This paper applies the GARCH-MIDAS (Mixed Data Sampling) model to examine whether information contained in macroeconomic variables can help to predict short-term and long-term components of the return variance. A principal component analysis is used to incorporate the information contained in different variables. Our results show that including low-frequency macroeconomic information in the GARCH-MIDAS model improves the prediction ability of the model, particularly for the long-term variance component. Moreover, the GARCH-MIDAS model augmented with the first principal component outperforms all other specifications, indicating that the constructed principal component can be considered as a good proxy of the business cycle.

**Keywords:** Mixed data sampling, long-term variance component, macroeconomic variables, principal component, variance prediction.

**THE KNUT WICKSELL CENTRE FOR FINANCIAL STUDIES**

The Knut Wicksell Centre for Financial Studies conducts cutting-edge research in financial economics and related academic disciplines. Established in 2011, the Centre is a collaboration between Lund University School of Economics and Management and the Research Institute of Industrial Economics (IFN) in Stockholm. The Centre supports research projects, arranges seminars, and organizes conferences. A key goal of the Centre is to foster interaction between academics, practitioners and students to better understand current topics related to financial markets.

**Importance of the macroeconomic variables for variance prediction**

A GARCH-MIDAS approach

HOSSEIN ASGHARIAN | AI JUN HOU | FARRUKH JAVED

KNUT WICKSELL WORKING PAPER 2013:4
Importance of the macroeconomic variables for variance prediction

A GARCH-MIDAS approach

Hossein Asgharian*: Department of Economics, Lund University and Knut Wicksell Center for Financial Studies
Ai Jun Hou: Department of Business and Economics, Southern Denmark University
Farrukh Javed: Department of Statistics, Lund University

Abstract

This paper applies the GARCH-MIDAS (Mixed Data Sampling) model to examine whether information contained in macroeconomic variables can help to predict short-term and long-term components of the return variance. A principal component analysis is used to incorporate the information contained in different variables. Our results show that including low-frequency macroeconomic information in the GARCH-MIDAS model improves the prediction ability of the model, particularly for the long-term variance component. Moreover, the GARCH-MIDAS model augmented with the first principal component outperforms all other specifications, indicating that the constructed principal component can be considered as a good proxy of the business cycle.

Keywords: Mixed data sampling, long-term variance component, macroeconomic variables, principal component, variance prediction.

* Tel.: +46 46 222 8667; fax: +46 46 222 4118; E-mail address: Hossein.Asgharian@nek.lu.se; Department of Economics, Lund University, Box 7082, S-22007 Lund, Sweden.
We are very grateful to the Jan Wallanders and Tom Hedelius Foundation (P2009-0041:1 and P2010-0077:1) for funding this research.
Importance of the macroeconomic variables for variance prediction

A GARCH-MIDAS approach

Abstract
This paper applies the GARCH-MIDAS (Mixed Data Sampling) model to examine whether information contained in macroeconomic variables can help to predict short-term and long-term components of the return variance. A principal component analysis is used to incorporate the information contained in different variables. Our results show that including low-frequency macroeconomic information in the GARCH-MIDAS model improves the prediction ability of the model, particularly for the long-term variance component. Moreover, the GARCH-MIDAS model augmented with the first principal component outperforms all other specifications, indicating that the constructed principal component can be considered as a good proxy of the business cycle.

Keywords: Mixed data sampling, long-term variance component, macroeconomic variables, principal component, variance prediction.
1. Introduction

A correct assessment of future volatility is crucial for asset allocation and risk management. Countless studies have examined the time-variation in volatility and the factors behind this time variation, and documented a clustering pattern. Different variants of the GARCH model have been pursued in different directions to deal with these phenomena. Simultaneously, a vast literature has investigated the linkages between volatility and macroeconomic and financial variables. Schwert (1989) relates the changes in the volatility of returns to the macroeconomic variables and addresses how bond returns, the short-term interest rate, producer prices or industrial production growth rate provide incremental information on monthly market volatility. Glosten et al. (1993) find evidence that short-term interest rates play an important role for the future market variance. Whitelaw (1994) finds statistical significance for a commercial paper spread and the one year treasury rate, while Brandt and Kang (2002) use the short-term interest rate, term premium, and default premium and find a significant effect. Other research including Hamilton and Lin (1996) and Perez-Quiros and Timmermann (2000) have found evidence that the state of the economy is an important determinant in the volatility of the returns.

Since the analyses of the time-varying volatility are mostly based on high frequency data, the previous studies are mostly limited to variables such as short-term interest rates, term premiums, and default premiums, for which daily data are available. Therefore, the impacts of variables such as unemployment rate and inflation on volatility have not been sufficiently examined. Ghysels et al. (2006) introduce a regression scheme, namely MIDAS (Mixed Data Sampling) which allows inclusion of data from different frequencies in the same model. This makes it possible to combine the high-frequency return data with macroeconomic data that are only observed at lower frequencies, such as monthly or quarterly. Engle et al. (2009) propose the GARCH-MIDAS model within the MIDAS framework to analyze time-varying market volatility. Within this framework, the conditional variance is divided into long-term and short-term components. The low frequency variables affect the conditional variance via the long-term component. This approach combines the component model suggested by Engle and Lee (1999)\footnote{For the component model see also Ding and Granger, 1996; Chernov et al., 2003.} with the MIDAS framework of Ghysels et al. (2006). The main advantage of the GARCH-MIDAS model is that it allows us to link the daily observations of stock returns...
with macroeconomic variables, sampled at lower frequencies, in order to examine directly the macroeconomic variables’ impact on the stock volatility.

In this paper, we apply the recently proposed GARCH-MIDAS methodology to examine the effect of the macroeconomic variables on stock market volatility. Departing from Engle et al. (2009), our investigation mainly focuses on variance predictability and aims to analyze if adding economic variables can improve the forecasting abilities of the traditional volatility models. Using GARCH-MIDAS we decompose the return volatility into its short-term and long-term component, where the latter is affected by the smoothed realized volatility and/or by macroeconomic variables. We examine a large group of macroeconomic variables which include unexpected inflation, term premium, per capita labour income growth, default premium, unemployment rate, short-term interest rate, per capita consumption. We investigate the ability of the GARCH-MIDAS models with economic variables to predict both short-term and long-term volatilities. The performances of these models are then compared with the GARCH (1, 1) model as a benchmark. In order to capture the information contained in different economic variables and investigate their combined effect, we perform a principal component analysis. The advantage of this approach is to reduce the number of parameters and increase the computational efficiency.

Our results show that including low-frequency macroeconomic information in the GARCH-MIDAS model improves the prediction ability of the model, particularly for the long-term variance component. Moreover, the GARCH-MIDAS model augmented with the first principal component outperforms all other specifications. Among the individual macroeconomic variables, the short-term interest rate and the default rate perform better than the other variables when included in the MIDAS equation.

To our knowledge this is the first study that investigates the out-of-sample forecast performance of the GARCH-MIDAS model. The paper also differs from Engle et al. (2009) and contributes to existing literature by investigating a number of macroeconomic variables in the GARCH-MIDAS framework.

The rest of the paper is organized as follows: Section 2 presents the empirical models, and the data and the econometric methods are described in Section 3, while section 4 contains the empirical results, and Section 5 concludes.
2. GARCH-MIDAS

In this paper, we use a new class of component GARCH model based on the MIDAS (Mixed Data Sampling) regression. MIDAS regression models are introduced by Ghysels et al. (2006). MIDAS offers a framework to incorporate macroeconomic variables sampled at different frequency along with the financial series. This new component GARCH model is referred to as MIDAS-GARCH, where macroeconomic variables enter directly into the specification of the long-term component.

This new class of MIDAS structure has gained much attention in recent years by Ghysels et al. (2004), Ghysels et al. (2006) and Andreaou et al. (2010). Chen and Ghysels (2007) extend the MIDAS setting to a multi-horizon semi-parametric framework. Chen and Ghysels (2009) provide a comprehensive study and a novel method to analyze the impact of news on forecasting volatility. Ghysels et al. (2009) discuss the Granger causality with mixed frequency data. Kotze (2007) uses the MIDAS regression with high frequency data on asset prices and low frequency inflation forecasts. In addition, a number of papers use MIDAS regression to obtain quarterly forecasts with monthly and daily data. For instance, Bai et al. (2009) and Tay (2007) use monthly data to improve quarterly forecasts. Alper et al. (2008) compare the stock market volatility forecasts across emerging markets using MIDAS regression. Clements and Galavaao (2006) study the forecasts of the US output growth and inflation in this context. Forsberg and Ghysels (2006) show, through simulation, the relative advantage of MIDAS over the HAR-RV (Heterogeneous Autoregressive Realized Volatility) model proposed in Anderson et al. (2007).

The GARCH-MIDAS model can formally be described as below. Assume the return on day $i$ in month $t$ follows the following process:

$$ r_{i,t} = \mu + \sqrt{\tau_t g_{i,t}} \varepsilon_{i,t}, \quad \forall i = 1, \ldots, N_t, $$

$$ \varepsilon_{i,t} | \Phi_{i-1,t} \sim N(0,1) $$

where $N_t$ is the number of trading days in month $t$ and $\Phi_{i-1,t}$ is the information set up to the $(i-1)^{th}$ day of period $t$. Equation (1) expresses the variance into a short-term component defined by $g_{i,t}$ and a long-term component defined by $\tau_t$. 

5
The conditional variance dynamics of the component $g_{i,t}$ is a (daily) GARCH (1,1) process, as:

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \left( \frac{\tau_{t-1} - \mu}{\tau_t} \right)^2 + \beta g_{i,t-1}$$  \hspace{1cm} (2)

and $\tau_t$ is defined as smoothed realized volatility in the spirit of MIDAS regression:

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k (w_1, w_2) RV_{t-k}$$  \hspace{1cm} (3)

where $RV_t = \sum_{i=1}^{N_t} \tau_{i,t}^2$

$K$ is the number of periods over which we smooth the volatility.\(^2\) We further modify this equation by involving the economic variables along with the $RV$ in order to study the impact of these variables on the long-run return variance:

$$\tau_t = m + \theta_1 \sum_{k=1}^{K} \varphi_k (w_1, w_2) RV_{t-k} + \theta_2 \sum_{k=1}^{K} \varphi_k (w_1, w_2) X_{t-k}^l + \theta_3 \sum_{k=1}^{K} \varphi_k (w_1, w_2) X_{t-k}^v$$  \hspace{1cm} (4)

where $X_{t-k}^l$ represents the level of a macroeconomic variable and $X_{t-k}^v$ represents the variance of that macroeconomic variable. The component $\tau_t$ used in our analysis does not change within a fixed time span (e.g. a month).

Finally, the total conditional variance can be defined as:

$$\sigma_{it}^2 = \tau_t g_{i,t}$$  \hspace{1cm} (5)

The weighting scheme used in equation (3) and equation (4) is described by beta lag polynomial, as:

$$\varphi_k (w) = \frac{\left( \frac{k}{K} \right)^{w_1-1} \left( 1 - \frac{k}{K} \right)^{w_2-1}}{\sum_{j=1}^{K} \left( \frac{j}{K} \right)^{w_1-1} \left( 1 - \frac{j}{K} \right)^{w_2-1}}$$  \hspace{1cm} (6)

\(^2\) A potential problem one might face is to ensure non-negative $\tau_t$. A possible solution can be the log form (log $\tau_t$) specification. However, we do not face the problem of negativity during our estimation. Therefore, we use the above formulation of $\tau_t$ throughout our analysis.
3. Data and Estimation Method

3.1. Data

We use the US, S&P 500 composite, daily price index to calculate stock returns. In our conditional variance model we use a number of financial and macroeconomic factors which have been found by previous studies to be important for return variance. The following variables are used:

- **Short-term interest rate** is a yield on the three-month US Treasury bill.
- **Slope of the yield curve** is measured as the yield spread between a ten-year bond and a three-month Treasury bill.
- **Default rate** is measured as the spread between Moody’s Baa and Aaa corporate bond yields of the same maturity.
- **Exchange rate** is the nominal major currencies dollar index from the Federal Reserve.
- **Inflation** is measured as the monthly changes in the seasonally adjusted consumer price index (CPI).
- Growth rate in the *Industrial Production index*.
- **Unemployment rate**.

Data cover the period from January 1991 to June 2008. All the items except the exchange rate are collected from DataStream®.
3.2. *Estimation Method*

3.2.1 *Various model specifications*

We use three different model specifications. The models differ with respect to the definition of the long-term variance component, $\tau_t$, while the equation for the short-term variance, $g_{\tau t}$, remains the same in all three cases. The three specifications are:

- The $RV$ model: In this specification, we solely use the monthly realized volatility ($RV$) in the long-term component of the variance, defined by the MIDAS equation, $\tau_t$, in equation (3). We have no economic variables in this model.

- The $RV + \chi^l + \chi^v$ model: Here, we augment the model by adding both the level and the variance of an economic variable to the MIDAS equation, $\tau_t$. This modification is supposed to capture the information explained by both the macroeconomic factor and the monthly $RV$.

- The $\chi^l + \chi^v$ model: In this specification, we only study the effect of macroeconomic variables, both level and variance, on the long-term variance component, i.e., the equation for $\tau_t$.

By analyzing these three alternatives, we can investigate to what extent the long-term variance can be explained by the past realized return volatility and the macroeconomic variables.3

3.2.2 *Estimation strategy*

Our estimations are based on the daily observations on returns, while we use monthly frequency in the MIDAS equation to capture the long-term component. The realized volatility is our preferred measure of the monthly variance, but since daily data are not available for most macroeconomic variables, it is not possible to use this measure. We select the squared first differences as the measure of the variance of the economic variables.

We estimate the models described above using an estimation window and then use the estimated parameters to make out-of-sample variance prediction.4 We use a ten-year

---

3 We have also estimated the model with only the level or the variance of the economic variables in the MIDAS equation. In order to save space, these results are not reported but are available upon request.

4 We use several alternative time spans for the estimation window, i.e. five, eight and ten years. Our results show that the estimation accuracy reduces as we decrease the length of the estimation window. We therefore
estimation window and keep the parameters over the subsequent year. The first estimation window starts in January 1994 and ends in December 2003. However, we also need three years lagged data before each time period to compute the historical realized volatility, which means that the realized volatility for January 1994 is estimated with data from January 1991 to December 1993. The estimation window is then moved forward by one year until December 2007. Our out-of-sample forecast covers the period January 2004 to June 2008. We chose not to use data after the start of the financial crisis of 2008, since the extreme outliers of the period of the crisis make it impossible to make any reliable and accurate out-of-sample comparisons of the models. One may address this issue by including jumps in the short-term component of the GARCH-MIDAS structure. However, it will significantly complicate the estimation procedure. Further, since we could only analyze the jump effects in the short-term movements, it does not improve the prediction of the long-term movements, which is one of the essences of the GARCH-MIDAS structure.

We use the estimated \( \tau_t \) from the MIDAS equation as the prediction of the long-term variance (see equations (3) and (4)). Since the values of \( \tau_t \) are on a daily basis, we multiply this value by the number of trading days within each month. The estimated daily total variance (\( \sigma_t^2 \)) is used as the prediction of short-term variance.

The forecasting ability of the GARCH-MIDAS model is compared with a simple GARCH (1,1) model,

\[
\begin{align*}
    r_t &= \mu + \eta_t, \quad \eta_t = \sigma_t z_t, \quad z_t \sim N(0,1), \\
    \sigma_t^2 &= \omega + \alpha \eta_{t-1}^2 + \beta \sigma_{t-1}^2
\end{align*}
\]

We predict the long-term volatility with the monthly observations and for the short-term forecast we use the daily observations.

We compare the out-of-sample predictions of the monthly variances from the GARCH-MIDAS and the GARCH models with the monthly realized volatility measured as the sum of the daily squared returns in month \( t \). To assess the short-term prediction ability of the models we compare the estimated daily total variance of the GARCH-MIDAS and the GARCH model with the realized daily volatility, measured as the squared returns.

elect to present only the results with a 10-year estimation window. The results for other estimation windows are available upon request.
We employ a number of measures to evaluate the variance prediction of a specific model by comparing the model predicted variance with the realized monthly volatility, estimated as the sum of the squared daily log returns within each month. We use two loss functions, the Mean Square Error (MSE) and the Mean Absolute Error (MAE), defined as

\[
MSE = \frac{1}{T} \sum_{t=1}^{T} \left( \sigma_{t+1}^2 - E_t \left( \sigma_{t+1}^2 \right) \right)^2
\]

\[
MAE = \frac{1}{T} \sum_{t=1}^{T} | \sigma_{t+1}^2 - E_t \left( \sigma_{t+1}^2 \right) |
\]

MSE is a quadratic loss function and gives a larger weight to large prediction errors compared to the MAE measure, and is therefore proper when large errors are more serious than small errors (see Brooks and Persand, 2003). We use the test suggested by Diebold and Mariano (1995), the DM-test, to compare the prediction accuracy of two competing models,

\[
DM = \frac{\overline{d}}{\sqrt{\text{var}(d_i)}} \sim N(0,1)
\]

where \( d_i = e_{A,t}^2 - e_{B,t}^2 \)

In addition to these measures we run the following regression of the realized variance on the predicted variance (see e.g., Andersen and Bollerslev, 1998 and Hansen, 2005).

\[
\sigma_{t+1}^2 = a + bE_t \left( \sigma_{t+1}^2 \right) + u_t
\]

If the predicted variance has some information about the future realized volatility, then the parameter \( b \) should be significantly different from zero. Furthermore, for an unbiased prediction we expect the parameter \( a \) to be zero and the parameter \( b \) to be equal to one. We also look at the \( R \)-square of this regression.

The maximum likelihood method is used to estimate the model parameters. The likelihood function of the GARCH-MIDAS model involves a large number of parameters, which do not always converge to a global optimum by the conventional optimization algorithms. We
therefore use the simulated annealing approach (see Goffe et al., 1994) for estimation. This method is very robust and seldom fails, even for very complicated problems.

3.2.3 Weights and number of lags in the MIDAS equation

During the estimation, we have chosen several strategies to simplify the estimation and to make the model work more efficiently.

First, we have to choose the weights ($w_1$ and $w_2$) in the beta functions specified in equation (6). We have three alternatives:

i) Taking both $w_1$ and $w_2$ as free parameters and estimating them within the model.

ii) Fixing $w_1$ a priori and letting $w_2$ be estimated within the model.

iii) Fixing both $w_1$ and $w_2$ a priori.

Figure 1 illustrates the plot of the weighting function for two choices of $w_1$ (1 and 2) and two choices of $w_2$ (4 and 8). It shows that the weight function is monotonically decreasing as long as $w_1$ is equal to one. Given $w_1$ equal to one, increasing $w_2$ will give a larger weight to the most recent observations. A $w_1$ larger than one gives a lower weight to the most recent observations. Alternative (i) sometimes results in very counterintuitive weighting patterns, e.g., a lower weight for more recent observations ($w_1$ larger than one). We therefore follow Engle et al. (2009) and fix the weight $w_1$ to one, which makes the weights monotonically decreasing over the lags. Since there are no a priori preferences for the choice of $w_2$, we let the model define $w_2$ (alternative (ii)) when estimating the $RV$ model. However, we keep the estimated weight from this model for the remainder of the specifications.

Second, we have to decide how many lags we should use in the MIDAS equation ($K$ in equations 3, 4 and 6). The total lags are determined by the number of years, or so-called MIDAS years, and by the time-span $t$ that will be used to calculate $\tau_t$ in equations (3) and (4). This time span can be a month, a quarter, or a half year. Regarding the length of the time-period used in our study and in order to have a sufficient number of out-of-sample predictions, we decide to use a monthly time span. In the lower graph of Figure 1, we plot the maximum values of the likelihood function using different lags in the MIDAS equation. It can be seen that the optimum value of the likelihood function increases with the number of lags and it converges to its highest level at around 36 lags. We therefore limit the number of lags in the MIDAS equation to 36, which results in three MIDAS years.
3.2.4. Principal components

GARCH-MIDAS is computationally complex and the inclusion of several macroeconomic variables in one model will result in identification and/or convergence problems. Therefore we use one variable at a time in the MIDAS equation. In order to incorporate the information contained in different variables in the same equation, we also construct principal components based on these variables. Stock and Watson (2002a; 2002b) construct principal component to extract the combined information from a large group of macroeconomic variables. They show that this method can be efficiently used instead of employing large number of variables in one regression. Since the macroeconomic variables have different scales, we use the correlation matrix to construct the principal components.

4. Results and Analyses

4.1. Descriptive analysis

Table 1 shows the correlation between monthly observations on the macroeconomic variables and the realized monthly volatility of the S&P 500 composite index ($RV$). The interest rate, as expected, has a high negative correlation with slope (-0.70). Further, the slope is higher when the unemployment rate is high. Unemployment and inflation are also highly correlated during the selected time-span.

Table 2 shows the correlations between the principal components and the macroeconomic variables. The first principal component ($PC_1$) has a high correlation with most of the variables, particularly with the interest rate, slope, default rate, and unemployment (average correlation is 0.48). Since most of these variables are commonly used as a measure for the business cycle we may consider the variable $PC_1$ as a proper proxy for the cycle. Similarly, we observe a relatively large correlation between some variables, i.e., inflation and interest rate with $PC_2$. Other principal components have either low correlations with the macroeconomic variables or are only related to one specific variable (such as $PC_3$ and industrial production). We choose therefore only to include $PC_1$ and $PC_2$ in the MIDAS equation. Figure 2 plots the monthly realized volatility of the return, the macroeconomic variables, and the first two principal components constructed based on the macroeconomic variables. A drastic fluctuation is observed in the realized volatility between 1997 and mid-2002. This may indicate the effect of the Asian crisis in 1998, the burst of the dot-com bubble
in 2000 and the September 11 incident in 2001. The last volatile period around 2007-2008 indicates the start of the recent financial crisis. We can find a similar pattern in the movements of the $PC_1$ series. It shows a declining trend in the beginning, followed by a sharp increase in the values after the financial turmoil in 2001, which remains until 2003. An increasing trend around the period of 2007-2008 signals the start of the recent financial crisis.

From the plot of $PC_2$, we can observe a continuously increasing trend throughout the sample period. The interest rate pattern is the reverse of that for $PC_1$, confirming the high negative correlation between them (-0.78). Similarly, the default rate is high during the financial crises of 1998, 2001 and 2007 compared to other time periods. The growth rate in industrial production is smooth apart from some peak points near 1998. The exchange rate changes slightly around 2001, but otherwise seems stable throughout the sample period. Inflation has an opposite behaviour to that of $PC_2$, supporting their highly negative correlation (-0.83). Similarly, the unemployment rate increases after the crisis of 2001 and remains high for the next couple of years. We can observe an increasing trend in the unemployment rate after the recent financial crisis of 2008.

### 4.2. In-sample estimations

In Table 3, we present the estimated parameters of the in-sample fit for the first estimation period, starting on January 1 1991 and ending on December 31 2003. The models are estimated with the first two principal components and with all the individual economic variables in the MIDAS equation. In order to save space we only report the results for $PC_1$ and $PC_2$. Most of the parameters in the equations for returns and the short-term variance component ($g_0$) are significant at the 5% level, indicating a clustering pattern in the short-term return variance. Turning to the long-term component, we can see that the $RV$ is significant at the 5% level in all three models, while the weight $w_2$ is only significant at the 10% level. In order to have the same degree of smoothness for all the variables, we use $w_2$ estimated from the model with only $RV$, when we augment the model with macroeconomic variables. The results show that the level of $PC_1$ is significant along with $RV$ but not its variance. However, if we exclude $RV$ from the equation of the long-term component, both the levels and the variance of $PC_1$ are significant. This shows that $RV$ captures the effect of the variance of $PC_1$. $RV$ is still significant at the 5% level when we use $PC_2$ as a macroeconomic variable. The parameter for the variance of $PC_2$ is also significant but at the 10% level. However, only the level of $PC_2$ is found to be significant if we exclude $RV$ from the model.
We may conclude that the joint effects of the economic variables, captured by $PC_1$ and $PC_2$, contain some information about the driving force of stock market return variance.

In Figure 3 we compare the estimated short-term, long-term and total variance from the GARCH-MIDAS model where we only use the realized volatility in the MIDAS equation ($RV$ model). In the first part of the estimation window, despite some large peaks in the short-term variance (possibly due to the Asian crisis) the long-term variance is quite low. After 2000 we observe a substantial increase in the long-term variance component, while the short-term component is below the long-term component most of the time.

Figure 4 illustrates the estimated long-term component of the return variance given by the MIDAS equation, for the first in-sample period. We compare the results from the $RV$ model with two alternative specifications, the $RV$ model augmented with a macroeconomic variable and a model which only includes the macroeconomic variable. In the first graph the macroeconomic variables are represented by $PC_1$, while in the second graph we present the estimated variances with $PC_2$. It shows that the estimated variance from the model $RV+PC_1$ follows mostly that from the $RV$ model, while the $PC_1$ model moves quite differently. Comparing all three models, it seems that the $RV+PC_1$ model combines the two other models, where $RV$ determines the variations and $PC_1$ affects mostly the level of the estimated variance. All three models give a relatively similar pattern, most of the time, when we use $PC_2$ as the macroeconomic variable.

### 4.3. Out-of-sample prediction

In this section, we analyze the ability of the GARCH-MIDAS model to forecast the long-term monthly variances (see equations (3) and (4)), and the total daily variances (see equation (5)). The parameters are obtained using a rolling ten-year estimation window and are held constant during the subsequent year. Our out-of-sample forecast covers the period from January 2004 to June 2008. We use three alternative MIDAS specifications: the $RV$ model that only includes the realized volatility of stock returns, the $RV+X_l+X_v$ model that includes the realized return volatility as well as the level and the variance of the economic variables, and finally the $X_l+X_v$ model with only the level and the variance of the economic variables. As our primary choice of the macroeconomic variables in the GARCH-MIDAS model, we use the two first principal components, $PC_1$ and $PC_2$. We use a ten-year estimation window and keep the parameters over the subsequent year. The first estimation window starts in January.
1994 and ends in December 2003. Table 4 reports the prediction performance of all the models using MSE and the DM test.

As a benchmark we estimate the GARCH (1, 1) model, where we use monthly observations for comparison with the GARCH-MIDAS long-term variance component and daily observations when we compare it with the GARCH-MIDAS total variance. The estimated MSE is based on the deviation between the variance forecasted and the realized variance, where the realized monthly variances are estimated as the sum of daily squared returns in each month, and the realized daily variances are the squared daily returns.

The left panel of Table 4 shows the results for the long-term variance component. The GARCH-MIDAS model with $RV + PC_1$ has the lowest MSE values for monthly predictions. This result is confirmed by the DM-test. (In order to save space, we only report the DM-test when using the traditional GARCH and GARCH-MIDAS as the benchmark models.) The model $RV + PC_1$ significantly outperforms both the GARCH model and the $RV$ model in the long-term variance prediction. The GARCH-MIDAS model without any economic variable performs better than GARCH but the difference between the models forecast is not statistically significant. The models with $PC_1$ and $PC_2$ alone, as a long-term variance driving factor, perform very poorly and are significantly worse than both GARCH and the $RV$ model.

In the right panel of the table, we display the findings from daily variance predictions. The $RV + PC_1$ model still performs better than the other models, but the differences are very small and statistically insignificant. In fact all the models perform better than the GARCH model.

In Figure 5, we plot the results of the regression of the realized volatility on the predicted variance. In general, if the predicted variance has some information about the future realized volatility, then the slope parameter should be significantly different from zero. Furthermore, for an unbiased prediction we expect the intercept parameter to be zero and the slope parameter to be equal to one. The first graph shows the $t$-statistics for the intercept for both daily and monthly variance predictions, and the slope parameters for daily and monthly variance predictions are presented in the second and third diagrams, respectively. According to the results above, the $RV + PC_1$ model shows a very strong ability in forecasting both long-term (monthly) and total (daily) variances; it has a very close to zero intercept and a close to one slope estimation in both predictions. None of the other models share these properties for both predictions; for example the $RV$ model performs well at the daily prediction but its slope is not significantly different from zero in the monthly prediction.
All in all, our out-of-sample analysis shows that adding proper macroeconomic information, measured by $PC_1$, to the long-term variance component of the GARCH-MIDAS model significantly enhances the prediction ability of the model. It is interesting to analyze the forecasting ability of the different macroeconomic variables separately. Figure 6 plots the DM-test result of the $RV+X_l+X_v$ model, using individual macroeconomic variables and the two principal components, and that of the $RV$ model. The GARCH (1, 1) model is used as the benchmark to compute the test statistics. According to the figure, all the statistics are negative, which implies that all the models give a lower forecast error than the GARCH model, in both monthly and daily predictions. However, the test is only significant for monthly predictions and for three cases, i.e., the specifications with $PC_1$, interest rate, and default. Since both the interest rate and default are highly correlated with $PC_1$, the strong out-of-sample performance of the model with $PC_1$ can, to a large extent, be related to these two variables.

5. Conclusion

In this paper, we have used the GARCH-MIDAS approach to forecast future variances. To estimate the long-term component of the variance, in addition to the smoothed realized volatility we use information from macroeconomic variables. A principal component approach is employed to combine the information from a large number of variables, which include interest rate, unemployment rate, term premium, inflation rate, exchange rate, default rate, and industrial production growth rate. We use a rolling window to estimate the parameters of the model and to make a forecast for out-of-sample variances. We compare the forecasting ability of GARCH-MIDAS models with the traditional GARCH model. Our findings show that the GARCH-MIDAS model constitutes a better forecast than the traditional GARCH model. We show that including the low-frequency (monthly) macroeconomic information not only significantly enhances the forecasting ability of the model for the long-term (monthly) variance, but also improves the prediction ability of the model for high-frequency (daily) variances. However, the latter result is not statistically significant based on the DM-test. The GARCH-MIDAS model that includes the first principal component outperforms all other specifications. The strong performance of the first principal component may be motivated by its close connection to the variables short term
interest rate and the default rate, which makes the first principal component a good proxy of the business cycle.

The paper contributes to existing literature by (1) augmenting the long-term component (MIDAS equation) with macroeconomic variables and (2) investigating the forecasting ability of the GARCH-MIDAS model.
References


Table 1. Correlation between variables

The table shows the correlation between monthly observations on the macroeconomic variables and the realized monthly volatility of the US stock return ($RV$). The macroeconomic variables are the yield on a three-month US Treasury bill ($Int.\ rate$), the yield spread between a ten-year bond and a three-month Treasury bill ($Slope$), the unemployment rate ($Unemp$), the growth rate in the industrial production ($Ind.\ prod$), the monthly changes in the consumer price index ($Inflation$), the monthly changes in the exchange rate ($Exch$) and the spread between Moody’s Baa and Aaa corporate bond yields ($Default$). Data cover the period from January 1991 to June 2008.

<table>
<thead>
<tr>
<th></th>
<th>$RV$</th>
<th>$Int.\ rate$</th>
<th>$Slope$</th>
<th>$Unemp$</th>
<th>$Ind.\ prod$</th>
<th>$Inflation$</th>
<th>$Exch$</th>
<th>$Default$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RV$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Int.\ rate$</td>
<td>-0.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Slope$</td>
<td>-0.18</td>
<td>-0.70</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Unemp$</td>
<td>-0.33</td>
<td>-0.37</td>
<td>0.80</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Ind.\ prod$</td>
<td>-0.15</td>
<td>0.12</td>
<td>0.04</td>
<td>0.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Inflation$</td>
<td>-0.17</td>
<td>0.39</td>
<td>0.16</td>
<td>0.56</td>
<td>-0.02</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Exch$</td>
<td>0.43</td>
<td>-0.07</td>
<td>0.13</td>
<td>-0.19</td>
<td>-0.02</td>
<td>-0.09</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$Default$</td>
<td>0.30</td>
<td>-0.48</td>
<td>0.23</td>
<td>0.12</td>
<td>-0.25</td>
<td>-0.01</td>
<td>0.05</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 2. The correlation of principal components with the macroeconomic variables

The table shows the correlation between the macroeconomic variables and the principal components (PC) constructed based on these variables. The macroeconomic variables are the yield on a three-month US Treasury bill (Int. rate), the yield spread between a ten-year bond and a three-month Treasury bill (Slope), the spread between Moody’s Baa and Aaa corporate bond yields (Default), the monthly changes in the exchange rate (Exch), the monthly changes in the consumer price index (Inflation), the growth rate in the industrial production (Ind. prod) and the unemployment rate (Unemp). Data cover the period from January 1991 to June 2008.

<table>
<thead>
<tr>
<th></th>
<th>Int. rate</th>
<th>Slope</th>
<th>Unemp</th>
<th>Ind. prod</th>
<th>Inflation</th>
<th>Exch</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pc1</td>
<td>-0.78</td>
<td>0.93</td>
<td>0.83</td>
<td>-0.10</td>
<td>0.21</td>
<td>0.02</td>
<td>0.49</td>
</tr>
<tr>
<td>Pc2</td>
<td>-0.54</td>
<td>-0.06</td>
<td>-0.53</td>
<td>-0.28</td>
<td>-0.83</td>
<td>0.36</td>
<td>0.43</td>
</tr>
<tr>
<td>Pc3</td>
<td>0.14</td>
<td>-0.25</td>
<td>-0.02</td>
<td>-0.82</td>
<td>0.33</td>
<td>-0.27</td>
<td>0.44</td>
</tr>
<tr>
<td>Pc4</td>
<td>0.19</td>
<td>0.09</td>
<td>-0.03</td>
<td>-0.16</td>
<td>0.33</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>Pc5</td>
<td>0.10</td>
<td>-0.16</td>
<td>-0.09</td>
<td>0.46</td>
<td>0.17</td>
<td>0.01</td>
<td>0.61</td>
</tr>
<tr>
<td>Pc6</td>
<td>-0.19</td>
<td>-0.14</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.12</td>
<td>0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>Pc7</td>
<td>0.02</td>
<td>-0.12</td>
<td>0.17</td>
<td>0.00</td>
<td>-0.09</td>
<td>0.04</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 3. Estimated parameters of the GARCH-MIDAS model

The table shows the estimated parameters of the GARCH-MIDAS model with different specifications of the MIDAS equation. The first row of the table presents the results of the model with only the realized volatility ($RV$) of returns in the MIDAS equation, while the remaining rows present the estimated parameters when we also include the level and the variance of the economic variables $X_l$ and $X_v$ in the MIDAS equation. We only present the results obtained for the first and the second principal components constructed based on seven macroeconomic variables. The values reported in the table are in percentage. Data cover the first estimation period starting in January 1991 and ending in December 2003.

<table>
<thead>
<tr>
<th></th>
<th>mu</th>
<th>alpha</th>
<th>beta</th>
<th>m</th>
<th>RV</th>
<th>level</th>
<th>var</th>
<th>$w_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RV$</td>
<td>0.072**</td>
<td>0.086**</td>
<td>0.887**</td>
<td>-0.634**</td>
<td>0.031**</td>
<td></td>
<td></td>
<td>2.677*</td>
</tr>
<tr>
<td>$PC_1$</td>
<td>$RV+X_l+X_v$</td>
<td>0.075**</td>
<td>0.090**</td>
<td>0.861**</td>
<td>-0.814**</td>
<td>0.034**</td>
<td>-0.219**</td>
<td>-2.004</td>
</tr>
<tr>
<td></td>
<td>$X_l+X_v$</td>
<td>0.072**</td>
<td>0.071**</td>
<td>0.924**</td>
<td>0.848</td>
<td>-0.438**</td>
<td>-12.983**</td>
<td></td>
</tr>
<tr>
<td>$PC_2$</td>
<td>$RV+X_l+X_v$</td>
<td>0.075**</td>
<td>0.099**</td>
<td>0.860**</td>
<td>-1.143**</td>
<td>0.038**</td>
<td>0.107</td>
<td>2.917*</td>
</tr>
<tr>
<td></td>
<td>$X_l+X_v$</td>
<td>0.072**</td>
<td>0.082**</td>
<td>0.900**</td>
<td>-0.115</td>
<td>-0.295**</td>
<td>2.677</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Comparisons of the out-of-sample prediction errors

The table shows the results of the estimated mean square error (MSE) and DM-test for the out-of-sample performance of the different models in predicting daily and monthly variances. We use three alternative specifications in the MIDAS equation, a model that includes only the realized volatility of stock returns (RV model), a model that includes the realized return volatility as well as the level and the variance of the economic variables (RV+X_l+X_v), and finally a model with only the level and the variance of the economic variables (X_l+X_v). The left panel shows the results for the long-term variance component, \( \tau \) in equations (3) and (4), while the right panel shows the results for the conditional daily total variance (equation (5)). The GARCH-MIDAS results are compared with corresponding GARCH estimations. As the macro variables we use the two first principal components, \( PC_1 \) and \( PC_2 \), in the MIDAS equation. We use a ten-year estimation window and keep the parameters over the subsequent year. The first estimation window starts in January 1994 and ends in December 2003. The realized monthly variances are estimated as the sum of daily squared returns in each month, while for the realized daily variances we use the squared daily returns. Out-of-sample forecasts cover the period from January 2004 to June 2008. The minus (plus) sign in each cell indicates that the model given in the row performs better (worse) than the model given in the column. An asterisk implies a significant difference in the performance.

<table>
<thead>
<tr>
<th></th>
<th>Long-term variance</th>
<th>Total variance</th>
<th></th>
<th>Long-term variance</th>
<th>Total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>GARCH</td>
<td>RV model</td>
<td>MSE</td>
<td>GARCH</td>
</tr>
<tr>
<td>GARCH</td>
<td>174.18</td>
<td>+</td>
<td></td>
<td>1.71</td>
<td>+</td>
</tr>
<tr>
<td>RV model</td>
<td>171.53</td>
<td>-</td>
<td></td>
<td>1.69</td>
<td>-</td>
</tr>
<tr>
<td>RV+PC_1</td>
<td>133.19</td>
<td>-*</td>
<td>-*</td>
<td>1.68</td>
<td>-</td>
</tr>
<tr>
<td>RV+PC_2</td>
<td>225.28</td>
<td>+</td>
<td>+</td>
<td>1.69</td>
<td>-</td>
</tr>
<tr>
<td>PC_1</td>
<td>219.98</td>
<td>+*</td>
<td>+*</td>
<td>1.70</td>
<td>-</td>
</tr>
<tr>
<td>PC_2</td>
<td>233.32</td>
<td>+*</td>
<td>+*</td>
<td>1.70</td>
<td>-</td>
</tr>
</tbody>
</table>
The upper graph shows the behaviour of weights as the function of the number of lags using different values for $w_1$ and $w_2$. We select two alternative values for $w_1$ (1 and 2) and two values for $w_2$ (4 and 8). In the lower graph, we plot the maximized value of log likelihood function of the GARCH-MIDAS model with different lag values. The long term component (MIDAS equation) includes only the realized return volatility.
Figure 2. Plot of the realized volatility and the economic variables

The figure illustrates the monthly realized volatility of the return and movements of the selected macroeconomic variables, as well as the first principal component constructed based on the macroeconomic variables. The data range from January 1991 to June 2008.
Figure 3. Comparison of the long-term, short-term and total variance

The figure illustrates the long-term, short-term and total variances estimated by the GARCH-MIDAS model. The MIDAS equation only includes the realized volatility of stock returns \((RV\ model)\). The estimation period covers the period from January 1991 to December 2003, while a sample of 36 monthly observations is used to estimate the exponentially moving average of the realized volatility in the MIDAS equation.
Figure 4. Estimated long-term variance

The figure illustrates the estimated long-term variance, $\tau_t$, based on three alternative specifications of the MIDAS equation, a model that includes only the realized volatility of stock returns ($RV$), a model that includes the realized return volatility as well as the level and the variance of the economic variables ($RV + X_l + X_v$), and finally a model with only the level and the variance of the economic variables ($X_l + X_v$). We illustrate the results for the first two principal components constructed based on seven macroeconomic variables. The estimation period covers the period from January 1991 to December 2003, while a sample of 36 monthly observations is used to estimate the exponentially moving average of the included variables in the MIDAS equation.
Figure 5. Regression of the realized volatilities on the predicted variances

The figure plots the results of the estimated parameters from the regression of the realized volatility on the predicted variance. The first figure plots the $t$-statistics for the intercept and the second and third figures give the slope parameters for monthly and daily variance prediction, respectively, and the related 95% confidence intervals. We use three alternative MIDAS specifications: $RV$ includes only the realized volatility of stock returns, $RV+X_l+X_v$ includes the realized return volatility and the level and the variance of the economic variables, $X_l+X_v$ contains only the level and the variance of the economic variables. As economic variables, we use two first principal components, $PC_1$ and $PC_2$, in the MIDAS equation. The results of the GARCH-MIDAS are compared with corresponding GARCH estimations. The realized monthly volatility is estimated as the sum of daily squared returns in each month, while the realized daily volatility is computed as the squared daily return.
Figure 6. DM-test of the individual macrovariables

The figure shows $t$-values of the DM test for the out-of-sample performance of the different models in predicting daily and monthly variances. It indicates the contribution of each macroeconomic variable, $PC_1$ and $PC_2$, in order to improve the prediction of long-term variance. We use two alternative specifications in the MIDAS equation, a model that includes only the realized volatility of stock returns ($RV$ model), and a model that includes the realized return volatility as well as the level and the variance of the economic variables ($RV+X_l+X_v$).
Importance of the macroeconomic variables for variance prediction A GARCH-MIDAS approach

HOSSEIN ASGHARIAN | AI JUN HOU | FARRUKH JAVED

This paper applies the GARCH-MIDAS (Mixed Data Sampling) model to examine whether information contained in macroeconomic variables can help to predict short-term and long-term components of the return variance. A principal component analysis is used to incorporate the information contained in different variables. Our results show that including low-frequency macroeconomic information in the GARCH-MIDAS model improves the prediction ability of the model, particularly for the long-term variance component. Moreover, the GARCH-MIDAS model augmented with the first principal component outperforms all other specifications, indicating that the constructed principal component can be considered as a good proxy of the business cycle.

Keywords: Mixed data sampling, long-term variance component, macroeconomic variables, principal component, variance prediction.

THE KNUT WICKSELL CENTRE FOR FINANCIAL STUDIES
The Knut Wicksell Centre for Financial Studies conducts cutting-edge research in financial economics and related academic disciplines. Established in 2011, the Centre is a collaboration between Lund University School of Economics and Management and the Research Institute of Industrial Economics (IFN) in Stockholm. The Centre supports research projects, arranges seminars, and organizes conferences. A key goal of the Centre is to foster interaction between academics, practitioners and students to better understand current topics related to financial markets.