SAE./No.176/March 2021

Studies in Applied Economics

NOWCASTING QUARTERLY GDP GROWTH IN SURINAME WITH FACTOR-MIDAS AND MIXED-FREQUENCY VAR MODELS

By Sailesh Bhaghoe and Gavin Ooft

Johns Hopkins Institute for Applied Economics,
Global Health, and the Study of Business Enterprise

Nowcasting Quarterly GDP Growth in Suriname with Factor-MIDAS and Mixed-Frequency VAR Models

By Sailesh Bhaghoe and Gavin Ooft

About the Series

The *Studies in Applied Economics* series is under the general direction of Professor Steve H. Hanke, Founder and Co-Director of the Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise (hanke@jhu.edu).

About the Authors

Sailesh Bhaghoe (sbhaghoe@cbvs.sr) is an economic modeller at the Central Bank of Suriname. His research area is Empirical Economics with a focus on modelling, nowcasting and forecasting with high frequency economic data for small emerging economies, such as Suriname.

Gavin Ooft (gavin.ooft@uvs.edu) is a lecturer at the Anton de Kom University of Suriname. He completed an MS in Applied Economics at the Zanvyl Krieger School of Arts and Sciences at the Johns Hopkins University.

Abstract

We apply Factor-MIDAS (FaMIDAS) and Mixed-Frequency Vector Autoregression (MF-VAR and MF-Bayesian VAR) to nowcast quarterly GDP growth of Suriname. For this purpose, we use a set of 44 microeconomic indices over the sample period 2012Q1 - 2020Q2. In the target equation, we regress GDP growth upon its first lag and a beta coefficient. In the explanatory equations the first set of monthly regressors explain the variation of growth without lags while the second set of regressors are fitted with two-month lags. We apply three set of samples for model estimations: 2012Q1 – 2019Q3, 2012Q1 – 2020Q1 and 2012Q1 – 2020Q2. Model nowcast accuracy is benchmarked against GDP growth of 2019 and economic activity growth estimated by the monthly GDP indicator of March and June 2020. The models provide mixed results as compared to the benchmark indicators. We select the models with the lowest Root Mean Squared Error (RMSE) and based on own Judgment to nowcast. As the forecast horizon increases from 2019Q4 to 2020Q2, so do the RMSE. To hedge against high biases and variances, we combine the best nowcasts to produce a single nowcast. Furthermore, it appeared that the FaMIDAS and the MF-VAR models deliver adequate results for two nowcast horizons.

Keywords: FaMIDAS, MF-VAR, MF-BVAR, Nowcasting
JEL codes: C22, C53, E37

1. Introduction

Suriname is a small developing economy in South America. The Central Bank of Suriname plays a major role when it comes to nowcasting and forecasting GDP growth. This institution operates a macroeconomic model that forecasts GDP 5-years ahead. In addition, there is access to a monthly GDP indicator that provides real time information about the evolution of economic activity 2-months ahead after closing of the reference month, see Bhaghoe and Eckhorst (2020). The availability of mixed frequency data allows for developing FaMIDAS regressions and mixed frequency VAR type of models to nowcast GDP on quarterly basis in an efficient way.

In this paper, we consider FaMIDAS, MF-VAR, and MIDAS Factor Bayesian VAR (MF-BVAR) type of models to nowcast quarterly GDP growth. These types of models are proposed in the forecasting literature, see for example Frale and Monteforte (2009), Ghysels (2011), Kuzin, Marcellino and Schumacher (2011) and Franta, Havrlant and Rusnák (2016), and the references therein. We use 44 real time microeconomic indices for developing the models. Our data for the quarterly GDP range is 2012Q1 - 2019Q4, while the range of the monthly GDP indicator is available from January 2011 to June 2020. As it is a tedious journey to develop a large set of individual models with all these 44 variables, we therefore resort to principal component analysis (PCA) to extract useful common factors to construct the models and combine the best nowcasts (see Timmermann, 2006).

We select the best models by means of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), the Root Mean Squared Error (RMSE), and by own judgement. In order to evaluate the nowcasting performance, we compare the model accuracy with the output of two benchmark indicators: GDP growth of the General Bureau of Statistics[1] and economic activity growth of the monthly GDP indicator[2]. To ensure comparability and evaluation, we adjust the sample size.

[1] https://statistics-suriname.org/en/

[2] https://www.cbvs.sr/

The first set of models are estimated using 2012Q1-2019Q4 data, then the next set of models uses 2012Q1-2020Q1 data, and the last set of models uses 2012Q1-2020Q2 data. It appeared that each model has a weak predictive power. We therefore use a boosting approach to combine the multiple weak models to create one stronger model, which also reduces bias and variance.

This paper makes a significant contribution for modelling in the Caribbean with mixed-frequency data models. We construct a number of models and provide a comparison of the performance of these models for nowcasting GDP growth of Suriname. Our results indicate that these high frequency models are very useful for complementing existing traditional macroeconomic models of the Central Bank of Suriname and the National Planning Office.

The rest of the paper is organized as follows. Section 2 discusses factor regression models, section 3 discusses MIDAS and FaMIDAS models, section 4 outlines the mixed-frequency VAR models, and section 5 elaborates on model combinations. Sections 6 and 7 provide the data and results respectively, and in section 8, we provide our concluding remarks. Appendix I provides supporting materials.

2. Factor Regression Models

Factor regression models are useful when dealing with many independent variables. We resort to PCA for component extraction. These principal components consequently enter the regression models as independent variables (Scott and John, 1966). Stock and Watson (1998, 1999) provide insight into forecasting macroeconomic time series with common factors that significantly reduce forecast errors. Boivin and Ng (2004, 2005) compare factor models to traditional forecasting models and report that they tend to suffer less from model specification and considerably improve forecasting. Bernanke, Boivin and Eliasz (2005) demonstrate the use of factor-augmented VAR (FAVAR) to estimate the effect of monetary policy on the economy. The aggregate time series they use for this purpose are: real output, income, employment, consumption, housing sales, inventories and orders, exchange rates, interest rates, stock prices, price indices, credits, and average hourly earnings. The results show plausible effects of monetary policy on these macroeconomic variables.

Factor models have been successfully employed to forecast economic activity on a real time basis. Giannone et al. (2005) develops a factor regression model for tracking real-time flow of information released at different lags to nowcast quarterly GDP growth for the US. Each time when new macroeconomic data becomes available, the nowcasts are updated based on "jagged edge" data[3]. Forni et al. (2005) improve this methodology and construct a factor model, which successfully tracks the current state of the economy for the Euro area. Banerjee and Marcellino (2006) and Eickmeier and Zieglier (2008) utilize large multidimensional datasets for developing factor regression models for the U.S. and Euro-area. These authors' findings show that factor models increase forecast accuracy for GDP growth, inflation, and other macroeconomic aggregates as well.

Marcellino and Schumacher (2007, 2008) utilize "ragged edge" data[4] to develop different types of factor regression models for the German economy. Their findings show that the most parsimonious fitted models for nowcasting with "ragged edge" data are FaMIDAS regressions. Kuzin, Marcellino and Schumacher (2011) compare the accuracy of MIDAS and MF-VAR and report that these types of models complement rather than substitute each other.

Franta, Havrlant and Rusnák (2016) apply MF-VAR, MF-BVAR, MIDAS, and Dynamic Factor Model (DFM) to forecast Czech GDP growth and evaluate their precision against the Czech National Bank`s forecast. Their results suggest that the forecasts of these models compete successfully with the Czech National Bank`s forecast. Cimadomo et al. (2020) apply MF-BVAR to nowcast US economic activity growth. These authors' point of departure is the three "Vs"[5] and shows that the BVAR has similar nowcasting performance as that of the DFM, a powerful tool for policy analysis. Recently, Schorfheide and Song (2020) apply MF-VAR to forecast U.S. economic activity during the COVID-19 pandemic and report that the COVID-19 shock generates long-lasting reduction in real activity. Others such as Frale and Monteforte (2009) suggest that regression models developed with factors do well in nowcasting and forecasting.

---

[3] See António, Maximiano and Francisco (2009).

[4] See Bouwman and Jacobs (2011).

[5] Volume (large number of time series continuously released), Variety (data are published with different frequencies and precisions) and Velocity (incorporating new data in a timely fashion after their release).

3. MIDAS Regression

Ghysels, Santa-Clara and Valkanov (2002, 2004, and 2006) develop the MIDAS regression. MIDAS incorporates crucial high frequency data into regressions over lower frequency target variable. Another definition of MIDAS regression is that it is a sparsely parameterized reduced form regression over one explanatory variable, utilizing non-linear least squared method. Clements and Galvão (2007) and Kuzin, Marcellino and Schumacher (2009) report that MIDAS regressions suffer less from the curse of dimensionality for nowcasting.

A MIDAS regression takes the form:

$$y_t = \beta_0 + B\left(\theta, L^{\frac{1}{m}}\right) X_t^m + \epsilon_t \qquad (1)$$

where y_t is the target variable, β_0 is the intercept in the equation to control the error term ϵ_t. We assume that the expected value of the error term is uncorrelated with the set of regressors, $E(\epsilon_t) = 0$. The second term in the equation $B\left(\theta, L^{\frac{1}{m}}\right)$ can be written in the form of a summation operator $B\left(\theta, L^{\frac{1}{m}}\right) = \sum_{k=0}^{k} b(\theta, k) L^{k/m}$. This part of the equation is a polynomial of lag k and $L^{\frac{1}{m}}$ is an operator such that $L^{k/m} * X_t^m = X_{t-k/m}^m$. With this regression equation, we can now project y_t into a higher frequency series X_t^m with k lags.

To construct the FaMIDAS regression we need to depart from the DFM. The following is the representation of the DFM.

$$\begin{aligned} y_t &= \vartheta_o f_t + \vartheta_o f_{t-1} + \gamma_t + S_t \beta, \; t = 1, \ldots, n, && (2) \\ \Phi(L)\Delta f_t &= \eta_t, && \eta_t \sim NID\left(0, \sigma_\eta^2\right), \\ D(L)\Delta \gamma_t &= \delta + \zeta_{t,} && \zeta_t \sim NID\left(0, \Sigma_\zeta\right), \end{aligned}$$

The dependent variable,y_t, represents a vector of N time series with mixed frequency data. In essence, the DFM decomposes,y_t, into nonstationary components, f_t, and errors, γ_t, for each time series included in the estimation procedures. An important feature of this model is that f_t and γ_t follows autoregressive conditions. The parameter, ϑ_o, and the unobserved factors, γ_t, measure the effects on the dependent variable.

The variable S_t represents a regression matrix of exogenous components (e.g. calendar effects, Easter, length of the month and outliers) and the components of β are used in the model to possibly control for fixed effects and for the purpose of initializing parameters. $\Phi(L)$ is an autoregressive polynomial of order p with stationary roots and $D(L)$ is a diagonal matrix with autoregressive polynomials of order $p_i(i = 1, \dots, N)$. Furthermore, η_t and $\zeta_{t,}$ are disturbances and uncorrelated. The DFM is estimated in a linear State Space Form (for reference see Frale and Monteforte, 2009).

The FaMIDAS regression is a combination of the DFM with mixed frequency data and the MIDAS structure. We first partitioned the dependent variable in the DFM into two groups: $y_t = \left[y'_{1,t}, y'_{2,t}\right]'$. The right-hand side of the equation represents the target variable at lower frequency. Introducing the MIDAS structure for the high frequency indicators initiate the following equation: $y'_{1,t} = [b(L_k, \theta)x_t]'$. The FaMIDAS is defined as follows:

$$\begin{aligned}
\begin{bmatrix} b\left(L_{k,}\theta\right)x_t \\ y_{2,t} \end{bmatrix} &= \vartheta_o f_t + \gamma_t + S_t\beta, \quad && t = 1, \dots, n, \\
\Phi(L)\Delta f_t &= \eta_t, && \eta_t \sim NID\left(0, \sigma_\eta^2\right), \\
D(L)\Delta\gamma_t &= \delta + \zeta_{t,} && \zeta_t \sim NID\left(0, \Sigma_\zeta\right).
\end{aligned} \tag{3}$$

4. Mixed - Frequency VAR and Factor Bayesian VAR

Mariano and Murasawa (2003, 2010) develop the MF-VAR for forecasting U.S. quarterly GDP growth with high frequency explanatory variables. The following is the MF-VAR model.

$$y_{t_q}^Q = \frac{1}{3}\tilde{y}_{t_m}^M + \frac{2}{3}\tilde{y}_{t_{m-1}}^M + \tilde{y}_{t_{m-2}}^M + \frac{2}{3}\tilde{y}_{t_{m-3}}^M + \frac{1}{3}\tilde{y}_{t_{m-4}}^M \tag{4}$$

where $y_{t_q}^Q$ is the quarterly GDP growth and t_m = 3, 6, 9,…,T_m^y that represents the monthly explanatory variables. Combining GDP growth with monthly explanatory variables ($X_{t_m}^M$) into the VAR model yields the estimator:

$$z_{t_m} = \sum_{j=1}^{p^M} \beta_j z_{t_m} - j + u_{t_m}, \tag{5}$$

$$\text{where } z_{t_m} = \begin{pmatrix} \tilde{y}^M_{t_m} & - & \tilde{\mu}^M_y \\ X^M_{t_m} & - & \mu_X \end{pmatrix} \text{ and } \mu_{t_m} \sim N(0, \Sigma)^{6}.$$

We apply MF-BVAR to nowcast GDP growth as well. The MF-BVAR is estimated with Gibbs sampling (see Chiu et al., 2012 and the references therein). This model allows for the incorporation of priors compared to the MF-VAR and MIDAS. The following represents the MF-BVAR model:

$$w_{t_m} = A + \sum_{j=1}^{p^M} B_j w_{t_{m-j}} + \mu_{t_m}, \tag{6}$$

where w_{t_m} represents a vector of endogenous variables on monthly and quarterly basis. A represents a vector of intercept coefficients, B_j is an A x A matrix of lag coefficients and μ_{t_m} a vector of errors with the Gaussian assumption $\mu_{t_m} \sim N(0, \Sigma)$. The Bayesian estimator is defined as:

$$w_{t_m} = \begin{pmatrix} \tilde{y}^Q_{t_m} \\ X^M_{t_m} \end{pmatrix}, \tag{6}$$

where $\tilde{y}^Q_{t_m}$ is the quarterly GDP growth and $X^M_{t_m}$ the explanatory principal components.

5. Nowcast Combination

According to Chong and Hendry (1986), nowcasts obtained from single models may contain errors because of misspecifications. Timmermann (2006) reports that finding a best single model for shorter high frequency time-series is challenging and therefore combining the predicted values is highly suggested. Atiya (2020) argues that single models tend to have large biases. Diebold (1989) argues that model combinations are less justified when access to accurate data is possible. Yang (2004) argues that combining weights for pooling models could lead to estimation errors. Finally, Atiya and El-Shishiny (2011) argue that a major pitfall for model combination is the presence of outliers in the data and misspecification of the underlying data generating process.

[6] For further mathematical elaboration, see Mariano and Murasawa (2003, 2010).

On the other hand, Clemen (1989), Makridakis and Hibon (2000), Stock and Watson (2001, 2004) and Marcellino (2004) argue that model combinations on average, increase forecast accuracy compared to that of individual models. Pesaran and Timmermann (2005) indicate that in the presence of structural breaks, it is plausible across periods with varying degrees of stability to combine models. Timmermann (2006) and Franta, Havrla and Rusnák (2016) report that pooling models for shorter time series improves nowcast. Kourentzes and Petropoulos (2016) justify model combination for achieving diversity by using different models on the same data.

We depart from the following approach: $y = (y_1, \dots, y_H)^T$. This vector contains observations to be nowcasted H steps ahead with FaMIDAS regressions and MF-VAR type models. If the models produce nowcast vectors $u(1), \dots, u(N)$ then the nowcast combination is expressed as

$$u = \sum_{i=1}^{N} \omega_i \, u(i), \qquad (7)$$

with

$$0 \leq \omega_i \leq 1 \qquad \sum_{i=1}^{N} \omega_i = 1$$

where ω_i is the weight of each model. In the combined nowcast, the bias and variance of the mean squared error are reduced. For detail mathematical elaborations, see Atiya (2020).

6. Data

We use 44 monthly microeconomic time series available for the sample period January 2011 to June 2020. These time series are retrieved from surveys of the Central Bank of Suriname and broadly cover activities in the following sectors: agriculture, mining, manufacturing, utilities, construction, transport, banking, insurance, communications, wholesale and retail sales, and the government. The actual observations consist of company revenues and production volumes. These variables are listed in the Appendix.

For nowcasting and forecasting purposes, the dataset needs to be homogenized. For this purpose, the central bank utilizes various corresponding price indices obtained for the local economy and for the main import and export markets.

The rationale for including foreign price indices in the estimations is straightforward: the Surinamese economy is a small open economy and therefore is strongly affected by price changes in main markets. The statistical department of the central bank transforms the dataset into a set of Laspeyres indices (2011=100). From these indices, we extract principal components for modelling purposes.

The PCA extracts 9 factors of which the first 5 explains more than 70% of the variance in the data (Appendix: Table 6 and 7). These 5 factors enter the FaMIDAS regressions, estimated with polynomial distributed lag (PDL) and unrestricted polynomial lags (U-MIDAS) for modeling and nowcasting quarterly growth. The MF-VAR is estimated with U-MIDAS and the MF-BVAR with Bayesian sampling. For this class of models, we conduct the estimations with the first 3 factors which explain around 60% of the variance. Compared to the MIDAS regressions, adding more factors in the MF-VAR and MF-BVAR lead to weak model specifications. Model diagnostics reveal that adding 3 factors solves for the roots of the autoregressive polynomial. Figure 1 depicts the variance of the principal components and figure 2 the quarterly GDP growth.

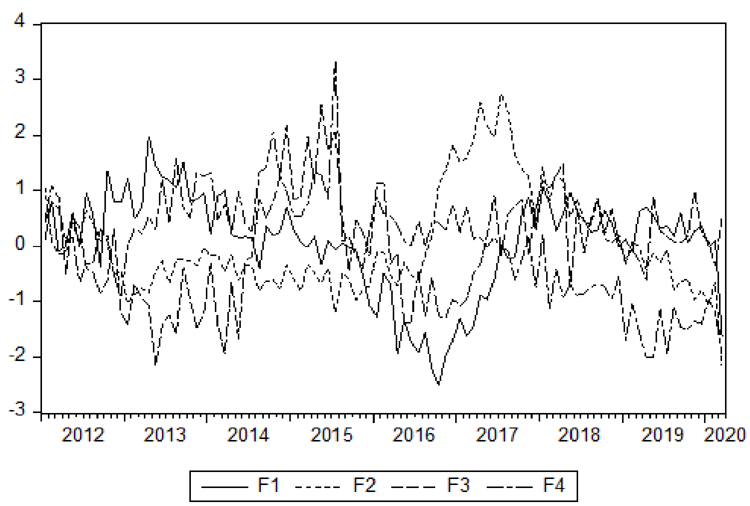

Figure 1. Principle components

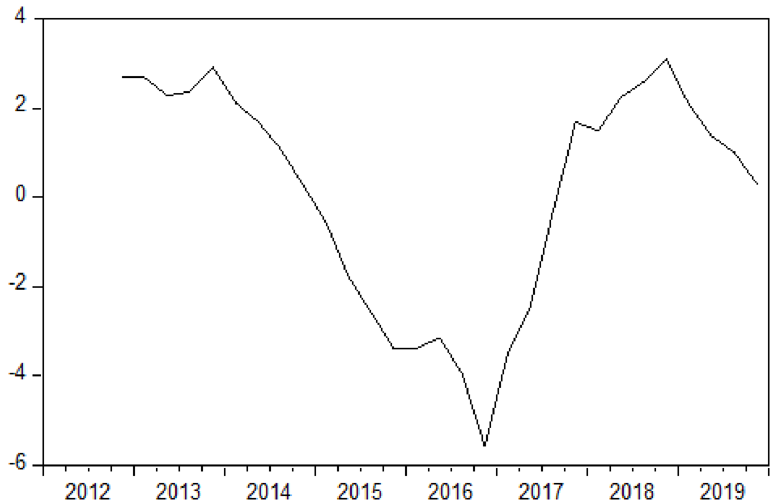

Figure 2. Quarterly GDP growth (see Bhaghoe, Ooft and Franses, 2019)

7. Results

We estimate 20 FaMIDAS regression models, which are parameterized with PDL and with U-MIDAS. We consider these lag polynomials, because they seem to be used frequently in MIDAS regressions for nowcasting short time series[7]. In addition, we estimate MF-VAR and MF-BVAR[8] models. We consider these models as complements for the FaMIDAS, as both of these approaches have their merits. Kuzin, Marcellino and Schumacher (2011) show that MIDAS are parsimoniously fitted models providing direct multiple forecasts, whereas the MF-VAR type models provide iterative forecasts.

The FaMIDAS equation regresses the quarterly GDP growth on its own lag and the 5 principal components (k). The explanatory variables of the MIDAS models (PDL 1-5) and (U-MIDAS 11-15) are not constructed with lagged terms, while we added two autoregressive terms to the remaining models (PDL 6 -10) and (U-MIDAS 16-20). We test for an optimal lag length using the AIC and the BIC to find the model that minimizes residual correlations.

[7] See Marcellino and Schumacher (2008) and the references therein.
[8] For differences between MIDAS and MF-VAR, see Kuzin, Marcellino and Schumacher (2009) and the references therein.

Hence, we produce different models to model the nature of the dynamic nowcast and to optimize forecast performance.

We parsimoniously fit the MF-VAR and MF-BVAR models only with intercepts and 3 common factors. For the Bayesian sampling, we optimize hyper-parameters iteratively to improve as much as possible forecast performance. For both the high frequency parameter as well as for the low frequency parameter, AR(1), we set $\rho = 0.4$. The frequency scale is set at $\upsilon = 0.5$.

To evaluate the dynamic nowcast performance of the FaMIDAS models, we carry out recursive estimations. The full sample is divided into an estimation part and an evaluation part. The estimation sample is 2012Q1 - 2019Q4, 2012Q1 - 2020Q1 and 2012Q1-2020Q2. While we perform dynamic nowcasting with the MIDAS, with the MF-VAR and the MF-BVAR, we perform stochastic nowcasting. We could not perform appropriate dynamic nowcast with the MF-VAR systems because of very few observations. Consequently, we only nowcast quarterly growth for 2019Q4 and 2020Q1.

To validate the model outcomes, we make use of benchmark indicators. In this context, the nowcast of 2019Q4 is benchmarked against annual GDP growth of 2019 (0.3%). For the nowcast figures of 2020Q1and 2020Q2, we use the economic activity growth of the monthly GDP indicator of March 2020 (0.0%) and of June 2020 (-7.3%). The result of benchmarking is expressed in the lowest RMSE.

We compute 22 nowcasts based on information available at time *t*. Table 1 and 2 presents the FaMIDAS and the MF-VAR models. Included are intercept estimates, fitted coefficients, Adjusted R-squared values, AIC and BIC. We conduct model selection assuming high Adjusted R-squared values, lowest AIC and BIC, and own judgment, Table 3 and 4. Included are the nowcasts for 2019Q4, 2020Q1, 2020Q2 and their respective RMSEs.

The models provide mixed results. For the estimation sample 2012Q1- 2019Q4, the FaMIDAS regression 5, estimated with unrestricted coefficients (U-MIDAS), provides the best nowcast for 2019Q4 (RMSE = 0.019).

The second best performing individual model is the BVAR (RMSE of 0.028). The FaMIDAS regression 5, estimated with PDL structure, also delivers best in-sample nowcast for 2019Q4 (RMSE = 0.039).

Re-estimating the models with the sample range 2012Q1 - 2020Q1 indicates the following: in the class of the FaMIDAS, model 1 estimated with U-MIDAS (RMSE = 0.011) followed by model 8 estimated with PDL (RMSE = 0.044), deliver best in-sample nowcasts. The MF-VAR is not far behind yielding a low RMSE of 0.074.

As we re-estimate the models with the sample range 2012Q1- 2020Q2 to increase the in-sample nowcast horizon, the following picture emerges. The RMSEs of almost all the models increase significantly, indicating high bias and high variance in the nowcasts, except for model 3 estimated with PDL. This model reports the lowest RMSE of 0.657. The RMSEs of the remaining FaMIDAS models varies between 2.673 and 6.747.

We also perform out-of-sample nowcasts. For 2019Q4, the MF-VAR yields the lowest RMSE followed by model 4 and 6 estimated with U-MIDAS. In the class of FaMIDAS estimated with PDL, the best predictive model is 6. The two best predictive models for nowcasting 2020Q1 is model 3 and 4 estimated with U-MIDAS. The best predictive model for nowcasting 2020Q2 is model 4 estimated with U-MIDAS followed by model 6 estimated with PDL. The FaMIDAS forecasts are highly biased, since the RMSEs varies between 3.622 and 6.587. The available short time-series did not work well for utilizing the MF-VAR type models, yielding unreliable results for 2020Q2.

To hedge against the inaccurate nowcasts, Timmermann (2006), Kourentzes and Petropoulos (2016) and Atiya (2020) suggest combining the different individual models as a weighted sum of their nowcast to reduce the remaining biases and variances in the nowcast. Based upon this suggestion, we combine the single selected nowcasts into one nowcast by deriving the mean value. The best average nowcast of 2019Q4 from in-sample model estimation comes from MF-VAR models 0.36% vs. GDP growth 2019, 0.30%. The best average nowcast for 2020Q1 is from FaMIDAS estimated with PDL, 0.12% vs. 0.0% economic activity growth of March 2020.

The best individual nowcast of 2020Q2 is from model 3 estimated with PDL, -6.64% vs. -7.30% of June 2020. The best average nowcast is from U-MIDAS -3.74% vs. -7.30%.

For the out-of-sample estimation, the best average nowcast for 2019Q4 is from U-MIDAS 0.28% while for 2020Q1, it is the MF-VAR type models with an average nowcast of 0.03%. For the nowcast of 2020Q2, the FaMIDAS estimated with PDL function delivers the best average figure -3.45%.

Table 1 In-Sample Model Estimations

FaMIDAS	**β**	**ρ(-1)**	**Adj-R^2**	**AIC**	**BIC**
1	-0.118	0.756	0.750	3.699	3.937
2	-0.056	0.091	0.879	3.047	3.427
3	-0.164	-0.061	0.909	2.810	3.333
4	-0.234	0.894	0.929	2.366	2.889
5	-0.183	-0.270	0.989	0.477	1.286
6	-0.147	0.704	0.754	3.681	3.918
7	-0.095	0.097	0.870	3.121	3.502
8	-0.265	0.581	0.971	1.465	1.989
9	-0.300	0.502	0.977	1.233	1.899
10	-0.498	-0.097	0.964	1.875	2.684
11	-0.160	0.954	0.876	2.865	3.246
12	-0.215	0.901	0.949	2.058	2.724
13	-0.184	0.843	0.927	2.278	3.230
14	-0.471	0.829	0.917	2.116	3.181
15	-0.269	0.901	0.903	2.419	3.466
16	-0.232	0.970	0.887	2.770	3.151
17	-0.136	0.919	0.952	1.986	2.652
18	-0.100	0.778	0.946	1.975	2.927
19	-0.091	0.713	0.952	1.934	2.790
20	0.077	0.910	0.970	1.247	2.294
MF-VAR	-0.138	0.633	0.960	1.554	2.562
MF-BVAR	-0.269	0.624	0.900	NA	NA

Note: the models (1) - (5) are estimated with PDL function. The models (6) - (10) estimated with the U-MIDAS function. The explanatory variables of all these models are without lags. Models (11) - (15) are PDL and (16) - (20) U-MIDAS. The explanatory variables of these models are estimated with 2 months lag. Both approaches allow for a wide variety model estimations from which we select well-fitted models for nowcasting.

Table 2 Out-of-Sample Model Estimations

FaMIDAS	β	ρ(-1)	Adj-R2	AIC	BIC
1	-0.142	0.997	0.880	2.797	3.037
2	-0.153	0.856	0.955	1.886	2.270
3	-0.183	0.961	0.961	1.790	2.318
4	-0.207	0.450	0.975	1.391	2.063
5	-0.034	0.503	0.991	0.293	1.109
6	-0.168	0.982	0.891	2.703	2.943
7	-0.164	0.737	0.966	1.630	2.014
8	-0.314	0.601	0.973	1.424	1.952
9	-0.357	0.509	0.978	1.249	1.921
10	0.408	0.633	0.996	-0.544	0.271
11	-0.126	0.939	0.876	2.911	3.294
12	-0.238	0.904	0.947	2.122	2.794
13	-0.147	0.818	0.921	2.354	3.314
14	-0.125	0.787	0.932	2.305	3.169
15	-0.192	0.890	0.896	2.433	3.489
16	-0.228	0.969	0.886	2.828	3.212
17	-0.144	0.920	0.950	2.058	2.730
18	-0.157	0.719	0.945	1.982	2.942
19	0.040	0.716	0.955	1.898	2.762
20	0.240	0.855	0.977	0.913	1.969
MF-VAR	-0.141	0.630	0.953	1.652	2.668
MF-BVAR	-0.156	0.845	0.922	NA	NA

Authors’ estimates

Table 3 In-Sample Evaluation

	Nowcast			RMSE		
	2019Q4	2020Q1	2020Q2	2019Q4	2020Q1	2020Q2
Real growth	**0.30**	**0.00**	**-7.30**			
MIDAS-PDL						
Model 2	-0.87	-0.69	-4.63	1.171	0.690	2.673
Model 3	1.25	0.78	-6.64	0.953	0.784	0.657
Model 5	0.34	0.18	-0.55	0.039	0.184	6.747
Model 8	-0.18	0.04	-2.34	0.482	0.044	4.962
Model 9	0.24	0.28	-3.39	0.064	0.284	5.388
Average	0.16	0.12	-3.51	0.145	0.064	3.790
Median	0.24	0.18	-3.39	0.064	0.044	3.913
U-MIDAS						
Model 1	-0.87	0.01	-4.63	1.171	0.011	2.673
Model 3	0.01	-1.11	-4.32	0.289	1.112	2.980
Model 5	0.28	-1.57	-3.78	0.019	1.572	3.519
Model 6	-0.38	-0.62	-3.39	0.677	0.616	3.913
Model 7	-0.46	0.07	-2.57	0.757	0.067	4.731
Average	-0.28	-0.64	-3.74	0.583	0.523	3.563
Median	-0.38	-0.62	-3.78	0.68	0.616	3.519
VAR						
MF-VAR	0.46	-0.07	NA	0.156	0.074	NA
MF-BVAR	0.27	-0.45	NA	0.028	0.452	NA
Average	0.36	-0.26	NA	0.064	0.263	NA

Authors’ estimates

Table 4 Out-of-Sample Evaluation

	Nowcast			RMSE		
	2019Q4	2020Q1	2020Q2	2019Q4	2020Q1	2020Q2
Real growth	**0.30**	**0.00**	**-7.30**			
MIDAS-PDL						
Model 1	1.18	0.21	-2.03	0.883	0.205	5.268
Model 6	0.62	0.52	-2.59	0.317	0.517	4.714
Model 7	-0.21	-0.11	-1.72	0.513	0.106	5.580
Model 8	-0.77	-0.90	-3.25	1.073	0.897	4.051
Model 9	-0.22	-0.25	-2.66	0.515	0.252	4.638
Average	0.12	-0.11	-2.45	0.180	0.107	4.850
Median	-0.21	-0.11	-2.59	0.513	0.106	4.714
U-MIDAS						
Model 1	0.93	1.03	-3.12	0.627	1.069	4.181
Model 3	0.80	-0.03	-3.68	0.497	0.030	3.622
Model 4	0.37	0.05	-0.71	0.069	0.048	6.587
Model 6	0.37	0.11	-2.69	0.075	0.110	4.607
Model 8	-1.08	1.03	-0.79	1.376	1.026	6.514
Average	0.28	0.44	-2.20	0.044	0.055	5.102
Median	0.37	0.11	-2.69	0.075	0.110	4.607
VAR						
MF-VAR	0.27	0.34	NA	0.027	0.343	NA
MF-BVAR	0.44	-0.28	NA	0.145	0.281	NA
Average	0.36	0.03	NA	0.059	0.031	NA

Authors’ estimates

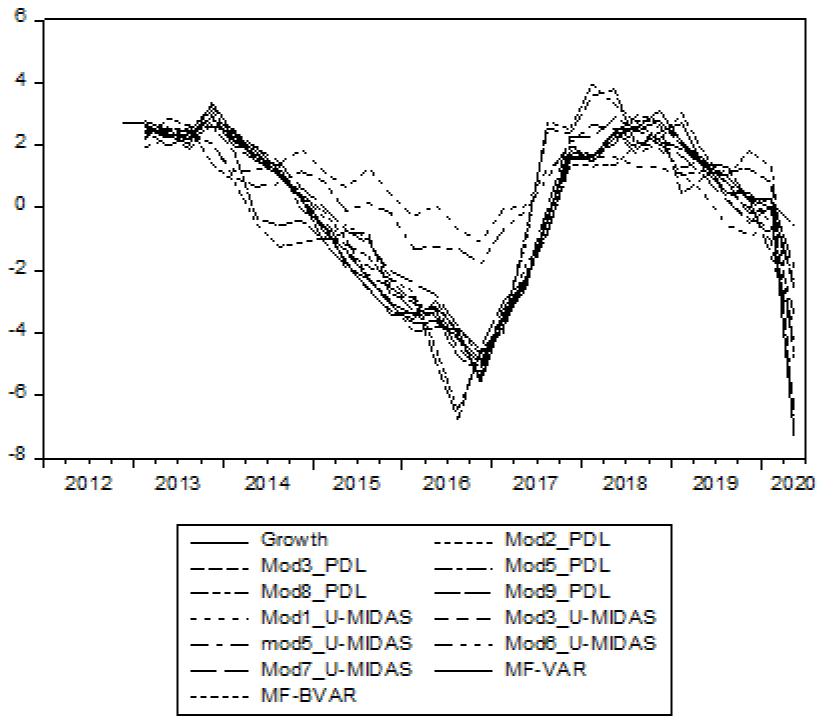

Figure 3. In-sample nowcasts

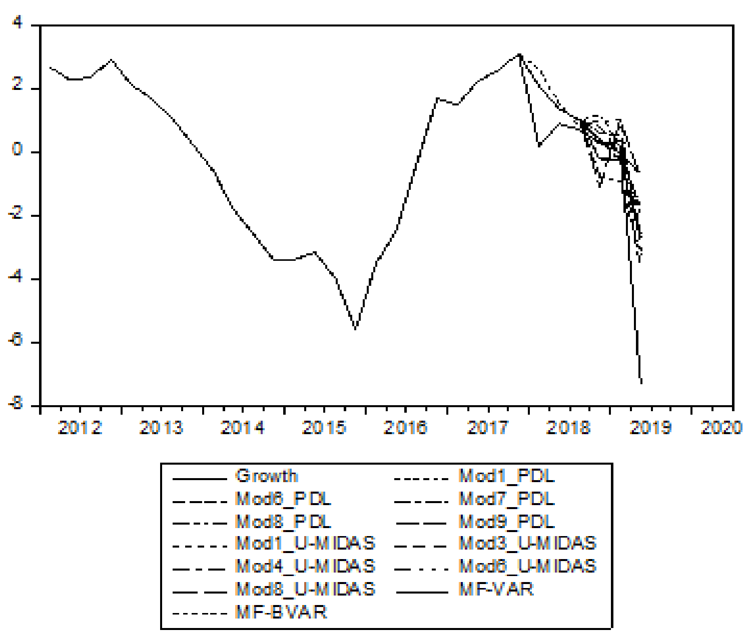

Figure 4. Out-of-sample nowcasts

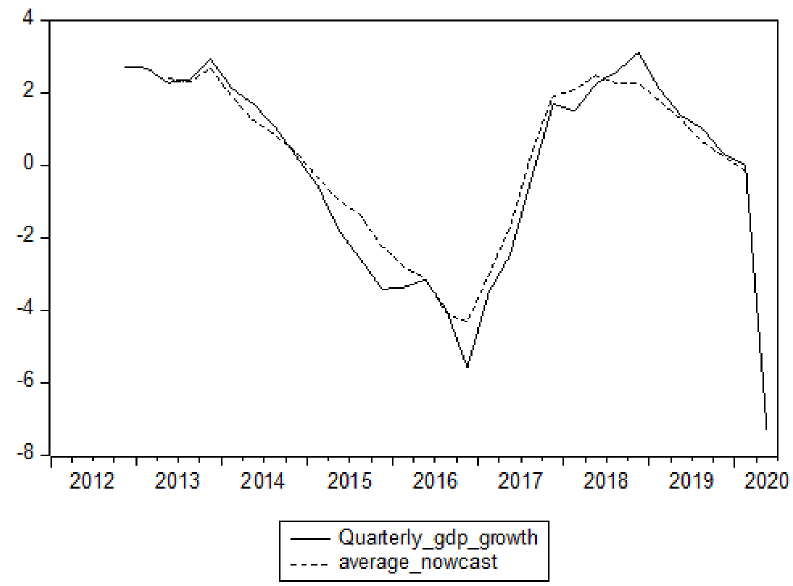

Figure 5. In-sample average nowcasts

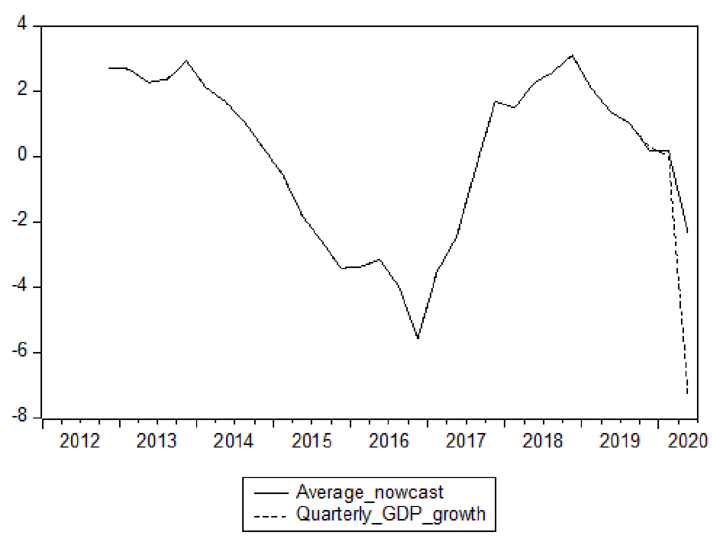

Figure 6. Out-of-sample average nowcasts

8. Concluding Remarks

This paper employs mixed frequency data models to nowcast GDP growth in real time for Suriname. For this purpose, we have full access to a set of 44 microeconomic Laspeyres indices constructed with ongoing monthly surveys from 2011. The sample consists of key enterprises in the whole economy. These enterprises are considered to be trend-setters in their respective sectors. We apply PCA to extract principal components from these indices to regress those components on quarterly GDP growth. For this purpose, we employ FaMIDAS, MF-VAR and MF-BVAR models. Different sample estimations are performed with these models to evaluate the nowcast precision against naïve benchmarks such as GDP growth of 2019Q4 and the volume growth of the monthly GDP indicator of March 2020 and June 2020.

The models provide mixed results. For the in-sample estimation, both the FaMIDAS and the MF-VAR models deliver significantly lower RMSEs. However, when forecasting 3 horizons further up to 2020Q2, only one model in the class of FaMIDAS estimated with PDL function delivers a lower RMSE. Consequently, we perform out-of-sample nowcast to further quantify our judgement. For one horizon (2019Q4) the MF-VAR models yield lowest RMSEs followed by U-MIDAS estimation. For two forecast horizons (2019Q4 - 2020Q1) the U-MIDAS estimation yields better results. For three forecast horizons (2019Q4 - 2020Q2) all the FaMIDAS regressions yield high RMSEs. We resort to model combination to reduce high biases and variances in the individual nowcasts. Well-fitted models are selected based on higher Adjusted R-squared values, lowest AIC and BIC and lowest RMSEs. We derive an adequate average nowcast from these models, which is acceptable.

Our analysis shows that the FaMIDAS, MF-VAR and MF-BVAR are important complementary models to the currently used macroeconomic forecast model of the Central Bank of Suriname (CBMOD4). This macro model draws heavily on historical trends and expert judgement for forecasting. As demonstrated, the mixed frequency data models are very useful to nowcast two quarters ahead. Adding one more quarter increases the RMSEs of selected models.

A limitation of modelling in this paper is the short length of the monthly explanatory variables (January 2011 to June 2020). As more survey data become available the nowcast accuracy will increase, especially with the MF-VAR and the MF-BVAR.

References

Aït-Sahalia, Y and Xiu, D (2015), Principal Component Analysis of High Frequency Data. Princeton University and the University of Chicago.

Aït-Sahalia, Y and Xiu, D (2017), Using principal component analysis to estimate a high dimensional factor model with high-frequency data. Journal of Econometrics, 2, 384 – 399.

Andrawis R, Atiya AF, El-Shishiny H (2011), Combination of long-term and short-term forecasts with application to tourism demand forecasting. International Journal of Forecasting, 27, 870-886.

António R, Maximiano P, Francisco CD (2009), Dynamic factor models with jagged edge panel data: Taking on board the dynamics of the idiosyncratic components, Working Papers, w200913, Banco de Portugal.

Atiya, AF (2020), Why does forecast combination work so well? International Journal of Forecasting, 36, 197-200.

Banerjee, A and Marcellino, M (2006), Are there any reliable leading indicators for US inflation and GDP growth? International Journal of Forecasting, 22, 137-151.

Bhaghoe, S and Eckhorst, K (2019), Constructing a monthly GDP indicator for Suriname. Journal of Economics Library, 6(4), 310 - 323.

Bhaghoe, S, Ooft, G, Franses, Ph.H.B.F (2019), Estimates of quarterly GDP growth using MIDAS regressions. Econometric Institute Research Papers. Retrieved from http://hdl.handle.net/1765/118667.

Bernanke, BS, Boivin, J, Eliasz, P (2005), Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. The Quarterly Journal of Economics, 120, 387- 422.

Boivin, J and Ng, S (2004), Are More Data Always Better for Factor Analysis? Forthcoming in Journal of Econometrics. Brillinger, David R. 1981. Time Series: Data Analysis and Theory. San Francisco: Wiley.

Boivin, J and Ng, S (2005), Understanding and comparing factor-based forecasts. International Journal of Central Banking, 3, 117 - 151.

Bouwman KE and Jan PAMJ (2011), Forecasting with real-time macroeconomic data: The ragged edge problem and revisions. Journal of Macroeconomics, 4, 784 - 792.

Chiu, CW, Eraker, B, Foerster, AT, Kim, TB, Seoane, HD (2012). Estimating VAR's sampled at mixed or irregular spaced frequencies: A Bayesian approach. The Federal Reserve Bank of Kansas City, Working Paper No. 11-11.

Chong, YY and Hendry, DF (1986), Econometric evaluation of linear macro-economic models. Review of Economic Studies, 53, 671–690.

Clemen, RT (1989), Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5, 559–581.

Clements, MP and Galvão, AB (2007), Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US Output Growth and Inflation. Queen Mary, University of London, Working Paper No. 616.

Diebold, FX (1989), Forecast combination and encompassing: Reconciling two divergent literatures. International Journal of Forecasting, 5, 589–592.

Eickmeier, S and Ziegler, C (2008), How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. Journal of Forecasting, 27, 237-265.

Franta, M, Havrlant, D, Rusnák, M (2016), Forecasting Czech GDP Using Mixed-Frequency Data Models. Journal of Business Cycle Research, 12, 165–185.

Frale, C and Monteforte, L (2009), FaMIDAS: A Mixed Frequency Factor Model with MIDAS Structure. Luiss Lab of European Economics Working Document no. 84.

Forni M, Hallin M, Lippi M, Reichlin L (2005), The Generalized Dynamic Factor Model: One-sided estimation and forecasting. Journal of the American Statistical Association 100, 830-840.

Giannone D, Reichlin L, Small D (2005), Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases. Finance and Economics Discussion Series 42.

Giannone, D, Reichlin, L, Small D (2008), Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases. Journal of Monetary Economics, 55, 665–676.

Ghysels E (2012), Mixed Frequency Vector Autoregressive Models. Department of Economics, University of North Carolina.

Ghysels E, Santa-Clara P, Valkanov R (2002), The MIDAS Touch: Mixed Data Sampling Regression Models. UNC and UCLA Working Papers.

Ghysels E, Santa-Clara P, Valkanov R (2004), The MIDAS touch: Mixed Data Sampling Regression Models, mimeo.

Ghysels E and Valkanov R (2006), Linear Time Series Processes with Mixed Data Sampling and MIDAS Regression Models, mimeo.

Hotelling, H (1933), Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417- 441.

Kourentzes, N and Petropoulos, F (2016), Forecasting with multivariate temporal aggregation: The case of promotional modelling. International Journal of Production Economics, 18, 145-153.

Kuzin V, Marcellino M, Schumacher C (2011), MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. International Journal of Forecasting, 27, 529‑542.

Kuzin V, Marcellino M, and Schumacher C (2009), Midas Versus Mixed-Frequency VAR: Nowcasting GDP in the Euro Area (2009). Bundesbank Series 1 Discussion Paper, available at SSRN: https://ssrn.com/abstract=2785336.

Makridakis, S and Hibon, M (2000), The M3-competition: Results, conclusions and implications. International Journal of Forecasting, 16, 451–476.

Mariano, RS and Murasawa, Y (2003), A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics, 18, 427–443.

Mariano, RS and Murasawa, Y (2010), A coincident index, common factors, and monthly real GDP. Oxford Bulletin of Economics and Statistics, 72(1), 27–46.

Marcellino, M (2004), Forecast pooling for short time series of macroeconomic variables. Oxford Bulletin of Economic and Statistics, 66, 91–112.

Marcellino, M and Schumacher, C (2007), MIDAS and Dynamic Linear Models, mimeo.

Marcellino, M and Schumacher, C (2008), Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP. CEPR Discussion Papers 6708.

Pesaran, MH and Timmermann, A (2005), Selection of estimation window in the presence of breaks. Mimeo, Cambridge University and University of California, San Diego.

Schorfheide F and Song D (2020), Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic. Federal Reserve Bank of Philadelphia Working paper 20-26.

Scott Jr and John T (1966), Factor analysis and regression. Econometrica: Journal of the Econometric Society, 552-562.

Sinharay, S (2010), An Overview of Statistics in Education. In: Peterson, P., et al., Eds., International Encyclopedia of Education, 3rd Edition, Elsevier Ltd., Amsterdam, 1-11. https://doi.org/10.1016/B978-0-08-044894-7.01719-X.

Stock, JH and Watson, M (1998), Diffusion Indexes, NBER Working paper 6702.

Stock, JH and Watson, M (1999), Forecasting Inflation. Journal of Monetary Economics, 44, 293 - 335.

Stock, JH and Watson, M (2001), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R.F., White, H. (Eds.), Festschrift in Honour of Clive Granger. Cambridge University Press, Cambridge, 1- 44.

Stock, JH and Watson, M (2004), Combination forecasts of output growth in a seven-country dataset. Journal of Forecasting, 23, 405 - 430.

Timmermann, A (2006), Forecast combinations. Elliott G., Granger C.W.J., Timmermann A. (Eds.), Handbook of economic forecasting, 135-196.

Yang, Y (2004), Combining forecasts procedures: Some theoretical results. Econometric Theory 20, 176 - 190.

Appendix 1

Polynomial Distributed Lag parametrization

The PDL parametrization restricts regression lag coefficients estimated as a p dimensional lag polynomial. The number of estimated coefficients, $\gamma_1, \gamma_2 \dots \gamma_p$, depends on the polynomial order (p) and not on the number of lags (j) as in other class of MIDAS regressions. The term k represents the number of lags:

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

Unrestricted MIDAS parametrization

U-MIDAS parameterizes the lag polynomials in a parsimonious fashion. The number of estimated coefficients $\gamma_1, \gamma_2 \dots \gamma_k$ are not subjected to restrictions, as is the case in other competing models.

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

Principal Component Analysis

The following authors have significantly contributed to PCA as we know it today: Hotelling (1933), Sinharay (2010) and Aït-Sahalia and Xiu (2015, 2017). In general, PCA estimates covariance matrices from large datasets. These covariance matrices explain a large portion of the variances, which are very useful for modelling. We apply PCA to extract k principle factors from the dataset.

$$y_{ij} = z_{ik}\beta_{1j} + z_{i2}\beta_{2j} + z_{i3}\beta_{3j} + \varepsilon_{ij}$$

where y_{ij} is the value of the ith observation on the jth variable, z_{ik} is the ith observation on the kth common factor, β_{1j} is the set of factor loadings and ε_{ij} is the residual term.

Table 5 Indicators

Industry	Type of Indicator	Number of Indicators
Agriculture	Production	7
Mining	Production	2
Manufacturing	Production and real turnover	6
Utilities	Production	2
Construction	Production and real turnover	4
Banking	Financial data	5
Hotel & Restaurants	Overnight stays and real turnover	2
Transport, Storage & Communication	Volumes and turnover	4
Wholesale & Retail Trade	Real turnover and imports	10
Government	Employees and expenditures	2

Source: Central Bank of Suriname

Table 6 Variance and Proportion of Extracted Factors

Factor	Variance	Cumulative	Proportion	Cumulative
F1	6.854	6.854	0.279	0.279
F2	4.185	11.039	0.171	0.450
F3	3.265	14.304	0.133	0.583
F4	2.126	16.430	0.087	0.670
F5	1.898	18.328	0.077	0.747
F6	1.871	20.199	0.076	0.823
F7	1.840	22.038	0.075	0.898
F8	1.265	23.303	0.052	0.950
F9	1.229	24.532	0.050	1.000

Estimated with software package EViews 11

Table 7 Goodness-of-fit Summary

Goodness-of-fit Summary			
	Model	**Independence**	**Saturated**
Parameters	394.000	43.000	946.000
Degrees-of-freedom	552.000	903.000	---
Parsimony ratio	0.611	1.000	---
Absolute Fit Indices			
Discrepancy	10.364	34.320	0.000
Chi-square statistic	1046.795	3466.287	---
Chi-square probability	0.000	0.000	---
Bartlett chi-square statistic	827.417	2945.772	---
Bartlett probability	0.000	0.000	---
Root mean sq. resid. (RMSR)	0.048	0.233	0.000
Akaike criterion	-0.561	16.277	0.000
Schwarz criterion	-14.767	-6.961	0.000
Hannan-Quinn criterion	-6.313	6.867	0.000
Expected cross-validation (ECVI)	18.166	35.171	18.733
Generalized fit index (GFI)	0.703	0.306	1.000
Adjusted GFI	0.491	-0.190	---
Non-centrality parameter	494.795	2563.287	---
Gamma Hat	0.479	0.151	---
McDonald Noncentralilty	0.086	0.000	---
Root MSE approximation	0.094	0.168	---
Incremental Fit Indices			
Bollen Relative (RFI)	0.506		
Bentler-Bonnet Normed (NFI)	0.698		
Tucker-Lewis Non-Normed (NNFI)	0.684		
Bollen Incremental (IFI)	0.830		
Bentler Comparative (CFI)	0.807		

Note: the parsimony ratio and the measures reported in the sections of the absolute and incremental fit indices indicate that the model fits the data adequately.

Evaluation Statistics

Root Mean Squared Error (RMSE):

$$\sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

where $\widehat{y_t}$ at time t is the nowcast and y_t is the actual observation.

Adjusted R-squared value:

$$1 - \frac{(1 - R^2)(N - 1)}{N - P - 1}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where R^2 quantifies the variance of y_i, p is the number of predictors and N is sample size.

Akaike Information Criterion (AIC):

$$-2\log(L) + 2k$$

Log-likelihood function (L):

$$L = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta x_i)^2$$

where k is the number of parameter estimates and L the maximum likelihood function of the model.

Bayesian Information Criterion:

$$2\log(L) + k\log(N)$$