The State and the Future of Cyprus Macroeconomic Forecasting

Elena Andreou and Andros Kourtellos

Department of Economics, University of Cyprus

Abstract

In this paper we discuss some state-of-the-art methods in the macroeconomic forecasting literature that can be adapted for macroeconomic forecasting in Cyprus emphasizing the Mixed Data Sampling (MIDAS) models. MIDAS models are reduced form parsimonious regression framework, which do not require modeling the dynamics of the individual high frequency predictor series. There are many advantages in macroeconomic forecasting from using mixed data frequency models. Take for example the situation that a high frequency financial variable (e.g., daily stock market returns), possibly together with other lower frequency macroeconomic indicators, are used to predict a low frequency macro variable (e.g., quarterly gdp growth). The choice is between using a midas model, which will use all the information in the sample by using the data at the higher frequencies or aggregate the data first (typically by taking an average) and then specify a predictive model at the lower frequency. Not using the readily available higher frequency series has two important implications: (1) one loses information through temporal aggregation which can lead to biased forecasts and (2) one foregoes the possibility of providing real-time daily, weekly or monthly updates of forecasts. The topic of mixing different sampling frequencies also emerges even when time series are available at the same frequency, but one is interested in multi-period forecasting. Multi-period forecasts can also be constructed using a mixed-data sampling approach. For example, a MIDAS model can use past quarterly data to produce directly multi-period forecasts. The MIDAS approach can be viewed as a middle ground between the direct and the iterated approaches. Namely, one preserves the past high frequency data, to directly produce multi-period forecasts. In addition, we review multivariate models and especially Vector Autoregressive (VAR) models that deal with mixed frequency variables in forecasting key macroeconomic variables (MIDAS-VAR). This is compared with the standard state-space approach of structural multivariate models which involves the Kalman filter. The state-space approach differs from the MIDAS-VAR approach in that the latter does not rely on latent processes/shocks

* Corresponding author, Address: Department of Economics, University of Cyprus, P.O. Box 537, CY 1678 Nicosia, Cyprus, e-mail: andros@ucy.ac.cy. Andros Kourtellos acknowledges support of the European Regional Development Fund of the EU under the Research Promotion Foundation grant ΑΝΘΡΩΠΙΣΤΙΚΕΣ/ΟΙΚΟΝ/0311(BIE)/03
representations and is formulated exclusively in terms of observable data. As a result it tends to be quite involved, as one must explicitly specify a linear dynamic model for all the series involved: the high-frequency data series, the latent high-frequency series treated as missing and the low-frequency observed processes. The system of equations in a structural Kalman filter approach therefore typically requires a lot of parameters, namely for the measurement equation, the state dynamics and their error processes. Therefore, state space models might be prone to specification errors.

Keywords: Macroeconomic forecasting, Mixed frequency, MIDAS regressions, State space models.

1. Introduction

Nowadays, a macroeconomic forecaster faces several challenges. First, there is a huge and increasing number of macroeconomic and especially financial time series that contain useful information about the future state of the economy. This raises the question how to summarize the information or extract the common components from the vast cross-section of macroeconomic and financial series. Second, these data are often observed at mixed frequencies and so the question arises how to match daily (or an arbitrary higher frequency such as potentially intra-daily) financial data with quarterly (or monthly) macroeconomic indicators when one wants to predict short as well as relatively long horizons, like one year ahead. Third, key macroeconomic data on the present state of the economy are available with substantial delay with unsynchronized data-release dates and the time of the last available observation differs from series to series, which results in a so-called ragged or jagged edge—“jagged edge” across the most recent months. So it is important that the forecasting method is capable to track the real-time flow of the type of information monitored by central banks and other agencies in order to provide accurate current-quarter forecasts (nowcasts) and short-term forecasts. Fourth, in macroeconomic forecasting we often need to consider interdependencies among macro variables such as GDP growth, inflation, monetary indicators and/or interest rates, which requires the use of multivariate forecasting methods. The problem is that even small dynamic systems face parsimony problems especially given that the real-time data flow is inherently high dimensional. So how does one deal with the problem of parameter proliferation but at the same time is able to capture the salient features of the data and interdependencies among the key macro variables?

Currently, the Cyprus macroeconomic forecasting has successfully dealt with first challenge; see Andreou, Kourtellos, and Pashourtidou (2012). In
the first paper we analyze how factor models can be useful for extracting the common component from a large cross section of macroeconomic and financial time series in Cyprus as well as other countries which are useful leading indicators of the state of the Cyprus economy. Moreover we discuss alternative forecast combination methods to deal with the alternative forecasts of different single equation models. In this paper we describe state-of-the-art methods in the macroeconomic forecasting literature than can be adapted for macroeconomic forecasting in Cyprus to address the remaining three challenges.

We consider two complementary approaches to deal with the above problems. The first approach is the Mixed Data Sampling (MIDAS) approach; see for example Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006) and Andreou, Ghysels, and Kourtellos (2010). It is a reduced form parsimonious regression framework, which does not require to model the dynamics of each and every daily predictor series. There are many advantages in macroeconomic forecasting from using mixed data frequency models. Take for example the situation involving a combination of past quarterly series and the choice between using past quarterly financial series - or instead using those same series sampled monthly or weekly or daily. Not using the readily available higher frequency (e.g. daily) series has two important implications: (1) one looses information through temporal aggregation which can lead to biased forecasts and (2) one foregoes the possibility of providing real-time daily, weekly or monthly updates of forecasts (see e.g., Andreou, Ghysels, and Kourtellos (2010, 2013)).

The idea of using models where the variables are of mixed data sampling frequencies was first introduced in Ghysels, Santa-Clara, and Valkanov (2005) and since then there is a large and growing literature. Empirical applications involve regression and quantile regression models for forecasting macroeconomic variables as well as volatility models for understanding and forecasting financial risk. A number of recent papers have documented the advantages of using such MIDAS regressions in terms of improving quarterly macro forecasts with monthly and daily data. For instance, Bai, Ghysels, and Wright (2010), Kuzin, Marcellino, and Schumacher (2011), Armesto, Hernandez-Murillo, Owyang, and Piger (2009), Clements and Galvao (2009), Clements and Galvao (2008), Galvao (2006), Schumacher and Breitung (2008), Tay (2007), use monthly data to improve quarterly forecasts. Similarly, quarterly and monthly macroeconomic predictions are improved by daily financial series, see e.g.

1See Armesto, Engemann, and Owyang (2010) for a user-friendly introduction to MIDAS regressions.

The topic of mixing different sampling frequencies also emerges even when time series are available at the same frequency, but one is interested in multi-period forecasting. Take the example of an annual forecast with quarterly data. The first approach is to estimate a model with past annual data, and hence collapse the original multi-period setting into a single step forecast. The second approach is to estimate a quarterly forecasting model and then iterate forward the forecasts to a multi-period annual prediction. The forecasting literature refers to the first approach as “direct” and the second as “iterated”; see for example Massimiliano, Stock, and Watson (2006). Multi-period forecasts can also be constructed using a mixed-data sampling approach. A MIDAS model can use past quarterly data to produce directly multi-period forecasts. The MIDAS approach can be viewed as a middle ground between the direct and the iterated approaches. Namely, one preserves the past high frequency data, to directly produce multi-period forecasts.

An alternative approach is a structural multivariate dynamic approach, which has state space representation. It consists of a system with two types of equations, the measurement equations which link observed series to a latent state process, and the state equations which describe the state process dynamics. The state-space approach allows one to capture the real-time flow of information from the perspective of market participants and policy makers much better than the reduced-form MIDAS approach. That is, it pays particular attention to the key features of macroeconomic data releases in real time including monitoring many data, forming expectations about them, and revising the assessment on the state of the economy whenever realizations diverge sizably from those expectations. To do so the Kalman filter generates projections for all the variables considered and therefore allows to compute for each data release “news”, that is, a model based surprise. Then the nowcast revision can be expressed as a weighted average of these news and the data releases with model based weights. An advantage of this approach is that one can evaluate the role of different categories of data (e.g., surveys, financial, macro, and labor market) in obtaining “news” about economic activity.

A number of recent papers documented the gains of real-time forecast updating, sometimes also nowcasting when it applies to current quarter assessments; see for Giannone, Reichlin, and Small (2008a), Doz, Giannone,
and Reichlin (2011, 2012), and Aruoba, Diebold, and Scotti (2009). The setup treats the low-frequency data as “missing data” and the Kalman filter is a convenient computational device to extract the missing data. The steady state Kalman filter gain, however, yields a linear projection rule to (1) extract the current latent state, and (2) predict future observations as well as states. The Kalman filter can then be used to predict low frequency macro series, using both past high and low frequency observations.

However, it has some serious drawbacks, especially when it is applied beyond nowcasting. The state-space approach differs from the MIDAS-VAR approach in that the latter does not rely on latent processes/shocks representations and is formulated exclusively in terms of observable data. As a result it tends to be quite involved, as one must explicitly specify a linear dynamic model for all the series involved: the high-frequency data series, the latent high-frequency series treated as missing and the low-frequency observed processes. The system of equations therefore typically requires a lot of parameters, namely for the measurement equation, the state dynamics and their error processes. State space models are therefore prone to specification errors.

The remainder of the paper is structured as follows. In section 2 we cover MIDAS regressions. Section 3 we cover multivariate MIDAS models. Section 4 discusses nowcasting in the context of both MIDAS regressions and state space models. Section 5 concludes.

2. MIDAS Regressions

Suppose we are interested in forecasting quarterly GDP growth rate, $Y_{t+1}^Q$, using daily stock returns, $X_{N_{D-j,t}}^D$, in the $j^{th}$ day counting backwards in quarter $t$. Hence, the last day of quarter $t$ corresponds to $j = 0$ and is therefore $X_{N_{D-0,t}}^D$. The conventional approach, in its simplest form, aggregates the data at the quarterly frequency by computing simple averages to obtain $X_j^Q = (X_{N_{D-j,t}}^D + X_{N_{D-1-j,t}}^D + ... + X_{1-j,t}^D)/N_D$ and then estimates a simple Distributed Lag (DL) model

$$Y_{t+1}^Q = \alpha + \beta X_j^Q + u_{t+1}, \quad (2.1)$$

---

Footnote:

2For notational brevity, we will be dealing with one-step ahead forecasts. All the models and methods we will be presenting can be easily extended to multi-step forecasting.
where $\alpha$ and $\beta$ are unknown parameters and $u_{t+1}$ is an error term. The implicit assumption in traditional models such as (2.1) is that temporal aggregation is based on an equal weighting scheme of the high frequency data.

An alternative naive approach to deal with mixed frequencies in a regression model would be to estimate an unrestricted model with all the high frequency lags (e.g. the daily lags referring to the number of trading days per quarter). This is an unappealing approach because of the parameter proliferation problem: when $N_D = 66$, we have to estimate not only 66 slope parameters, denoted by $\beta_j$, but also $\mu$ and $\alpha$, hence a total of 68 slope coefficients. However, this approach can be useful when the number of lags of the high frequency regressor is small. Indeed, the Unrestricted-MIDAS (U-MIDAS) approach does not restrict the number of lags/parameters to be estimated and is relevant for a small number of lags e.g. when the dependent variable is quarterly and the high frequency predictor is monthly. The advantages of U-MIDAS models are discussed in Foroni, Massimiliano, and Schumacher (2012).

In general the MIDAS regression models use a parsimonious and data-driven aggregation scheme based on a low dimensional high frequency lag polynomial, $W(L^{N_D}; \theta)$ such that $W(L^{N_D}; \theta^D)X^D_t = \sum_{j=0}^{N_D-1} w_j(\theta^D)X^D_{t-j}$. There are various alternatives for the polynomial specification. Two flexible specifications that parameterize the weights into a two parameter vector include the two parameter exponential Almon lag and the Beta lag. Ghysels, Sinko, and Valkanov (2006) provide a discussion on the two specifications as well as for step-functions. This approach yields a distributed lag model as a linear projection of high frequency data $X^D_t$ onto $Y^Q_t$ using a MIDAS filter, $DL-MIDAS(q^D_X)$

$$Y^Q_{i+1} = \mu + \beta \sum_{j=0}^{N_D-1} w_j(\theta^D)X^D_{N_D-j,i} + u_{i+1}, \quad (2.2)$$

where for simplicity we take lags in blocks of quarterly sets of daily data, $q^D_X$ but the daily lags can be extended beyond the last day of quarter $t$. Note that equation (2.2) nests the simple DL model in equation (2.1) under flat-weights. To see this, note that, if $\theta_1 = \theta_2 = 0$ then the exponential almon lag polynomial function yields equal.flat aggregation weights. We

---

3Typically we have about 66 observations for many daily financial data over a quarter since each month has 22 trading days.
assume that the weights are standardised and sum to one that allows the identification of the slope coefficient $\beta$ in the MIDAS regression model. The parameters $(\mu, \beta, \theta^D)$ are estimated by Nonlinear Least Squares (NLS).

Within the MIDAS regression setup, Andreou, Ghysels, and Kourtellos (2010) show that the traditional temporal aggregation approach, which imposes flat weights, $w_i = 1/N_D$, yields a nonlinear omitted variable term in the regression model (2.1). The nonlinearity of the omitted term is due to the nonlinear weighting schemes of MIDAS regression models. This implies that for a general, non-flat weighting scheme the conventional temporal aggregation approach may result in an omitted variable bias if the omitted term is correlated with the traditional flatly aggregated term. Hence the LS estimation of equation (2.1) will generally give rise to a bias, which depends on the type of the high frequency process and on the shape of the weighting scheme $W(L^{ND}; \theta)$. For instance, declining weights imply an omitted variable that exhibits memory decay or mean reversion, which will be associated with higher bias than an omitted variable with a near-flat weighting scheme. In addition, Andreou, Ghysels, and Kourtellos (2010) show that when the omitted term is correlated with the equally weighted aggregated term in (2.2), then the Asymptotic Mean Squared Error (AMSE) of the LS estimator of $\beta$, is relatively larger than the AMSE of the NLS estimator of $\beta$ in (2.2). In this case, one can easily show that the DL-MIDAS model in equation (2.2) based on NLS yields more accurate forecasts than forecasts based on LS estimation of the simple DL model in equation (2.1) assuming flat weights.

The above discussion applies to other related forecasting models such the Augmented Distributed Lag (ADL) type models. When $Y_{t+1}$ is serially correlated, as it is typically the case for time series variables, the simple model in equation (2.1) is extended to a dynamic linear regression or autoregressive distributed lag (ADL) model. Again the conventional approach, in its simplest form, aggregates the high frequency data at the low frequency by computing simple averages and estimates a simple linear regression of $Y_{t+1}$ on $X_t$. In this case the traditional ADL model can be extended to the $ADL-MIDAS(p_y^Q, q_x^Q)$:

$$Y_{t+1} = \mu + \sum_{j=0}^{p_y^Q-1} \alpha_j Y_{t-j} + \beta \sum_{i=0}^{N_D-1} w_i (\theta^D) X_{N D-j}^D + u_{t+1}$$

(2.3)
2.1 Factors and other regressors in ADL-MIDAS models

A large body of recent work has developed factor model techniques that are tailored to exploit a large cross-sectional dimension; see for instance, Bai and Ng (2002), Bai (2003), Forni, Hallin, Lippi, and Reichlin (2000), Forni, Hallin, Lippi, and Reichlin (2005), Stock and Watson (1989), Stock and Watson (2003), among many others. These factors, which are usually estimated at quarterly frequency using a large cross-section of time-series are used as predictors in ADL models. Following this literature Andreou, Ghysels, and Kourtellos (2013) investigate whether one can improve quarterly factor model forecasts by augmenting such models with daily financial variables and in particular daily financial factors. Such factors (at either low or high frequency) can be obtained by the Factor Model (DFM) (static or dynamic) extracted from a panel of $X_t = (X_{1t},...,X_{nt})'$. Let $F_t$ be the $r$-vector of factors, $\Lambda$, the $n \times r$ matrix of factor loadings. Note that factor models can allow for the possibility that the factor loadings change over time (compared to the standard DFMs) which may address potential instabilities during our sample period. The extracted common factors could be robust to instabilities in individual time series, if such instability is small and sufficiently dissimilar among individual variables; see Stock and Watson (2002) for formal conditions. Following the above assumptions the time-varying DFM can be estimated using principal components, which delivers consistent estimates of the common factors if $N \to \infty$ and $T \to \infty$. The number of factors can be chosen by the information criteria proposed; see for example Bai and Ng (2002).

These factors are then employed to augment the aforementioned MIDAS regression models. For instance, in the case of quarterly factors equation (2.3) generalizes to the FADL-MIDAS model given by

$$Y_{t+1}^Q = \mu + \sum_{k=0}^{p_Q-1} \alpha_k Y_{t-k}^Q + \sum_{k=0}^{p_E-1} \beta_k F_{t-k}^Q + \gamma \sum_{i=0}^{N_D-1} w_j (\theta_X^D) X_{n_{D,-j,t}}^D + u_{t+1}$$

(2.4)

Additionally, Andreou, Ghysels, and Kourtellos (2013) construct daily factors, denoted by $F_t^D$, which pool information from a large cross-section of daily financial data. This approach allows us to specify ADL-MIDAS models with both quarterly and daily factors that incorporate information across different frequencies while at the same time retain parsimony. For example, consider the FADL-MIDAS model in equation (2.4) using the

---

4Although the parametric AR assumption for $F_t$ and $e_{it}$ is not needed to estimate the factors, such assumptions can be useful when discussing forecasts using factors.
daily factor as the daily predictor, $X_t^D = F_t^D$. Using a similar approach Clements and Galvao (2009) use leading indicators as predictors for quarterly macroeconomic variables, which is estimated using DL-MIDAS models with AR errors.

### 2.2 Multivariate MIDAS models

Sims (1980) introduced the VAR model to study the dynamic relationship between macroeconomic time series and in particular key economic indicators like GDP growth and inflation. In many cases these low frequency macro variables interact with other economic indicators such as interest rates, credit spreads e.t.c. which are observed at a higher frequency. Therefore in specifying the class of mixed frequency VAR models we can characterize the timing of information releases for a mixture of sampling frequencies as well as the real-time updating of predictions caused by the flow of high frequency information (see Appendix A).

In the traditional forecasting setup we can predict the high and low frequency data given previous quarter’s high and low frequency observations. In the mixed frequency VAR consider the high frequency data sampled $m$ times more often than the low frequency series. For example, in the quarterly/monthly MIDAS-VAR we stack for the high frequency variable the months of January, February and March together with the first quarter of the low frequency variable. Similarly, we stack the April, May, June monthly data with the second quarter of the low frequency variable, e.t.c.. The last equation of the MIDAS-VAR model is related to the single equation MIDAS regression model in the special case of the single equation ADL-MIDAS model in equation (2.3). In addition to the remaining dynamics the MIDAS-VAR also contains the impact of the low frequency shock onto both future high and low frequency series. The same applies to the high frequency shocks on future high and low frequency series.

The real time prediction of the MIDAS-VAR allows the leads of the high frequency data to provide real time updating. The simplicity of the restricted or unrestricted MIDAS-VAR model for real-time prediction is particularly appealing as it is based on estimating the mixed frequency VAR and subsequently computing the Choleski factorization of the errors and taking the $m-1$ lower triangular truncations of the original factorization. Various estimation procedures for mixed frequency VAR models can be pursued both classical and Bayesian that solves the parameter proliferation problem of unrestricted MIDAS-VAR models.
3. Nowcasting

As we noted in the introduction key statistics such as the Gross Domestic Product (GDP) are available with a significant delay. In Cyprus the first official estimate (flash) is published about 45 days after the end of the reference quarter. Can one exploit surveys, macro, and financial data series, which are published earlier and possibly at higher frequencies than the GDP or other variables of interest in order to obtain a current estimate before the official figure becomes available? The answer is yes using nowcasting.

Nowcasting refers to within-period updates of forecasts. An example would be weekly updates of current quarter GDP forecasts. MIDAS with leads can be viewed as - say again weekly updates - of not only current quarter GDP forecasts, but any future horizon GDP forecast (i.e., over several future quarters). Two alternative approaches for nowcasting have emerged in the literature.

The first approach employs a state space model developed by Giannone, Reichlin, and Small (2008b) originally for monthly data and recently extended by Banbura, Giannone, Modugno, and Reichlin (2013) to also allow daily data. This approach can be viewed as a structural approach as it involves a set of measurement and transition equations (see Appendix B). A measurement equation describes the relationship between observed variables and latent common factors while a transition equation describes the dynamics of the latent factors, typically in the form of a VAR process. The estimates of the factors are obtained by Kalman filter and maximum likelihood based on the EM algorithm using initial parameter values based on principal components. The principal components are extracted by filling-in the missing observations using splines and then the data are filtered to reflect different aggregation intervals. Nowcasts are then obtained by regressing the low frequency variable (e.g., GDP growth) on temporally aggregated factor estimates. Notably, Doz, Giannone, and Reichlin (2011, 2012) show consistency of the two-step and maximum likelihood estimates and the asymptotic properties are analyzed under different sources of misspecification: omitted serial and cross-sectional correlation of the idiosyncratic components, and non-normality. Interestingly, they found that the effects of misspecification on the estimation of the common factors is negligible for long time-series and large cross-sections.

---

5The Kalman filter is a set of recursive equations that determine the optimal estimates of the state parameter vector conditional on an information set.
The second approach is a reduced form that employs approach MIDAS regressions with leads to update current quarter forecasts. Suppose we are one or two months into the current quarter. Then, MIDAS models with leads can naturally incorporate real-time information, which is mainly available on financial variables. Our objective is to forecast quarterly economic activity and in practice we often have a monthly release of macroeconomic data within the quarter and the equivalent of at least 44 trading days of financial data observed with no measurement error. This means that if we stand on the first day of the last month of the quarter and wish to make a forecast for the current quarter we could use and around 44 leads of daily data for financial markets that trade on weekdays.\(^6\)

Bai, Ghysels, and Wright (2010) and Kuzin, Marcellino, and Schumacher (2011) discuss in detail the connections between the Kalman filter and MIDAS regressions. It is the purpose of this subsection to summarize their findings. The first important observation is that a MIDAS regression can be viewed as a reduced form representation of the linear projection that emerges from a state space model approach - where by reduced form we mean that the MIDAS regression does not require the specification of a full state space system of equations. More importantly, the Kalman filter requires to specify a complete system of equations, which we kept to an absolute minimum representation in the above motivating example. In some cases the MIDAS regression is an exact representation of the Kalman filter, in other cases it involves approximation errors that are typically small.\(^7\) The Kalman filter, while clearly optimal as far as linear projections goes, has two main disadvantages First, it is more prone to specification errors as a full system and latent factors is required and second, it requires a lot more parameters to achieve the same goal.

### 4. Conclusion

In this paper we review forecasting models that involve data sampled at different frequencies. The main advantage of these models are the real-time forecast updating, nowcasting. We discuss two approaches of univariate and multivariate models with mixed frequency. The first approach is the Mixed Data Sampling (MIDAS) approach which is a

---

\(^6\) MIDAS with leads differs from the MIDAS regressions involving “leading indicator” series, as in Clements and Galvao (2009). The latter use MIDAS regressions without leads appearing in the MIDAS polynomial, but with (monthly) leading indicator series aligned with quarterly GDP growth data.

\(^7\) Bai, Ghysels, and Wright (2010) discusses both the cases where the mapping is exact and the approximation errors in cases where the MIDAS does not coincide with the Kalman filter.
reduced form parsimonious regression framework, that does not require to model the dynamics of each and every high frequency predictor. The second method is a structural multivariate dynamic approach, which has state space representation. It consists of a system with two types of equations, the measurement equations which link observed series to a latent state process, and the state equations which describe the state process dynamics. The state-space approach allows one to capture the real-time flow of information from the perspective of market participants and policy makers.

One problem that we did not consider is the effect of data revisions. As noted by Croushore (2011) and others this could affect the results of forecast evaluation and comparison. The main difficulty however is that currently real-time data vintages for many series are not readily available in Cyprus but we hope that in the future such data will be available by the Cyprus statistical agency.

Appendix

A. Multivariate MIDAS models

In a mixed frequency setup consider the high frequency data samples \( m(\tau_L) \) times more often than the low frequency series where either \( m(\tau_L) = m \), a constant or \( m(\tau_L) \) has a deterministic time path. For example the quarterly/monthly or annual/monthly mixed frequencies refer to a fixed \( m \) whereas the quarterly/daily or quarterly/weekly involve featuring predetermined calendar effects. Let a \( K \)-dimensional process with the first \( K_L < K \) elements, collected in the vector process \( x_L(\tau_L) \), observed every fixed period \( m \). The remaining series, \( K_H = K - K_L \), are observed at a higher frequency \( k_H = 1, ..., m \) during the \( \tau_L \) period are denoted by the double-indexed vector process \( x_H(\tau_L, k_H) \). The MIDAS-VAR is related to periodic models and is constructed using stacked skip-sampled processes. For example consider all the low frequency \( \tau_L \) series appear at the end of the (low frequency) period with the following finite order VAR representation of the stacked vector:

\[
\begin{bmatrix}
  x_H(\tau_L, 1) \\
  \vdots \\
  x_H(\tau_L, m) \\
  x_L(\tau_L)
\end{bmatrix} = A_0 + \sum_{j=1}^{p} A_j \begin{bmatrix}
  x_H(\tau_L - j, 1) \\
  \vdots \\
  x_H(\tau_L - j, m) \\
  x_L(\tau_L - j)
\end{bmatrix} + \varepsilon(\tau_L)
\]
which is a $K_L + mK_H$ dimensional VAR with $p$ lags. In the quarterly/monthly MIDAS-VAR example we stack for the high frequency variable the months of January, February and March together with the first quarter of the low frequency variable. Similarly we stack the April, May, June monthly data with the second quarter of the low frequency variable, e.t.c. The last equation of the MIDAS-VAR model is related to the single equation MIDAS regression model in the special case where $K_H = K_L = 1$ which is the ADL-MIDAS models in Andreou, Ghysels and Kourtellos (2010). In addition to the remaining dynamics the MIDAS-VAR also contains the impact of the low frequency shock, $\varepsilon(\tau_L)^{m+1.1}$, which is the last element in the innovation vector in the above model, onto both future high and low frequency series. The same applies to the high frequency shocks $\varepsilon(\tau_L)^{i.1}$ $i = 1,...,m$ on future high and low frequency series.

In the traditional forecasting setup predict the high and low frequency data given previous quarter’s high and low frequency observations. Inspired by structural VAR models where we pre-multiply the vector $[x_H(\tau_L,1),...,x_H(\tau_L,m),x_L(\tau_L)]'$ with a matrix $A_c$ which pertains to contemporaneous relationships we obtain:

$$
A_c \begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} = A_0 + \sum_{j=1}^{p} A_j \begin{bmatrix} x_H(\tau_L-j,1) \\ \vdots \\ x_H(\tau_L-j,m) \\ x_L(\tau_L-j) \end{bmatrix} + \varepsilon^*(\tau_L)$$

(A.1)

which can also be re-written for the $K_L + mK_H$ dimensional vector $x^*(\tau_L)$ as

$$A(L_L)(x^*(\tau_L) - \mu_x) = \varepsilon^*(\tau_L)$$

where $L_L$ is the low frequency lag operator, i.e. $L_L x^*(\tau_L) = x^*(\tau_L-1)$ and:

$$A(L_L) = I - \sum_{j=1}^{p} A_j L_L^j \text{ and } \mu_x = \left(I - \sum_{j=1}^{p} A_j \right)^{-1} A_0$$

where assuming that the VAR is covariance stationary and $E[\varepsilon^*(\tau_L)\varepsilon^*(\tau_L)'] = CC'$. The real time prediction of the MIDAS-VAR allows the leads of the high frequency data to provide real time updating. The simplicity of the restricted or unrestricted MIDAS-VAR model for real-time prediction is particularly appealing as it is based on estimating the mixed frequency VAR and subsequently computing the Choleski factorization of the errors and taking the $m-1$ lower triangular truncations of the original factorization. Various estimation procedures for mixed frequency VAR models can be pursued both classical and Bayesian that solves the parameter proliferation problem of unrestricted MIDAS-VAR models.
B. State space models with daily data

Following Banbura, Giannone, Modugno, and Reichlin (2013) we consider a multivariate version of equation (4.2)-(4.3). To fix notation, denote \( y^m_t \) a variable which is observed \( m \) times within the low frequency period \( t \). Assuming that frequency is constant across time \( y^m_t \) is sampled for \( t = m, 2m, 3m, ... \) and for \( k > 1 \) the observations of \( y^m_t \) will be periodically missing. Let

\[
Y^M_t = \left( y^m_{t,1}, y^m_{t,2}, \ldots, y^m_{t,N} \right)'
\]

be a vector of \( N \) variables possibly defined at different frequencies. Note that for each variable the high frequency is possibly unobserved.

\[
Y_t = \mu + \Lambda F_t + E_t, \quad E_t \sim N(0, \Sigma_E) \quad (B.1)
\]

\[
F_t = \Phi(L) F_t + U_t, \quad U_t \sim N(0, \Sigma_U), \quad (B.2)
\]

where \( \Sigma_E \) is assumed to be diagonal but it is generally robust to violations of this assumption.

The above model nests the model of Giannone, Reichlin, and Small (2008a) who only used a large set of monthly indicators produce nowcasts of GDP. In their application \( Y_t \) contains only monthly (observed) variables, hence, equations (1)- (2) constitute a state space representation. The estimates of the factors are obtained by Kalman filter and smoother and nowcasts are then obtained by regressing GDP on temporally aggregated factor estimates

\[
y^m_{t,i} = \alpha + \beta F^m_{t,i} + e^m_{t,i}, \quad t = m, 2m, ... \quad (B.3)
\]

where \( \Omega_v = \{ y^m_{t,n}, t = m, 2m, ..., T_n(v), n = 1, 2, ..., N \} \) denotes the available information set of data vintage \( v \). \( T_n(v) \) is a multiple of \( m_n \) and refers to the last observation of variable \( n \) in the data vintage \( v \).

Consider the \( n^{th} \) flow variable with \( m_n \) frequency

\[
y^m_{t,n} = \sum_{i=0}^{2m_n-2} \omega^m_{n,f} y^m_{t-i,n} = \sum_{i=0}^{2m_n-2} \omega^m_{n,f} (\Lambda^m_{n,f} F_{t-i} + e_{t-i,n}) \quad (B.4)
\]

and the \( n^{th} \) stock variable with \( m_n \) frequency

---

8Traditionally if the variable is stock then it is sampled by taking the last observation in each \( t \) (end-of-period sampling). Alternatively, if the variable is flow then flat sampling (or average sampling).

9Mixed frequency of the data-set and non-synchronous releases imply that \( T_n(v) \neq T_m(v) \) for some \( n \neq m \).
\[ y_{t,n}^m = \sum_{i=0}^{2m-2} \omega_{i} n \cdot y_{t-i,n} = \sum_{i=0}^{2m-2} \omega_{i} n \cdot (\Lambda_{n,f} F_{t-i} + e_{t-i,n}) \]  

(B.5)

where \( \Lambda_{n,f} \) and \( \Lambda_{n,s} \) are the corresponding \( n_{th} \) rows of \( \Lambda \). Also, consider separate aggregators for each frequency and for stock and flow variables, \( F_{t}^{m,f} \) and \( F_{t}^{m,s} \) for \( m = m_q, m_m, m_d \), where \( m = m_q, m_m, \) and \( m_w \) refer to the number of days in a quarter, month, and week, respectively. The idea is that we employ these variables recursively so that at the end of period \( t \) we get

\[ F_{t}^{m,f} = \sum_{i=0}^{2m-2} \omega_{i} n \cdot F_{t-i}, \ t = m, 2m, ..., m = m_q, m_m, m_w. \]  

(B.6)

and

\[ F_{t}^{m,s} = \sum_{i=0}^{2m-2} \omega_{i} n \cdot F_{t-i}, \ t = m, 2m, ..., m = m_q, m_m, m_w. \]  

(B.7)

Similar, the idiosyncratic error in the measurement equation will be a moving average of the daily \( e_{t,n} \). Denote both stock and flow variables sampled at frequency \( m \) by \( Y_{t}^{m} \). Then the measurement equation can be written as

\[
\begin{pmatrix}
Y_{t}^{m,q,f} \\
Y_{t}^{m,m,f} \\
Y_{t}^{m,m,s} \\
Y_{t}^{m,w,f} \\
Y_{t}^{m,w,s} \\
Y_{t}^{m,d}
\end{pmatrix}
= \begin{pmatrix}
\tilde{\Lambda}_{q,f} & 0 & 0 & 0 & 0 & 0 \\
0 & \tilde{\Lambda}_{m,f} & 0 & 0 & 0 & 0 \\
0 & 0 & \Lambda^{m,s} & 0 & 0 & 0 \\
0 & 0 & 0 & \tilde{\Lambda}_{w,f} & 0 & 0 \\
0 & 0 & 0 & 0 & \Lambda^{w,s} & 0 \\
0 & 0 & 0 & 0 & 0 & \Lambda^{d}
\end{pmatrix}
\begin{pmatrix}
\tilde{F}_{t}^{m,q,f} \\
\tilde{F}_{t}^{m,m,f} \\
\tilde{F}_{t}^{m,m,s} \\
\tilde{F}_{t}^{m,w,f} \\
\tilde{F}_{t}^{m,w,s} \\
F_{t}
\end{pmatrix}
\]  

(B.8)

where \( \tilde{F}_{t}^{m,q,f} = (F_{t}^{m,q,f}, F_{t}^{m,q,f})', \tilde{F}_{t}^{m,m,f} = (F_{t}^{m,m,f}, F_{t}^{m,m,f})', \) \( \tilde{F}_{t}^{m,w,f} = (F_{t}^{m,w,f}, F_{t}^{m,w,f})' \) and \( \tilde{\Lambda}_{q,f} = (\Lambda_{q,f},0), \tilde{\Lambda}_{m,f} = (\Lambda_{m,f},0), \) and \( \tilde{\Lambda}_{w,f} = (\Lambda_{w,f},0) \).

Then coefficients of the transition equation are time-varying and take the form

\[ ^{10} \text{However, in the estimation one assumes that it is a white noise.} \]
\[
\begin{pmatrix}
I_{2r} & 0 & 0 & 0 & 0 & W_{t}^{m_{q,f}} & \tilde{F}_{t}^{m_{q,f}} \\
0 & I_{2r} & 0 & 0 & 0 & \tilde{W}_{t}^{m_{m,f}} & \tilde{F}_{t}^{m_{m,f}} \\
0 & 0 & I_{r} & 0 & 0 & W_{t}^{m_{q,s}} & F_{t}^{m_{q,s}} \\
0 & 0 & 0 & I_{2r} & 0 & W_{t}^{m_{w,f}} & \tilde{F}_{t}^{m_{w,f}} \\
0 & 0 & 0 & 0 & I_{r} & W_{t}^{m_{w,s}} & F_{t}^{m_{w,s}} \\
0 & 0 & 0 & 0 & 0 & \Phi & F_{t-1} \\
\end{pmatrix}
\]

= 

\[
\begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\end{pmatrix}
\]

\[
(B.9)
\]

where \(W_{\cdot}^{\cdot}\) and \(I_{\cdot}^{\cdot}\) are suitable aggregation weights and matrices of zeros and ones, respectively.

References


Bai, J., and Ng S., (2002), 'Determining the number of factors in approximate factor models', *Econometrica* 70: 191-221.


