EL COLEGIO DE MÉXICO
CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE MAESTRO EN ECONOMÍA

Macroeconomic forecasting with mixed frequency data: Evidence for Mexico

Luis Manuel Gómez Zamudio

Promoción 2012-2014

ASESOR:

RAÚL IBARRA

Julio 2014

Contents

Abstract

This article evaluates the role of using data sampled at high frequencies to forecast quarterly GDP in Mexico. The model used integrates both sampling frequencies whilst remaining parsimonious. In particular, the MIDAS (mixed data sampling) regression model is introduced which helps us tackle the multifrequency problem and then, to retain parsimony, factor analysis and forecast combination techniques are used to summarize the 392 daily financial series. Our findings suggest that the MIDAS model is more accurate than traditional models when forecasting with macroeconomic indicators as exogenous regressors plus the financial data. Furthermore, we explore the ability of the MIDAS model for nowcasting. We conclude based on the results obtained that the MIDAS model is more accurate than a simple flat aggregation scheme, although the nowcasts do not improve significantly over the forecasts without leads.

We conclude that this methodology seems to to improve the forecasts for the Mexican case even though this economy presents higher volatility than developed countries. Our main results are consistent with those of recent literature that forecasts the GDP from developed nations.

1. Introduction

Forecasting influences the economy as a whole, as individuals and policy makers rely upon predictions to reach decisions. With this in mind, it is fundamental that the predictions are accurate in the sense that they are a good approximation to the observed values of the variable of interest. In turn, accuracy of the forecasts is related to the information and the model fitted to the data.

Financial data, such as indexes or futures, contain a wealth of useful information for making predictions due to its forward looking nature. There are however, some challenges that must be addressed to render financial data usable.

The first one is the fact that financial information is sampled at a much higher frequency than the usual macroeconomic variables of interest (e.g. GDP). These macro variables usually contain quarterly information whereas many of the financial variables are sampled on a daily basis. The usual approach is to average the high frequency data in the quarter and proceed normally with a regression. This method, however, fails to utilize the information properly.

One possible way to overcome this difficulty is to use the Mixed Data Sampling (MIDAS) approach proposed by Ghysels et al. (2004, 2006). This family of models has been used in recent literature, such as Clements & Galvão (2008) or Marcellino & Schumacher (2010), to improve the accuracy of predictions of macro variables. More importantly, the specific usage of financial data paired with the MIDAS model has been recently explored in Androu et al. (2013). All of these articles have concluded that the usage of mixed frequency data improves the forecasts.

A second difficulty that needs to be addressed is how to use all the information in such a way that the model remains parsimonious. In this regard, there are a few options that could be used such as factor analysis and forecast combinations, not to mention, the wide variety of model parameterization options that considerably reduce the number of estimated coefficients.

To the extent of our knowledge the abovementioned methodology has not been applied yet to developing economies. It is relevant to do so, because the volatility of economic variables in these countries tends to be higher thus affecting the accuracy of the forecasts. We would like to verify if the proposed methodologies are also a viable improvement on accuracy for these countries.

This article focuses on forecasting the Mexican GDP since it is one of the most important economic activity indicators of a country. First a large set of financial variables was obtained from Bloomberg, these variables are categorized as: commodities, equities, corporate risk, foreign exchange or fixed income. This set will be used as the main information source. As such, factor analysis is performed to obtain the 5 most important factors from around 400 financial variables. With these calculated factors the MIDAS model is estimated and several forecast profiles are obtained for different models whose performance is then compared to traditional benchmark models. Finally, forecast combinations are carried out to improve the accuracy of the previously mentioned models.

The main result is that the inclusion of financial data to forecast macro variables does improve accuracy over more traditional models. Furthermore, according to the literature on the topic, it is found that forecast combinations provide an effective medium to improve the forecasting performance of a set of models. . The methodologies described herein strive to incorporate additional information, whilst preserving parsimony. As such, we conclude that they are successful.

The last part of the article presents statistical comparisons of the forecasting prowess of the MIDAS model. First, we would like to find out if the employment of financial data is useful when compared to models that use common (in the literature) macroeconomic regressors. Second, we would like to find out how the MIDAS model compares against a flat aggregation weighting scheme (more on this on section 3). The results show that the model with financial data has the same predictive ability as a traditional model with macro variables, hence, we favor the MIDAS model with financial and macro data over the traditional model with macro data. Furthermore, we find that MIDAS model outperforms the flat aggregation scheme accuracy wise.

The rest of the article is organized in the following way: section 2 consists of a brief review of the literature, section 3 explains the MIDAS model, Factor analysis and forecast combination, section 4 gives an overview of the data used, section 5 presents the primary results, section 6 concludes the article and the appendix presents detailed information on the code and data used.

2. Literature review

Clements & Galvão (2008) follow a similar proposal as the one in this work but for the US economy. Their goal is to improve GDP forecasts utilizing an extension to the original MIDAS model from Ghysels et al. (2004, 2006). The original model does not include AR terms of the dependent variable; the authors proposed extension is to include these terms.

The high frequency data employed by the authors consists of three monthly macroeconomic time series: industrial production, unemployment and capacity utilization. The authors pay specific attention to the use of real time data as opposed to revised data.

Clements & Galvão conclude that the MIDAS model has a good performance relative to the benchmark models proposed by them. On another similar paper that propose extensions to the model Marcellino & Schumacher (2010) strive to forecast the German GDP using a relatively big set of monthly indicators. The way the authors propose to mitigate regressor proliferation follows a two-step procedure: first, a set of factors that summarize the information of the indicators is estimated; second, the MIDAS model is used to forecast the GDP using the estimated factors instead of the high frequency indicators.

The authors propose a total of 9 models for comparison. These differ due to varying structural characteristics and methodologies employed to estimate factors. The authors find that more parsimonious MIDAS models are relatively more accurate and that there are no substantial differences with respect to the way the factors are estimated.

Complementary to the previous 3 papers, Arnesto et al. (2010) propose 3 methods of mixed frequency forecasting: averaging, missing value estimation and MIDAS. They set these 3 types of models to forecast 4 low frequency variables (GDP, Inflation, industrial production and employment growth). The novelty is that the quarterly dependent variable has monthly regressors whilst the 3 monthly ones are estimated and forecasted using daily data. The goal of this set up is to assess the differences in performances for the 3 selected models when the frequencies of both the independent and dependent variables vary.

Results are inconclusive. In general, Arnesto et al., find that more parsimonious models tend to be more accurate with the MIDAS model slightly improving over the averaging model. However for longer horizons these three models are almost the same.

Kuzin et al. (2011) compares the MIDAS model with a mixed frequency VAR model. The VAR model proposes is less restrictive than the MIDAS models but may suffer from high dimensionality. The authors conclude that for the euro area, both models are complemetary as MIDAS performs better for shorter horizons and, in turn, the VAR model yields more accurate forecasts for longer ones. The relationship between these two models is further explained in Bai et al. (2013) where state space models and MIDAS are studied. Reinforcing the findings from Kuzin et al. (2011), Bai et al. find that MIDAS and state space models are equal under ideal circumstances but that, even though state space models can be more accurate, they are more prone to parameter estimation errors.

In yet another paper that employs the MIDAS model, Kuzin et al. (2012) propose several methodologies to tackle the so called ragged edge as well as the mixed frequency sampling problems. The authors focus on nowcasting as an important way to validate policy decisions which requires both problems solved. Furthermore, they also claim that due to the fact that model selection is complicated (because of the wide variety of viable regressors and model specifications) model pooling is proposed as a more reliable method to improve nowcasting accuracy.

Finally, in a somewhat different application of the model, Çelik & Ergin (2013) use the mixed frequency data to predict stock market volatility in Turkey. The dataset contains information about futures sampled daily and intra-daily.

The conclusion is that compared to traditional volatility forecasting methods (GARCH), the usage of intra-day data improves the forecasts. Particularly, the use of the MIDAS model produces more accurate predictions during the recent crisis period.

The rest of the relevant papers, mainly Ghysels et al. (2004, 2006) and Andreou et al. (2013) are discussed in greater detail throughout this article and will consequently not be mentioned in this section.

3. The model

MIDAS

To explain the MIDAS model se shall resort, as an example, to two of the variables used in this article: as the dependable variable Mexican GDP growth and as the independent variable used to explain the GDP, the Goldman Sacks Commodity Index (GSCI) of silver. The GDP growth is a sampled quarterly and the GSCI index is sampled daily.

With this in mind suppose $Y_t^Q = GDP$ and $X_{m,t}^D = GSCI\ index$ and Q stands for quarterly, D for daily and m is the number of trading days in a quarter. Using this notation, a prediction of the GDP growth *h* periods in the future with the model proposed by Ghysels et al. (2004, 2006), has the following form:

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{j=0}^{p_Y^Q - 1} \rho_{j+1}^h Y_{t-j}^Q + \beta^h \sum_{j=0}^{q_X^D - 1} \sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} X_{m-i,t-j}^D + u_{t+h}^h$$

This model has a constant, the traditional AR terms with p_Y^Q quarterly lags of the dependent variables and a term that introduces q_X^D times m daily lags for the independent variable. In addition, special attention must be paid to the term multiplying the daily variable $w_{i+j*m}^{\theta^h}$. This term is the weighting scheme that will reduce the needed parameters to estimate and will therefore maintain a parsimonious model. Some of the weighting schemes available are presented shortly as explained in Ghysels et al. (2006). In all the definitions *N* is the number of high frequency lags used in the regression.

These weighting methodologies besides the U-MIDAS and the Almon lag Polynomial have to be estimated using non-linear LS, however, the advantages are clear since they reduce the number of parameters to estimate greatly while generally taking into account the generally accepted idea that the higher the lag, the smaller the relationship with the present value of the series i.e. the smaller the weight.

a) U-MIDAS: this is an unrestricted version in the sense that every high frequency lag has its own coefficient to estimate. It can be useful when *m* is small. A good characteristic of this weights is that they can be calculated using traditional OLS.

b) Normalized Beta probability function: this scheme calculates three parameters and has the following form:

$$w_i(\theta_1, \theta_2, \theta_3) = \frac{x_i^{\theta_1 - 1}(1 - x_i)^{\theta_2 - 1}}{\sum_{i=1}^{N} x_i^{\theta_1 - 1}(1 - x_i)^{\theta_2 - 1}} + \theta_3$$

Where $x_i = \frac{(i-1)}{(N-1)}$

This scheme can be made more parsimonious by restricting parameter 1 to be one and/or parameter three to be zero.

c) Normalized exponential Almon lag polynomial: this one estimates two parameters

$$w_i(\theta_1, \theta_2) = \frac{\exp(\theta_1 i + \theta_2 i^2)}{\sum_{i=1}^{N} \exp(\theta_1 i + \theta_2 i^2)}$$

As noted above the second parameter can be restricted to be zero.

d) Almon lag Polynomial: Is unable to identify the parameter β, therefore:

$$\beta w_i(\theta_0, \dots, \theta_P) = \sum_{p=0}^{P} \theta_p i^p$$

e) Step functions: As above it does not allow for the separate estimation of β:

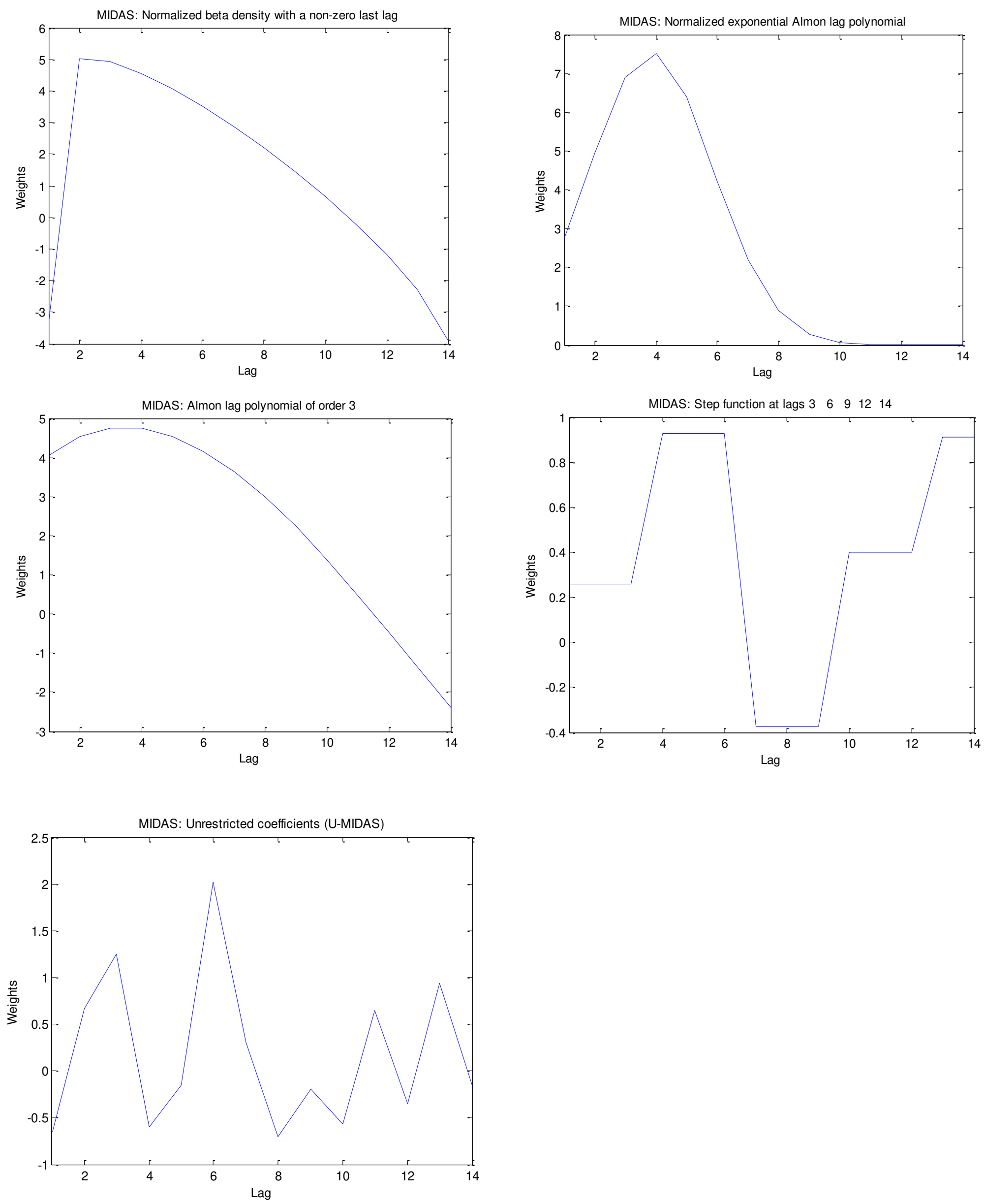
$$\beta w_i(\theta_0, \dots, \theta_P) = \theta_1 I_{i \in [a_0, a_1]} + \sum_{p=2}^{P} \theta_p I_{i \in [a_{p-1}, a_p]}$$

$$a_0 = 1 < a_1 < \dots < a_p = N$$

Where *I* is an indicator function and thus is 1 whenever *i* is between the specified interval and zero elsewhere

Next we show in figure 1 the calculated weights for different weighting schemes on a random model specification with one lag of the dependent variable and 14 of the independent high frequency variable.

Figure 1: Weights



Note: Calculated weights of different schemes for a model with 14 lags of the high frequency variable

In general we see that the first lags have bigger weights however it is not a monotonic decay as it responds to the optimal weights obtained after the non-linear LS problem is solved. It might seem mysterious that for some weighting schemes the last lag's weights become negative. We believe this is because the most flexible models that are bound by a functional form try to capture a monthly seasonal effect in the financial variable. If a model using an Almon polynomial with more lags and a higher polynomial degree is estimated, the computed weights oscillate roughly every 20 lags.

As a comparison, the more traditional way of using high frequency data is to make an average. In our case, that would mean to average the GSCI index info for each quarter and obtain 4 readings per year, what this actually means is that the weighting scheme used is assigning the same weight to all the lags in a quarter thus contradicting time series stylized facts even though parsimony is maintained.

Factor Analysis

This part is mainly based on Stock & Watson (2002). The idea behind their paper is to find a way to condense the information of a lot of variables into what they call "factors". The goal is to obtain a subset of these factors that explains the majority of the variation of the whole set of variables; the factor subset being much smaller than the whole set of variables.

Formally suppose there is a large set of independent variables that are going to be used for forecasting. This set X contains T variables and each variable has N observations. It is entirely possible that $T>N$. Therefore, the goal is to find a set of factors F and a set of parameters Λ that best explain X. That is:

$$X_t = \Lambda \mathrm{F}_t + e_t$$

Another way to look at a factor is to think of it as an unobservable variable that explains the variability of several of the observed variables.

To estimate the factors, the authors propose the use of the Principal Components Analysis which consists in minimizing the following expression:

$$V(\tilde{F}, \tilde{\Lambda}) = (NT)^{-1} \sum_i \sum_t (x_{it} - \tilde{\lambda}_i \widetilde{F_t})^2$$

Once factors are calculated, they are incorporated in the MIDAS regression instead of just one high frequency variable as before, that is:

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{j=0}^{p_Y^Q-1} \rho_{j+1}^h Y_{t-j}^Q + \beta^h \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} F_{m-i,t-j}^1 + u_{t+h}^h$$

In our case, the first factor accounts for 23% of the variability of the more than 250 daily time series used. The first 5 factors explain 42.7% of underlying variation.

An important caveat is that there are 2 types of factor estimation proposed in Stock & Watson (2002): static and dynamic. Andreou et al. (2013) utilize the dynamic method, whilst the static method is used in this article. The reason for this difference is that, following the conclusions of Stock & Watson (2002), the use of the dynamic method has a very modest effect on the subsequent forecasting.

An important aspect to consider is that the series have to be standardized before the factors can be obtained. This is necessary as a wide variety of series are employed and they differ in their units of measurement, and, even if they did share the same units of measurement, we would have wanted for all of them to have the same weight in calculating the forecasts.

The question now arises, as to what to do with the other estimated factors. How could we benefit from them without increasing the parameter count?

Forecast Combination

There are several cases when forecast combinations are advised [Timmermann (2006)]. For this study, it is clear that at least a few of these cases apply, for example, even though all the information sets for all the forecasts done in this exercise are known, it would still be unadvisable to create the so called “super model” that encompasses all the explicative variables in each information set. If such a thing was done the number of parameters would grow considerably; besides losing parsimony in the model, the cumulative parameter estimation error would make this model less reliable than a forecast combination of individually biased forecasting models.

Furthermore, an important reason to use combinations is the relatively large number of possible parameterizations of the MIDAS model. The flexibility of the model would pose a difficulty when deciding which one to use. Luckily, combination theory of allows us to circumvent this complication.

As a general result in the literature, forecast combinations are beneficial in terms of accuracy, but there is still great insufficient knowledge, as to why [Timmermann (2006)]. Nonetheless, following Androu et al. (2013), we present a few combinations that improve the Root Mean Squared Forecast Error (RMSFE) of the original predictions.

Formally:

$$\hat{Y}_{CM,t+h}^{Q,h} = \sum_{i=1}^{n} w_{i,t}^{h} \hat{Y}_{i,t+h}^{Q,h}$$

It is readily apparent that a forecast combination is nothing more than a weighted sum of the *n* forecasts for the horizon *h* of *n* models. Again, the important thing to select is the weighting scheme. However this is not a trivial issue; to get more intuition on how to assign the proper weights it is necessary to think on some type of loss function that will be helpful to pick the optimum combination. Formally, a combination of N forecasts is preferred to a single forecast if:

$$E\left[\mathcal{L}\left(\hat{Y}_{i,t+h}^{Q,h}, Y_{t+h}\right)\right] > \min_{C(\cdot)} E\left[\mathcal{L}\left(C\left(\hat{Y}_{1,t+h}^{Q,h}, \hat{Y}_{2,t+h}^{Q,h}, \dots, \hat{Y}_{N,t+h}^{Q,h}\right), Y_{t+h}\right)\right]$$

for $i = \{1,2, \dots, N\}$.

In the inequality above: $\mathcal{L}$ is a loss function that relates the forecasted and the observed value. Intuitively it is expected for the loss function to grow as the forecasted value grows farther apart from the actual value. C on the other hand, is the combination function that will relate all the individual forecasts. Thus we would like to select the function C that minimizes the expected loss, and the forecast combination would be preferred if the expected value of the loss function for such combination is smaller than each of the expected losses for each of the individual forecasts.

To finish this derivation let us suppose $\widehat{\boldsymbol{Y}}_{t+h}^{Q,h}$ is a vector containing all the individual forecasts and $\boldsymbol{w}_{t+h}^{h}$ is a vector of parameters, then the combination function can be rewritten as $C\left(\widehat{\boldsymbol{Y}}_{t+h}^{Q,h}\,;\boldsymbol{w}_{t+h}^{h}\right)$. The last part is to define a loss function, following Androu et al. (2013), the Mean Squared Forecast Error (MSFE) is used as it is, they claim, the one that gives the highest improvement in forecasts. Given the above mentioned suppositions, the solution is a linear combination of the individual forecasts as stated in the beginning.

The MSFE weights are selected by analyzing the forecasting performance of the model and assigning a weight inversely related to the MSFE calculated for each model.

Nowcasting

To close this section an interesting topic is discussed. The MIDAS models have the ability of adding recent information to improve the forecasts. To better understand this characteristic suppose that next quarter GDP growth needs to be predicted. If a month since the beginning of the present quarter has passed, there is a plethora of financial information generated during the course of the present quarter. Using the information up to date to forecast the next data point of a variable of interest is called *nowcasting*.

Formally the MIDAS model is extended in the following way:

$$Y_{t+h}^{Q,h} = \mu^{h} + \sum_{j=0}^{p_Y^Q-1} \rho_{j+1}^{h} Y_{t-j}^{Q} + \beta^{h}\left[\sum_{i=(3-Jx)*m/3}^{m-1} w_{i-m}^{\theta^h} X_{m-i,t+1}^{D} + \sum_{j=0}^{q_X^D-1}\sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} X_{m-i,t-j}^{D}\right] + u_{t+h}^{h}$$

The new term has two noticeable things, first, the sub index *t+1* says that it is high frequency information generated during the present quarter. The other important thing to notice is the value of *i* and the variable *Jx*. Let's suppose *m=60*, that means there are 60 trading days in a quarter, if the first month of the quarter has just finished there are 20 days of data available thus *Jx=1* needs to be selected to obtain the appropriate limits of the sum.

4. Data

The high frequency data is divided basically into 5 different categories of financial information: commodities (166 series), equities (94 series), foreign exchange (27 sereis), corporate risk (53 series) and fixed income (52 series) and as previously stated, the dependent variable is the Mexican GDP. The period of estimation is 1999Q1-2009Q4 and the period of forecasting is 2010Q1-2013Q4 i.e. 16 quarters.

The time series of the Mexican GDP though not as long as the one from developed countries, does reach back to 1993. Nevertheless, the forecasting and estimation periods were shorter because a lot of useful financial information is available from 1999 onwards. The author is aware that this might be a short period for forecasting purposes, however this allowed for the inclusion of useful information and the usefulness of it outweighs the limitations of the short estimation period.

Another limitation of the information comes from the fact that the fixed income financial information for the Mexican bonds was, to say the least, hard to obtain because it had a different sampling frequency or because the records started on a date well after the beginning of the estimation period. The same happened for commodity and equity information specific to Mexico. That is why the database constructed is primarily a subset of the time series suggested by Androu et al. (2013). Nonetheless, there are a few notable mentions regarding the Mexican data:

A) The CETES 28 days rate is part of the dataset as a part of the Fixed Income group. It was especially important to have information on the risk-free rate for Mexico since it is the tool of policy makers rely on to alter the business cycle, this is even more important considering the fact that the financial crisis lies within our estimation window. It would have been desirable to obtain data from other CETES maturities however US bonds and bills should compensate this lack of information somewhat

B) The foreign exchange rates are solely in terms of the Mexican peso.

C) In the equity section there are 2 indicators of relevance: MEXBOL which is an IPC index of the Mexican Stock Exchange and VIMEX which is a measure of the expected volatility in the Mexican Stock Market.

For detailed information concerning the series used please refer to the appendix for a complete list of the 392 series used.

To close this section it is necessary clarify out that some of the series employed here were transformed because they were not stationary. An augmented Dickey Füller (ADF) test with 12 lags was utilized to check for unit roots. Those series not deemed stationary were transformed in a similar fashion: the log of the original series was differenced. This procedure removes the stationarity problem and also smooths the series. To conclude a second ADF text was executed to verify the series were indeed stationary.

It is worth noting that the GDP growth was not modified so as to remove seasonality, but instead regressions were run with seasonal dummy variables in order to improve forecast accuracy.

All the financial information was retrieved from Bloomberg and the access to this resource was obtained through Banco de Mexico (BANXICO). The CETES 28 rate was obtained from BANXICO's webpage and the GDP was obtained from INEGI's.

Another important set of information is the quarterly macro data; this set comprises 20 macro variables found to have good explicative and forecasting power [Andreou et al. (2013)]. The set contains information such as CPIs, international trade variables and economic activity indexes for Mexico and it also includes USA's GDP and inflation rate. Some of the data was monthly so a flat aggregation scheme was used to transform it into quarterly data i.e. the monthly data was averaged for every quarter.

The Mexican data was obtained from INEGI and BANXICO and the USA data was obtained from the Federal Reserve Bank of St. Louis webpage.

5. Results

Before showcasing results of the forecasting a few points require further clarification. The first one is the windowing as there are 3 options as described below:

a) Fixed Window: The estimation of the model begins in the estimation start date 1999Q1 and ends in 2009Q4. The estimation is done only once and the forecasts are multistep.

b) Rolling Window: It begins in 1999Q1+i and ends in 2009Q4+i. The model is estimated each time the window changes and the forecast is done only one step ahead. The new window includes the next actual value of the dependent variable. The window stays constant in size throughout the whole forecasting process.
c) Running Window: The start date is fixed at 1999Q1 however the end date changes with each forecasted value, which is 2009Q4+i. Similar to the Rolling Window, the model is estimated each time the window changes and the forecasts are computed one step ahead. This window grows with each forecasted point as it includes the next observed value.

Windowing is important as it affects the forecasts, intuitively, it would be expected that the rolling window and recursive window perform better because they re-estimate the model for each forecast and because they forecast just one period ahead. The caveat is the computational time that increases considerably.

Employing a fixed window is required to be able to obtain the predictions for different forecasting horizons.

The second important aspect when forecasting is to specify whether the exercise is in real time or not as results may change. The main goal in this work is to test the hypothesis that high frequency financial data can improve forecast accuracy over traditional models that is why this model is not a real time exercise and utilizes revised data.

However, an interesting capability of the MIDAS model is nowcasting as stated above so let us extend this notion and the application to a real time exercise. As described previously, nowcasting permits the model to forecast using up to date information. Suppose that every month has exactly 20 trading days and that all the daily information used is available at the end of the day, it would be theoretically feasible to obtain 60 progressive forecasts of the next quarter's GDP.

The problem is that the quarterly GDP figures are published with a lag of two months on average and this is why nowcasting becomes more relevant as it would be able to incorporate new information to the forecast up to the time when the value is officially published.

In the case where Q1GDP has been published, that would mean it is 2 months deep into the next quarter. The model allows the user to use this extra information to forecast Q2 and it does more, because it is also capable of using data up to the date the Q2 GDP is officially published; if we suppose the publishing lag of two months applies to next quarter too then up to 5 months of useful information could be used to estimate and forecast.

A small exercise is presented in table 1 to illustrate this line of thought. 2 weighting schemes are used to show the RMSFE of nowcasting for up to 5 months of updated information on a monthly basis (i.e., forecasts are performed every month after the last published quarter). Only 2 models are chosen because as time passes, the less parsimonious models can no longer accommodate the increasing amount of lags, therefore, only the Beta NZ and the exp. Almon lag schemes are shown for an arbitrary model with one lag of the dependent variable and 15 lags of the independent variable and a fixed window to save time:

Table 1: RMSFE comparison for models with leads

	0m	1m	2m	1q	4m	5m
Beta NZ	1.349	1.476	1.984	1.210	1.609	1.898
Exp. Almon	1.715	1.211	2.011	1.057	1.179	1.185

Note: This table presents the root mean square forecast errors (RMSFE) for 2 model specifications with leads. The GDP is forecasted and nowcasted for the sample 1999Q1-2013Q4: Estimation period: 1999Q1-2009Q4. Forecasting period:2010Q1-2013Q4.

Each column shows the RMSFE with the information available farther into the next quarter and beyond. The first column shows the forecast with no updated information beyond current quarter, (i.e., traditional forecasting).

In general, the accuracy improves compared to traditional forecasting for each passing month of new information except for month 2. However, between the two models it is found that the Exp Almon Lag performs better in the RMSFE for the later months; this seems to be because it is better at incorporating a higher order of lags of the high frequency variable to the estimation since we would be talking of more than 100 lags for 5 months into the next quarter. It also helps that it is the most parsimonious model available.

Model Selection

The first step is the model selection, since there is a wide variety of options available it is necessary to resort to the Akaike information criteria (AIC) and the Bayesian information criteria (BIC) for the aforementioned weighting schemes, a recursive window and different combinations of lags for both the dependent and the independent variable. From now on all the MIDAS models will be using the first factor as the high frequency variable unless otherwise specified.

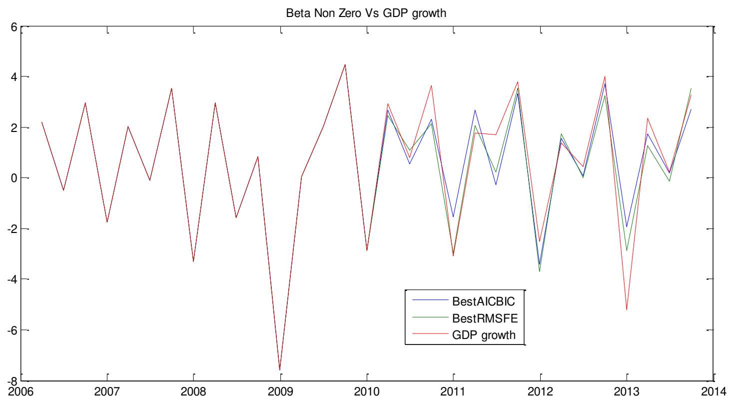
The tests to identify the best models were done using 1-7 lags of the dependent variable and 5-30 of the independent factor. These tests favor one model for each weighting scheme, the one chosen for all the following comparisons and tests is the a Beta Non Zero last lag weight with 1 lag for the dependent variable and 10 for the independent factor (p=1, q=10). An important reason behind this election is that the variance of the RMSFE of this weighting scheme is smaller.

This model was chosen because of two reasons, first, even though other models had the exact same lag selection, this one in particular presented the smallest RMSFE, the second reason is that both the AIC and BIC concluded in favor of the same lag structure, this is convenient because this way it is almost certain that the model is parsimonious and its residuals are uncorrelated and that its parameters are significant.

It is important to note that even though this model had a fairly low RMSFE amongst all the obtained, it was not the best in its family. The smallest RMSFE was found to be achieved by a bigger model with 7 lags of the low frequency variable and 5 of the high frequency one (p=7, q=5). One must not forget that regardless of the high frequency lags specified, the model estimates only 3 because of the weighting scheme.

To conclude this part a graph is presented that compares the GDP growth and the two models mentioned above

Figure 2: MIDAS Beta Non Zero forecasts

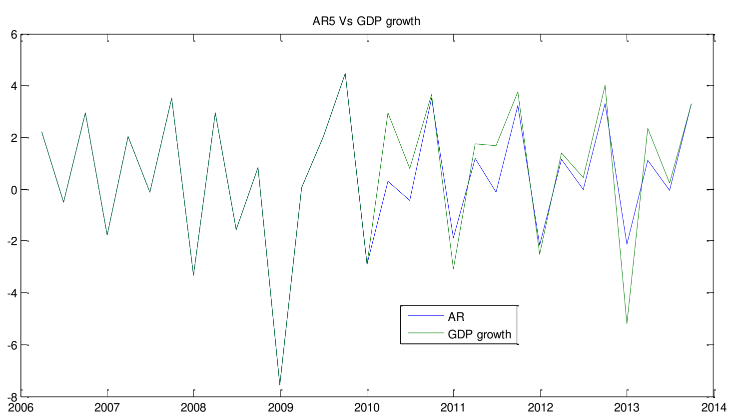


Note: The graph shows the difference of the logarith of the observed GDP vs the forecasted values for the period 2010Q1-2013Q4 using the models described in the figure

The Benchmark and some graphs

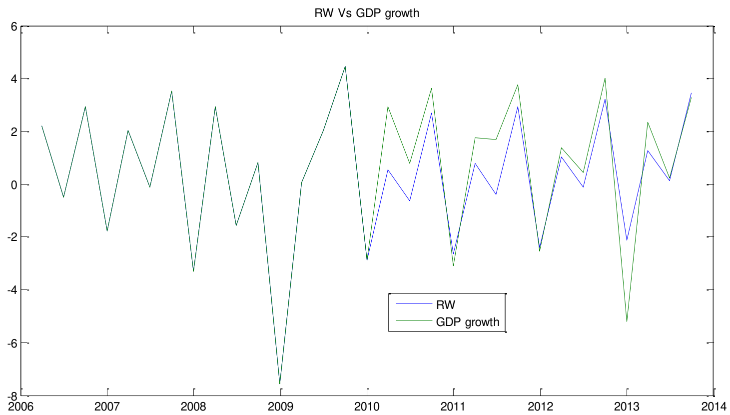
To be certain of the accuracy of the chosen model in the previous subsection, a benchmark is needed. Two "traditional" models are estimated and used to forecast the same period: an AR and a random walk model. The AR was chosen using the AIC/BIC criteria as well and it is an AR1. Figures 2 and 3 show their respective graphs

Figure 3: AR forecasts



Note: The graph shows the difference of the logarith of the observed GDP vs the forecasted values for the period 2010Q1-2013Q4 using the models described in the figure

Figure 4



Note: The graph shows the difference of the logarith of the observed GDP vs the forecasted values for the period 2010Q1-2013Q4 using the models described in the figure

The next step is table 2 that summarizes the results for all the estimations.

Table 2: RMSFE comparison for models with no leads

	Model	h=1 RMSFE	h=1 RMSFE as % of AR	h=4 RMSFE	h=4 RMSFE as % of AR
Benchmark	AR	1.135	1	1.114	1
	RW	1.289	1.137	3.233	2.903
Factor 1	Beta NZ Best AICBIC (p=1, q=10)	1.031	0.909	1.047	0.940
	Beta NZ Best RMSFE (p=7, q=5)	1.018	0.897	1.095	0.984
	Step Func Best RMSFE (p=6, q=5)	0.999	0.881	0.969	0.870
commodities F1	Beta NZ (p=1, q=14)	1.072	0.945	1.052	0.945
equities F1	Beta NZ (p=1, q=16)	1.277	1.125	1.262	1.134
corporate F1	Beta NZ (p=1, q=5)	1.456	1.283	1.423	1.278
FX F1	Beta NZ (p=1, q=7)	1.511	1.331	1.540	1.383
Fixed Income F1	Beta NZ (p=1, q=9)	1.621	1.428	1.806	1.622
Factor 2	Beta NZ Best AICBIC (p=1, q=17)	1.066	0.939	1.129	1.014
	Beta NZ Best RMSFE (p=5, q=17)	1.030	0.908	1.126	1.011

Forecast Combinations		h=1		h=4	
Beta NZ Best AICBIC (p=1, q=10)	Beta NZ Best AICBIC Factor 2	0.988	0.870	1.038	0.932
Beta NZ Best AICBIC (p=1, q=10)	Step Func Best RMSFE Factor 1	0.954	0.841	0.948	0.851

Note: The table shows the root mean square forecast error (RMSF) for 1 and 4 step ahead horizons of the GDP for the sample 1999Q1-2013Q4: Estimation period: 1999Q1-2009Q4. Forecasting period: 2010Q1-2013Q4. It also presents an autoregresive (AR) and a random walk (RW) model as benchmark and the RMSFE are also presented as a percentage of the AR. The forecasts are calculated for the first daily factor of the 392 financial variables and the Beta NZ weighting scheme. It is also calculated for the first factor of the individual variable groups. Finally a few forecast combinations between different model specifications and different information sets. A recursive window is used for all the calculations.

Table 2 presents forecasts computed with two different forecasting horizons: 1 quarter ahead (h=1) and 1 year ahead (h=4). As suggested by the previously summarized literature, the forecasting power of the MIDAS model decreases for most of the specifications as the forecasting horizon increases.

Out of the two traditional models selected for this exercise the random walk model performs poorly compared to the AR1. The MIDAS RMSFE outperforms both of the traditional models in both horizons thus proving the usefulness of the high frequency data for forecasting. This model employs the first factor that explains around 20% of total variation of the larger set. The 2 best RMSFE models for 2 different weighting schemes illustrate that the model chosen by an information criterion is not too distant in terms of predictive power.

A factor decomposition identic to the one applied to the whole dataset is then applied to each group of financial variables. From this decomposition, 5 explicative factors are extracted for each of the 5 groups of financial variables and it is shown here the results of forecasting with the first factor of each group. The weighting scheme is a Beta non zero and the lags are selected using Akaike and Bayesian criteria.

Even though this is a parsimonious weighting specification, predictive power for all variable groups except commodities, is worse than the traditional models. In other words, parameter estimation error for these model specifications outweighs the additional information incorporated through g financial series from two otherwise equal models (1 lag of the dependent variable and seasonal dummies).

There are two plausible explanations: the first is that the commodities dataset is the most complete in terms of how much it resembles the one used by Andreou et al. (2013) and further work needs to be done on the other 4 sets of data or perhaps it is because the other sets are not as adequate for Mexico as they are for US.

From the dataset it is apparent that corporate risk and fixed income are two groups that focus on US and even though there are some variables such as risk free interest rates that are without doubt good proxies for the analogous Mexican variables, they could be failing to provide sufficient information. The equities might also suffer from a similar problem.

The foreign exchange group of variables is related to Mexican trade solely however the lack of predictive power might be explained to the inconclusive positive relation between commerce and growth. There is a long standing debate in the literature, for example Rodriguez & Rodrik (2001) and Dollar & Kraay (2004), to determine whether commerce is good for growth. The most reasonable answer is that this needs to be determined for each individual country [Wacziarg & Horn-Welch(2003)] and Mexico is one of the uncertain ones. However this group's power might be improved by the introduction of other financial instruments like futures and forward contracts other than just spot prices, the caveat is that they might be difficult to come by for the periods of interest in this article.

Nonetheless, when all the variables are put together and the Factors contain mixed information it is clear that they are successful in improving the accuracy of the model.

The last section of the table presents two combined forecasts

Best AIC BIC from Factor 1 y Factor 2: uses MSFE combination to mix two Beta non-zero model, each optimal in the AICBIC sense but for 2 different factors. The first is our friend (p=1, q=10) for the first factor and the second one (p=1, q=17) is for the second calculated factor.

BETA Non-zero & Step function): uses MSFE combination to produce a better forecast by mixing the results obtained using a Beta non-zero model (p=1, q=10) and Step Function model (p=6, q=5) which happened to be the second best predictive model.

As expected both combinations yield a lower RMSFE. The first combination's improvement is explained because of the combination of the information set of each estimation and prediction. The combined forecast manages to extract the best qualities from both information sets using the MSFE as explained above. Notably, the RMSFE is even smaller than the best RMSFE obtained for the Beta Non Zero and the Step Function specifications: in a real time forecast this method tends to greatly improve on the accuracy of otherwise difficult to select models.

Conversely, the last combination might be explained in of functional form of the models. The different weighting schemes impose different biases on the estimations; a combination of both specifications manages to reduce this bias by extracting the more desirable qualities of each model.

The goal of the final part of this section is to investigate whether MIDAS regression model with daily financial data is useful in forecasting GDP beyond macroeconomic data. We also compare the forecasting accuracy of the MIDAS model with the traditional models that take a simple average of daily financial data.

Table 3 contains a summary of several model's RMSFE and they require some explanation:

a) The macro data incorporated to the models is analyzed and synthesized using factor analysis in a likewise fashion to the one applied to the financial variables. As a result, the first 3 factors resume close to 76% of overall variation.

a) *Flat* is used to denote the family of models that use high frequency data alongside quarterly data by using a flat aggregation scheme, in other words, all trading days of the daily financial assets within the quarter were averaged to obtain a single value per quarter.

b) Combined MIDAS is used to refer to a combination of 5 MIDAS specifications: one for each of the 5 information sets available in the 5 factors calculated

Table 3: RMSFE comparison for models with no leads

	h1		h4	
Model	RMSFE	RMSFE as % de AR	RMSFE	RMSFE as % de AR
Benchmark				
AR	1.135	1	1.113	1
RW	1.289	1.137	3.233	2.903
Benchmark + macro data				
AR	0.741	0.652	0.730	0.656
RW	1.658	1.461	3.117	2.799

Financial Data				
Flat	1.818	1.602	1.553	1.395
Beta NZ (p=1, q=10)	1.031	0.909	1.047	0.940
Combined MIDAS	1.024	0.902	1.042	0.936
Financial Data + macro data				
Flat	0.716	0.631	0.668	0.600
Beta NZ (p=1, q=10)	0.660	0.582	0.709	0.637
Combined MIDAS	0.571	0.504	0.607	0.545

Note: The table shows the root mean square forecast error (RMSF) for 1 and 4 step ahead horizons of the GDP for the sample 1999Q1-2013Q4: Estimation period: 1999Q1-2009Q4. Forecasting period: 2010Q1-2013Q4. It also presents an autoregresive (AR), a random walk (RW) and a flat aggregation model (FLAT) as benchmarks and the RMSFEs are also presented as a percentage of the AR. The 5 MIDAS forecasts calculated from each one of the daily factors are combined to obtain the Combined MIDAS RMSFE. A recursive window is used for all forecasts.

Table 4: Diebold-Mariano comparisons for models with no leads

	h=1	h=4
Model	DM	DM
Financial data Vs AR		
Flat	0.073	0.215
Combined MIDAS	0.327	0.430
Financial data Vs Benchmark + macro data (AR)		
Flat	0.015	0.032
Combined MIDAS	0.246	0.312
Financial + macro data Vs Benchmark + macro data (AR)		
Flat	0.750	0.477
Combined MIDAS	0.043	0.122
MIDAS Vs Flat		
Finance data	0.039	0.128
Finance + macro data	0.059	0.411

These table reports p-values of a test to compare predictive ability; the comparison is based on a Diebold-Mariano test. Sample 1999Q1-2013Q4:Estimation period: 1999Q1-2009Q4. Forecasting period:2010Q1-2013Q4. Recursive window

The actual comparisons can be found in Table 4, these table shows the p-values obtained from a Diebold and Mariano (1995) test that consists of checking a null hypothesis that states that 2 different models have the same forecasting ability. Formally we define a forecast loss function for model i as:

$$g(e_{i,t}) = e_{i,t}^2$$

Under the null both models have equal forecasting ability, that is:

$$H_o: g(e_{1,t}) = g(e_{2,t})$$

Diebold and Mariano (1995) first define the difference between the loss functions for 2 alternative models: $d_t = g(e_{1,t}) - g(e_{2,t})$ and then propose the following statistic to verify the null:

$$DM = \frac{\bar{d}}{\sqrt{var(\bar{d})}}$$

Where $\bar{d}$ is the sample mean of d_t and $\sqrt{var(\bar{d})}$ is defined as $\sqrt{var(\bar{d})} = \frac{\gamma_0+2\gamma_1+\cdots+2\gamma_q}{H-1}$: H is the number of forecasted periods and $\gamma_j = cov(d_t, d_{t-j})$. The statistic has a t-student distribution with H-1 degrees of freedom. The p-values on table 4 come from running a regression with robust errors of d_t on a constant and testing whether the constant is statistically significant.

If we analize the results from table 3, it is clear that the combined MIDAS model with financial data has, in general, better RMSFE than the benchmark models. For example, the use of financial data improves over the univariate AR benchmark by 10%. However it is also important to notice that the macroeconomic regressors make a difference when they are included into both models. This is not surprising as they are highly correlated to the GDP, however the combined MIDAS improves upon the AR with macro regresors by 15%.

Table 4 however gives us an important lesson: we cannot distinguish statistically between the MIDAS model with financial data and the benchmark with macroeconomic regressors.

In short, from table 3 and 4 the 2 main things we learn are that the MIDAS model with financial data possesses, statistically, the same predictive ability as a traditional model with macro data and that using quarterly macro data and financial data in the MIDAS model improves forecasting ability over traditional models with macro data.

The results suggest that the inclusion of financial data provides the model with useful information to forecast the GDP.

It is also possible to observe that the flat aggregation with just financial factors has poor predictive ability and that MIDAS has gains over the flat aggregation with financial data and with financial and macro data that are significant at least at the 5% and 10% respectively. We conclude that MIDAS is superior to a simple flat aggregation scheme.

Lastly, table 5 and 6 show the same results for *nowcasting* with information one month farther into the quarter. From these tables our main findings are that *nowcasting* and forecasting with MIDAS are statistically indistinguishable prediction-wise. We note, however, that for the same model presented above, this result is dependent on the weighting schemes used.

Importantly, we find that MIDAS is statistically superior to the 10% in its predictive ability over the flat aggregation scheme.

Most of the results for h=1 are comparable to the ones obtained in Andreou et al. (2013). Contrary to our results for h=4 Andreou et al. find statistically significant differences in predictive power that favor MIDAS model, whereas, we find similar results as the ones found by Marcellino & Schumacher (2010) and Arnesto et al. (2010), that conclude that for higher forecasting horizons the forecasting power differences between the models are similar or, at least, that choosing one among them is. This might be related to the level of adequacy of the financial variables to the Mexican economy relative to the US in terms of their predictive power.

Table 5: RMSFE comparison for models with leads

	h1		h4	
Model	RMSFE	RMSFE as % de AR	RMSFE	RMSFE as % de AR
Financial Data				
Flat	1.379	1.215	1.319	1.184
Combined MIDAS	1.062	0.936	1.089	0.979
Financial Data + macro data				
Flat	0.688	0.607	0.713	0.640
Combined MIDAS	0.550	0.485	0.588	0.528

The table shows the root mean square forecast error (RMSF) for 1 and 4 step ahead horizons of the GDP for the sample 1999Q1-2013Q4: Estimation period: 1999Q1-2009Q4. Forecasting period: 2010Q1-2013Q4. It also presents an autoregresive (AR), a random walk (RW) and a flat aggregation model (FLAT) as benchmarks and the RMSFEs are also presented as a percentage of the AR. The 5 MIDAS forecasts calculated from each one of the daily factors are combined to obtain the Combined MIDAS RMSFE. A recursive window is used for all forecasts.

Table 6: Diebold-Mariano comparisons of models with leads

	h=1	h=4
Model	DM	DM
nowcasting Vs forecasting		
Flat (financial data)	0.081	0.208
Combined MIDAS (financial data)	0.488	0.413
Flat (financial+macro data)	0.411	0.235
Combined MIDAS (financial+macro data)	0.154	0.170
Flat Vs MIDAS		
Financial data	0.080	0.255
Financial+macro data	0.082	0.158

These table reports p-values of a test to compare predictive ability; the comparison is based on a Diebold-Mariano test. Sample 1999Q1-2013Q4:Estimation period: 1999Q1-2009Q4. Forecasting period:2010Q1-2013Q4. Recursive window

6. Conclusions

Along the lines proposed by Ghysels et al. (2004, 2006) and applied in Andreou et al. (2013), a MIDAS model which incorporates a large dataset of financial variables is estimated using factors and then used to provide out of sample forecasts of the Mexican GDP. We found that the use of this methodology and dataset improves on the accuracy beyond quarterly macroeconomic data in more traditional models.

Furthermore, the use of the MIDAS model suggests a method to circumvent the problems initially found when dealing with data of different sampling frequencies whilst maintaining the requirements of an adequate model. The use of factor analysis and forecast combination is chosen as a valid way to reinforce the positive characteristics of the MIDAS model and of the use of High Frequency data.

The model comparisons in section 5 favor the use of the MIDAS model against one with flat aggregation but they also suggest that a MIDAS model is more efficient than the AR model extended to include macro variables as regressors. When nowcasting, the results favor the MIDAS model over the flat aggregation scheme even though they are not significantly more accurate than traditional forecasting. As suggested before, the nowcasting prowess is affected by the weighting scheme used, therefore, further testing is required compare the results of different model specifications.

More specifically, we conclude that this methodology is a viable way to improve the forecasts for developing countries; which show a higher volatility in their macro variables.

In order to improve or extend this work, a more solid group of financial variables could be chosen that are more directly related to Mexico. The lack of historic data on useful variables might be a hindrance that could be solved simply by waiting and repeating the exercise in a few years from now.

There is also the option of employing this same dataset to predict other monthly or quarterly macro variables such as unemployment or inflation.

In Andreou et al. (2013), forecasts using financial data are compared against those calculated employing a different dataset that contains monthly indicators and macro variables. Perhaps for Mexico a similar set of variables is more readily available and would improve forecasting accuracy for at least for some time, until there is a longer historic time series and/or other index and indicators are developed.

Appendix A: Data

Finacial Data

	Type	Description
1	Commodity	RJ CRB
2	Commodity	SILVER
3	Commodity	brent oil
4	Commodity	PL-NYD
5	Commodity	ZINC
6	Commodity	XPD-D
7	Commodity	WHEAT
8	Commodity	C-US2D
9	Commodity	SOYB
10	Commodity	COTTON
11	Commodity	SUGAR
12	Commodity	COFFEE
13	Commodity	COCOA
14	Commodity	CATTLE
15	Commodity	HOGS
16	Commodity	GOLD
17	Commodity	ALUMINUM
18	Commodity	WTI OIL
19	Commodity	LEAD
20	Commodity	NICKEL
21	Commodity	TIN
22	Commodity	ALUM FUT
23	Commodity	LEAD FWD
24	Commodity	NICKEL FWD
25	Commodity	TIN FWD
26	Commodity	wti first
27	Commodity	heating oil first
28	Commodity	gas oil first
29	Commodity	nat gas first
30	Commodity	corn first
31	Commodity	soybean first
32	Commodity	wheat first
33	Commodity	rough rice first
34	Commodity	lumber first
35	Commodity	sugar first
36	Commodity	gold
37	Commodity	copper stock
38	Commodity	nickel stock
39	Commodity	aluminum stock
40	Commodity	zink stock
41	Commodity	lead stock
42	Commodity	tin stock

43	Commodity	wti spot midland
44	Commodity	european brent crude
45	Commodity	copper closing plrice
46	Commodity	aluminum closing price
47	Commodity	nickel closing price
48	Commodity	zink closing price
49	Commodity	tin closing price
50	Commodity	copper fwd price
51	Commodity	aluminum fwd price
52	Commodity	nickel fwd price
53	Commodity	zink fwd price
54	Commodity	lead fwd price
55	Commodity	tin fwd price
56	Commodity	fiber all items
57	Commodity	fiber metal
58	Commodity	fiber textiles
59	Commodity	fiber oil
60	Commodity	gsci heating oil
61	Commodity	gsci crude oil
62	Commodity	gsci gasoline
63	Commodity	gsci gasoil
64	Commodity	gsci gas
65	Commodity	gsci metals
66	Commodity	gsci aluminum
67	Commodity	gsci copper
68	Commodity	gsci lead
69	Commodity	gsci nickel
70	Commodity	gsci zink
71	Commodity	gsci precious metals
72	Commodity	gsci gold
73	Commodity	gsci silver
74	Commodity	gsci agriculture
75	Commodity	gsci wheat
76	Commodity	gsci soy
77	Commodity	gsci cotton
78	Commodity	gsci sugar
79	Commodity	gsci coffee
80	Commodity	gsci cocoa
81	Commodity	gsci energy
82	Commodity	gsci livestock
83	Commodity	gsci cattle
84	Commodity	gsci hogs
85	Commodity	gsci softs
86	Commodity	gsci light energy
87	Commodity	gsci energy metals
88	Commodity	gsci non livestock
89	Commodity	gsci grains
90	Commodity	gsci all wheat
91	Commodity	gsci all crude
92	Commodity	gsci biofuel
93	Commodity	gsci 1 m fwd
94	Commodity	gsci 3m fwd
95	Commodity	gsci 2m fwd
96	Commodity	gsci 4m fwd

97	Commodity	gsci 5m fwd
98	Commodity	excess return total
99	Commodity	excess return crude oil
100	Commodity	excess return brent
101	Commodity	excess return gasoline
102	Commodity	excess return heating oil
103	Commodity	excess return gasoil
104	Commodity	excess return nat gas
105	Commodity	excess return metals
106	Commodity	excess return aluminum
107	Commodity	excess return copper
108	Commodity	excess return lead
109	Commodity	excess return zinc
110	Commodity	excess return precious metals
111	Commodity	excess return gold
112	Commodity	excess return silver
113	Commodity	excess return agri and livestock
114	Commodity	excess return soybean
115	Commodity	excess return corn
116	Commodity	excess return cotton
117	Commodity	excess return sugar
118	Commodity	excess return coffee
119	Commodity	excess return cocoa
120	Commodity	excess return livestock
121	Commodity	excess return hogs
122	Commodity	excess return non energy
123	Commodity	excess return light energy
124	Commodity	excess return ultra energy
125	Commodity	excess return energy metals
126	Commodity	excess return petroleum
127	Commodity	excess return grains
128	Commodity	excess return all wheat
129	Commodity	excess return all crude
130	Commodity	excess return biofuel
131	Commodity	total return total
132	Commodity	total return crude oil
133	Commodity	total return energy
134	Commodity	total return gasoline
135	Commodity	total return heating oil
136	Commodity	total return gasoil
137	Commodity	total return nat gas
138	Commodity	total return metals
139	Commodity	total return aluminum
140	Commodity	total return copper
141	Commodity	total return zinc
142	Commodity	total return precious metals
143	Commodity	total return gold
144	Commodity	total return silver
145	Commodity	total return agri and livestock
146	Commodity	total return agriculture
147	Commodity	total return soybean
148	Commodity	total return corn
149	Commodity	total return cotton
150	Commodity	total return sugar

151	Commodity	total return coffee
152	Commodity	total return cocoa
153	Commodity	total return livestock
154	Commodity	total return cattle
155	Commodity	total return hogs
156	Commodity	total return non energy
157	Commodity	total return light energy
158	Commodity	total return ultra energy
159	Commodity	total return energy metals
160	Commodity	total return petroleum
161	Commodity	total return grains
162	Commodity	total return all crude
163	Commodity	philadelphia semiconductor
164	Commodity	corn spot price
165	Commodity	palladium
166	Commodity	platinum
167	Equity	INMEX
168	Equity	S&P500
169	Equity	S&P 500 Industrials Sector Index GICS Level 1
170	Equity	Dow Jones Industrial Average - DJI
171	Equity	NASDAQ
172	Equity	NASDAQ 100
173	Equity	VIX
174	Equity	IPC
175	Equity	Dow Jones Industrial Goods and Services Titans 30 Index Euros
176	Equity	Dow Jones Transportation Average
177	Equity	Dow Jones Utilities Average
178	Equity	Dow Jones Composite Average
179	Equity	Dow Jones Internet Commerce Index
180	Equity	Dow Jones Internet Composite Index
181	Equity	S&P 500 Industrials Sector TR Index
182	Equity	S&P 500 Financials Sector Index GICS Level 1
183	Equity	S&P Smallcap 600 Index
184	Equity	NASDAQ Industrial Index
185	Equity	Russell 2000 Index
186	Equity	Value Line Arithmetic
187	Equity	Value Line Geometric
188	Equity	Dow Jones Equity REIT Total Return Index
189	Equity	Dow Jones US Completion Total Stock Market Total Return Index
190	Equity	CBOE Equity Put/Call Ratio
191	Equity	CBOE US Put/Call Ratio Composite Intraday
192	Equity	PUT/CALL RATIOS COMPOSITE
193	Equity	CBOE US SPX Put/Call Ratio Intraday
194	Equity	Put/Call Volume Ratios on SPX Pt/Cl
195	Equity	CBOE US Index Put/Call Ratio Intraday
196	Equity	Put/Call Ratios OEX Pt/Cl
197	Equity	CBOE US OEX Put/Call Ratio Intraday
198	Equity	Put/Call Ratios RUT Pt/Cl
199	Equity	US Option Call Volumes on CBOE
200	Equity	US Option Put Volumes on CBOE
201	Equity	New York Stock Exchange Advancing Stocks
202	Equity	New York Stock Exchange Declining Stocks
203	Equity	NASDAQ Total Volume Composite
204	Equity	NASDAQ Advancing Stocks Index

205	Equity	Nasdaq Declining Stocks Index
206	Equity	FTSE 100 Index
207	Equity	FTSEurofirst 300 Index
208	Equity	Dow Jones Islamic Market World Index
209	Equity	Dow Jones Sustainability World Composite Index
210	Equity	Dow Jones World Technology Index
211	Equity	Dow Jones Sustainability World Total Return Index Composite USD
212	Equity	Dow Jones Sustainability World Total Return Index Composite Euro
213	Equity	Dow Jones Islamic Market World Developed Index
214	Equity	Dow Jones Islamic Market World Total Return Index
215	Equity	Dow Jones World Financials Index
216	Equity	Dow Jones Sustainability World Index in EUR
217	Equity	Dow Jones World Consumer Goods Index
218	Equity	Dow Jones Sustainability World Total Return Index Ex AGTAF Euro
219	Equity	Dow Jones World Consumer Services Index
220	Equity	Dow Jones Sustainability World Ex Alcohol Tobacco Gambling Armaments & Firearms
221	Equity	Dow Jones World Oil & Gas Index
222	Equity	Dow Jones Islamic Market World Malaysia Index USD
223	Equity	Dow Jones Islamic Market World Developed Total Return Index
224	Equity	Dow Jones World Basic Materials Index
225	Equity	Dow Jones Islamic Market World Excluding US Index
226	Equity	Dow Jones Sustainability World Excluding Tobacco Index
227	Equity	Dow Jones World Healthcare Index
228	Equity	Dow Jones Islamic Market World Emerging Markets Total Return Index
229	Equity	Dow Jones Sustainability World Index in CHF
230	Equity	Dow Jones Sustainability World Total Return Index Ex Gambling USD
231	Equity	Dow Jones Sustainability World Total Return Index Ex Tobacco USD
232	Equity	Dow Jones Sustainability World Total Return Index Ex Tobacco Euro
233	Equity	Dow Jones Islamic Market World Developed Excluding US Total Return Index
234	Equity	Dow Jones Islamic Market World Excluding US Total Return Index
235	Equity	Dow Jones Sustainability World ex Australia Index ex Tobacco AUD
236	Equity	Dow Jones Sustainability World ex Australia Total Return Index ex Tobacco AUD
237	Equity	Dow Jones World Industrials Index
238	Equity	Dow Jones Sustainability World Total Return Index in CHF
239	Equity	Dow Jones Sustainability World Total Return Index Ex AGTAF USD
240	Equity	Dow Jones Sustainability World Excluding Gambling Index
241	Equity	Dow Jones Sustainability World Total Return Index Ex Gambling Euro
242	Equity	Dow Jones World Excluding US Technology Index
243	Equity	Dow Jones World Excluding US Utilities Index
244	Equity	Dow Jones World Developed - Ex. U.S. Index
245	Equity	Dow Jones World Ex Asia Basic Materials Index
246	Equity	Dow Jones World - Ex Asia/Pacific Consumer Services Index
247	Equity	Dow Jones World - Ex Asia/Pacific Oil & Gas Index
248	Equity	Dow Jones World - Ex Asia/Pacific Financials Index
249	Equity	Dow Jones World Ex Asia Healthcare Index
250	Equity	Dow Jones World - Ex Asia/Pacific Industrials Index
251	Equity	Dow Jones World - Ex Asia/Pacific Consumer Goods Index
252	Equity	Dow Jones World Ex Asia Technology Index
253	Equity	Dow Jones World Ex Asia Telecommunications Index
254	Equity	Dow Jones World Ex Asia Utilities Index
255	Equity	EURO STOXX Index
256	Equity	Merrill Lynch Option Volatility Estimate MOVE Index
257	Equity	Merrill Lynch Option Volatility Estimate 3-Month
258	Equity	Merrill Lynch Swaption Option Volatility Estimate 3 Month

259	Equity	Merrill Lynch Swaption Option Volatility Estimate 6 Month
260	Equity	Morgan Stanley Cyclical Index
261	Corporate	1MLIBOR
262	Corporate	3MLIBOR
263	Corporate	6MLIBOR
264	Corporate	1YLIBOR
265	Corporate	Fed Funds
266	Corporate	Federal Reserve 30 Day A2 P2 Nonfinancial Commercial Paper Interest Rate
267	Corporate	Federal Reserve 30 Day A Commercial Paper
268	Corporate	Federal Reserve 30 Day AA Financial Commercial Paper Interest Rate
269	Corporate	Federal Reserve US Treasury Note Constant Maturity Not Averaged 3 Month
270	Corporate	Moody's Bond Indices Corporate AAA
271	Corporate	Moody's Bond Indices Corporate BAA
272	Corporate	US Generic Govt 10 Year Yield
273	Corporate	BlackRock Corporate Bond Fund (London)
274	Corporate	BlackRock Corporate Bond Fund (London)
275	Corporate	BlackRock Total Return Fund (Trade Reporting Facility LLC)
276	Corporate	BlackRock Global Funds - US Dollar High Yield Bond Fund (Luxembourg)
277	Corporate	BlackRock Global Funds - US Dollar High Yield Bond Fund (Luxembourg)
278	Corporate	BlackRock Global Funds - Global High Yield Bond Fund (Luxembourg)
279	Corporate	Merrill Lynch 10-year U.S. Treasury Futures Total Return
280	Corporate	Merrill Lynch 5-year U.S. Treasury Futures Excess Return
281	Corporate	Merrill Lynch 2-year U.S. Treasury Futures Total Return
282	Corporate	Merrill Lynch 30-year U.S. Treasury Futures Total Return
283	Corporate	Merrill Lynch 5-year U.S. Treasury Futures Total Return
284	Corporate	Merrill Lynch 10-year U.S. Treasury Futures Excess Return
285	Corporate	Merrill Lynch 2-year U.S. Treasury Futures Excess Return
286	Corporate	Merrill Lynch 30-year U.S. Treasury Futures Excess Return
287	Corporate	ICE LIBOR USD 1 Week
288	Corporate	Federal Reserve Commercial Paper Financial Discount Basis 1 Day
289	Corporate	Federal Reserve 7 Day AA Financial Commercial Paper Interest Rate
290	Corporate	Federal Reserve Commercial Paper Financial Yield Basis 1 Day
291	Corporate	Federal Reserve Commercial Paper Non-Financial Yield Basis 1 Day
292	Corporate	Federal Reserve 60 Day AA Financial Commercial Paper Interest Rate
293	Corporate	Federal Reserve Overnight AA Financial Commercial Paper Interest Rate
294	Corporate	Federal Reserve Commercial Paper Non-Financial Discount Basis 1 Day
295	Corporate	Federal Reserve 7 Day AA Nonfinancial Commercial Paper Interest Rate
296	Corporate	Federal Reserve 15 Day A2 P2 Nonfinancial Commercial Paper Interest Rate
297	Corporate	Federal Reserve 15 Day AA Nonfinancial Commercial Paper Interest Rate
298	Corporate	Federal Reserve 7 Day A2 P2 Nonfinancial Commercial Paper Interest Rate
299	Corporate	Federal Reserve Overnight A2 P2 Nonfinancial Commercial Paper Interest Rate
300	Corporate	Mtge Current Cpns Fnma 30 Year
301	Corporate	Fannie Mae Commitment Rates 30 Year Fixed Rate 30 Day

302	Corporate	Mtge Current Cpns FNMA 30 Year Spread
303	Corporate	Fannie Mae Commitment Rates 30 Year Fixed Rate 60 Day
304	Corporate	Fannie Mae Commitment Rates 15 Year Fixed Rate 30 Day
305	Corporate	Fannie Mae Commitment Rates 30 Year Fixed Rate 10 Day
306	Corporate	Fannie Mae Commitment Rates 20 Year Fixed Rate 30 Day
307	Corporate	Fannie Mae Commitment Rates 30 Year Fixed Rate 90 Day
308	Corporate	Fannie Mae Commitment Rates 10 Year Fixed Rate 30 Day
309	Corporate	DBIQ MBS TBA: FNMA: 30 Years FNCL
310	Corporate	Fannie Mae Commitment Rates 15 Year Biweekly Fixed Rate 30 Day
311	Corporate	Citigroup Mortgage 30 Year FNMA Sector Local Currency
312	Corporate	Fannie Mae Commitment Rates 30 Year Biweekly Fixed Rate 10 Day
313	Corporate	Fannie Mae Commitment Rates 30 Year Biweekly Fixed Rate 30 Day
314	FX	USA cross rate
315	FX	suiza cross rate
316	FX	Fwd USD
317	FX	Canada
318	FX	Japan
319	FX	Costa Rica
320	FX	Dominican Republic
321	FX	Iceland
322	FX	Israel
323	FX	Kazakhstan
324	FX	United Arab Emirates
325	FX	Australia
326	FX	Switzerland
327	FX	Chile
328	FX	Euro
329	FX	Great Britain
330	FX	Guatemala
331	FX	New Zealand
332	FX	USA
333	FX	Peru
334	FX	Paraguay
335	FX	Romania
336	FX	Saudi Arabia
337	FX	Slovakia
338	FX	Thailand
339	FX	Turkey
340	FX	South Africa
341	Fixed Income	3MTB
342	Fixed Income	6MTB
343	Fixed Income	6 M Treasury Bill

344	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 1 Year
345	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 2 Year
346	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 3 Year
347	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 5 Year
348	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 10 Year
349	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 20 Year
350	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity 2 Year
351	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity 3 Year
352	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity 3 Month
353	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity 7 Year
354	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity 20 Year
355	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity 6 Month
356	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 7 Year
357	Fixed Income	US Treasury Yield Curve Rate T Note Constant Maturity Composite Over 10 Year
358	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged 6 Month
359	Fixed Income	Federal Reserve US Treasury Note Constant Maturity Not Averaged Composite 30+ Ye
360	Fixed Income	US Generic Govt 30 Year Yield
361	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 10 Year
362	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 10 Year
363	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 3 Month
364	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 1 Year
365	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield
366	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 1 Year
367	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 30 Year
368	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 6 Month
369	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 5 Year
370	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 7 Year
371	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 30 Year
372	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 5 Year
373	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 2 Year
374	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 2 Year
375	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 3 Year
376	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 9 Year
377	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 20 Year
378	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 6 Month

379	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 20 Year
380	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 4 Year
381	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 6 Year
382	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 7 Year
383	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 4 Year
384	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 3 Year
385	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 15 Year
386	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 8 Year
387	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 6 Year
388	Fixed Income	USD Treasury Actives (IYC 25) Zero Coupon Yield 9 Year
389	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 8 Year
390	Fixed Income	USD US Treasury Bonds/Notes (FMC 82) Zero Coupon Yield 15 Year
391	Fixed Income	Cetes 28
392	Fixed Income	Cetes 91

Macro data

	Description
1	Private consumption
2	Government consumption
3	Imports
4	Net investment
5	Non-oil exports
6	Oil exports
7	CPI: food, drinks and tobacco
8	CPI: housing
9	CPI: clothing, shoes and accesories
10	CPI: furniture and domestic appliances
11	CPI: health
12	CPI: transport
13	CPI: education and recreation
14	CPI: other services
15	Industrial activity
16	M1
17	M2
18	IGAE
19	USA GDP
20	USA CPI

Appendix B – Matlab Code

```
clear all
format long

%Initial data processing
[A, fechas] = xlsread('test1.xlsx');
[X,fechasgdp]=xlsread('mexgdp.xlsx');
fechasgdp=fechasgdp(1:83,1);
[o,p]=size(A);

for i=1:p
    h(i,:)=[i adftest(A(:,i),'lags',12)];
end

for i=1:p
    if h(i,2)==0
        difs(:,i)=(log(A(2:end,i))-log(A(1:end-1,i)))*100;
    else
        difs(:,i)=A(2:end,i);
    end
end

difsgdp=(log(X(2:end))-log(X(1:end-1)))*100;
save('commo.mat', 'fechas', 'difs','fechasgdp','difsgdp');

%%%%%%%%%%%%%%%%%%
%Factor Estimation Totals
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[o,p]=size(difs);
difsnorm=zeros(3818,p);
for i=1:p
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm);
[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:5))

Fact=difsnorm*lambda;
for i=0:4
    F(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end
save('FTotales.mat','F','percexp')

%%%%%%%%%%%%%%%%%%
%Factor Estimation Commodities
```

```
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[o,p]=size(difs);
difsnorm=zeros(3818,p);
for i=1:p
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm(:,1:166));
[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:5))

Fact=difsnorm(:,1:166)*lambda;
for i=0:4
    F(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end
save('FCom.mat','F','percexp')

%%%%%%%%%%%%%%%%%%
%Factor Estimation Equities
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[o,p]=size(difs);
difsnorm=zeros(3818,p);
for i=1:p
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm(:,167:260));
[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:5))
[w,q]=size(eigvals);
Fact=difsnorm(:,167:260)*lambda;
for i=0:4
    F(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end
save('FEqu.mat','F','percexp')

%%%%%%%%%%%%%%%%%%
%Factor Estimation Corporate
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[o,p]=size(difs);
difsnorm=zeros(3818,p);
for i=1:p
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm(:,261:313));
[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:5))
[w,q]=size(eigvals);
Fact=difsnorm(:,261:313)*lambda;
for i=0:4
    F(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end
```

```
save('FCor.mat','F','percexp')
%%%%%%%%%%%%%%%%%%
%Factor Estimation Foreign Exchange
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[o,p]=size(difs);
difsnorm=zeros(3818,p);
for i=1:p
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm(:,314:340));
[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:5))
[w,q]=size(eigvals);
Fact=difsnorm(:,314:340)*lambda;
for i=0:4
    F(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end
save('FFX.mat','F','percexp')

%%%%%%%%%%%%%%%%%%
%Factor Estimation Fixed Income
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[o,p]=size(difs);
difsnorm=zeros(3818,p);
for i=1:p
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm(:,341:392));
[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:5))
[w,q]=size(eigvals);
Fact=difsnorm(:,341:392)*lambda;
for i=0:4
    F(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end
save('FFix.mat','F','percexp')

%%%%%%%%%%%%%%%%%%
%Factor Estimation Macro variables
%%%%%%%%%%%%%%%%%%
clear all
load('commo.mat')
[q,r]=size(difs);
difsnorm=zeros(q,r);
for i=1:r
    difsnorm(:,i)=(difs(:,i)-mean(difs(:,i)))/std(difs(:,i));
end
N = cov(difsnorm);


[lambda, eigvals, percexp] = pcacov(N);
explicado=sum(percexp(1:3))
```

```
Fact=difsnorm*lambda;
for i=0:2
    Fmacro(:,i+1)=Fact(:,i+1)/eigvals(i+1);
end

save('Fmacro.mat','Fmacro','percexp')


%%%%%%%%%%%%%%%%%%
%Model estimations
%%%%%%%%%%%%%%%%%%

clear all
load('commo.mat')
load('FTotales.mat')
load ('Fmacro.mat')
% load('FCom.mat')
% load('FEqu.mat')
% load('FCor.mat')
% load('FFX.mat')
% load('FFix.mat')

% reng=1;
% for y=1:10
%     for x=5:20
Xlag = 10;
Ylag = 1;
Horizon = '1q';
EstStart = '1999-06-30';
EstEnd = '2009-12-31';
Method = 'Recursive';
dummy=[1 0 0;0 1 0; 0 0 1; 0 0 0];
dum=repmat(dummy,21,1);
ExoReg=dum(1:end-1,:);
ExoRegDate=fechasgdp;

        [OutputForecast2,OutputEstimate2]=
MIDAS_ADL(difsgdp,fechasgdp,F(:,1),fechas,...

'Xlag',Xlag,'Ylag',Ylag,'Horizon',Horizon,'EstStart',EstStart,'EstEnd',..
.

EstEnd,'ExoReg',ExoReg(:,1:3),'ExoRegDate',ExoRegDate,'Polynomial','betaN
N','Method',Method,'Display','full');


        Forecast=OutputForecast2.Yf;
        pronostico1(i)=Forecast(4);

         [BZNZForF2,BZNZEstF2]=
MIDAS_ADL(difsgdp,fechasgdp,F(:,2),fechas,...

'Xlag',17,'Ylag',Ylag,'Horizon',Horizon,'EstStart',EstStart,'EstEnd',...
```

```
EstEnd,'ExoReg',ExoReg(:,1:3),'ExoRegDate',ExoRegDate,'Polynomial','betaN
N','Method',Method,'Display','full');

        Forecast=BZNZForF2.Yf;
        pronostico2(i)=Forecast(4);


         [BZNZForF3,BZNZEstF3]=
MIDAS_ADL(difsgdp,fechasgdp,F(:,3),fechas,...

'Xlag',19,'Ylag',Ylag,'Horizon',Horizon,'EstStart',EstStart,'EstEnd',...

EstEnd,'ExoReg',ExoReg(:,1:3),'ExoRegDate',ExoRegDate,'Polynomial','betaN
N','Method',Method,'Display','full');

        Forecast=BZNZForF3.Yf;
        pronostico3(i)=Forecast(4);


         [BZNZForF4,BZNZEstF4]=
MIDAS_ADL(difsgdp,fechasgdp,F(:,4),fechas,...

'Xlag',18,'Ylag',Ylag,'Horizon',Horizon,'EstStart',EstStart,'EstEnd',...

EstEnd,'ExoReg',ExoReg(:,1:3),'ExoRegDate',ExoRegDate,'Polynomial','betaN
N','Method',Method,'Display','full');

        Forecast=BZNZForF4.Yf;
        pronostico4(i)=Forecast(4);


         [BZNZForF5,BZNZEstF5]=
MIDAS_ADL(difsgdp,fechasgdp,F(:,5),fechas,...

'Xlag',18,'Ylag',Ylag,'Horizon',Horizon,'EstStart',EstStart,'EstEnd',...

EstEnd,'ExoReg',ExoReg(:,1:3),'ExoRegDate',ExoRegDate,'Polynomial','betaN
N','Method',Method,'Display','full');

        Forecast=BZNZForF5.Yf;
        pronostico5(i)=Forecast(4);


% RES(reng,:)=[y 1];
% RES(reng+1,:)=[x OutputForecast2.RMSE];
% reng=reng+2;

% RES(reng,:)=[y OutputForecast1.aic OutputForecast2.aic
OutputForecast3.aic OutputForecast4.aic OutputForecast5.aic
OutputForecast6.aic 1 2 3 4 5 6];
% RES(reng+1,:)=[x OutputForecast1.bic OutputForecast2.bic
OutputForecast3.bic OutputForecast4.bic OutputForecast5.bic
OutputForecast6.bic OutputForecast1.RMSE OutputForecast2.RMSE
```

```
OutputForecast3.RMSE OutputForecast4.RMSE OutputForecast5.RMSE
OutputForecast6.RMSE];
% reng=reng+2;

% RES(reng,:)=[y OutputForecast1.aic OutputForecast2.aic
OutputForecast3.aic OutputForecast4.aic OutputForecast5.aic
OutputForecast6.aic];
% RES(reng+1,:)=[x OutputForecast1.bic OutputForecast2.bic
OutputForecast3.bic OutputForecast4.bic OutputForecast5.bic
OutputForecast6.bic];
% reng=reng+2;
    end
end

YMSFEC1 =
ForecastCombine(OutputForecast2,BZNZForF2,BZNZForF3,BZNZForF4,BZNZForF5);
RMSEcomb1=sqrt(mean(power(difsgdp(end-14:end)-YMSFEC1,2)))
```

*Note: Diebold and Mariano p-values where estimated using STATA and are not shown here.

REFERENCES

- Andreou, Elena , Ghysels, Eric & Kourtellos, Andros (2013): Should Macroeconomic Forecasters Use Daily Financial Data and How?, Journal of Business & Economic Statistics, 31:2, 240-251
- Arnesto, Michelle T., Engerman, Kristie M. & Owyang, Michael T. (2010): Forecasting with Mixed Frequencies, Federal Reserve Bank of St. Louis Review, 92(6), 521-536
- Bai, Jennie, Gysels, Eric & Wright, Jonathan H. (2013): State Space Models and MIDAS Regressions, Econometric Reviews, Volume 32, Issue 7, 779-813
- Clements, Michael P. & Galvão, Ana Beatriz (2008): Macroeconomic Forecasting With Mixed-Frequency Data, Journal of Business & Economic Statistics, 26:4, 546-554
- Çelik, Sibel & Ergin, Hüseyin (2013): Volatility forecasting using high frequency data: Evidence from stock markets, Economic Modelling, 36, 176-190
- Diebold, F. X., & Mariano, R. S. (1995): Comparing Predictive Accuracy, Journal of Business and Economic Statistics, 13, 253–265.
- Dollar, David & Kraay, Aart (2004): Trade, Growth and Poverty, Economic Journal, 114(493), F22-F49
- Ghysels, Eric, Santa-Clara, Pedro & Valkanov, Rossen (2004): The MIDAS Touch: Mixed Data Sampling Regression Models, Working Paper
- Ghysels, Eric, Sinko, Arthur, & Valkanov, Rossen. (2006): MIDAS Regressions: Further Results and New Directions, Econometric Reviews, 26, 53–90. [241]
- Kuzin, Vladimir, Marcellino, Massimiliano & Schumacher, Christian (2011): MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area, International Journal of Forecasting 27 (2011) 529–542
- Kuzin, Vladimir, Marcellino, Massimiliano & Schumacher, Christian (2012): Pooling Versus Model Selection for Nowcasting GDP with Many Predictors, Journal of Applied Econometrics, Volume 28, Issue 3, pages 392–411, April/May 2013
- Marcellino, Massimiliano & Schumacher, Christian (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
- Ibarra, Raul (2012): Do disaggregated CPI data improve the accuracy of inflation forecasts?, Economic Modelling, 29, 4, 1305-1313
- Rodriguez, Francisco & Rodrik, Dani (2001): Trade Policy and economic growth: a skeptic's guide to the cross- national evidence, Macroeconomics Annal 2000, NBER, Cambridge, MA.
- Stock, James H. & Watson, Mark W. (2002): Forecasting Using Principal Components From a Large Number of Predictors, Journal of the American Statistical Association,97,1167–1179
- Timmermann, A. (2006): Forecast Combinations, in Handbook of Economic Forecasting (Vol. 1), eds. G. Elliott, C. Granger, and A. Timmermann, Amsterdam: Elsevier-North Holland, pp. 136–196. [245]
- Wacziarg, Romain & Horn-Welch, Karen (2003): Trade Liberalization and growth: new evidence, NBER, working paper.