Nowcasting Finnish GDP growth using financial variables: a MIDAS approach*

Olli-Matti Laine
Bank of Finland and Tampere University
olli-matti.laine@bof.fi

Annika Lindblad
Bank of Finland
annika.lindblad@bof.fi

Abstract

We analyse the performance of financial market variables in nowcasting Finnish quarterly GDP growth. In particular, we assess if prediction accuracy is affected by the sampling frequency of the financial variables. Therefore, we apply MIDAS models that allow us to nowcast quarterly GDP growth using monthly or daily data without temporal aggregation in a parsimonious way. We find that financial market data nowcasts Finnish GDP growth relatively well: nowcasting performance is similar to industrial production, but financial market data is available much earlier. Our results suggest that the sampling frequency of financial market variables is not crucial: nowcasting accuracy of daily, monthly and quarterly data is similar.

Keywords: *MIDAS, Nowcasting, Financial markets, GDP*
JEL codes: *E44, G00, E37*

**We thank Eleonora Granziera and two anonymous referees for useful and insightful comments on this research. We also thank Rasmus Rannikko for helping us with the data. This paper represents the views of the authors which are not necessarily those of the Bank of Finland. Any remaining errors are solely our own responsibility.*

1. Introduction

Financial markets provide high-frequency information about investors' expectations. This information may be useful in predicting or nowcasting GDP growth because asset prices are based on expected future cash flows, which in turn are linked to macroeconomic conditions.[1] In situations where the economic environment is changing quickly, daily or weekly updates on economic conditions can be crucial for forming an accurate and timely view of the economy.

The previous literature has shown that including financial market variables in short-term forecasting or nowcasting models for economic activity is useful (e.g., Friedman and Kuttner, 1993; Estrella and Mishkin, 1998; Henry, Olekalns and Thong, 2004; Chionis, Gogas and Pragidis, 2010; Nyberg, 2010). Junttila and Korhonen (2011) show that dividend yields and short-term interest rates are relevant for forecasting output growth in multiple countries. According to their results, financial variables are especially useful in turbulent times. Kuosmanen and Vataja (2014) find similar results for Finland. Although many studies provide evidence of the usefulness of financial market information in forecasting, Alessi, Ghysels, Onorante, Peach and Potter (2014) argue that central banks have not utilised this information in the best possible way.

Despite many studies analysing the predictive power of financial market variables, the optimal sampling frequency for financial market data in a nowcasting model for GDP has not been widely studied. Should we care about, for example, daily fluctuations, or is it better to consider temporally aggregated data? Typically, if the variables in a model are measured at different frequencies, the high-frequency variables are aggregated to the level of the lowest frequency variable. However, temporal aggregation inevitably ignores some of the information in the daily financial market data. Some of this information could, nevertheless, be important for nowcasting GDP growth and ignoring it could confound the relationship between the variables (see, e.g., Lütkepohl (2010) for a discussion on forecasting with aggregated data). Time series data can have a decaying memory structure, and thus giving the same weight to each daily observation within a quarter is not necessarily optimal. Higher frequency data also enables nowcasts to be updated frequently and during the on-going quarter, which is naturally a crucial feature of nowcasting models. This could be especially useful during the beginning of a recession, when traditional macroeconomic variables are slow to react and additionally suffer from publication lags. On the other hand, daily data tends to be noisier than temporally aggregated data. As a lower signal-to-noise ratio could obscure the relationship between the variables and therefore lead to a deterioration in forecast accuracy, high-frequency data does not automatically improve predictions. It is therefore clear that the sampling frequency could impact the accuracy of a nowcasting model. Analysing the effect of the sampling frequency is also important for understanding the relationship between variables. Due to the availability of daily data and mixed-frequency methods, it is possible to include high-frequency data in a quarterly model for GDP growth and provide policy makers with (close to) real-time GDP nowcast updates.

With the development of modelling frameworks being able to handle data sampled at different frequencies, the question of whether using higher frequency data directly to nowcast and forecast lower frequency data improves accuracy has naturally arisen. Ghysels, Santa-Clara and Valkanov (2004) propose a Mixed Data Sampling (MIDAS) regression framework to overcome the issues related to different sampling frequencies without relying on temporal aggregation. This framework has turned out to be useful in forecasting returns and volatility in financial markets. In these applications, the sampling frequency of the explanatory variable seems to have an impact on the results (Ghysels, Sinko and Valkanov, 2007). For example, Ghysels, Santa-Clara and Valkanov (2006) conclude that intra-daily data does not outperform daily data (aggregated from the high-frequency data) in forecasting realized volatility. They also show that the most recent days are the most important when predicting monthly or weekly financial market volatility.

MIDAS models have also been widely used to nowcast and forecast quarterly GDP growth using mainly monthly predictors. For example, Clements and Galvão (2008, 2009) (using US data), Marcellino and Schumacher (2010) (using German data) and Kim and Swanson (2017) (using Korean data) show that monthly macroeconomic data, and especially monthly data on the current quarter, improves quarterly output growth forecasts in different MIDAS specifications. In addition,

[1] *Nowcasting means the prediction of the current state of the variable of interest. Many economic time series are published with a substantial lag. Therefore, economists do not only predict the future but also the present and the past.*

Armesto, Engemann and Owyang (2009) concludes that using a MIDAS weighting scheme for monthly employment growth improves forecast accuracy over an equal-weighted quarterly average when forecasting US GDP growth, especially over a short forecasting horizon. There is, therefore, some evidence that increasing the sampling frequency from quarterly to monthly improves nowcasting performance.

The literature closest to our paper is that which considers the relationship between financial market data and output growth in a mixed frequency setting. For example, Andreou, Ghysels and Kourtellos (2013) (using US data), Ferrara and Marsilli (2013) (using euro area data and data for the four largest euro area countries), Ferrara, Marsilli and Ortega (2014) (using US, UK and French data) and Marsilli (2014) (using US data) have shown using MIDAS models that daily or monthly financial market variables include useful information for GDP nowcasts and forecasts not included in macroeconomic data. However, these papers do not focus on assessing the relevance of the sampling frequency or discuss in detail the relationship between individual financial variables and GDP growth. In other words, most of the earlier literature does not consider whether, for example, daily or monthly variation in asset prices includes useful information for nowcasting GDP growth not included in the same data aggregated to a lower frequency. For example, a decline in asset prices mid-quarter could provide a useful indication of near-term economic activity, which could be obscured if a quarterly average was used. The choice of sampling frequency is, however, discussed in Tsui, Xu and Zhang (2018), who conclude that the best frequency of stock market returns is weekly when forecasting output growth for Singapore with a MIDAS model. In addition, Gómez-Zamudio and Ibarra (2017) (using Mexican data) and Doğan and Midiliç (2019) (using Turkish data) conclude that increasing the frequency of financial market data from quarterly to daily improves the accuracy of output growth forecasts using MIDAS and factor MIDAS models. On the other hand, Tay (2007) finds that using daily stock market returns aggregated over a year generally leads to better output growth nowcasts for the US than using a non-parametric MIDAS specification. The evidence regarding the optimal sampling frequency of different financial variables is thus scarce. In addition, most of the previous literature does not explore whether any specific days within a quarter are more important than others. Thus, based on the previous literature it is not possible to conclude whether an equal-weighted average is a suitable aggregation method for financial data when nowcasting GDP growth.

In this paper we study the usefulness of including daily financial market variables in a nowcasting model for quarterly Finnish GDP growth in MIDAS framework. We also consider whether the choice of sampling frequency matters for nowcasting Finnish GDP growth. Determining the impact of the sampling frequency amounts to studying whether any important information available in the higher frequency data is lost in the aggregation. However, as seen in the previous literature, there could be a point beyond which increasing the frequency of the explanatory data is not beneficial. Determining this point is an empirical question and therefore specific to the country and variable under consideration. Here we consider the question for Finnish GDP growth using a relatively broad range of both domestic and foreign financial variables. MIDAS models allow us to study how temporal aggregation of variables affects prediction accuracy and it allows us to determine whether a simple average is a satisfactory aggregation method, or whether a more flexible weighting scheme is necessary to correctly characterise the relationship.

We also assess different ways to utilise financial market variables in the MIDAS framework. The earlier literature has considered, for example, individual variables, forecast combinations and principal components (Andreou et al., 2013; Ferrara et al., 2014). The previous papers discussing the gains of high-frequency data for nowcasting GDP growth focus on the US or other large economies. In order to broaden the literature and the applicability of previous results, our aim here is to provide evidence of these potential gains from the perspective of a small open economy. Therefore, we also include foreign variables, namely from Germany and the US, as the international financial markets might include important information for nowcasting growth in a small open economy.

Our results suggest that it does not matter significantly and systematically for nowcast accuracy whether one uses financial market data at the daily, monthly or quarterly frequency when nowcasting Finnish GDP growth. However, there may be some practical reasons to prefer higher frequency data, such as the ability to update the nowcast on a monthly, weekly or even daily frequency. On the other hand, increasing the frequency also brings some challenges, in particular by increasing the noisiness of the data. Ultimately, our results suggest that the choice of frequency can be made based on data availability and the needs of the forecaster, without significantly compromising nowcast accuracy.

To gauge the importance of financial market variables for nowcasting GDP growth we compare their nowcasting ability to that of industrial production growth, which is a traditional predictor of GDP growth. Our results imply that financial market variables predict GDP growth as accurately as industrial production. Because industrial production is observed with more than a one-month lag, the results suggest that we can, without loss of accuracy, nowcast GDP earlier using financial variables. Different kinds of financial ratios – like the dividend yield – nowcast Finnish GDP growth well, which is in line with the results in Junttila and Korhonen (2011). Our results also provide some evidence that nowcast accuracy can be improved by combining a financial market based nowcasts to a nowcast based on macroeconomic data. However, these improvements are not statistically significant.

The rest of the article is organised as follows. Section 2 introduces the MIDAS regression framework. Section 3 summarises the data. Section 4 studies the nowcasting performance of individual financial market variables and considers the effects of increasing the frequency of the financial market data. Section 5 discusses the best way of utilising financial market data to nowcast Finnish GDP growth. Section 6 concludes.

2. The MIDAS framework

MIDAS models were introduced by Ghysels et al. (2004), Ghysels et al. (2005), Ghysels et al. (2006) and Ghysels et al. (2007). The central idea of the MIDAS approach is to explain a low-frequency variable by variables sampled at higher frequencies. The MIDAS framework is used in this paper because it is a simple nowcasting framework which allows the parsimonious inclusion of several lags of the explanatory data and enables data sampled at different frequencies to be included into the same model.[2] Although the modelling framework is linear, the weight functions allow complex relationships between the dependent and independent variables. An important and useful feature of the MIDAS framework is that the number of parameters to be estimated does not depend on the sampling frequency of the data or the length of the sample period. The chosen framework therefore enables us to compare in a straightforward way the nowcasting performance of several variables and compare their individual performance when sampled at different frequencies.

The standard MIDAS model with one explanatory variable can be written as follows:

$$y_t = \beta_0 + \beta_1 \sum_{h=0}^{d} \theta_h \, x_{tm-h} + u_t, \tag{1}$$

where y_t is a low-frequency variable (GDP growth in our models), x_{tm-h} is a high-frequency variable (a financial market variable or industrial production growth in our models) and d is the number of lags of the explanatory data included in the model.[3] There are m observations of the high-frequency variable to one observation of the low-frequency variable. For example, if we explained a quarterly variable by a monthly variable, m would be 3. If the number of lags, d, was 2, then we would explain the quarterly variable by all the monthly observations of the high-frequency variable from the given quarter. If $\beta_1 \neq 0$, there is a connection between the low-frequency and the high-frequency variables. The function θ_h is a polynomial that weights the contemporaneous observation of the high-frequency variable and its lags in a parsimonious way. In this paper we use the (normalised) exponential Almon lag polynomial:

[2] *The MIDAS model is not the only mixed-frequency method available (see, for example, Foroni and Marcellino (2013) for a survey on various mixed-frequency methods). For example, Kuzin, Marcellino and Schumacher (2011) compares the MIDAS model and a mixed-frequency VAR (MF-VAR) approach for nowcasting and forecasting monthly GDP in the euro area. They find in their empirical application that the MIDAS model tends to perform better for shorter horizons, while the MF-VAR model performs better for longer horizons. On the other hand, Franta, Havrlant and Rusnák (2016) finds that the dynamic factor model with mixed frequency data performs better than MIDAS models for nowcasting GDP in the Czech Republic. The results are similar in Galli, Hepenstrick and Scheufele (2019) for Switzerland, except that the MIDAS models slightly outperform the dynamic factor model after the financial crisis.*

[3] *The AR-MIDAS model by Andreou et al. (2013), which includes lagged values of the dependent variable, could be used for determining whether financial market data include useful information for nowcasting GDP in addition to lagged GDP. However, in this paper we concentrate on evaluating the usefulness of different financial market variables and the impact of varying their sampling frequency in a MIDAS model for GDP. Thus, exploring the AR-MIDAS model in this context is left for future work.*

$$\theta_h = \frac{e^{\lambda_1(h+1)+\lambda_2(h+1)^2}}{\sum_{s=0}^{d} e^{\lambda_1(s+1)+\lambda_2(s+1)^2}}. \tag{2}$$

The normalisation ensures that the weights sum up to one, which allows the separate identification of β_1. The exponential Almon lag polynomial allows flexible, such as decaying or hump-shaped, weighting schemes. Parameters λ_1 and λ_2 are estimated simultaneously with the other model parameters, and together with the number of lags they govern the shape of the weighting scheme. If $\lambda_1 = \lambda_2 = 0$, the lags have equal weights ($1/d$) and we essentially include a moving average of the past d lags in the MIDAS model. Note that in this case using temporally aggregated data, by taking an average over the lowest frequency, yields the same result. The benefit of the MIDAS framework over temporal aggregation is thus that the MIDAS model allows the data to decide the weights and therefore enables taking into account the decaying memory structure often present in time series data. The MIDAS model is estimated using non-linear least squares (NLS).

Due to the lag polynomial, MIDAS models are especially useful when the number of lags is large, as they allow including, for example, many daily lags without increasing the parameter space. However, when only a few lags are included an unrestricted MIDAS (U-MIDAS) model, which does not include a weighting scheme but estimates a separate regression parameter for each lag, can be used (Foroni, Marcellino and Schumacher, 2015). The U-MIDAS regression model can be written as:

$$y_t = \beta_0 + \sum_{h=0}^{d} \beta_h x_{tm-h} + u_t. \tag{3}$$

In this case the model can be estimated using OLS.

3. Data

We use daily, monthly and quarterly data from Q2/2002 to Q3/2019. The beginning of the sample period is, firstly, restricted by the availability of daily data. For example, many financial ratios are available to us from 2002 onwards. Secondly, using data only from the early 21st century reduces the risk of structural breaks. This sample period does, however, include two recessions for Finland and thus covers, in our opinion, enough business cycle variation for drawing conclusions. The Finnish GDP data is from Statistics Finland. We use the latest vintage of the quarterly real GDP growth rate.[4]

We use a comprehensive set of 28 financial market variables, covering both domestic and foreign data. All the financial data is obtained via Bloomberg. Andreou et al. (2013) uses a much wider data set of 991 daily series, but, for example, Ferrara et al. (2013), Ferrara et al. (2014) and for Finland Kuosmanen and Vataja (2014) use a significantly narrower set of financial market data. We assess the predictive power of stock indices and interest rates from Finland (OMX Helsinki), Germany (DAX) and the USA (S&P 500). In addition, we use the average price-to-earnings ratios, price-to-book ratios and dividend yields from the same countries. The averaging utilises the same weights that have been used in the construction of the stock market indices. The ratios are calculated using past information about earnings and dividends from the past 12 months.[5] The last available information is used for the book value. The predictive power of the oil price, expected stock market volatility implied by Eurostoxx 50 index options and the EUR/USD exchange rate are also considered.

The stock indices and the oil price are in log-differences. Interest rates and the exchange rate are in differences. We also calculate the spread between the German 10-year yield and the 12-month yield. Financial ratios, stock market volatility and the interest rate spread are in levels. For comparison, we nowcast GDP growth also using the monthly growth rate

---

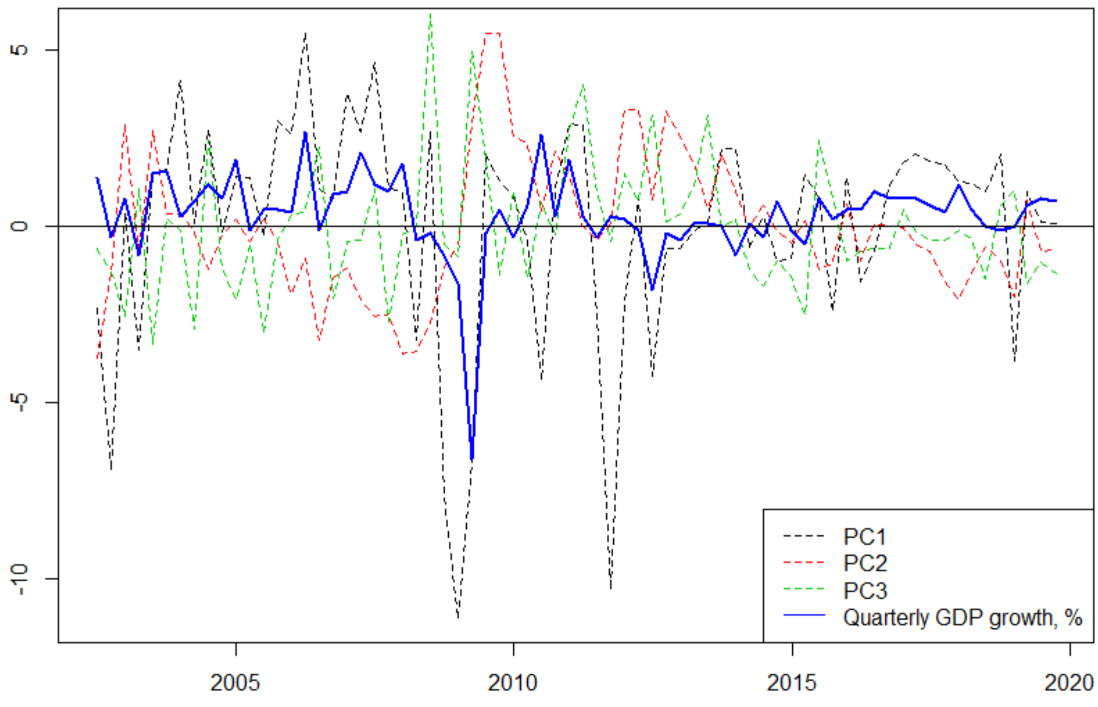
[4] *GDP statistics are typically revised substantially. We use the latest available vintage at the time of writing (data collected on December 31st, 2019) as we consider it the best available estimate for the actual, final growth rate. However, we recognize that there are also arguments for using an earlier vintage, such as first release data.*

[5] *This calculation method causes some minor revisions to financial ratios. The series are recalculated every day using the latest information about the past equity level fundamentals. This may potentially overstate the nowcasting performance of financial ratios.*

of industrial production, for which the latest vintage is used.[6] The transformations are done to achieve stationarity and are based on previous studies (for example, Bernanke, Boivin and Eliasz, 2005; Becker, Lee and Gup, 2012; Lewellen, 2004; Koustas and Serletis, 2005; Marcellino and Schumacher, 2010). The time series included in the data set are listed (together with their sources and transformations) and plotted in Appendix A.

To summarise the financial market information, we use principal component analysis to extract common factors from the financial market data. Figure 1 plots the quarterly GDP growth rate together with the first three principal components (PCs) based on standardised data (excluding industrial production). The principal components are here calculated from quarterly data for visual reasons. The principal components capture the common variation on the financial markets over time. Figure 1 shows that the relationship between financial markets and GDP growth seems to be time-varying and at its strongest during turbulent times.[7] Especially the first PC seems to capture well some of the (negative) spikes in GDP growth. Corresponding figures for the individual financial market variables can be found in Appendix A.

Figure 1. *This figure shows the development of the first three principal components together with quarterly GDP growth. The principal components are calculated from the data summarised in Appendix A (excluding industrial production). The sample is from Q2/2002 to Q3/2019 and the frequency is quarterly.*



6 *Using revised data may potentially overstate the nowcasting performance of industrial production.*

7 *The first PC has a correlation of 0.48 with GDP growth, the second PC a correlation of -0.24, and the third PC a correlation of -0.35.*

4. Nowcasting GDP growth using financial market data

In this section we assess how well different variables sampled at different frequencies nowcast GDP growth. We conduct a rolling window analysis, where the first estimation sample is from Q2/2002 to Q4/2011 and the first out-of-sample observation is Q1/2012. We have chosen 2011 to be the end of the first estimation sample in order to see how well the models predict the sharp decline in Finnish quarterly GDP growth in Q2 2012.[8] Altogether we produce 31 pseudo out-of-sample nowcasts, which, although a relatively small number of out-of-sample observations, is of similar length as the out-of-sample period in, for example, Andreou et al. (2013). To plausibly estimate the parameters for the first out-of-sample observation, the length of the out-of-sample period cannot be significantly increased, as increasing the number of out-of-sample observations naturally decreases the number of observations used in the estimation. As a robustness check Appendix E reports the results when 40 out-of-sample observations are used.

In models using daily data we include 62 daily lags. In our sample, every quarter includes at least 63 working days. By choosing 62 daily lags we avoid technical issues relating to "missing data" in the estimation.[9] In models with monthly data the number of lags is two and in models with quarterly data the number of lags is zero, which means that only the contemporaneous quarter is included. Regardless of the frequency used, all models therefore rely on the same information set and utilise data from one quarter only. We include additional lags in Section 5.

The models using quarterly data are naturally unrestricted as there are no lags. For the monthly models we consider both the restricted and the unrestricted versions as the number of lags is small. The daily models are restricted, as it is infeasible to estimate 62 individual regression parameters. The parameters are re-estimated every period in the rolling window analysis. Appendix B reports some estimation results for the whole sample (Q2/2002 to Q3/2019).

[8] As Q2/2012 can be considered an outlier, Appendix E reports the results excluding the out-of-sample forecast for Q2/2012. These results are discussed briefly in Section 5.

[9] We use the R package midasr by Virmantas Kvedaras and Vaidotas Zemlys-Balevicius to estimate the models (see also Ghysels, Kvedaras and Zemlys, 2016).

Table 1: *Out-of-sample RMSEs of MIDAS regression models. The models are ordered based on the RMSEs of the quarterly models. The abbreviations for the variables are explained in Appendix A. The RMSEs are calculated from a rolling window analysis, in which the first estimation sample is from Q2/2002 to Q4/2011 and the first out-of-sample nowcast is Q1/2012. Thus, the results are based on 31 out-of-sample observations.*

Explanatory variable	Quarterly	Monthly	Monthly (unrestricted)	Daily
SP500 p/e	0.52	0.55	0.56	0.55
DAX Dividend yield	0.58	0.58	0.60	0.58
OMX Hels Dividend yield	0.60	0.66	0.66	0.61
OMX Hels p/b	0.60	0.63	0.72	0.64
PC1	0.63	0.61	0.61	0.64
SP500 p/b	0.64	0.66	0.76	0.67
OMX Hels p/e	0.65	0.69	0.64	0.68
FI_10y	0.66	0.85	0.95	0.67
OIL	0.66	0.66	0.70	0.67
DE 10y-1y	0.66	0.66	0.65	0.66
DE_10y	0.67	0.77	0.91	0.66
DE_5y	0.67	0.72	0.72	0.68
DE_7y	0.67	0.77	0.83	0.68
EURUSD	0.67	0.77	0.78	0.67
DAX p/e	0.68	0.67	0.85	0.68
Industrial production	0.68	0.62	0.66	–
OMX Hels Telec	0.69	0.73	0.77	0.83
DE_1y	0.70	0.64	0.71	0.65
FI_5y	0.70	0.76	0.79	0.69
OMX Hels Industrials	0.71	0.73	0.75	0.70
OMX Hels Utilities	0.71	0.66	0.71	0.77
SP500 Dividend yield	0.71	0.70	0.73	0.70
DAX p/b	0.72	0.67	0.80	0.78
OMX Hels Hlth Care	0.72	0.81	0.99	0.67
OMX Hels Technology	0.72	0.70	0.85	0.75
SP500	0.74	0.77	0.72	0.64
OMX Hels	0.75	0.71	0.79	0.67
Eurostoxx 50 volatility	0.77	0.74	0.78	0.77
DAX	0.84	0.78	0.84	0.65
OMX Hels Basic Metal	0.84	0.82	0.87	0.72

Table 1 reports average root-mean-square errors (RMSEs) for the whole out-of-sample period. The predictors are ordered based on the RMSEs of the quarterly models. The average price-to-earnings ratio in the United States has, perhaps surprisingly, been the best predictor for Finnish GDP growth, regardless of the frequency used. The price-to-earnings and price-to-book ratios especially for the Finnish and the US stock markets have also performed well. The dividend yields for the Finnish and German stock markets produce accurate nowcasts, while implied volatility, the EUR/USD exchange rate, and the stock indices tend to perform worse. However, some of the results are sensitive to the sampling frequency of the explanatory variables. For example, industrial production performs relatively better on the monthly frequency, while the S&P 500, OMX Helsinki and DAX indices seem to improve as predictors as the sampling frequency increases. The relative performance of the various interest rates also varies depending on the sampling frequency.

In addition to the individual financial market variables, we also include the first principal component based on the financial market data. It also produces accurate nowcasts, indicating that combining information from a large set of financial market indicators provides useful information for nowcasting Finnish GDP growth. At least the average price-to-earnings ratio in the United States, the German dividend yield and the first PC produce more accurate nowcasts on the monthly frequency than the benchmark model, where industrial production growth is used as a predictor. Thus, regardless of the frequency, some financial variables seem useful for nowcasting Finnish GDP growth, in particular considering that they are available much earlier than industrial production.

In a MIDAS model the impact of an explanatory variable is determined by β_1 in equation (1) as well as by the weighting scheme (θ_h) in equation (2). Their combined effect (β_1* θ_h) is shown in Appendix B for the daily and monthly models reported in Table 1. For several daily financial variables most of the weight is on just the last few days of the quarter. This implies that the daily intra-quarterly variation is not a significant factor for explaining GDP growth and that all the relevant economic information is included in the recent values of the variables. For some variables, such as the 5-year Finnish interest rate, all weight is on the first days of the quarter, implying that potentially a longer lag length would be needed. For series with counterintuitive weighting schemes there could also be an issue with too much noise in the data for the Almon polynomial to be able to discern a stable relationship between the financial variable and GDP growth. Aggregating to a lower frequency could aid in separating the signal from the noise. For some variables, such as the price-to-book ratios and the dividend yields, the weighting schemes are hump-shaped, implying that these variables affect GDP with a lag. The largest weight is attained in all these cases for lags between 10 and 20 days. As expected, the relationship between GDP growth and the price-to-book ratios is positive while the relationship between GDP growth and the dividend yield is negative. For the price-to-earnings ratio we would expect the relationship to be positive, and this is true for the German and US price-to-earnings ratios. Financial market volatility has a negative relationship to GDP growth and the largest weight is on approximately the 50th daily lag. The relationship between the first PC and GDP growth is positive and most of the weight is on lags between approximately 15 and 30 days.

For the monthly models two monthly lags is generally not enough to make the weights decay to zero and thus the daily models seem better specified. On the other hand, for example, for the price-to-book ratios most weight is on the first monthly lag, which is roughly in line with the daily data. For the dividend yields most weight is on the current and first lags, also in line with the daily results. The weighting schemes for the restricted and unrestricted monthly models are mostly similar.

Overall, looking at the weighting schemes it seems that equal weights (such as in the quarterly models) do not represent the relationship between GDP growth and financial market data accurately. Whether this impairs nowcasting performance depends on the variable. As the differences in nowcast accuracy are, however, mostly small, the impact of the sampling frequency on nowcasting performance seems minor.

The results regarding the importance of individual financial market variables are largely in line with the previous literature. For example, the dividend yields perform well, as concluded by Junttila and Korhonen (2011). As a small open economy Finland is significantly affected by global fluctuations, which might explain why foreign variables, such as the average price-to-earnings ratio of the S&P 500, nowcast well. Forecasting US GDP growth Andreou et al. (2013) finds no differences between the importance of various asset classes. Ferrara et al. (2013) use stock returns, oil prices and the term spread on a monthly frequency to nowcast and forecast GDP growth in the euro area, France, Germany, Italy and Spain with mixed results. For nowcasting, stocks and oil prices receive some support in some countries when compared to the benchmark MIDAS model with confidence data.

Some caution is needed when interpreting the results for the models using financial ratios. It is reasonable to assume that, for example, dividends and stock prices have equal growth rates in the long run. Nevertheless, the ratio may deviate from its long run mean very persistently, and there has been some discussion about the stationarity of these variables (e.g., Koustas and Serletis, 2005; Lewellen, 2004; Becker et al., 2012). Therefore, one could argue that financial ratios should not be used in levels but in differences. Table 2 shows that the nowcasting accuracy of these variables tends to be worse when they are considered in log-differences.

One should also note that in this section we only use information from the contemporaneous quarter to nowcast GDP growth. However, some variables (for example, interest rate spreads) could be correlated to GDP growth with a lag. We discuss including further lags in a more aggregated setting in the next section, but for individual variables we do not consider nowcasts using additional lags, as in their case we concentrate on the choice of sampling frequency.

Table 2: *Out-of-sample RMSEs of regression models which have financial ratios or implied volatility as an explanatory variable. The frequency is quarterly. The abbreviations for the variables are explained in Appendix A. The RMSEs are calculated from a rolling window analysis, in which the first estimation sample is from Q3/2002 to Q4/2011 and the first out-of-sample nowcast is Q1/2012. Thus, the results are based on 31 out-of-sample observations.*

	level	**log-difference**
OMX Hels p/e	0.66	0.71
DAX p/e	0.69	0.67
SP500 p/e	0.53	0.71
OMX Hels Dividend yield	0.59	0.73
DAX Dividend yield	0.59	0.80
SP500 Dividend yield	0.72	0.69
OMX Hels p/b	0.60	0.88
DAX p/b	0.71	0.82
SP500 p/b	0.64	0.72
Eurostoxx 50 volatility	0.77	0.71

All in all, using a higher sampling frequency does not seem to clearly and consistently improve nowcasts. The differences in the RMSEs also tend to be small, and whether performance improves or deteriorates as the sampling frequency increases depends on the variable. This contrasts to some extent the results in, for example, Gómez-Zamudio and Ibarra (2017) and Doğan and Midiliç (2019), where on an aggregate level daily financial data improved forecasts compared to quarterly financial data. The average RMSE of our quarterly models is 0.68, of the restricted monthly models 0.70, of the unrestricted monthly models 0.76 and of the daily models 0.68. Importantly, despite the small number of monthly lags the MIDAS model utilising a weighting scheme performs generally better than the unrestricted model.

5. Does financial market data improve GDP nowcasts?

The results in Section 4 showed that financial market variables can be useful for nowcasting Finnish GDP growth. In this section our aim is to find the best nowcasting specification when using a combination of financial market variables in a MIDAS framework. In addition to the first PC, we combine the information in financial market variables by simple forecast averaging, i.e., we take an average of the MIDAS model nowcasts from Section 4. However, as some of the models in Section 4 produced consistently inferior nowcasts, we also combine only the nowcasts produced by MIDAS models driven by financial ratios (price-to-earnings, price-to-book, dividend yield), which were the most accurate class of predictors in Section 4. We also consider how increasing the lag length in the MIDAS models impacts nowcasting performance. As benchmark models we consider the MIDAS model based on monthly growth in industrial production as well as a simple AR(1) model for GDP growth.

We use monthly data as a compromise between daily and quarterly data because our previous results did not suggest a preference for any specific frequency. Monthly data provides relatively timely information, allowing the nowcast to be updated within the quarter, while being less noisy than daily data. In addition, technical issues limited our use of daily data to within-quarter data. As some variables, such as interest rate spreads or stock indices, could affect GDP with a significant lag, it could be important to include lags beyond the current quarter in a nowcasting model. With monthly data we now consider models with 2, 5 and 11 monthly lags. Since we allow very flexible weights in the weighting scheme, and the weights are estimated based on the data, the estimated weight functions will take into account the leading properties of the explanatory data up to eleven months.

Table 3 reports the RMSEs of these models.[10] The average nowcast based on only financial ratios produces the most accurate nowcasts during the out-of-sample period from Q1/2012 to Q3/2019 regardless of the number of lags included. Summarising the information by principal component analysis leads to slightly lower RMSEs than combining all the MIDAS model nowcasts, and the PC based nowcast thus provides a viable alternative if one does not wish to preselect the models to be combined.[11] This finding is in line with Andreou et al. (2013), whose results also show a slight preference to using factors instead of forecast combinations. However, none of the three purely financial market based nowcasts in Table 3 outperforms the nowcast based on industrial production growth in a statistically significant way.[12] In this sample, all the models nowcast GDP roughly as accurately as an AR(1) model (RMSE 0.58).[13]

---

[10] *All the weight functions are plotted in Appendices B, C, D and G.*

[11] *As pointed out by an anonymous referee, machine learning methods could be used to improve nowcasting performance. For example, Babii et al. (2021) introduces a sparse-group LASSO estimator which enables a large number of predictors to be simultaneously included in a MIDAS model. ML methods could also be used to estimate the weights of a U-MIDAS model. The application of this to the Finnish case is left for future research and an even larger number of explanatory variables, including financial variables which we have in this paper shown to be useful for nowcasting Finnish GDP growth. Here we concentrate on analysing the performance of individual financial market variables in Finland and the importance of the sampling frequency.*

[12] *However, we are aware that tests for statistical significance can be unreliable for small sample sizes, and thus we focus on the ranking of the models. This follows the approach taken in other studies with small sample sizes, such as Andreou et al. (2013) (for their small sample), Ferrara et al. (2013) and Ferrara et al. (2014).*

[13] *It should be noted that, first of all, the latest vintage of the GDP data is used, i.e., the data includes revisions. Thus, the performance of the AR(1) model might be better than it would have been in real time. Secondly, GDP is released with a roughly two-month lag in Finland.*

Table 3: *RMSEs of MIDAS regression models. In the model using industrial production the only explanatory variable is the MoM growth rate of the volume of industrial production. In the PC model, the only explanatory variable is the first principal component of the financial market variables (see Appendix A). 'Average nowcast' is the simple average of all the financial variable based nowcasts (nowcasts produced using the financial market variables listed in Appendix A one at a time). 'Average nowcast based on financial ratios only' is the average of the models in which the explanatory variable is the price-to-earnings ratio, the price-to-book ratio or the dividend yield. The RMSEs are calculated from a rolling window analysis, in which the first estimation sample is from Q1/2003 to Q4/2011 and the first out-of-sample nowcast is Q1/2012. Thus, the results are based on 31 out-of-sample observations. To test the statistical significance of the RMSE differences between the industrial production-based nowcast and the other nowcasts we use the Diebold-Mariano test assuming no heteroscedasticity or autocorrelation because the forecast horizon is zero. Asterisks *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.*

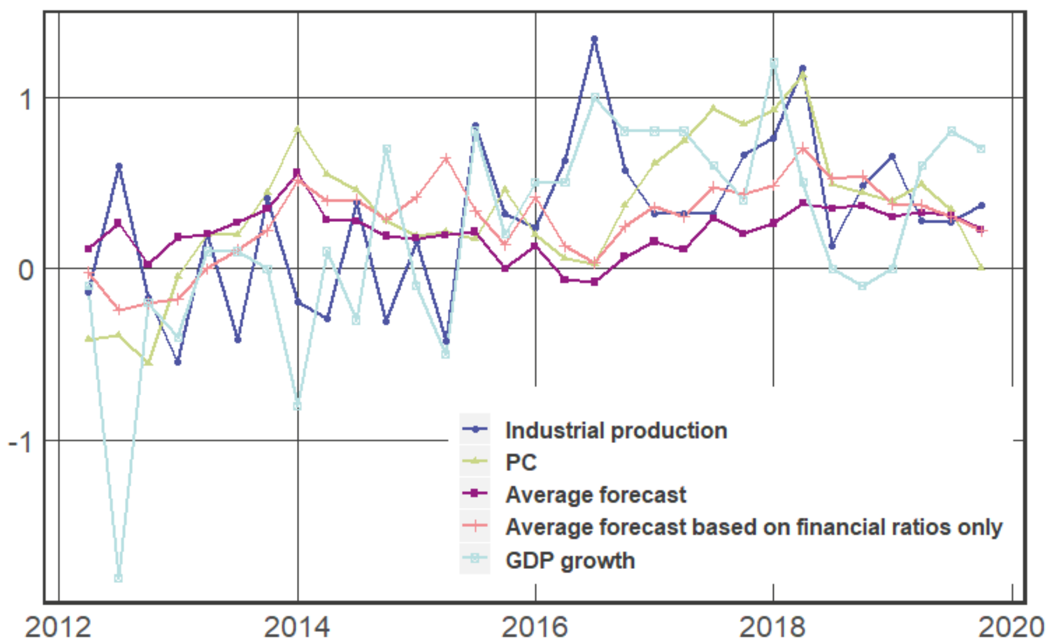
	2 lags	**5 lags**	**11 lags**
Industrial production	0.71	0.57	0.60
PC	0.60	0.62	0.59
Average nowcast	0.63	0.66	0.65
Average nowcast: financial ratios only	0.55	0.56	0.58
Average nowcast: financial ratios only and industrial production	0.60*	0.51	0.54

Regarding the choice of lag length, including only two monthly lags of industrial production growth leads to inferior nowcasting performance. From the weight functions (see Appendix G) we can see that when eleven lags are included most of the weight is on lags three and four and the lags after lag seven are virtually zero. This might also explain why increasing the lag length from two to five improves nowcast accuracy, while increasing the lag length from five to eleven does not: lags three and four receive large weights while lags past the fifth lag do not impact the nowcast much anymore compared to the first five lags. For most of the other models in Table 3 nowcasting performance is similar regardless of the lag length chosen. A longer lag length can be generally preferred based on the idea that the exponential Almon polynomial should be able put the weight of any excess lags to zero, and therefore it can be argued that including too many lags is less detrimental than including too few lags. By comparing the weighting schemes plotted in Appendices B, C and D we can clearly see that two monthly lags tends to be too few for the weights to decay to zero, whereas for many variables, such as the first PC, eleven lags seems to be the most suitable lag length. Using elven lags also leads to the lowest RMSE in Table 3 for the model including the first PC. The weight functions are in many cases hump-shaped, implying that the variables are leading indicators compared to GDP. On the other hand, it could be argued that the importance of financial data lies in it being able to quickly reflect the current economic situation and changes to it, which would indicate that putting at least most of the weight on near-term data is desirable. This would advocate including only two lags, as nowcasting performance is similar regardless of the number of included lags.

Financial market data and macroeconomic variables might include different types of information, useful for nowcasting GDP during different time periods, depending, for example, on the origin of the downturn or upturn. Therefore, it could be useful to combine the nowcasts produced by the MIDAS model based on industrial production growth and the MIDAS models based on financial data. This nowcast results in the lowest RMSEs when using at least 5 lags, while the improvement compared to the nowcast based on only industrial production is weakly statistically significant when using two lags. This is in line with the results in Andreou et al. (2013), who find that combining forecasts from models including macroeconomic and financial data tend to improve short term forecasts over models using only macroeconomic data. Ferrara et al. (2014) find that including financial market volatility and macroeconomic data in a MIDAS model leads to lower RMSEs for nowcasts of GDP growth in France, the UK and the US.

Finally, in Figure 2 we compare the MIDAS model nowcasts to realised GDP growth, in order to assess whether there are differences in model performance over time. It is clear that none of the models capture the sharp decline in GDP in 2012 or the recent slowdown in growth in 2018 very well. Especially in the second quarter of 2012 all the models produce very large nowcast errors. The model driven by industrial production growth performed particularly badly while financial variables were relatively more useful for nowcasting GDP. However, especially due to the limited number of observations available for evaluation, we also consider the RMSE excluding the forecast error for Q2/2012 in Appendix F. The results confirm that in particular the RMSE of the industrial production driven model improves by the exclusion of Q2/2012, but the average nowcast combining financial ratios and industrial production growth still outperforms this nowcast in a weakly statistically significant way when two lags are used. Industrial production growth nowcasts GDP growth particularly well around 2015-2016. It also seems more volatile than the nowcasts based on financial variables. As expected, especially the two combination nowcasts are relatively stable, and thus fail to account for the strong variation in quarterly GDP growth.

Figure 2. *Out-of-sample nowcasts of the different models (number of lags 11) and realised GDP growth. A rolling window, in which the first estimation sample is from Q2/2002 to Q4/2011 (GDP data from Q1/2003), is used to produce the nowcasts.*



Overall, financial market variables and more traditionally used predictors, here represented by industrial production, seem to nowcast Finnish GDP approximately equally well. Therefore, the usefulness of financial variables for nowcasting Finnish GDP growth arises from the fact that they are available earlier than macroeconomic variables. In fact, many of them are available almost immediately and can even be updated on a daily frequency. This issue is highlighted by the fact that the analysis above is conducted using the final, revised vintage of industrial production data, which could give an advantage to industrial production as a predictor of GDP. According to Kuosmanen and Vataja (2014) financial market variables are especially useful during turbulent times. Combined with the fact that when uncertainty is high or the economy is plunging into a recession, timely nowcasts are extremely valuable for economic agents, this highlights the benefits of our MIDAS models based on financial market data.

6. Conclusion

We conclude that financial market variables are useful for nowcasting Finnish quarterly GDP growth. The main advantage of financial market data is its immediate availability. Unlike many other variables that have been traditionally used for forecasting or nowcasting GDP growth, nowcasts based on financial market data may be updated daily. This is particularly useful at the time of crisis, when financial markets might react immediately, but many macroeconomic variables are published with a delay. Using financial market variables in combination with macroeconomic data when nowcasting Finnish GDP growth thus seems beneficial.

On the other hand, our results show that one cannot improve nowcast accuracy for Finnish GDP growth by increasing the sampling frequency of financial market data from quarterly to monthly or daily frequency. This indicates that high-frequency fluctuations of financial markets do not contain additional useful information for nowcasting GDP, which is not already included in corresponding lower frequency data. On the other hand, using higher frequency data does not overall lead to a deterioration in forecast accuracy either, allowing the choice of frequency to be determined by the needs of the researcher.

This paper focuses on the results for Finland. One may argue that this could limit the generalizability of our results. Ultimately, finding the best nowcasting model or the optimal sampling frequency is an empirical question, which depends, among other things, on the sample period used, the explanatory variables chosen as well as the structure of the financial markets and the economy in general. As the previous literature shows, the benefits of individual financial market variables for nowcasting GDP growth varies across countries. For example, Ferrara and Marsilli (2013) shows that the forecasting accuracy of financial market variables differs across four large euro area countries, implying that results are not easily generalisable across countries. However, as discussed earlier in the paper, some of our findings lend support to earlier results, conducted using data mostly on large economies and smaller or more aggregated data sets. Thus, our results broaden the applicability of these results. We also confirm that foreign financial market variables can be important for nowcasting GDP growth in a small open economy using a MIDAS framework. If the results vary between countries as much as the still relatively limited previous literature suggest, it would be interesting to consider the theoretical underpinnings of these results. There could, for example, be some structural differences in the economies driving the variations in the results. This interesting question is left for future research to explore.

Regarding the optimal sampling frequency, which we consider the most interesting research question of our paper, one could argue that the results should not vary significantly across countries where the structure of the economy and the financial markets are similar to Finland. However, factors which could impact the choice of sampling frequency include, for example, persistent differences in the noisiness and the information content of financial market data. One limitation of our analysis, complicating also the comparison to previous papers on other countries, is that these factors could also vary over time. For example, using a higher sampling frequency could improve (by providing more information) or weaken (by increasing noise) forecasting performance during turbulent periods. There is some evidence that increasing the sampling frequency improves forecast accuracy for output growth, but to our knowledge there does not yet exist broad-based evidence of this on an international scale. The relatively short time series of daily data available in many countries complicate the analysis. Overall, we believe determining whether the optimal sampling frequency varies between countries and over time systematically is a highly interesting avenue for future research. If variation of this kind exists, determining the reasons behind this variation is also important.

One issue that may impact the generalizability of our results is the chosen sample period. The results by, for example, Junttila and Korhonen (2011) suggest that financial market variables are especially relevant for forecasting macroeconomic developments during crisis periods. Our in-sample period includes the financial crisis, but the out-of-sample period is less turbulent. Therefore, it is possible that our results understate the nowcasting accuracy of financial market variables, compared to studies where the financial crisis or other turbulent time periods are included.

Although we have used a relatively broad set of variables which we believe to represent the Finnish financial markets well, it is possible that an even more comprehensive set of variables would reveal different results. Another interesting extension left for future research is the use of machine learning methods, which would allow including a larger number of explanatory variables in the same model or enable more flexible weight estimation in the U-MIDAS model.

References

Alessi, L., Ghysels, E., Onorante, L., Peach, R., & Potter, S. (2014). "Central bank macroeconomic forecasting during the global financial crisis: The European Central Bank and Federal Reserve Bank of New York experiences." *Journal of Business & Economic Statistics*, 32(4), 483–500.

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). "Should macroeconomic forecasters use daily financial data and how?" *Journal of Business & Economic Statistic*s, 31(2), 240–251.

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

Babii, A., Ghysels E., and Striaukas, J., (2021). "Machine Learning Time Series Regressions with an Application to Nowcasting." Unpublished manuscript. Available at SSRN 3503191

Becker, R., Lee, J., & Gup, B. E. (2012). "An empirical analysis of mean reversion of the S&P 500's P/E ratios." *Journal of Economics and Finance*, 36(3), 675–690.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach." *The Quarterly Journal of Economics*, 120(1), 387–422.

Chionis, D., Gogas, P., & Pragidis, I. (2010). "Predicting European Union recessions in the euro era: The yield curve as a forecasting tool of economic activity." *International Advances in Economic Research*, 16(1), 1–10.

Clements, M. P., & Galvão, A. B. (2008). "Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States." *Journal of Business & Economic Statistics*, 26(4), 546–554.

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.

Doğan, B. Ş., & Midiliç, M. (2019). "Forecasting Turkish real GDP growth in a data-rich environment." *Empirical Economics*, 56(1), 367–395.

Estrella, A., & Mishkin, F. S. (1998). "Predicting US recessions: Financial variables as leading indicators." *Review of Economics and Statistics*, 80(1), 45–61.

Ferrara, L., & Marsilli, C. (2013). "Financial variables as leading indicators of GDP growth: Evidence from a MIDAS approach during the Great Recession." *Applied Economics Letters*, 20(3), 233–237.

Ferrara, L., Marsilli, C., & Ortega, J. P. (2014). "Forecasting growth during the Great Recession: is financial volatility the missing ingredient?" *Economic Modelling*, 36, 44–50.

Foroni, C., & Marcellino, M. G. (2013). "A survey of econometric methods for mixed-frequency data." *Norges Bank Working Paper* 06/2013. Available at SSRN 2268912.

Foroni, C., Marcellino, M., & Schumacher, C. (2015). "Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1), 57–82.

Franta, M., Havrlant, D., & Rusnák, M. (2016). "Forecasting Czech GDP using mixed-frequency data models." *Journal of Business Cycle Research*, 12(2), 165–185.

Friedman, B. M., & Kuttner, K. (1993). "Why does the paper-bill spread predict real economic activity?" In Business cycles, indicators and forecasting (pp. 213–254). University of Chicago Press.

Galli, A., Hepenstrick, C., & Scheufele, R. (2019). "Mixed-frequency models for tracking short-term economic developments in Switzerland." *International Journal of Central Banking*, 58.

Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). "Mixed Frequency Data Sampling Regression Models: The R Package midasr." *Journal of Statistical Software*, 72(4), 1–35.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). "The MIDAS touch: Mixed data sampling regression models." Discussion paper, UCLA.

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

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). "Predicting volatility: getting the most out of return data sampled at different frequencies." *Journal of Econometrics*, 131(1–2), 59–95.

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

Gomez-Zamudio, L. M., & Ibarra, R. (2017). "Are daily financial data useful for forecasting GDP? Evidence from Mexico." *Journal of the Latin American and Caribbean Economic Association*, 17(2), 173–203.

Henry, O. T., Olekalns, N., & Thong, J. (2004). "Do stock market returns predict changes to output? Evidence from a nonlinear panel data model." *Empirical Economics*, 29(3), 527–540.

Junttila, J., & Korhonen, M. (2011). "Utilizing financial market information in forecasting real growth, inflation and real exchange rate." *International Review of Economics & Finance*, 20(2), 281–301.

Kim, H. H., & Swanson, N. R. (2018). "Methods for backcasting, nowcasting and forecasting using factor-MIDAS: With an application to Korean GDP." *Journal of Forecasting*, 37(3), 281–302.

Koustas, Z., & Serletis, A. (2005). "Rational bubbles or persistent deviations from market fundamentals?" *Journal of Banking & Finance*, 29(10), 2523–2539.

Kuosmanen, P., & Vataja, J. (2014). "Forecasting GDP growth with financial market data in Finland: Revisiting stylized facts in a small open economy during the financial crisis." *Review of Financial Economics*, 23(2) 90–97.

Lewellen, J. (2004). "Predicting returns with financial ratios." *Journal of Financial Economics*, 74(2), 209–235.

Lütkepohl, H. (2010). "Forecasting aggregated time series variables: A survey." *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2010(2), 1–26.

Marcellino, M., & Schumacher, C. (2010). "Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP." *Oxford Bulletin of Economics and Statistics* 72(4), 518–550.

Marsilli, C. (2014). "Variable selection in predictive MIDAS models." *Banque de France Working Paper*, 520, Available at SSRN: https://ssrn.com/abstract=2531339.

Nyberg, H. (2010). "Dynamic probit models and financial variables in recession forecasting." *Journal of Forecasting*, 29(1-2), 215–230.

Tay, A. S. (2007). "Financial variables as predictors of real output growth." Singapore Management University Economics and Statistics Working Paper No. 14-2007.

Tsui, A. K., Xu, C. Y., & Zhang, Z. (2018). "Macroeconomic forecasting with mixed data sampling frequencies: Evidence from a small open economy." *Journal of Forecasting*, 37(6), 666–675.

Appendix

A

The following table lists all the used variables, their abbreviations, sources and transformations. All the data are obtained via Bloomberg.

Table A 1

Variable	Category	Abbreviation	Source	Transformation
Average ratio of the price of a stock and the company's earnings per share in OMX Helsinki	Financial ratio	OMX Hels p/e	NASDAQ OMX Helsinki	No transformation
Average ratio of the stock price to the book value per share in OMX Helsinki.	Financial ratio	OMX Hels p/b	NASDAQ OMX Helsinki	No transformation
Average dividend yield in OMX Helsinki	Financial ratio	OMX Hels Dividend yield	NASDAQ OMX Helsinki	No transformation
Average ratio of the stock price to the book value per share in DAX	Financial ratio	DAX p/e	Deutsche Börse	No transformation
Average ratio of the stock price to the book value per share in DAX.	Financial ratio	DAX p/b	Deutsche Börse	No transformation
Average dividend yield in DAX	Financial ratio	DAX Dividend yield	Deutsche Börse	No transformation
Average ratio of the stock price to the book value per share in S&P 500	Financial ratio	SP500 p/e	Standard and Poor's	No transformation
Average ratio of the stock price to the book value per share in S&P 500	Financial ratio	SPX p/b	Standard and Poor's	No transformation
Average dividend yield in S&P 500	Financial ratio	SP500 Dividend yield	Standard and Poor's	No transformation
The yield of Finland government bond with maturity of 10 years	Interest rate	FI_10y	Bloomberg	Difference
The yield of Finland government bond with maturity of 5 years	Interest rate	FI_5y	Bloomberg	Difference
The yield of Germany government bond with maturity of 12 months	Interest rate	DE_1y	Bloomberg	Difference
The yield of Germany government bond with maturity of 5 years	Interest rate	DE_5y	Bloomberg	Difference
The yield of Germany government bond with maturity of 7 years	Interest rate	DE_7y	Bloomberg	Difference
The yield of Germany government bond with maturity of 10 years	Interest rate	DE_10y	Bloomberg	Difference
The spread between German 10 year yield and 12 month yield	Interest rate	DE 10y-1y	Bloomberg	No transformation

Variable	Category	Abbreviation	Source	Transformation
Implied volatility on Eurostoxx 50 index options with a rolling 30 day expiry	Other	Eurostoxx 50 volatility	Deutsche Börse	No transformation
The price of oil (brent)	Other	OIL	Deutsche Börse	Log-difference
The price of euro in dollars	Other	EURUSD	Deutsche Börse	Difference
Finland Industrial Production Volume, MoM growth rate, SA	Real economy	Industrial production	Statistics Finland	No transformation
OMX Helsinki, price index	Stock index	OMX Hels	NASDAQ OMX Helsinki	Log-difference
OMX Helsinki Technology, price index	Stock index	OMX Hels Technology	NASDAQ OMX Helsinki	Log-difference
OMX Helsinki Utilities, price index	Stock index	OMX Hels Utilities	NASDAQ OMX Helsinki	Log-difference
OMX Hels Industrials, price index	Stock index	OMX Hels Industrials	NASDAQ OMX Helsinki	Log-difference
OMX Hels Telecommunication, price index	Stock index	OMX Hels Telec	NASDAQ OMX Helsinki	Log-difference
OMX Helsinki Basic Materials, preice index	Stock index	OMX Hels Basic Matl	NASDAQ OMX Helsinki	Log-difference
OMX Helsinki Health Care, price index	Stock index	OMX Hels Hlth Care	NASDAQ OMX Helsinki	Log-difference
S&P 500, price index	Stock index	SP500	Standard and Poor's	Log-difference
DAX, price index	Stock index	DAX	Deutsche Börse	Log-difference

The following figures plot all the (transformed) variables (solid lines) aggregated to a quarterly frequency together with quarterly GDP growth (dashed lines). All variables have been demeaned and divided by their standard deviations (in the figures) to make interpretation easier.

Figure A 1

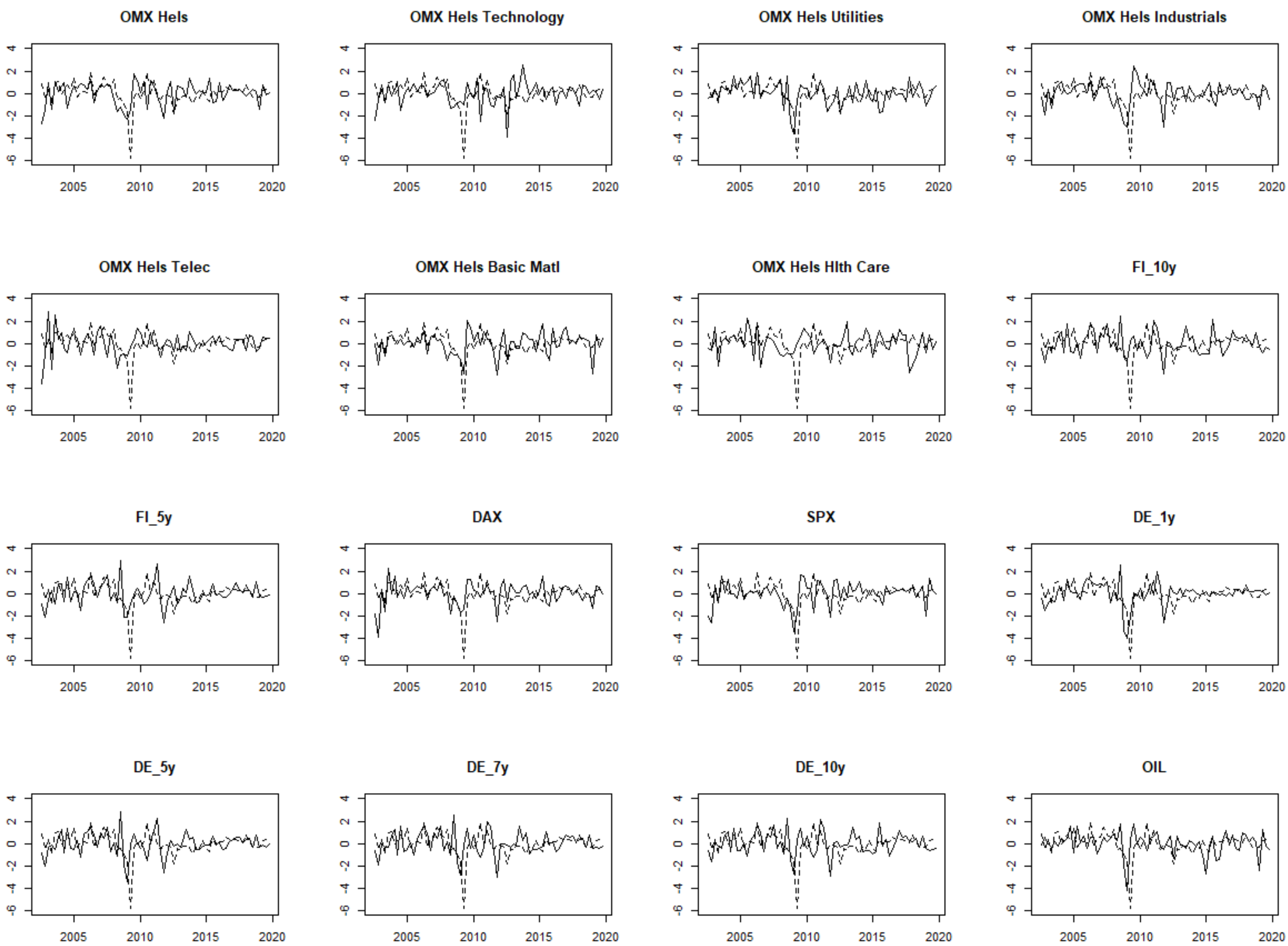
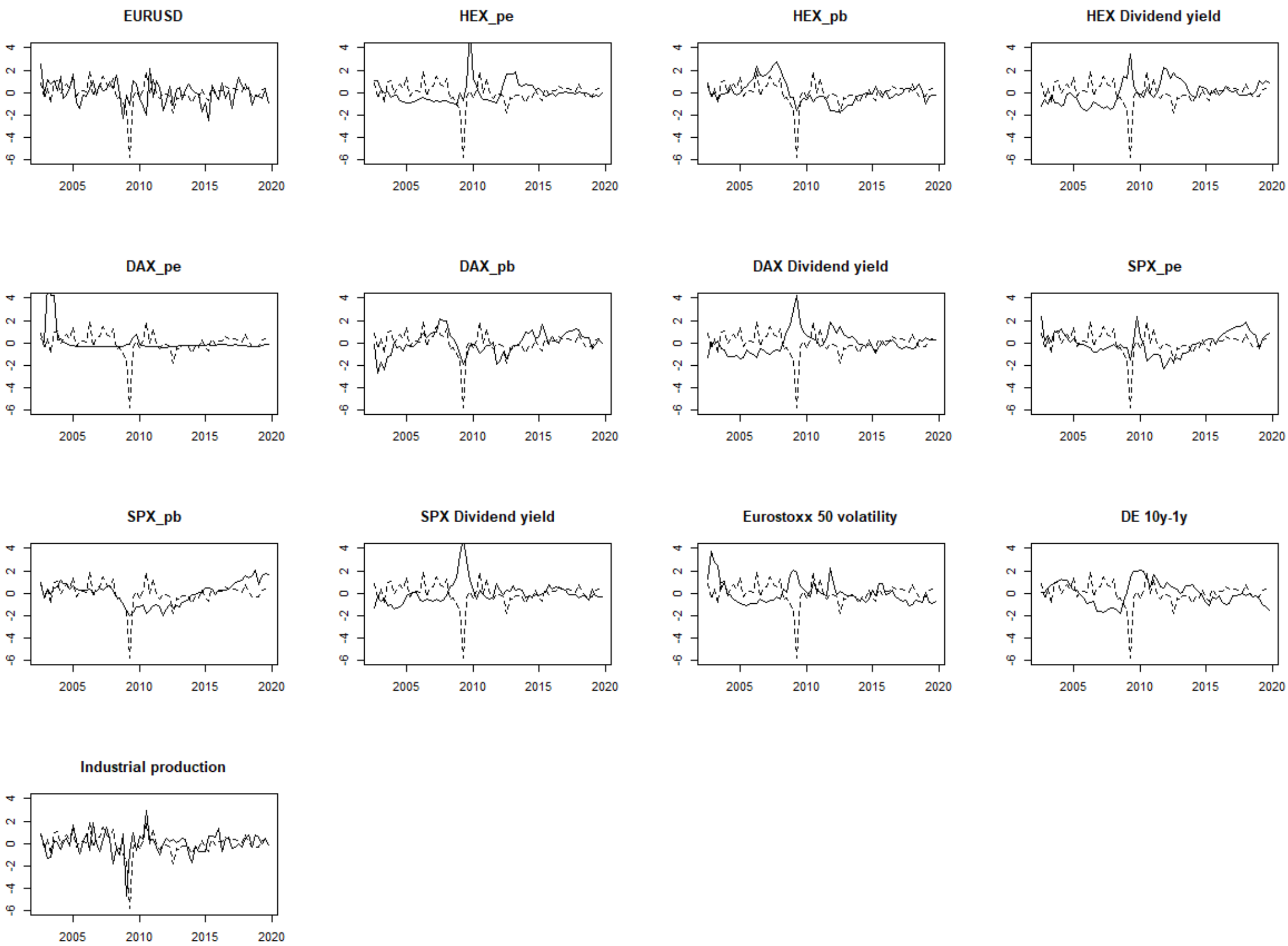


Figure A 2


EURUSD
HEX_pe
HEX_pb
HEX Dividend yield
DAX_pe
DAX_pb
DAX Dividend yield
SPX_pe
SPX_pb
SPX Dividend yield
Eurostoxx 50 volatility
DE 10y-1y
Industrial production

B

The following figures show the weighting schemes that are estimated using daily data from Q2/2002 to Q3/2019.

Figure B 1

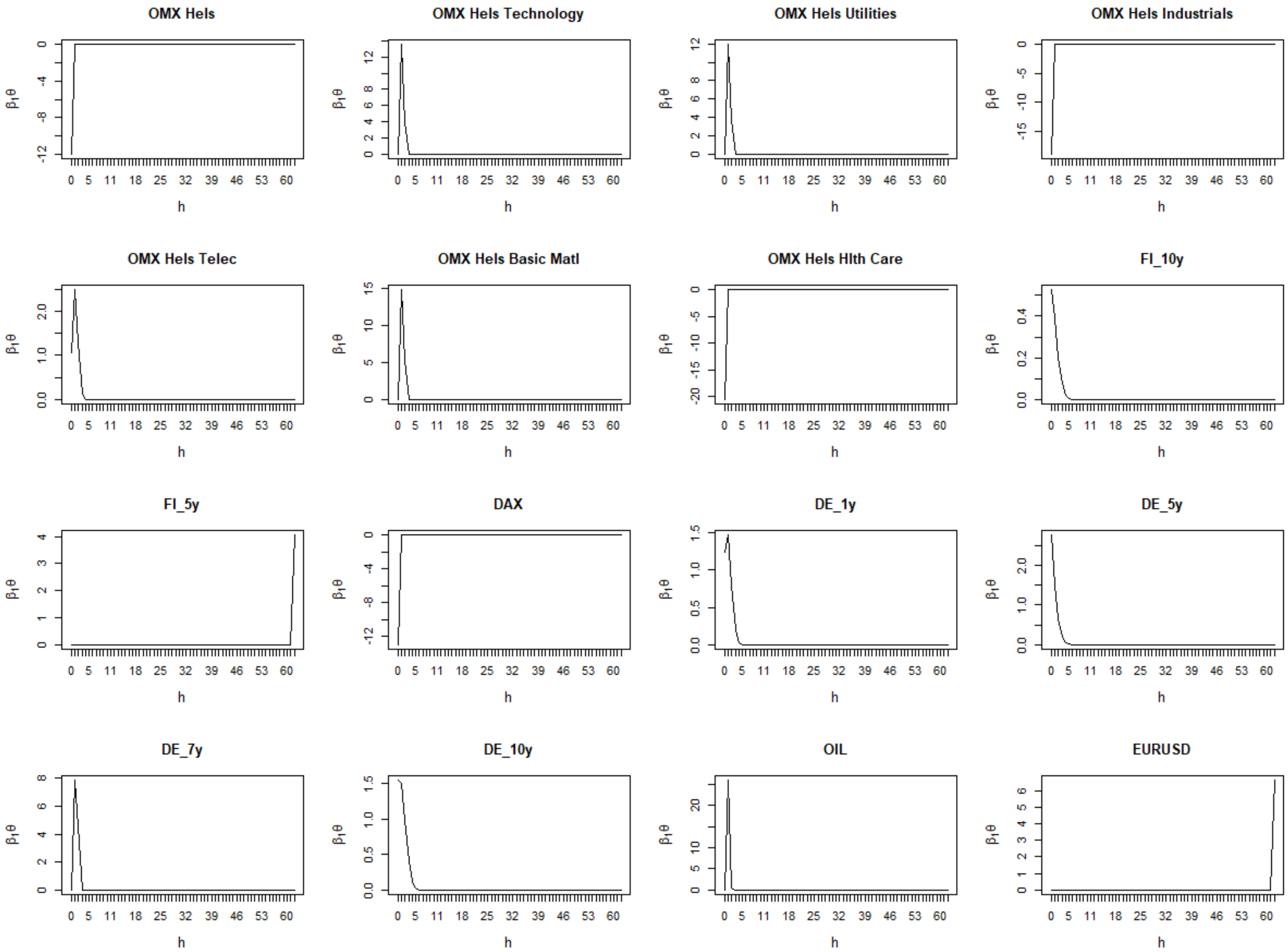
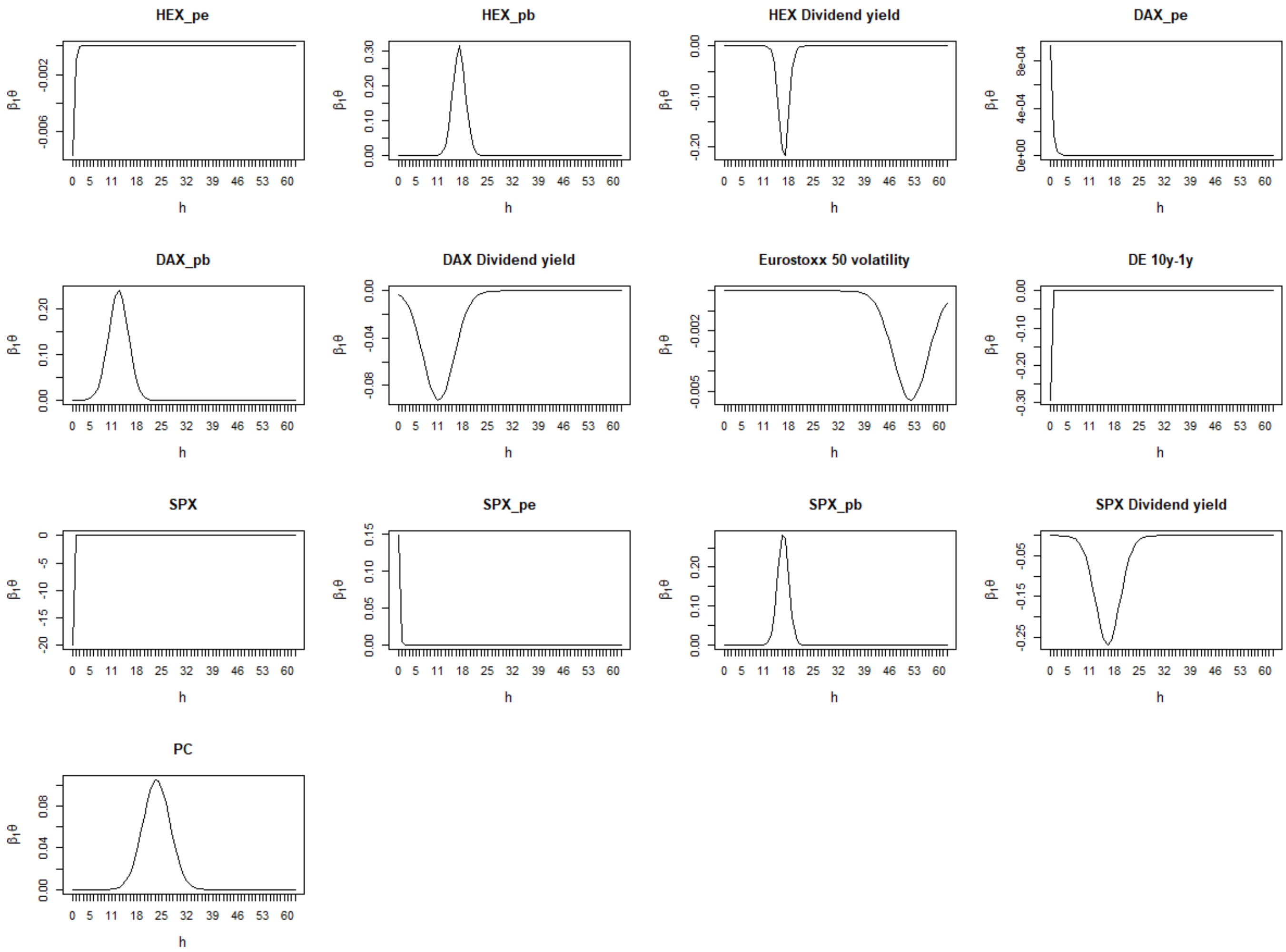


Figure B 2


HEX_pe
HEX_pb
HEX Dividend yield
DAX_pe
DAX_pb
DAX Dividend yield
Eurostoxx 50 volatility
DE 10y-1y
SPX
SPX_pe
SPX_pb
SPX Dividend yield
PC
$\beta_1\theta$
h
0 5 11 18 25 32 39 46 53 60

The following figures show the weighting schemes that are estimated using monthly data from Q2/2002 to Q3/2019.

Figure B 3

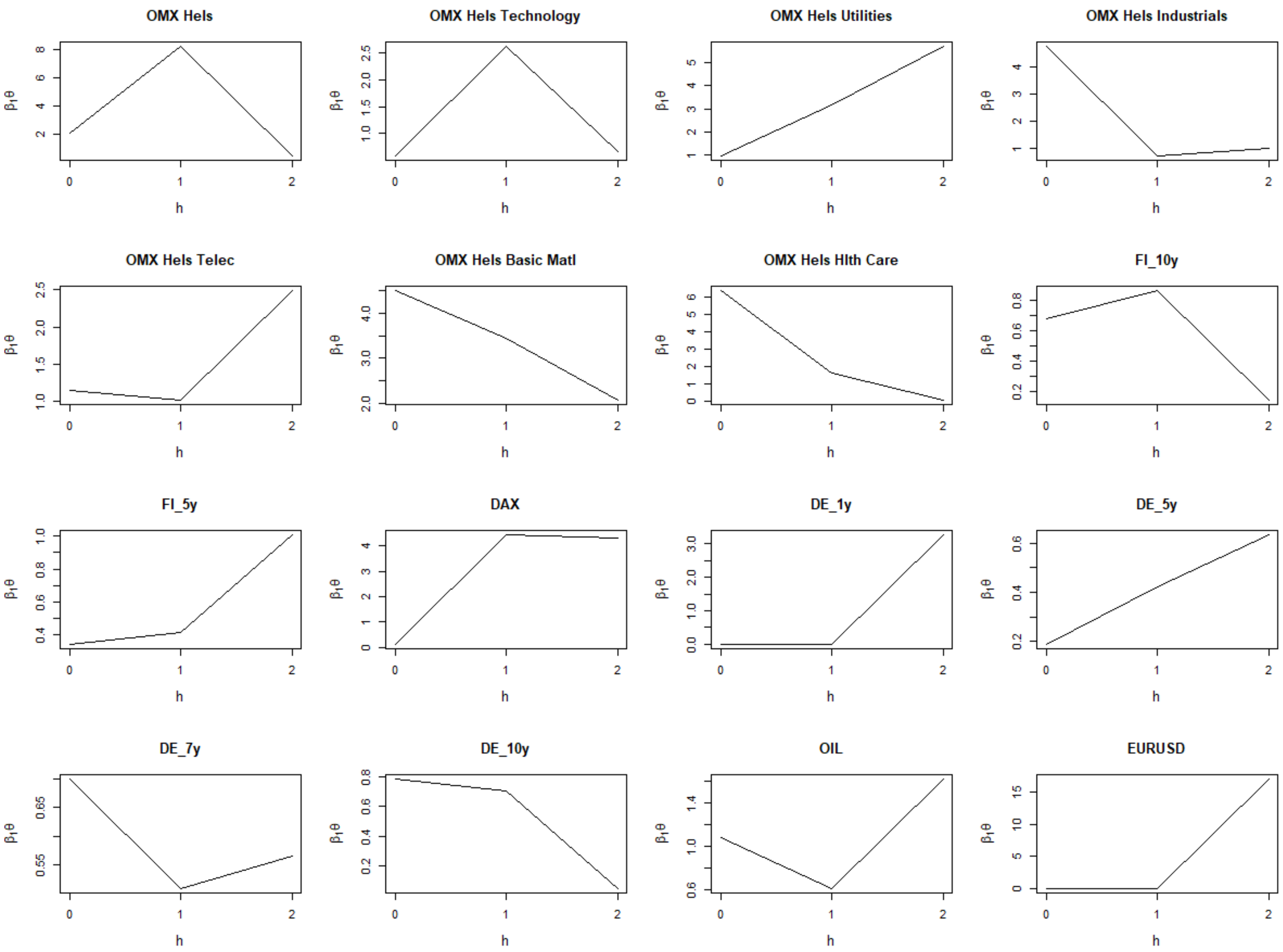
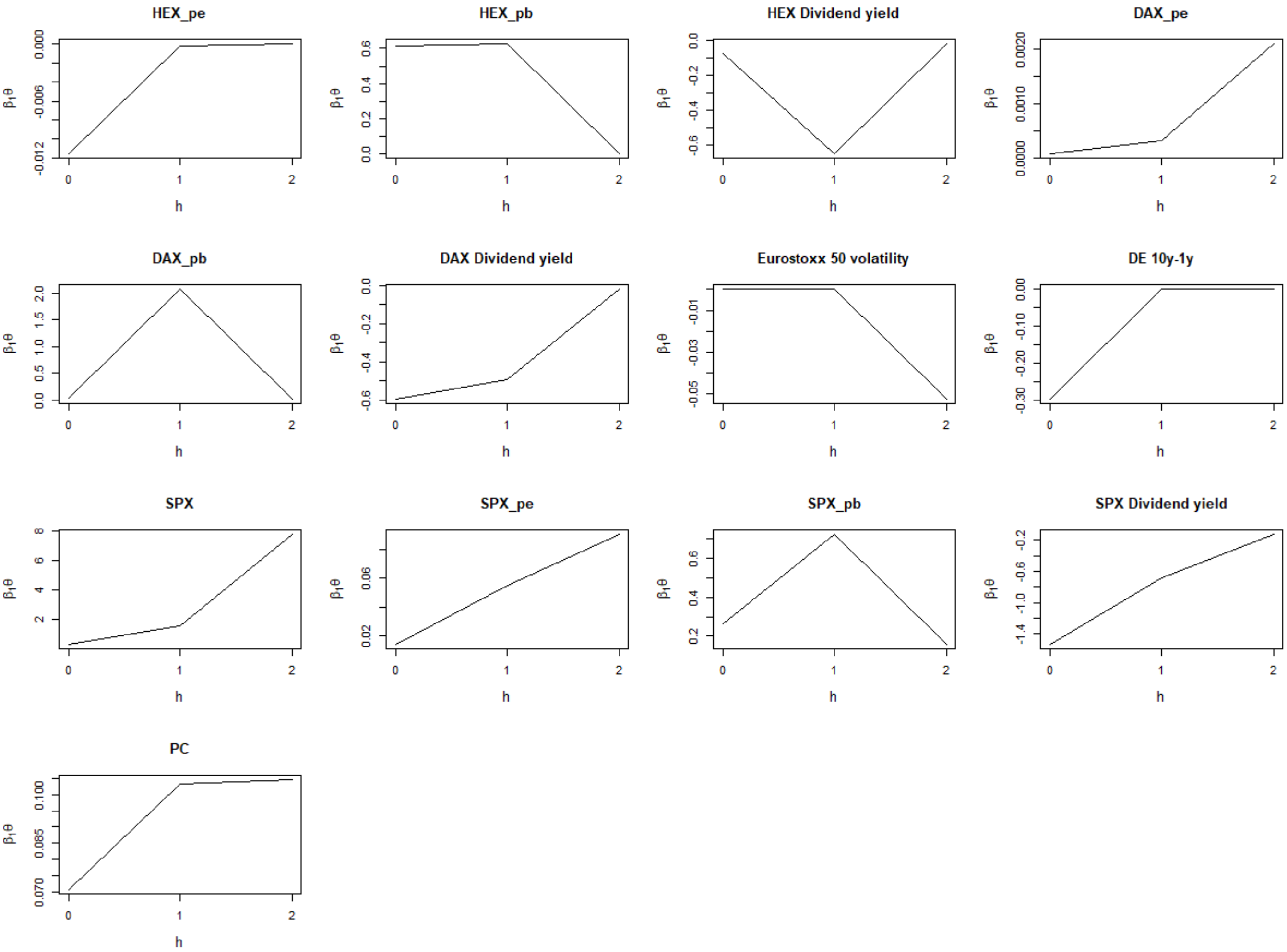


Figure B 4


HEX_pe
HEX_pb
HEX Dividend yield
DAX_pe
DAX_pb
DAX Dividend yield
Eurostoxx 50 volatility
DE 10y-1y
SPX
SPX_pe
SPX_pb
SPX Dividend yield
PC
β1θ
h

The following figures show the coefficient estimates of the unrestricted model that is estimated using monthly data from Q2/2002 to Q3/2019.

Figure B 5

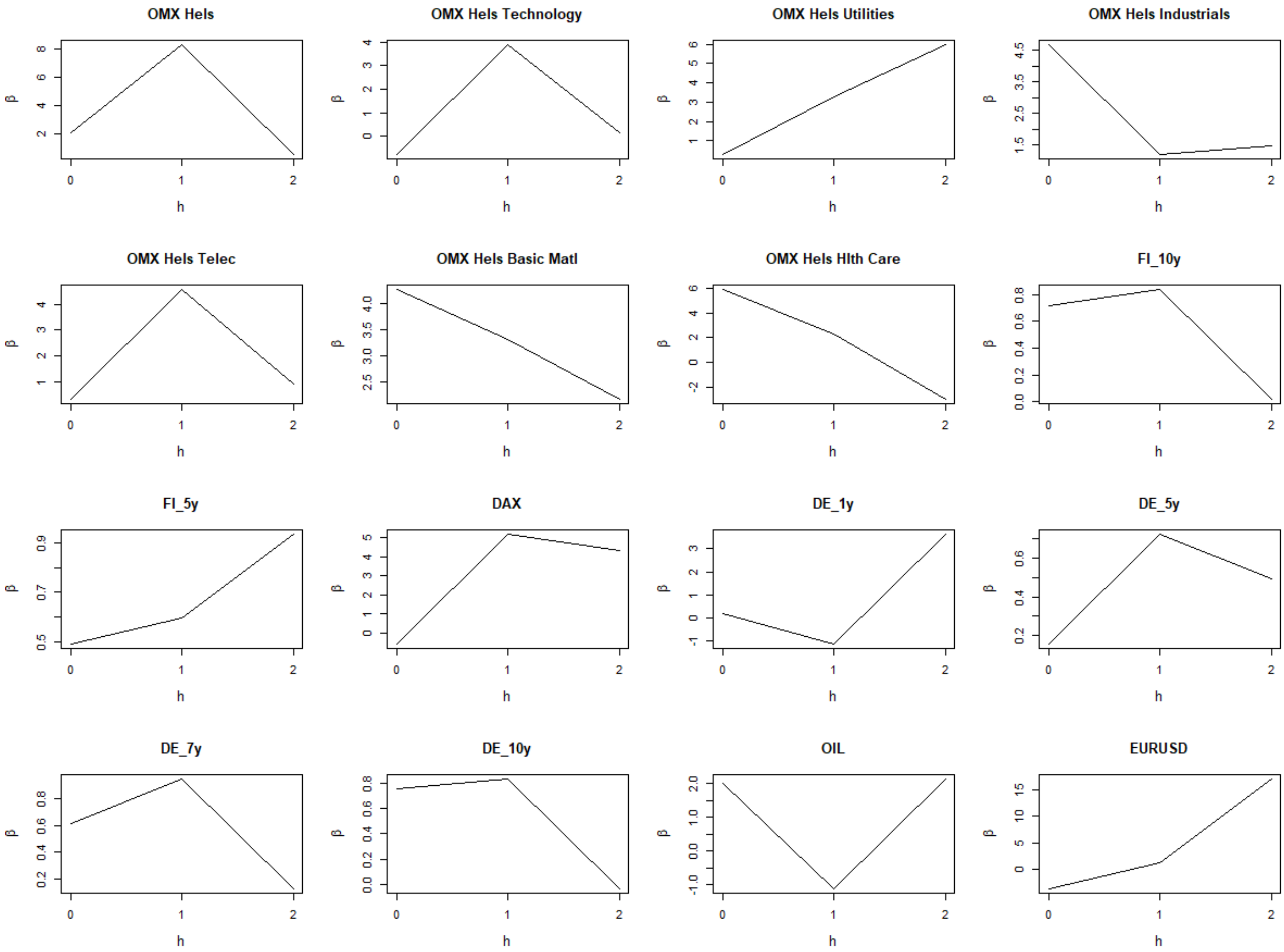


Figure B 6

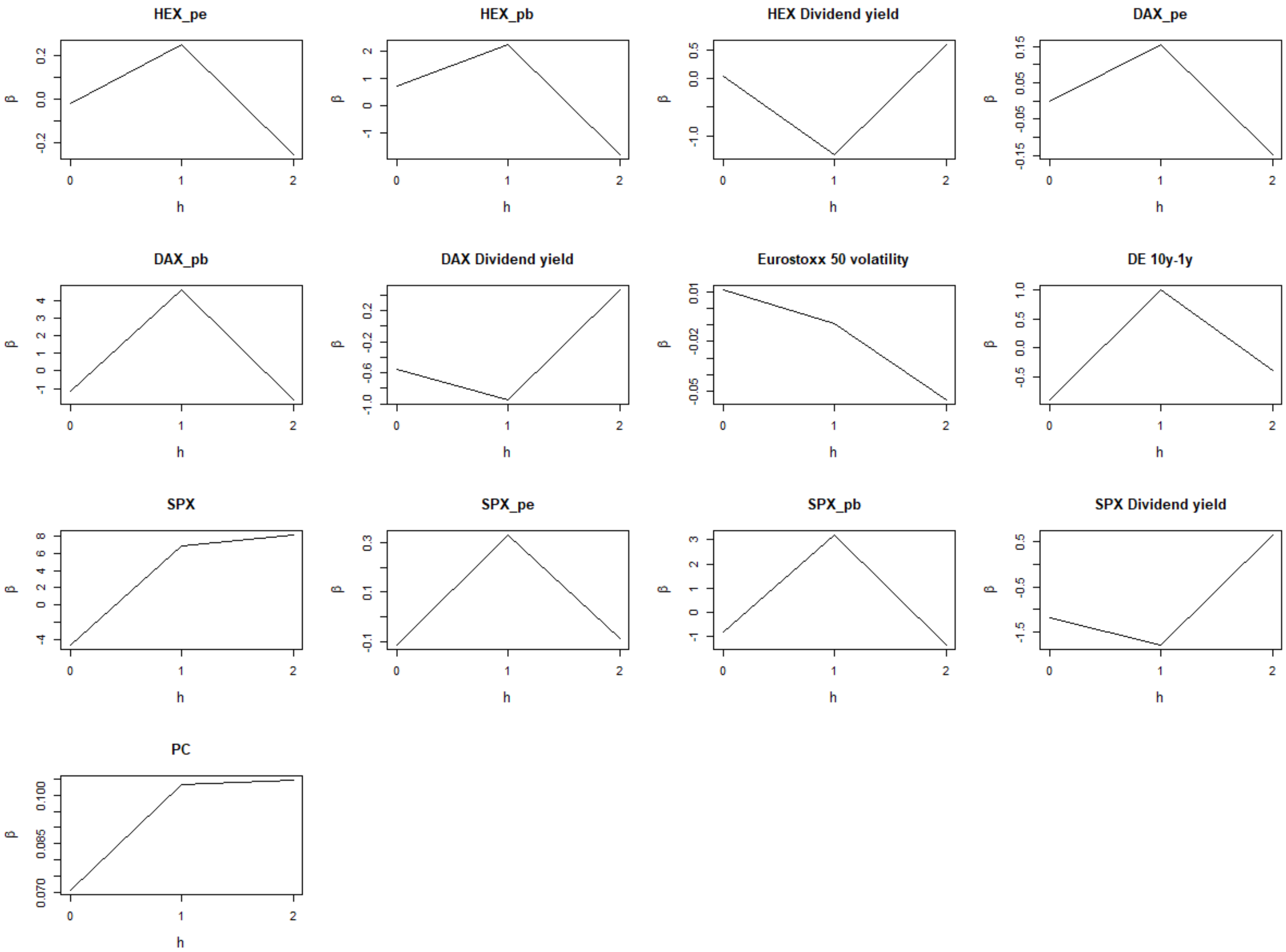
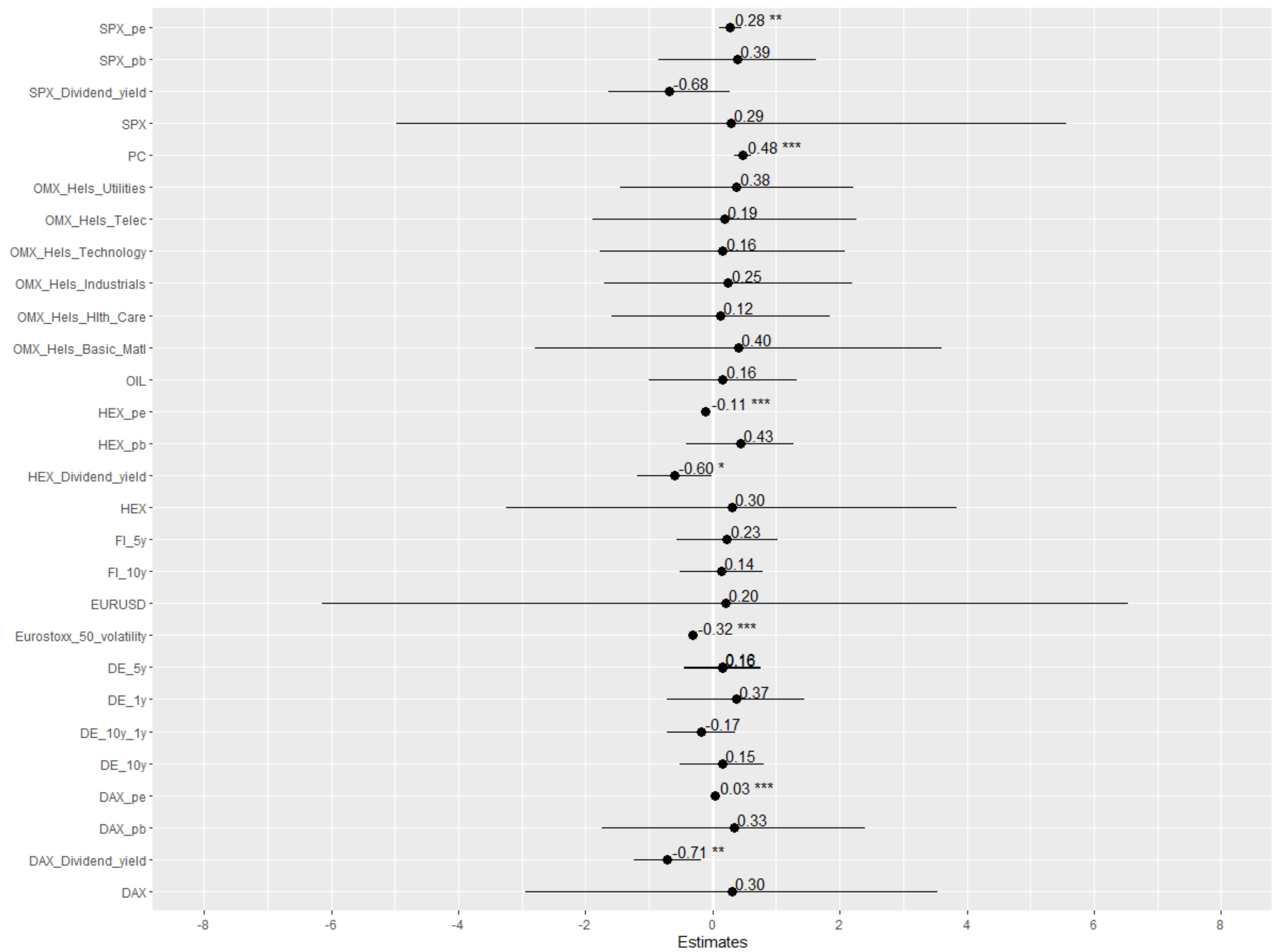

HEX_pe
HEX_pb
HEX Dividend yield
DAX_pe
DAX_pb
DAX Dividend yield
Eurostoxx 50 volatility
DE 10y-1y
SPX
SPX_pe
SPX_pb
SPX Dividend yield
PC
β
h

Figure B 7 reports standardised coefficient estimates for the quarterly models. The lines represent 95 per cent confidence intervals based on heteroscedasticity and autocorrelation robust standard errors.

Figure B 7



C

The following figures show the weighting schemes that are estimated using monthly data from Q3/2002 to Q3/2019 assuming 5 lags.

Figure C 1

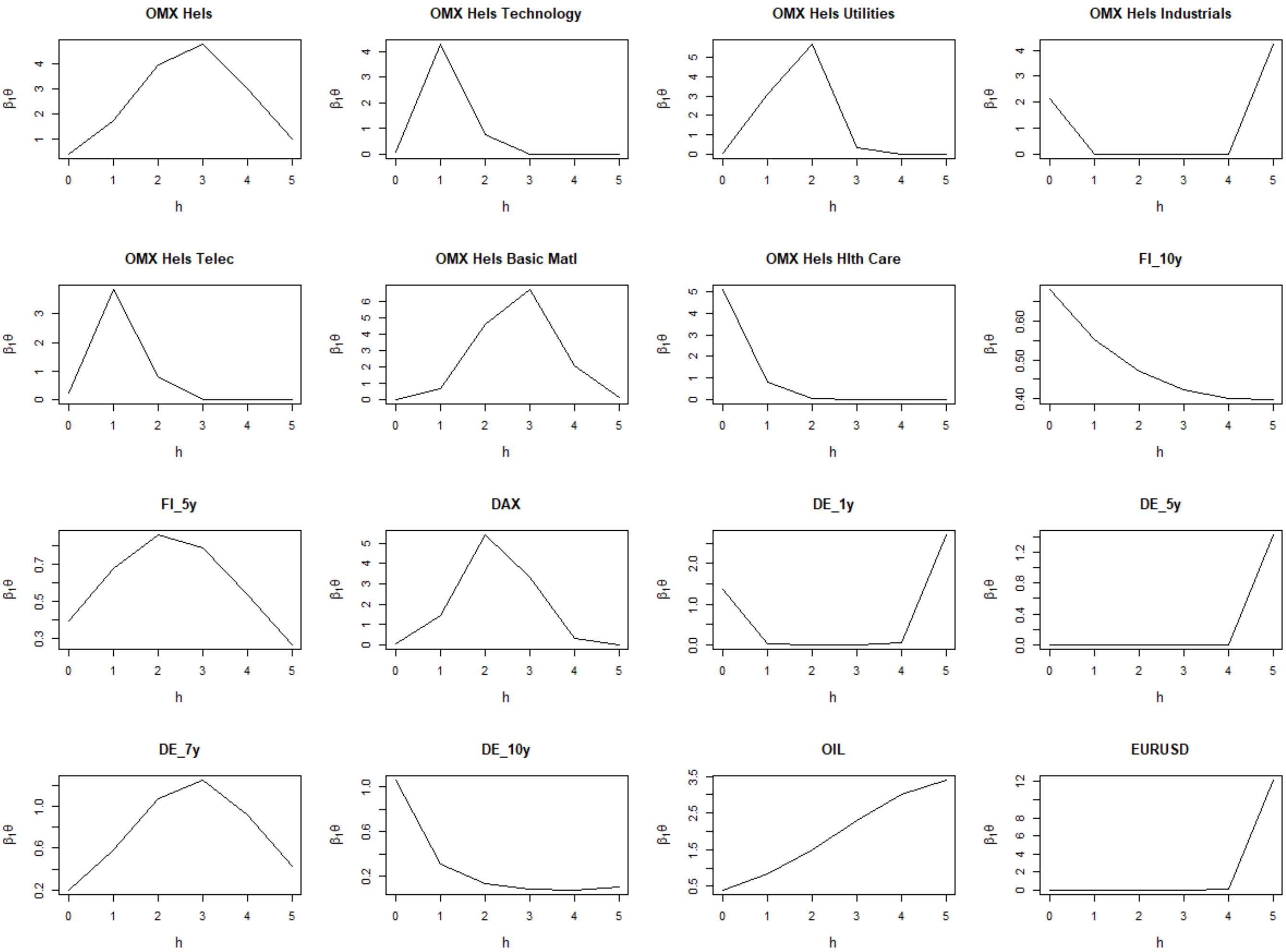
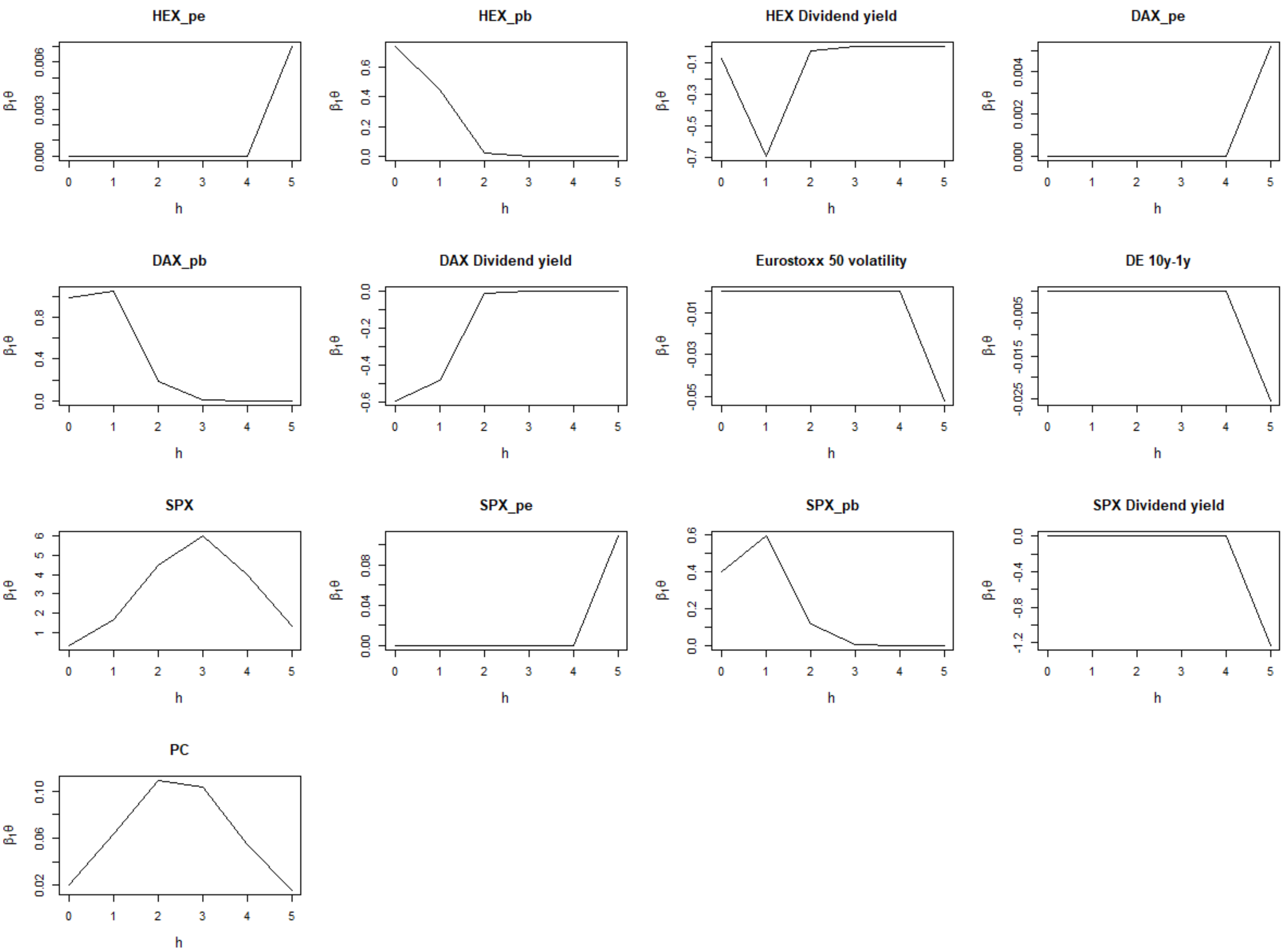


Figure C 2


HEX_pe
HEX_pb
HEX Dividend yield
DAX_pe
DAX_pb
DAX Dividend yield
Eurostoxx 50 volatility
DE 10y-1y
SPX
SPX_pe
SPX_pb
SPX Dividend yield
PC

D

The following figures show the weighting schemes that are estimated using monthly data from Q1/2003 to Q3/2019 assuming 11 lags.

Figure D 1

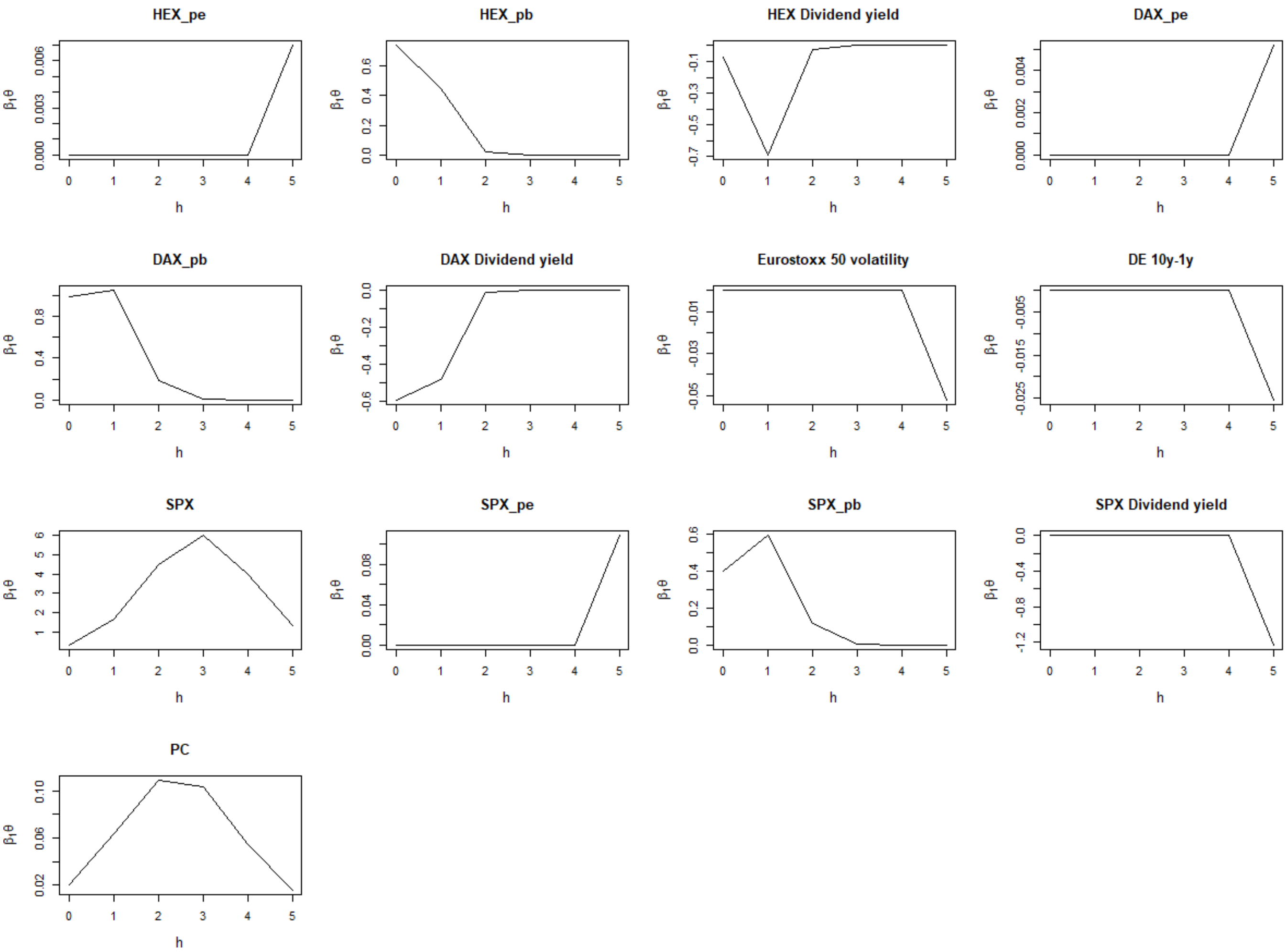
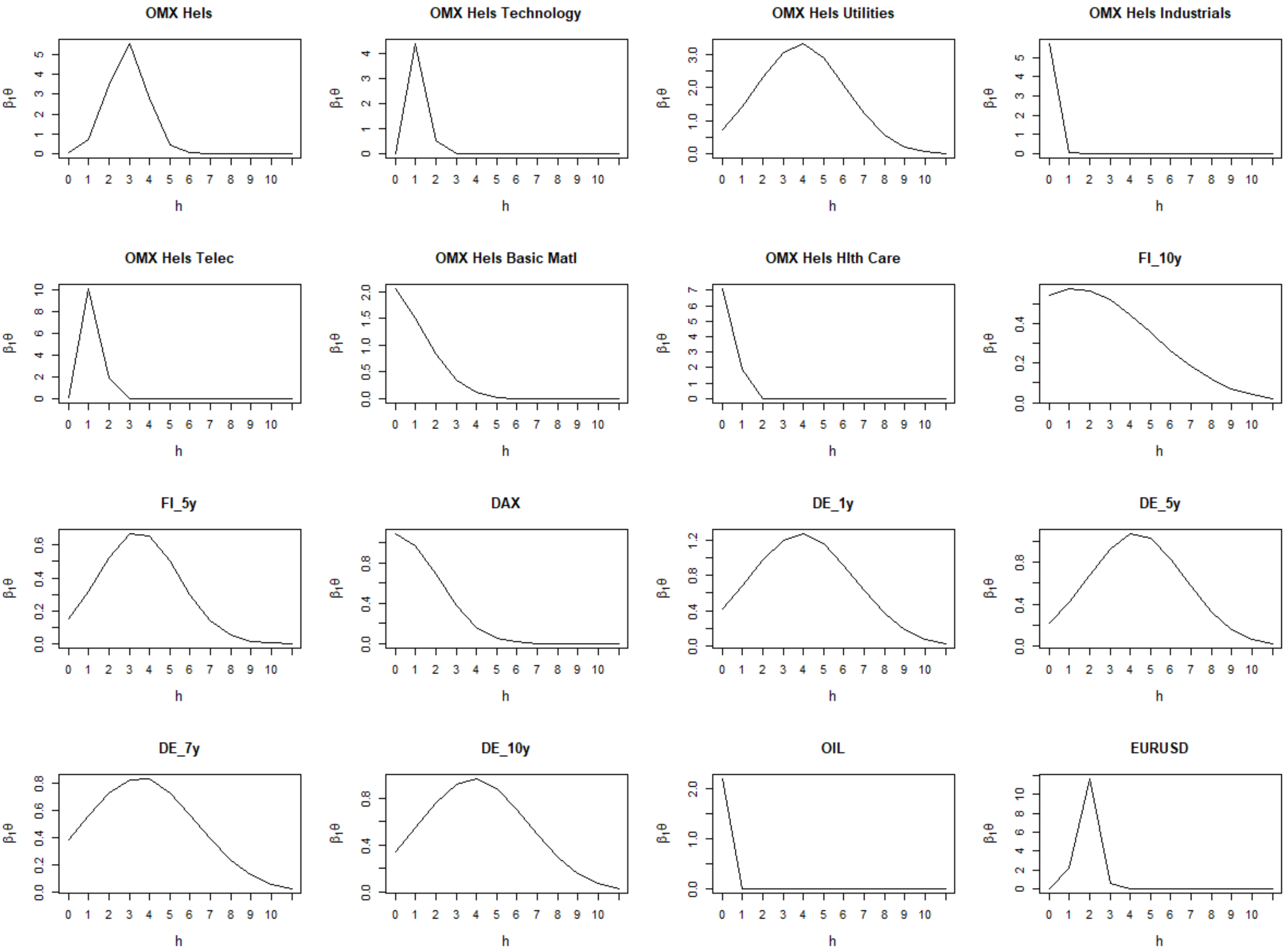


Figure D 2


OMX Hels
OMX Hels Technology
OMX Hels Utilities
OMX Hels Industrials
OMX Hels Telec
OMX Hels Basic Matl
OMX Hels Hlth Care
FI_10y
FI_5y
DAX
DE_1y
DE_5y
DE_7y
DE_10y
OIL
EURUSD
$\beta_1\theta$
h

E

Table E 1 replicates Table 1 excluding the out-of-sample forecast for the period Q2/2012. The forecast error in Q2/2012 is clearly larger than at other times and is excluded here in order to control for the effect of this potential outlier.

Table E 1: *Out-of-sample RMSEs of MIDAS regression models. The models are ordered based on the RMSEs of the quarterly models. The abbreviations for the variables are explained in Appendix A. The RMSEs are calculated from a rolling window analysis, in which the first estimation sample is from Q2/2002 to Q4/2011 and the first out-of-sample forecast is Q1/2012. The out-of-sample forecast for the period Q2/2012 is excluded. Thus, the results are based on 30 out-of-sample observations.*

Explanatory variable	Quarterly	Monthly	Monthly (unrestricted)	Daily
SP500 p/e	0.42	0.47	0.49	0.45
OMX Hels p/e	0.53	0.56	0.64	0.55
DE 10y-1y	0.54	0.54	0.57	0.54
Industrial production	0.54	0.46	0.50	–
DE_10y	0.55	0.59	0.77	0.54
FI_10y	0.55	0.69	0.83	0.54
FI_5y	0.55	0.57	0.59	0.53
DAX Dividend yield	0.56	0.57	0.60	0.57
DAX p/e	0.56	0.56	0.78	0.56
DE_5y	0.56	0.60	0.60	0.55
DE_7y	0.56	0.62	0.71	0.55
OIL	0.56	0.56	0.56	0.62
HEX p/b	0.57	0.61	0.72	0.63
DE_1y	0.58	0.54	0.63	0.52
EURUSD	0.58	0.69	0.69	0.54
PC	0.59	0.58	0.58	0.65
SP500 p/b	0.59	0.63	0.74	0.63
OMX Hels Dividend yield	0.60	0.66	0.65	0.61
OMX Hels Telec	0.61	0.68	0.72	0.59
OMX Hels Industrials	0.62	0.60	0.61	0.58
OMX Hels Hlth Care	0.62	0.72	0.91	0.62
SPX Dividend yield	0.62	0.61	0.66	0.61
SP500	0.66	0.68	0.69	0.57
DAX p/b	0.67	0.62	0.79	0.73
Eurostoxx 50 volatility	0.67	0.65	0.67	0.67
OMX Hels Utilities	0.69	0.64	0.71	0.62
OMX Hels	0.72	0.71	0.80	0.58
OMX Hels Technology	0.72	0.71	0.86	0.61
DAX	0.79	0.75	0.82	0.59
OMX Hels Basic Matl	0.82	0.79	0.84	0.64

Table E 2 is otherwise identical to Table 1, but the number of out-of-sample observations is increased from 31 to 40.

Table E 2: *Out-of-sample RMSEs of MIDAS regression models. The models are ordered based on the RMSEs of the quarterly models. The abbreviations for the variables are explained in Appendix A. The RMSEs are calculated from a rolling window analysis, in which the first estimation sample is from Q2/2002 to Q4/2011 and the first out-of-sample forecast is Q3/2009. Thus, the results are based on 40 out-of-sample observations.*

Variable	Quarterly	Monthly	Monthly (unrestricted)	Daily
Industrial production	0.67	0.75	0.72	0.75
OMX Hels p/e	0.73	0.98	0.92	1.26
DAX p/e	0.75	0.76	0.97	0.76
OIL	0.75	0.75	0.81	0.75
DE_10y	0.76	0.95	1.06	0.74
FI_10y	0.76	0.99	1.14	0.85
HX6000PI	0.77	0.85	0.93	0.73
HX7000PI	0.77	0.81	0.85	0.74
DE 10y-1y	0.77	0.77	0.81	0.83
DE_5y	0.78	0.86	0.89	0.82
DE_7y	0.78	0.92	1.00	0.87
HX2000PI	0.78	0.78	0.83	0.76
HX4000PI	0.80	0.84	1.01	0.75
SP500 Dividend yield	0.82	0.80	0.90	0.81
SP500 p/e	0.82	0.74	1.01	0.88
FI_5y	0.83	0.86	0.91	0.90
DE_1y	0.84	0.83	0.90	0.76
HEX p/b	0.84	0.85	0.99	0.84
DAX Dividend yield	0.85	0.85	0.92	0.85
DAX p/b	0.86	0.80	1.03	0.88
EURUSD	0.89	0.97	1.04	0.74
OMX Hels	0.90	1.02	1.03	0.72
PC1	0.93	1.02	1.00	0.83
SP500	0.94	0.93	1.03	0.73
HX1000PI	0.95	0.94	1.07	0.77
OMX Hels Dividend yield	0.97	0.92	0.91	0.91
DAX	0.99	0.90	1.16	0.74
HX9000PI	0.99	0.96	1.22	0.83
Eurostoxx 50 volatility	1.00	0.83	1.02	1.06
SP500 p/b	1.07	1.09	1.19	1.07

F

Table F 1 replicates Table 2 excluding the out-of-sample forecast for the period Q2/2012. The forecast error in Q2/2012 is clearly larger than in other periods and is excluded here in order to control for the effect of this potential outlier. Comparing Table 2 to Table F 1 reveals that especially the RMSE of the industrial production driven model improves by the exclusion of Q2/2012. However, the average forecast using the financial ratios and industrial production still outperforms the industrial production driven forecast in a weakly statistically significant way when including two monthly lags and performs equally well when using more lags. Using only financial ratios for forecasting still produces competitive forecasts as well.

Table F 1: *RMSEs of MIDAS regression models. In the model using industrial production the only explanatory variable is the MoM growth rate of the volume of industrial production. In the PC model, the only explanatory variable is the first principal component of the financial market variables (see Appendix A). 'Average forecast' is the simple average of all the financial variable based forecasts (forecasts produced using the financial market variables listed in Appendix A one at a time). 'Average forecast based on financial ratios only' is the average of the models in which the explanatory variable is the price-to-earnings ratio, the price-to-book ratio or the dividend yield. The RMSEs are calculated from a rolling window analysis, in which the first estimation sample is from Q1/2003 to Q4/2011 and the first out-of-sample forecast is Q1/2012. The out-of-sample forecast for the period Q2/2012 is excluded. Thus, the results are based on 30 out-of-sample observations. To test the statistical significance of the RMSE differences between the forecast based on industrial production and the other forecasts, we use the Diebold-Mariano test assuming no heteroscedasticity or autocorrelation because the forecast horizon is zero. Asterisks *,** and*** indicate statistical significance at the 10%, 5% and 1% levels, respectively.*

	2 lags	**5 lags**	**11 lags**
Industrial production	0.56	0.39	0.42
PC	0.57	0.56	0.54
Average forecast	0.54	0.55*	0.55
Average forecast: financial ratios only	0.50	0.50	0.52
Average forecast: financial ratios only and industrial production	0.50*	0.39	0.42

G

Figure G 1 shows the weighting schemes of industrial production that are estimated using monthly data from Q2/2002 to Q3/2019 with 3, 5 and 11 lags.

Figure G 1