Estimates of Quarterly GDP Growth using MIDAS regressions

Sailesh Bhaghoe
Central Bank of Suriname
sbhaghoe@cbvs.sr

Gavin Ooft
Central Bank of Suriname
gooft@cbvs.sr

Philip Hans Franses
Erasmus School of Economics
franses@ese.eur.nl

August 2019

EI2019-29

Abstract

This paper provides new estimates of year-to-year quarterly real GDP growth in Suriname for 2013Q1 to 2018Q4. The methodology to arrive at these estimates consists of the following steps. Using the familiar Chow and Lin method, the available annual data are disaggregated into a first round of quarterly data. The quarterly data are then included in a MIDAS model, which links the quarterly observations with a new but well established monthly observed indicator of economic activity. The best-performing MIDAS model is then used to update the initial estimates of quarterly GDP growth to final estimates, which in turn can be used in macro-economic modelling and analysis.

Key Words: Quarterly real GDP growth; Disaggregation; MIDAS Regression Models; Monthly indicator of economic activity.

JEL Classifications: C32, C52, C53

1. Introduction

For many countries in the world, data on real GDP growth are available only at the annual level. For various reasons, it seems preferable to have such growth data at a higher frequency, like quarterly, as this may allow for example to monitor the effects of economic policy and to make proper decisions. At the same time, national statistical offices usually provide annual data a long time after the year ended, and one would want to have more timely estimates of quarterly growth, perhaps already during the quarter, or at least not long after the quarter has ended. In this paper we provide a methodology to provide timely estimates of quarterly GDP growth.

Our methodology is based on three steps. The first step amounts to an interpolation of annual data into quarterly data, which is necessary for the next step to create degrees of freedom. The second step involves linking the first round estimates of quarterly data with a monthly observed variable that measures economic activity. For many countries in the world such an indicator can be created using data on production, on inflation, interest rates, stock market returns, employment figures, all data that can be observed at the monthly frequency. This second step involves a regression of quarterly data on an explanatory variable that is observed at a higher frequency, that is, the month. Such a regression is called a MIDAS regression, and we will evaluate various versions of such a regression. The final step involves taking the fit of the MIDAS regressions as the final round estimates of quarterly GDP growth.

In this paper we apply our methodology to Suriname. For this country we have a monthly indicator of economic activity (MEAI[1]) for 2013M1 to 2018M12. We thus translate annual GDP growth for 2013 to 2018, which is just 6 observations, into a quarterly variable using a standard technique, creating data for 2013Q1 to 2018Q4, providing us with 24 initial estimates, and as such a larger sample to analyze various MIDAS regression models.

---

[1] See for example https://www.cbvs.sr/72-highlights-nl/1691-meai-bulletin-final-april-final-25juli2019 for the April 2019 report on the MEAI.

The outline of our paper is as follows. Section 2 outlines our methodology, whereas Section 3 illustrates it for Suriname. We present the quarterly GDP growth rates in a separate table. Section 4 concludes with limitations and an outlook for further research.

2. Methodology

In this section we outline the various steps in our methodology. The first step involves a first round of estimates of quarterly observations based on annual data. We simply rely on a well-established method for that. Next, we link these initial data with a monthly observed variable in a MIDAS regression.

Interpolation

Chow and Lin (1971) use a statistical relationship between low frequency and high frequency variables to derive a new time series by means of univariate regression and multivariate regression. The method of Denton (1971) is well-known for benchmarking to achieve consistency between time series on the same target variables measured at different frequencies. This method is often applied for the compilation of quarterly national accounts, see for example Bloem, et al. (2001). Fernández (1981) and Litterman (1983) extended the optimal procedures of Chow and Lin (1971). Here we apply the Chow and Lin method to derive a new series by combining annual real GDP growth and the MEAI growth series, when appropriately scaled downwards to quarterly data. We select this method because it is easy to apply for shorter time series and produces robust results.

Suppose that annual series up to N years are available for disaggregation into quarterly series. If the quarterly series is related to k time series, the following relationship may hold:

$$y = X\beta + \mu \tag{1}$$

The term y is an $nx1$ vector, where $n = 4N$ of the quarterly series to be estimated, X represents the matrix nxk of k variables at the higher frequency, β is an $kx1$ vector of coefficients and μ is the $nx1$ vector of stochastic disturbances with mean E(μ) $= 0$ and variance E($\mu\mu'$) $= V$, where V is an nxn matrix and takes the following identity:

$$V = \frac{\sigma_{\varepsilon}^{2}}{(1-\rho^{2})} = \begin{bmatrix} 1 & \rho & \rho^{2} & \dots & \rho^{n-1} \\ \rho & 1 & \rho^{2} & \dots & \rho^{n-2} \\ \rho^{2} & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \dots & \dots & 1 \end{bmatrix} \quad (2)$$

For the complete derivation process, we refer the reader to European Communities (2003) and Islam (2009). Application of this well-known method gives us a first round of estimates of y.

MIDAS regressions

MIDAS models have gained popularity in recent years, as these models have the ability to connect a lower frequency variable with high frequency variables. Conveniently, MIDAS models are often tightly parameterized and reduced form regressions that consider macroeconomic variables at different frequencies. MIDAS models exhibit favorable forecasting and nowcasting abilities as the higher-frequency variable is often more timely available in contrast to the lower-frequency indicator.

Ghysels, Santa-Clara, and Valkanov (2002), Andreou, Ghysels, and Kourtellos (2010), Armesto, Engemann, and Owyang (2010) and Ghysels, Kvedaras and Zemlys (2016) shed light on simple and parsimonious MIDAS regression models and their applicability. Ghysels and Valkanov (2012), among others, use these models for forecasting purposes. Others who have used MIDAS regressions for more elaborated econometric analysis are Wohlrabe (2009), Kvedaras and Račkauskas (2010), Rodriguez and Puggioni (2010), Andreou, Ghysels, and Kourtellos (2011), Bai, Ghysels and Wright (2013) and Foroni, Marcellino and Schumacher (2015).

Ghysels, Santa-Clara and Valkanov (2002) and others show that MIDAS models solve the following issues. MIDAS reduces the loss of information and bias in modeling and forecasting. Conventional models estimate different coefficients for each high-frequency observation and the number of coefficients tend to be quite large, whereas MIDAS models can reduce the coefficients by placing higher weights on closer lags than those that are far from the contemporaneous value.

Let a set of high frequency data, $X^{(m)}$, be sampled m times faster than low frequency data. For example for annual data $m = 4$ and we write that quarterly data are sampled at the frequency $X^{(4)}$. Using this notation, Ghysels, Santa-Clara and Valkanov (2002) derived the MIDAS model as follows:

$$Y_t = \beta_0 + \beta_1 B\left(L^{\frac{1}{m}}; \gamma\right) X_t^{(m)} + \epsilon_t^{(m)} \tag{3}$$

In our case, the dependent variable Y_t, which is the interpolated variable, is sampled at the quarterly frequency. The term t defines the time for model estimation up to $1, \dots, T$. The regressors $X_t^{(m)}$ are sampled m times faster than Y_t, here $m = 3$. The term $B(L^{\frac{1}{m}}) = \sum_{j=0}^{jmax} B(J) L^{j/m}$ is a polynomial of length j^{max} in the $L^{1/m}$ lag operator, which is defined as $L^{j/m} X_t^{(m)} = X_{t-j/m}^{(m)}$. The operator $L^{j/m}$ produces the value of $X_t^{(m)}$ lagged by j/m periods.

An annual/quarterly example would imply that the above equation is a projection of yearly Y_t onto quarterly data $X_t^{(m)}$ using up to j^{max} quarterly lags.

A parsimonious regression model is now formulated as:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda f\left(\gamma, X_{j,t}^H\right) + \varepsilon_t \tag{4}$$

The term Y_t^L is the dependent variable sampled at low frequency. The W_t^L is the set of regressors sampled at the same frequency as the dependent variable. $X_{j,t}^H$ is the set of explanatory variables sampled at a higher frequency. The β_i, λ, and γ are the parameters to be estimated. The *f( )* is a function translating the higher frequency data into lower frequency, and ε_t is a white noise process with mean zero and constant variance.

A following step can be to take the average of the high frequency data that occur between samples of the lower frequency variable.

$$X_t^L = \frac{1}{m} \sum_{j=0}^{m-1} X_{t-j}^H \tag{5}$$

The term m is the number of periods in the higher frequency corresponding to a single period in the lower frequency. The X_t^H are the high-frequency variables corresponding to the last observation in period t of the lower frequency. The regression model in equation (4) can be re-estimated with the time aggregation in (5), that is,

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda X_t^L + u_t \tag{6}$$

One main issue, however, with this regression model is that the estimated coefficients of each high frequency regressor of X at time t are equal. However, it is not difficult to allow for non-equal coefficients for each lag of high frequency regressor $(t-j)$. For that, (6) can be updated using different coefficients.

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{j=0}^{m-1} \gamma_{t-j} X_{t-j}^H + u_t \tag{7}$$

The issue with this model is that the number of coefficients can be quite large. However, it makes sense, certainly in our application, to put higher weights on closer lags than those that are far from the contemporaneous value through implementing a parsimonious MIDAS model. The model in (7) can be updated to a MIDAS model, like

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda \sum_{j=0}^{m-1} w_{t-j}(\gamma) X_{t-j}^H + u_t \tag{8}$$

where the weighting function $w(..)$ transforms high-frequency parameters into low frequency parameters.

The parametrization of the lagged coefficients $w_{t-j}(\gamma)$ in a parsimonious fashion is one of the key features of MIDAS models. In our application below, we will estimate various MIDAS regression models with different lag specifications, and these are Almon Lag Weighting, Exponential Almon Lag Weighting, Step Weighting, Beta Weighting, and Unrestricted Lag Polynomials. In Appendix A we provide various details of the specifications.

3. Quarterly GDP in Suriname

The data that we will work with are the annual real GDP growth rates, as they are displayed in Figure 1, and the monthly observations on the MEAI as they are pictured in Figure 2.

We start with the Chow-Lin temporal disaggregation method, using the MEAI as indexation variable, to transform annual GDP to a rolling four-quarter GDP series, maintaining the match of the last observation in a respective year. The choice for this indicator of quarterly output is to match the last quarters' growth with the annual growth.

Next, the lower-frequency variable is the first-round estimated quarterly real GDP growth, while the higher-frequency variable is the MEAI. To evaluate the quality of the MEAI we also consider the monthly mining activity index. As economic activity is trending in levels, the variables enter the MIDAS models in growth rates to avoid spurious regression results. We will compare different in-sample and out-of-sample forecasts through various forecast evaluation statistics (see Appendix B) and we test for statistical significance in forecast differences using the Diebold-Mariano (1995) forecast evaluation test.

We estimate the MIDAS models, both with the (monthly growth rates of) MEAI and of the mining activity index, with different lag functions and we also consider the Unrestricted MIDAS (U-MIDAS) model. The effective sample period runs from 2013Q1 to 2018Q4, as the growth rate of the higher-frequency variable, the MEAI, is available in the corresponding period.

MIDAS Models with MEAI

First we estimate a set of MIDAS models with the MEAI as the independent variable. The estimated MIDAS models with a beta lag distribution, the Almon lag distributions and the unrestricted MIDAS (U-MIDAS) models explain the growth of quarterly GDP by more than 95%, see Table 1.

All models include an autoregressive term of the lower-frequency variable, that is, quarterly GDP growth, to account for possible persistence and to improve model fit. We check for homoscedasticity of the residuals, normality of residuals and for no serial correlation, and we do not have evident signs of misspecifications.

The results point at significant persistence in quarterly economic growth, as the lag of quarterly GDP growth is significant and positive, while the higher-frequency variable also improves the fit. Out of the four models in Table 1, the U-MIDAS model seems to be the preferred, when comparing the R^2 and minimum values of AIC and BIC.

MIDAS Models with Mining Activity Index

Table 2 presents the estimation results for the MIDAS model when the explanatory variable is the 12- month moving average index in mining activity, which is compiled by the Central Bank of Suriname. In Suriname, the Mining Activity Index is readily available as the mining sector comprises of only a few key companies, and does not have to be constructed like the MEAI. The mining companies also report timely which makes it extremely relevant to use this indicator to predict quarterly GDP.

The MIDAS models in Table 2 show that the monthly mining activity index as an independent high-frequency variable also yields useful results. However, when we compare the models in Table 2 with those in Table 1, we see that the MEAI variable provides models with better fit, so we continue with the models 1 to 4.

Forecast Evaluation

Table 3 reports on the forecast accuracy. The model with the lowest forecast evaluation statistics (RMSE, MAE, MAPE, and Theil-U) is preferred. Based on the forecast evaluation statistics, the model with PDL/Almon lags (model 3) and the Unrestricted MIDAS (model 4) yield the best in-sample forecast for the models that employ the total MEAI as regressor.

When we split our sample in an estimation period (2012Q4 to 2017Q4) and an out-of-sample evaluation period which runs from 2018Q1 to 2018Q4, we obtain the results presented in the second panel of Table 3. The models are estimated with similar lag specifications as the in-sample forecasts. Now, we see that model 3 is clearly preferable.

Table 4 presents the Diebold-Mariano tests, and there we see again that model 3 significantly outperforms the models 1, 2, and 4.

Finally, Table 5 and Figure 3 present our final estimates of year-to-year real GDP growth, where we compare it with the initial Chow and Lin estimates. We see that our new estimates provide a more positive look on economic growth in 2015Q1 to 2015Q3.

To provide some further validation of the quality of the newly estimated quarterly growth series, we use the annual real GDP growth as published by the General Bureau of Statistics (GBS) and the Planning Office. Based on our model 3, the annual quotes would be 2.4, 0.5, -2.7, -5.1, 0.1, and 2.0 for 2013 to 2018, where GBS reports 2.9, 0.3, -3.4, -5.6, 1.7 and 1.9, respectively. Apart for 2017, where our data show less growth, we believe that these differences are small, and add to the face validity of our new quarterly growth series.

4. Conclusions

This paper presented a methodology to construct and nowcast quarterly GDP growth for the economy of Suriname. Hence, this study demonstrated that nowcasting quarterly GDP in a low-data environment is possible when representative leading indicators are available in a timely fashion. We utilize the Monthly Economic Activity Index of the Central Bank of Suriname and the Mining Activity Index to estimate MIDAS models with various lag specifications and a U-MIDAS model to forecasts quarterly growth. This methodology enabled us to regress a lower frequency indicator, that is, quarterly GDP growth on a higher frequency indicator, that is, a Monthly Economic Activity Index. The estimated models performed well in fitting the dependent variable.

The fit of the best performing model is used as a newly created quarterly GDP growth variable. Of course, a limitation is that this variable is an estimate. On the other hand, it does provide new insights as the negative growth in 2015 seems to be attributable to 2015Q4 only. We believe that our method can be applied to many more countries, for which it would be possible to have a monthly indicator of economic activity.

Table 1: MIDAS estimation results, where the Chow-Lin based interpolated quarterly GDP growth series is the dependent variable (year-on-year growth) and the lags of the growth in the **Monthly Economic Activity Index** is the explanatory variable. Effective sample size is 2013Q1 to 2018Q4. Standard errors are in parentheses.

	Model			
	1	2	3	4
Variable	Beta lags	Exponential Almon lags	PDL Almon lags	UMIDAS
Intercept	-0.475 (0.242)	-0.757 (0.355)	-0.406 (0.214)	-0.406 (0.214)
Growth t-1	0.458 (0.166)	0.195 (0.404)	0.529 (0.143)	0.529 (0.143)
Lag 0	1.563	0.564	1.720	
Lag 1	-0.595		-0.939	1.720
Lag 2	-0.595		-0.463	-0.939
Lag 3				-0.463
R^2	0.952	0.866	0.955	0.956
AIC	2.660	3.604	2.546	2.492
BIC	2.955	3.850	2.793	2.738

Table 2: MIDAS estimation results, where the Chow-Lin based interpolated quarterly GDP growth series is the dependent variable (year-on-year growth) and the lags of the growth in the **Mining Activity Index** is the explanatory variable. Effective sample size is 2013Q1 to 2018Q4. Standard errors are in parentheses.

	Model			
	5	6	7	8
Variable	Beta lags	Exponential Almon lags	PDL Almon lags	UMIDAS
Intercept	-0.140 (0.247)	-0.126 (0.309)	-0.139 (0.235)	-0.139 (0.235)
Growth t-1	0.711 (0.090)	0.735 (0.112)	0.712 (0.087)	0.712 (0.087)
Lag 0	0.144		-0.562	
Lag 1	1	0.131	0.958	-0.562
Lag 2	1		-0.280	0.958
Lag 3	-0.333			-0.280
R^2	0.904	0.841	0.903	0.904
AIC	3.362	3.781	3.319	3.279
BIC	3.657	4.027	3.566	3.524

Table 3: In-sample and out-of-sample forecast accuracy of the models in Tables 1 and 2, 2013Q1-2018Q4. Lowest values are underlined

	Model							
	1	2	3	4	5	6	7	8
In-sample: 2013Q1-2018Q4								
RMSE	0.761	1.353	<u>0.700</u>	<u>0.700</u>	1.356	1.785	1.360	1.360
MAE	0.637	1.108	<u>0.565</u>	<u>0.565</u>	1.150	1.530	1.156	1.156
MAPE	80.00	112.4	<u>75.86</u>	<u>75.86</u>	144.0	213.1	143.8	143.8
Theil U2	0.989	1.479	<u>0.947</u>	<u>0.947</u>	2.007	2.852	2.002	2.002
Out-of-sample: 2018Q1-2018Q4, when models are estimated for 2013Q1-2017Q4								
RMSE	1.143	0.744	<u>0.729</u>	1.143	1.255	1.033	1.255	1.255
MAE	0.917	<u>0.573</u>	0.609	0.917	0.926	0.875	0.926	0.926
MAPE	29.40	<u>18.99</u>	19.47	29.40	29.73	28.03	29.73	29.73
Theil U2	1.757	1.322	<u>1.088</u>	1.757	2.200	1.742	2.200	2.200

Table 4: Diebold-Mariano tests

	In-sample		Out-of-sample	
	Test value	p value	Test value	p value
Model 1 versus Model 2	-2.978	0.003	2.053	0.040
Model 1 versus Model 3	2.385	0.017	1.762	0.078
Model 1 versus Model 4	1.102	0.270	-2.953	0.003
Model 2 versus Model 3	3.015	0.003	0.089	0.929
Model 2 versus Model 4	3.015	0.003	-2.053	0.040
Model 3 versus Model 4	-0.486	0.637	1.762	0.078

Table 5: The Chow-Lin based quarterly GDP growth, 2013Q1-2018Q4, and the final nowcasted values using the MIDAS regression model 3.

Quarter	Chow-Lin estimate	Our estimate
2013Q1	2.39	2.67
2013Q2	4.25	3.87
2013Q3	3.98	3.36
2013Q4	2.92	2.36
2014Q1	2.65	2.23
2014Q2	0.13	0.60
2014Q3	-0.42	-0.62
2014Q4	0.26	0.51
2015Q1	-0.88	0.12
2015Q2	-0.16	0.90
2015Q3	-1.47	0.01
2015Q4	-3.40	-2.70
2016Q1	-4.52	-3.80
2016Q2	-6.48	-6.33
2016Q3	-6.40	-6.39
2016Q4	-5.57	-5.08
2017Q1	-4.71	-4.83
2017Q2	-0.70	-1.45
2017Q3	-1.70	-1.20
2017Q4	1.69	0.14
2018Q1	3.21	3.33
2018Q2	2.47	2.09
2018Q3	3.09	3.28
2018Q4	3.11	1.95

Figures

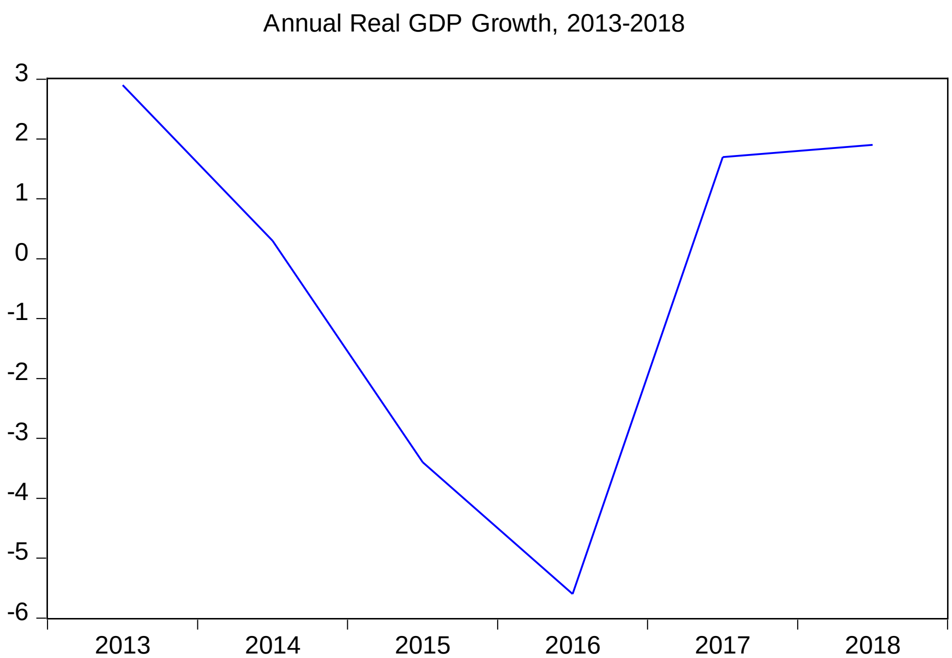

Figure 1: Real GDP growth in Suriname, for the years 2013-2018

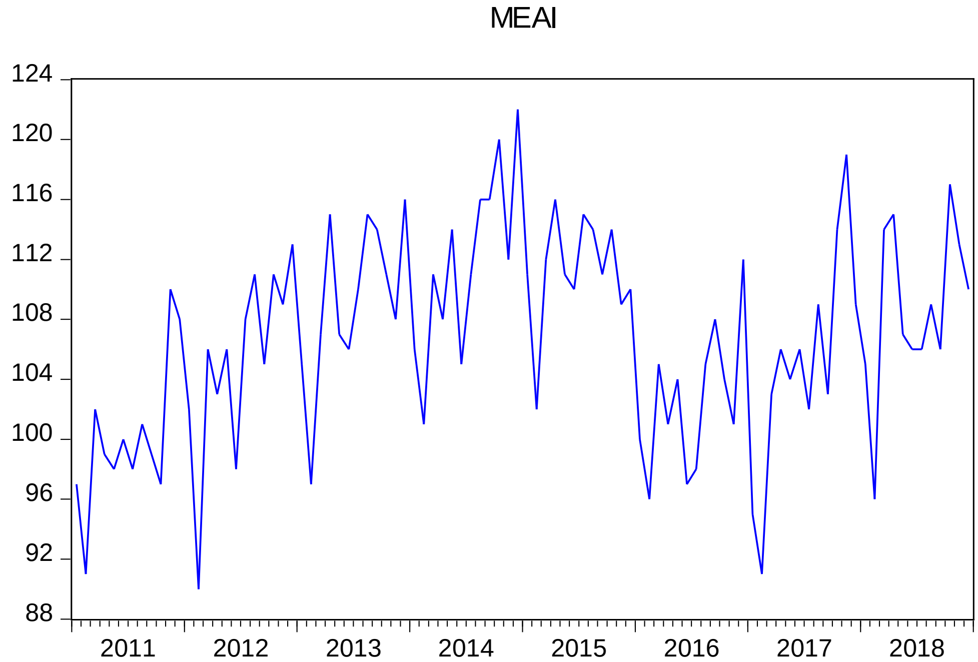

Figure 2: The Monthly Economic Activity Index for the months 2011M01-2018M12

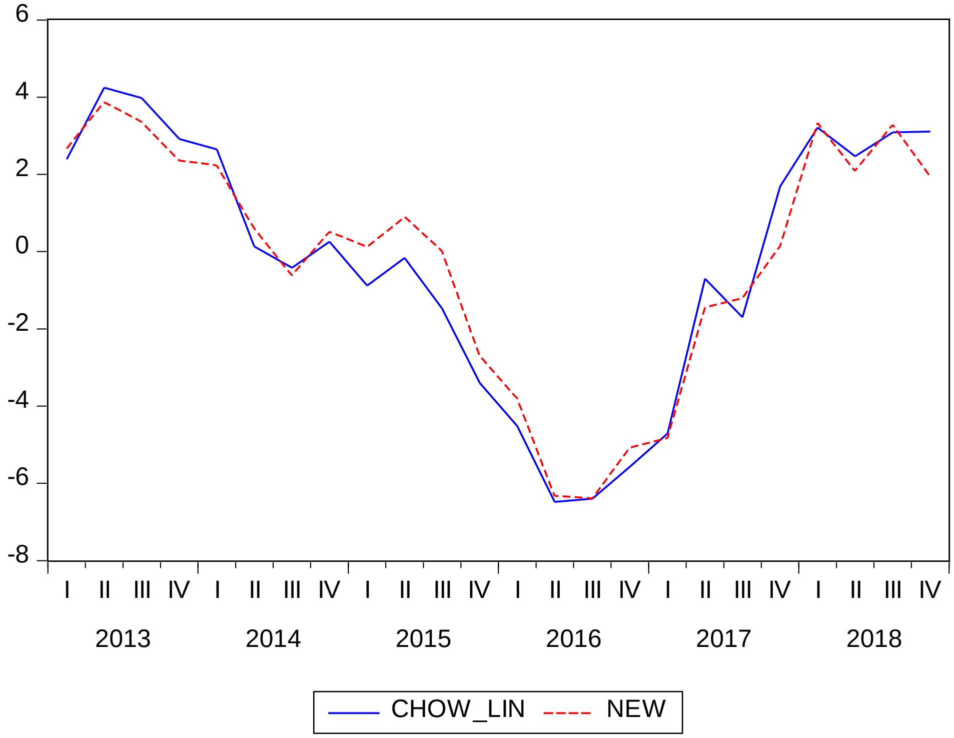

Figure 3: The Chow-Lin based quarterly GDP growth, 2013Q1-2018Q4, and the final nowcasted values using the MIDAS regression

Appendix A - Polynomial Lag Specifications

Polynomial Distributed Lag or Almon Weighting parametrization

The PDL or Almon Weighting is widely used to put restrictions on lag coefficients in the class of autoregressive models. This weighting scheme is considered as natural candidate for mixed frequency weighting. For each high frequency lag up to j, the regression coefficients are modeled as a p dimensional lag polynomial in the MIDAS parameters. The number of the coefficients to be estimated depends on the polynomial order (p) and not on the number of lags *(j)* chosen. It uses the functional form with up to p coefficients of γ_1, γ_2,…, γ_p. Equation (8) can be updated to a MIDAS with PDL structure:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{i=1}^{p} \gamma_i \sum_{j=0}^{k} j^{i-1} X_{t-j}^H + u_t$$

The term *k* is the chosen number of lags; *p* is the order of the polynomial.

The Exponential Almon Weighting

An alternative method to Almon Lags is the Exponential Almon Lag Weighting. This method restricts the number of polynomials to the second order $[\gamma_1; \gamma_2]$ but takes exponents (see Almon (1965). The Exponential Almon Lag Weighting is known to be quite flexible. In the Exponential Almon Lag structure $\gamma_1 = \gamma_2 = 0$, meaning equal weights. The weight can decline slowly or fast with the lag. A declining weight is guaranteed as long as $\gamma_2 \leq 0$. The rate of decline determines how many lags are included in the regression. Once the functional form $[\gamma_1; \gamma_2]$ is specified, the lag length selection is purely data driven, leading to a MIDAS model with Exponential Almon Lag structure.

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda \sum_{j=0}^{k} Z_{j,t} + u_t$$

$$Z_{j,t} = \frac{\exp(j\gamma_1 + j^2\gamma_2)}{\sum_{i=0}^{k} exp(i\gamma_1 + i^2\gamma_2)} X_{t-j}^{H}$$

Beta Weighting Parametrization

The Beta function involves the estimation of three coefficients compared to the Exponential Almon Lag Weighting with two coefficients. The coefficients can be restricted by imposing either: $\gamma_1 = 1$ or $\gamma_3 = 0$ or $\gamma_1 = 1$ and $\gamma_3 = 0$. The number of parameters estimated can therefore be either $1, 2$ or 3 (depending on the types of restrictions imposed). This weighting scheme does not allow the number of parameters to increase with the number of lags, but the estimation involves a highly non-linear estimation procedure as that of the Exponential Almon Lag. Both these polynomial specifications have two important characteristics, that is, they provide positive coefficients and they sum up to unity. They are categorized as finite polynomials (Ghysels, Sinko and Valkanov, 2006). MIDAS with Beta parametrization is defined as follows:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda \sum_{j=0}^{k} Z_{j,t} + u_t$$

$$Z_{j,t} = \left(\frac{{\psi_j}^{\gamma 1-1}(1-\psi_j)^{\gamma 2-1}}{\sum_{i=0}^{k} {\psi_i}^{\gamma 1-1}(1-\psi_i)^{\gamma 2-1}} + \gamma_3 \right) X_{t-j}^{H}$$

$$\psi_j = \frac{j-1}{k-1}$$

Step Weighting Parametrization

MIDAS with step-functions was introduced by Forsberg and Ghysels (2007). The step-function approximates the distributed lag pattern by a number of discrete steps (η). The more steps appear in the regressions the less parsimonious is the regression model. The number of coefficients (γ_1, γ_2,…, γ_η) to be estimated depends on the functional form up to (η) steps. Step weighting lowers the number of estimated coefficients since it restricts consecutive lags to have the same coefficient. For example, if $j = 12$ and $\eta = 4$, the first 4 lags have the same coefficient, the next four lags have the same coefficient and so on, all the way up to $j = 12$. MIDAS with step weighting is defined as follows:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{j=0}^{k} \phi_{t-j} X_{t-j}^H + u_t$$
$$\phi_j = \gamma_k$$

Unrestricted MIDAS (U-MIDAS)

Unrestricted Lag Polynomials are useful in macroeconomic application, when differences in sampling frequencies are small (for instance monthly and quarterly). When the difference in sampling frequencies between the regressand and the regressors is large, distributed lag functions are typically employed to model dynamics avoiding parameter proliferation. The number of the coefficients to be estimated $\gamma_1, \gamma_2, \ldots, \gamma_k$ does not depend on lag structures as in the case of Almon, Exponential Almon, Beta and Step. The U-MIDAS has the following identity:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{j=0}^{m-1} \gamma_{t-j} X_{t-j}^H + u_t$$

The MIDAS specification captures a very rich dynamic of the high frequency process in a very simple and parsimonious fashion. The parametrization of the polynomial is similar in spirit to the distributed lag models, see for example Griliches (1967), Dhrymes (1971), and Sims (1972).

Appendix B - Forecast evaluation criteria

Root Mean Square Error (RMSE)

$$\sqrt{\sum_{t=T+1}^{T+h} \frac{(\widehat{y_t} - y_t)^2}{h}}$$

Mean Absolute Error (MAE):

$$\sum_{t=T+1}^{T+h} \frac{|\widehat{y_t} - y_t|}{h}$$

Mean Absolute Percentage Error (MAPE):

$$100 \sum_{t=T+1}^{T+h} \frac{\left|\left(\widehat{y_t} - y_t\right)/y_t\right|}{h}$$

Theil-U Statistic:

$$\frac{\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2/h}}{\frac{\sqrt{\sum_{t=T+1}^{T+h} \hat{y}_t{}^2}}{h} + \frac{\sqrt{\sum_{t=T+1}^{T+h} y_t{}^2}}{h}}$$

References

Almon, S (1965), The distributed lag between capital appropriations and expenditures, *Econometrica*, 33, 178-196.

Andreou E, Ghysels E, Kourtellos A (2010), Regression models with mixed sampling frequencies, *Journal of Econometrics*, 158, 246–261.

Andreou E, Ghysels E, Kourtellos A (2011), Forecasting with mixed-frequency data. In MP Clements, DF Hendry (eds.), *Oxford Handbook of Economic Forecasting*, pp. 225–245.

Armesto MT, Engemann KM, Owyang MT (2010), Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, 92, 521–536.

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

Bloem, A M, Dippelsman, RJ, and Mæhke, NØ (2001), Quarterly National Accounts Manual - Concepts, Data Sources, and Compilation, International Monetary Fund, Washington DC.

Chow, GC and Lin, A (1971), Best linear unbiased interpolation, distribution and extrapolation of time series by related series, *Review of Economics and Statistics*, 53, 372-375.

Denton, FT (1971), Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association*, 66, 99–102.

Dhrymes, P (1971), Distributed lags: Problems of estimation and formulation. San Francisco: Holden-Day.

Diebold, FX and Mariano, RS (1995), Comparing predictive accuracy, *Journal of Business & Economic Statistics*, 13, 253-263.

European Communities (2003), Temporal disaggregation of economic time series: towards a dynamic extension, European Communities, Luxembourg.

Fernández, R (1981), A methodological note on the estimation of time series, *Review of Economics and Statistics*, 63, 471–478.

Forsberg, L and Ghysels E (2007), Why do absolute returns predict volatility so well?, *Journal of Financial Econometrics*, 5, 31-67.

Foroni C, Marcellino M, Schumacher C (2015), Unrestricted Mixed Data Sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials, *Journal of the Royal Statistical Society A*, 178(1), 57–82.

Ghysels E, Santa-Clara P, Valkanov R (2002), The MIDAS touch: Mixed data sampling regression models. Working paper, UNC and UCLA.

Ghysels E, Sinko A, Valkanov R (2006), MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26, 53–90.

Ghysels E, Valkanov R (2012). Forecasting volatility with MIDAS. In L Bauwens, C Hafner, S Laurent (eds.), *Handbook of Volatility Models and Their Applications*, pp. 383–401, New York: Wiley.

Ghysels E, Kvedaras V, Zemlys V (2016), Mixed frequency data sampling regression models: The R Package midasr. R package version 0.6, URL https://CRAN.R-project.org/ package=midasr.

Griliches, Z (1967), Distributed lags: A survey, *Econometrica*, 35, 16-49.

Islam, MR (2009), Evaluation of different temporal disaggregation techniques and an application to Italian GDP. *BRAC University Journal* VI, No. 2, 21–32.

Kvedaras V, Račkauskas A (2010), Regression models with variables of different frequencies: The case of a fixed frequency ratio, *Oxford Bulletin of Economics and Statistics*, 72(5), 600–620.

Litterman, R (1983), A random walk, Markov model for the distribution of time series,.*Journal of Business and Economic Statistics*, 1, 169–173.

Rodriguez A, Puggioni G (2010). Mixed frequency models: Bayesian approaches to estimation and prediction. *International Journal of Forecasting*, 26, 293–311.

Sims, CA (1972), Money, income, and causality, *American Economic Review*, 62, 540-552.