDAS model by optimizing the log-likelihood function.

GARCH-AMIDAS model

1. **garchamidas_frv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the GARCH-AMIDAS model with fixed span RV by optimizing the log-likelihood function.

2. **garchamidas_rrv.m**: this function collects the daily stock returns, after which it returns the optimal parameters for the GARCH-AMIDAS model with rolling window RV by optimizing the log-likelihood function.

3. **garchamidas_level.m**: this function collects the leveled macroeconomic variable of interest, after which it returns the optimal parameters for the GARCH-AMIDAS model by optimizing the log-likelihood function.

4. **garchamidas_var.m**: this function collects the macroeconomic variable of interest which is expressed by its variance, after which it returns the optimal parameters for the GARCH-AMIDAS model by optimizing the log-likelihood function.

Forecasts

1. **main_forecast_rv.m**: accepts user input for the number of steps ahead that should be performed and runs the forecasting functions that deal with realized variances.

2. **forecast_fix.m**: this function performs the forecast for any volatility component model using its fixed span specification, and returns the statistical measures as well as the forecast errors.

3. **forecast_roll.m**: this function performs the forecast for any volatility component model using its rolling window specification, and returns the statistical measures as well as the forecast errors.

4. **main_forecast_macro.m**: accepts user input for the number of steps ahead that should be performed and runs the forecasting functions that deal with macroeconomic variables.

5. **forecast_level.m**: this function performs the forecast for any volatility component model using the leveled macroeconomic variables, and returns the statistical measures as well as the forecast errors.

6. **forecast_var.m**: this function performs the forecast for any volatility component model using those macroeconomic variables that are expressed as variances, and returns the statistical measures as well as the forecast errors.

Log-likelihood functions

1. **logl_garch.m**: calculates and returns the negative log-likelihood function of the GARCH-MIDAS model using a normal distribution.

2. **logl_lgarch.m**: calculates and returns the negative log-likelihood function of the log-specification of the GARCH-MIDAS model using a normal distribution.

3. **logl_tgarch.m**: calculates and returns the negative log-likelihood function of the GARCH-MIDAS model using a Student's t-distribution.

4. **logl_macro.m**: calculates and returns the negative log-likelihood function of the GARCH-MIDAS model with macroeconomic variables using a normal distribution.

5. **logl_tmacro.m**: calculates and returns the negative log-likelihood function of the GARCH-MIDAS model with macroeconomic variables using a Student's t-distribution.

6. **logl_combo.m**: calculates and returns the negative log-likelihood function of the GARCH-MIDAS model with both levels and variances as macroeconomic variables using a normal distribution.

7. **logl_gas.m**: calculates and returns the negative log-likelihood function of the GAS-MIDAS model using a Student's t-distribution.

8. **logl_amidas.m**: calculates and returns the negative log-likelihood function of the GARCH-AMIDAS model using a normal distribution.

Utils

1. **get_descriptives.m**: returns the time series plots and descriptive statistics for a (macroeconomic) variable of interest.

2. **get_subsample.m**: returns a subsample of a table or array based on two dates.

3. **unit_garch.m**: performs a GARCH(1,1) process based on the historical returns and estimates of a period and returns a vector of short-term volatilities.

4. **unit_gas.m**: performs a GAS(1,1) process based on the historical returns and estimates of a period and returns a vector of short-term volatilities.

5. **weight_scheme.m**: calculates and returns the weights for a K-lagged MIDAS model based on either a Beta or exponential weighting function.

6. **weights_amidas.m**: calculates and returns the weights for a K-lagged asymmetric MIDAS model based on either a Beta or exponential weighting function.

Additional

1. **get_macros.m**: returns the quarterly macroeconomic growth rates based on the original data set.

C. Additional Results

Table 10: Parameter estimates for the log-specification of the GARCH-MIDAS with realized variance.

				Fixed Realized Variance				
Sample	μ	α	β	ϑ	ω	ξ	LLF	BIC
1890-2010	0.00063	0.10508	0.87278	48.42501	2.60889	-9.65944	113,526.8	-6.7787
	(16.70)	(5.19)	(43.86)	(7.06)	(17.98)	(-65.10)		
1890-1919	0.00053	0.15394	0.78829	49.01209	28.61624	-9.80716	31,786.3	-6.9358
	(3.47)	(2.27)	(9.98)	(3.62)	(3.02)	(-51.29)		
1920-1952	0.00075	0.10403	0.86597	43.89966	3.92198	-9.58690	31,260.1	-6.4218
	(5.56)	(4.18)	(35.65)	(9.06)	(11.57)	(-50.91)		
1953-1984	0.00059	0.09056	0.89363	108.33486	4.31592	-10.04408	28,697.4	-7.1293
	(0.93)	(10.47)	(71.55)	(1.95)	(3.07)	(-41.20)		
1985-2010	0.00073	0.09318	0.88777	43.34235	2.61410	-9.50914	21.851,7	-6.6561
	(2.47)	(12.96)	(95.22)	(3.88)	(1.38)	(-69.46)		
1953-2010	0.00064	0.08764	0.89995	81.63492	0.90728	-9.77094	50,536.5	-6.9184
	(12.03)	(18.54)	(159.09)	(5.10)	(2.35)	(-72.82)		
				Rolling Realized Variance				
Sample	μ	α	β	ϑ	ω	ξ	LLF	BIC
1890-2010	0.00062	0.10548	0.87118	55.44253	3.03702	-9.69020	113,989.1	-6.7810
	(16.92)	(28.41)	(197.49)	(33.33)	(86.42)	(-155.32)		
1890-1919	0.00052	0.15470	0.77641	79.74407	22.12404	-9.95525	32,251.3	-6.9426
	(5.42)	(14.55)	(66.75)	(16.06)	(12.47)	(-172.24)		
1920-1952	0.00074	0.10376	0.86725	49.76224	3.37309	-9.56545	31,629.0	-6.4205
	(6.72)	(4.01)	(25.44)	(6.36)	(9.84)	(-54.50)		
1953-1984	0.00062	0.09410	0.88527	137.51987	6.85352	-10.21339	28,877.3	-7.1341
	(8.55)	(1.60)	(14.59)	(5.28)	(4.39)	(-16.15)		
1985-2010	0.00072	0.09412	0.88442	41.01723	3.40478	-9.54751	21,633.4	-6.6504
	(6.71)	(4.77)	(33.28)	(8.27)	(1.17)	(-59.63)		
1953-2010	0.00065	0.08777	0.90035	93.07176	0.64364	-9.81065	50,713.6	-6.9213
	(11.03)	(19.71)	(178.21)	(17.88)	(7.16)	(-88.01)		

Notes. The GARCH-MIDAS estimates are created using four MIDAS lag years. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

Table 11: Parameter estimates for different component models with RV for the period 1890-2010.

Model	μ	α	β	ϑ	$\omega_{1,+}$	$\omega_{2,+}$	$\omega_{1,-}$	$\omega_{2,-}$	φ	ξ	ν	LLF/BIC
Fixed Realized Variance												
GARCH	0.00063	0.10985	0.85782	0.00894	4.48675					0.00003		113,551.4
	(16.97)	(28.92)	(168.11)	(25.89)	(154.54)					(25.75)		/ -6.7802
GARCH-t	0.00072	0.06941	0.86417	0.00310	6.5822					0.00001	6.19036	114,488.0
	(20.16)	(24.88)	(158.24)	(24.90)	(17.93)					(24.42)	(38.62)	/ -6.8358
GAS-t	0.00072	0.06897	0.88311	0.00270	4.33835					0.00001	6.87881	114,390.7
	(16.39)	(13.23)	(100.05)	(21.03)	(15.81)					(9.91)	(118.73)	/ -6.8300
AMIDAS	0.00058	0.10913	0.85310	0.10561	4.77498	1.23744	18.83888	1.92471	1.44588	0.00002		113,555.8
	(16.07)	(40.88)	(181.95)	(43.92)	(39.51)	(9.73)	(10.46)	(18.06)	(4.67)	(20.38)		/ -6.7809
Rolling Realized Variance												
Model	μ	α	β	ϑ	$\omega_{1,+}$	$\omega_{2,+}$	$\omega_{1,-}$	$\omega_{2,-}$	φ	ξ	ν	LLF/BIC
GARCH	0.00063	0.11397	0.84306	0.01043	9.23857					0.00003		114,023.3
	(6.18)	(4.25)	(20.63)	(8.94)	(29.53)					(8.43)		/ -6.7830
GARCH-t	0.00071	0.07096	0.85488	0.00401	9.14604					0.00001	6.23364	114,954.5
	(7.95)	(4.23)	(30.78)	(14.44)	(13.13)					(16.04)	(12.28)	/ -6.8381
GAS-t	0.00072	0.07142	0.87191	0.00379	7.95121					0.00001	6.92228	114,853.4
	(10.96)	(11.89)	(49.65)	(6.13)	(79.14)					(2.33)	(69.07)	/ -6.8321
AMIDAS	0.00064	0.11176	0.84978	0.01116	5.44933	1.23340	22.43331	2.24405	1.19916	0.00002		114,008.2
	(16.12)	(29.01)	(58.14)	(7.96)	(2.04)	(1.46)	(11.76)	(2.80)	(3.61)	(118.73)		/ -6.7809

Notes. The first column contains the GARCH-MIDAS with normal distribution (GARCH) and with Student's t-distribution (GARCH-t), GAS-MIDAS with Student's t-distribution (GAS-t) and the GARCH-AMIDAS model with normal distribution (AMIDAS). The estimates are created using four MIDAS lag years. The parameter $\omega_{1,+}$ is comparable to ω in Table 3 for GARCH-MIDAS and GAS-MIDAS (irrespective of their distribution). The remaining $\omega_{i,j}$ are those from equation (18). The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.

Table 12: Parameter estimates for different component models with variance PPI and IP for the period 1920-2010.

Producer Price Index												
Model	μ	α	β	ϑ_v	$\omega_{1,+}$	$\omega_{2,+}$	$\omega_{1,-}$	$\omega_{2,-}$	φ	ξ	ν	LLF/BIC
GARCH	0.00066	0.09191	0.89820	0.11642	0.68187	0.36456					-9.60176	76,554.4
	(2.10)	(24.96)	(222.26)	(6.86)	(1.42)	(1.09)					(-71.73)	/ -6.7037
GARCH-t	0.00078	0.05496	0.91020	0.08612	4.27222	1.05676				-11.37877	5.90289	77,259.0
	(18.83)	(14.22)	(124.16)	(3.85)	(1.40)	(3.98)				(-141.28)	(28.78)	/ -6.7650
GAS-t	0.00079	0.05706	0.91388	0.09459	2.93315	0.98414				-11.45244	6.43769	77,220.9
	(2.74)	(3.53)	(63.05)	(1.82)	(0.81)	(1.21)				(-19.22)	(3.09)	/ -6.7616
AMIDAS	0.00070	0.10046	0.87114	0.10985	3.27752	0.95246	1.64512	3.25295	0.99360	-10.01727		76,512.3
	(16.53)	(30.53)	(263.13)	(10.02)	(3.18)	(3.45)	(4.54)	(11.31)	(8.83)	(-168.99)		/ -6.6987
Industrial Production												
Model	μ	α	β	ϑ_v	$\omega_{1,+}$	$\omega_{2,+}$	$\omega_{1,-}$	$\omega_{2,-}$	φ	ξ	ν	LLF/BIC
GARCH	0.00063	0.11397	0.84306	0.01043	9.23857					0.00003		114,023.3
	(6.18)	(4.25)	(20.63)	(8.94)	(29.53)					(8.43)		/ -6.7830
GARCH-t	0.00078	0.05470	0.91149	0.02955	30.18449	3.95894				-11.25546	5.88363	77,259.4
	(17.40)	(3.15)	(43.56)	(6.68)	(1.39)	(1.60)				(-20.71)	(18.61)	/ -6.7650
GAS-t	0.00079	0.05704	0.91402	0.48507	3.57772	0.89464				-11.41252	6.43615	77,225.4
	(2.75)	(23.86)	(227.66)	(6.21)	(13.12)	(2.26)				(-106.38)	(24.85)	/ -6.7620
AMIDAS	0.00065	0.09361	0.90488	0.03254	0.77430	1.23601	3.52404	5.99200	0.66779	-7.89808		76,529.3
	(13.91)	(25.64)	(242.63)	(3.86)	(7.00)	(9.23)	(23.15)	(23.85)	(9.19)	(-65.32)		/ -6.7002

Notes. The first column contains the GARCH-MIDAS with normal distribution (GARCH) and with Student's t-distribution (GARCH-t), GAS-MIDAS with Student's t-distribution (GAS-t) and the GARCH-AMIDAS model with normal distribution (AMIDAS). The estimates are created using four MIDAS lag years. That is why the first sample starts in 1924 instead of 1920. The parameter ϑ_v is multiplied by 10^{-4} to represent percentages. The values in parentheses are robust t-statistics which are determined using HAC standard errors. LLF represents the log-likelihood function and BIC the Bayesian Information Criterion.