Policy Research Working Paper 9247

# Nowcasting Economic Activity in Times of COVID-19

## An Approximation from the Google Community Mobility Report

*James Sampi*
*Charl Jooste*

World Bank Group
Macroeconomics, Trade and Investment Global Practice
May 2020

## Abstract

This paper proposes a leading indicator, the "Google Mobility Index," for nowcasting monthly industrial production growth rates in selected economies in Latin America and the Caribbean. The index is constructed using the Google COVID-19 Community Mobility Report database via a Kalman filter. The Google database is publicly available starting from February 15, 2020. The paper uses a backcasting methodology to increase the historical number of observations and then augments a lag of one week in the mobility data with other high-frequency data (air quality) over January 1, 2019 to April 30, 2020. Finally, mixed data sampling regression is implemented for nowcasting industrial production growth rates. The Google Mobility Index is a good predictor of industrial production. The results suggest a significant decline in output of between 5 and 7 percent for March and April, respectively, while indicating a trough in output in mid-April.


This paper is a product of the Macroeconomics, Trade and Investment Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at jsampibravo@worldbank.org and cjooste@worldbank.org.

# Nowcasting Economic Activity in Times of COVID-19: An Approximation from the Google Community Mobility Report

**James Sampi and Charl Jooste***

*We are extremely grateful to Jorge Araujo for his invaluable support in this project. We thank to Pablo Saavedra and Stefano Curto for valuable advice, and Julio Velasco, Barbara Cunha, Marek Hanusch, Gabriel Zaourak and Luigi Butron for helpful discussions.
Corresponding authors: jsampibravo@worldbank.org and cjooste@worldbank.org

# 1 Introduction

The economic impact of COVID-19 has been severe. The combined impact of government policies with the health implications is leading to a sharp contraction in economic activity. The extent of output losses is yet to be determined. Output losses will vary by country, the rate of infection, and the extent of policy interventions coupled with behavioral responses.

Economic forecasts for 2020 will have to be conditioned on the effects of COVID-19. This is not an easy task - the latest forecasts of the IMF (World Economic Outlook April 2020) and World Bank (Macro Poverty Outlook April 2020) show significant variations in growth outcomes within and across regions compared to the 2020 outlooks prepared in October 2019 (see Figure 1). Since we have only surpassed the first quarter of 2020 at the time of writing, the annual forecast estimates remain very uncertain.

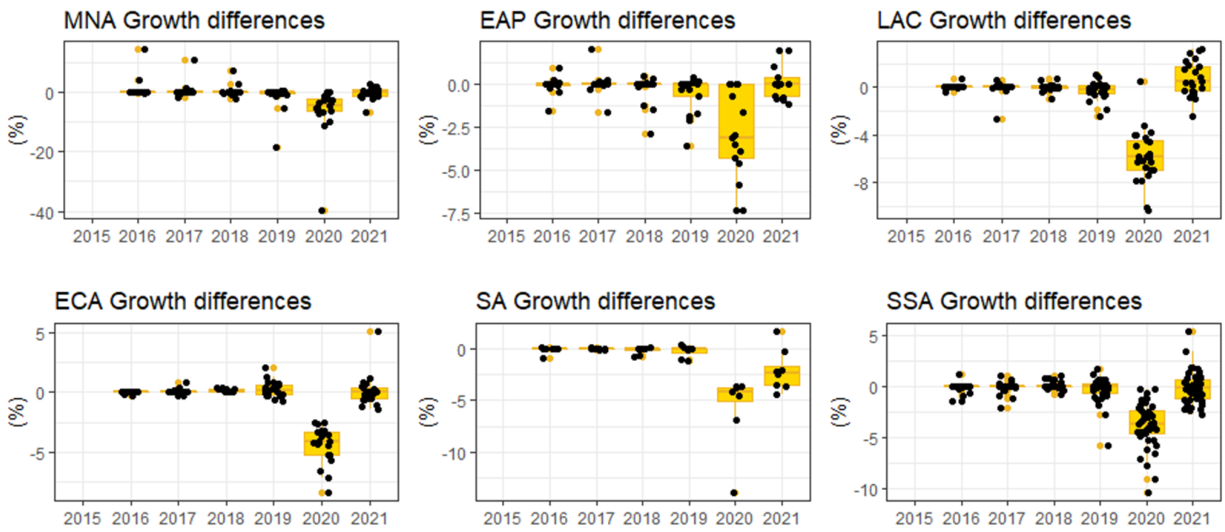


Figure 1: WB growth revisions by region. Note: MNA: Middle East and North Africa; EAP: East Asia and the Pacific; LAC: Latin America and the Caribbean; ECA: Europe and Central Asia; SA: South Asia and SSA: Sub-Saharan Africa.

To reduce some of the uncertainty, we utilize high-frequency data that proxy the COVID-19 economic activity responses. The Google mobility data summarize by country various mobility trends (e.g. Retail and recreation, Grocery and pharmacy, Parks, among others). From Figure 2, it is quite clear that global mobility indicators have slowed. We use these data to test the correlation with industrial production . Analysts can use the change in industrial production to back out estimates for annual GDP growth.

Unfortunately, some of these data go back only to February 15, 2020. To increase the degrees of freedom in the analysis, we backcast the mobility data using daily weather and pollution data. The assumption is that pleasant weather and low pollution are correlated with an increase in mobility.

The rest of the paper is structured as follows: In Section 2 we describe recent nowcasting literature. The overall methodology is discussed in Section 3, which is followed in Section 4 by a discussion of estimating and approximating the models via a Kalman filter. In Section 4.2 we describe the approximation done using pollution data, while Section 4.3 discusses the links to industrial production. In Section 5 we present the results and Section 6 concludes.

## 2 Literature review

The seminal papers of Geweke (1989), Stock and Watson (1989), and Bai and Ng (2002) have placed the dynamic factor model (DFM) as the predominant framework for research on macroeconomic forecasting using high-frequency indicators. Overall, this framework allows us to study large panels of time series through a few common factors, especially, when the data series are strongly collinear.

The available methodologies for estimating DFMs can be divided into two groups. The first group of estimators entails nonparametric estimation with large N using cross-sectional averaging methods, primarily principal components. Principal components analysis (PCA) is the most popular factor extraction method in the treatment of dynamic factors models. PCA is appealing because of its computational advantages and asymptotic properties in large data sets, see Bai (2003). Unfortunately, for many empirical applications the PCA assumptions are arguably not realistic, see Onatski (2012).

The second group consists of parametric models estimated in the time domain using maximum likelihood estimation (MLE) and the Kalman filter. MLE has been used successfully to estimate the parameters of low-dimensional DFMs. However, there are significant computational requirements to maximize the likelihood function with many parameters. In order to deal with the dimensionality problem associated with the likelihood function, further estimators have been implemented. The main idea behind these methods is to use the consistent parameters estimated by the first group methods for computing the factors required by the second one, see Doz and Reichlin (2011) and Doz, Giannone and Reichlin (2012).

Regardless of the method for extracting a common factor, increasingly the literature is suggesting mixing sampling frequencies aimed at improving the accuracy of nowcasting techniques. The challenges of mixed data frequency are reviewed in the context of econometric analysis by Ghysels and Marcellino (2016) and discussed in the context of forecasting by Armesto, Engemann and Owyang (2010) and Andreou, Ghysels and Kourtellos (2010). A widely used method for incorporating high-frequency data to produce forecasts of low-frequency variables is the Mixed Data Sampling (MIDAS) method of Ghysels, Santa-Clara and Valkanov (2004).

MIDAS is a regression-based method that transforms the high-frequency variables into low-frequency indicators via a weighting scheme. The weights reflect the relative importance of recent observations as opposed to older ones as information to predict future values of the low-frequency variable.

In this paper we compress the six Google mobility indicators: Retail & recreation, Grocery & pharmacy, Parks, Transit stations, Workplaces, and Residential into one common factor to capture the economic effects of COVID-19 in Latin America and the Caribbean (LAC) economies. In this exercise the dimensionality of variables is not a big concern, therefore, the parametric methods embedded in the second group are adequate. Meanwhile, we select the MIDAS approach for nowcasting the industrial production growth rate, which performs significantly better when using DFM compared to the PCA methods, see Gorgi, Koopman and Mengheng (2018).

# 3 Approximate factor model

Let $y_{it}$ be the observed data for the $i$th variable at time $t$. In total we have $N$ variables indexed by $i = 1, \ldots, N$. Also, we have $T$ time periods and $t = 1, \ldots, T$. The approximate factor model decomposes $N$ dimensional vectors $y_t = (y_{1t}, \ldots, y_{Nt})'$, for $t = 1, \ldots, T$, as follows

$$y_t = \Lambda f_t + \varepsilon_t \tag{1}$$

where $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is the $N \times r$ matrix of factor loading with $r$ as the number of factors, $f_t = (f_{1t}, \ldots, f_{rt})'$ is the $r \times 1$ vector of factors and $\varepsilon_t$ is the $N \times 1$ idiosyncratic disturbance term.

In approximate factor settings, the consistency and asymptotic normality of the estimators when both $N$ and $T$ go to infinity have been recently shown by Bai (2003), Bai and Ng (2002) and Doz et al. (2012). In order to prove these properties, Bai (2003) makes a strong assumption related to the eigenvalues of the population covariance matrix of the data. Specifically, it requires that the ratio between the $r - th$ largest and the $r + 1 - th$ largest eigenvalues, $d_r$, increase proportionately to $N$. Asymptotically, this implies that the cumulative effects of the normalized factors strongly dominate the idiosyncratic disturbances.

Recently, Onatski (2012) and Onatski (2015) show that the strong factor assumption requires one of the following two scenarios. Either, an overwhelming domination of the factors represented by higher values of $d_r$ for all $r$, or $\varepsilon\varepsilon'/T$ needs to be close to the identity matrix, where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$ is the $N \times T$ disturbances matrix. It implies that all the commonalities across variables occur through the factors and that the individual elements of

$\varepsilon_t$ are purely shocks which are idiosyncratic to each variable. However, the former scenario is unwanted as long as we do not assume an overwhelming domination of factors over the idiosyncratic disturbances. The latter scenario does not hold as typically the expected covariance matrix of the disturbances is not the identity, $E(\varepsilon_t \varepsilon_t') = \Omega \neq I_N$.

Notice that the Google mobility information is composed of six indicators, which configures $N = 6$, and those became available from February 15, 2020, making the time dimension roughly $T = 60$. Notably, with short $N$ and $T$ it becomes difficult to assume that the strong factor assumption holds, and we would need to consider a more consistent approach besides the standard Principal Component Analysis (PCA).

# 4 Estimation procedure

This section provides a detailed explanation of the empirical procedure for estimating the leading factor, the "Google Mobility Index", and extending back the resultant index by using air quality-related information with the overall objective of nowcasting the effects of COVID-19 on the industrial production growth rates. Because $T$ and $N$ dimensions are small for consistency of Principal Component Analysis (PCA) or the standard Kalman Filter methods, the econometric procedure relies on the two-step approach introduced by Doz et al. (2012) or typically known as the "quasi-maximum likelihood approach ", where asymptotic properties perform significantly better for small $T, N$ when compared to standard methods. In addition, the construction of one single Google leading indicator requires that $r = 1$.

## 4.1 The two-step approach for estimating the leading factor

The first stage proceeds to obtain consistent estimates of the parameters $\Omega$ and $\Lambda$ for estimating the unobservable factor, $f_t$, using Maximum Likelihood approaches. Specifically, the first stage uses Principal Component Analysis (PCA), while the second stage involves the Kalman filter.

The first stage solves the following PCA optimization problem

$$V = \min_{\Lambda, f} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \lambda_i f_t)^2 \tag{2}$$

subject to the normalization of either $\Lambda'\Lambda/N = 1$ or $ff'/T = 1$. We use the notation $\lambda_i$ as the $i^{th}$ row of $\Lambda$ for $i = 1, \ldots, N$. The optimization problem is identical to maximizing $\text{tr}(f(y'y)f')$ where $y = (y_1, \ldots, y_N)'$ is the $N \times T$ matrix of the observed data. Here tr() denotes the trace operator. Let $Q$ be the largest eigenvalue of the sample covariance matrix $S = \frac{1}{T} \sum_{t=1}^{T} y_t y_t'$. The solution to the above minimization problem is not unique, even though the sum of squared residuals $V$ is unique, see Bai and Ng (2002). The estimated parameters of interest can be expressed as

$$\begin{aligned} \hat{\Lambda}^{PCA} &= PQ^{1/2} \\ \hat{\Omega} &= (y_t - PP'y_t)'(y_t - PP'y_t) \end{aligned} \tag{3}$$

where $P$, the eigenvector associated with $Q$.

For the second stage we need to make an assumption about the stochastic process of the factor, such that the model can be written in state space form. In particular, the factor is assumed to follow a vector auto-regressive model of order one. We have,

$$f_t = \alpha f_{t-1} + \eta_t \qquad \eta_t \sim IID(0, \sigma_\eta^2) \tag{4}$$

where $\alpha$ is the scalar transition parameter and $\eta_t$ is the $1 \times 1$ factor error term that has mean zero and variance $\sigma_\eta^2$. This specification can easily be extended to allow for higher order vector auto-regressions. Together with the observation of equation (1), the model can be viewed as a state space model.

The parametric MLE method is well documented in Durbin and Koopman (2012) and Ghahramani and Hinton (1996). The method relies on the Kalman filter. They start by defining the conditional moments as $a_{t|s} = E(f_t|y_1, \ldots, y_s; \psi^{MLE})$ and $P_{t|s} = Var(a_{t|s} -$

$f_t|y_1,\ldots,y_s;\psi^{MLE})$ for $t,s = 1,\ldots,T$, where $\psi^{MLE} = \{\alpha,\sigma^2_\eta\}$ contains the parameters that pertain to the distribution of the factor. Notice that $\Lambda$ and $\Omega$ are estimated in the first stage. Moreover, the initial factor has density $N(0,P_1)$ where $P_1 = \text{inv}(1-\alpha\alpha')$ and $\varepsilon_t \sim NID(0,\Omega)$ is the $N\times 1$ disturbance term.

The estimation of the parameter vector $\psi^{MLE}$ is based upon maximizing the log-likelihood function associated with (1) and (4). Meanwhile the estimated factor, $f_t^{MLE}$, is obtained through a recursive procedure. Specifically, the log-likelihood function associated to the Gaussian density is given by

$$\log L\Big(y;\psi^{MLE}\Big) = -\frac{NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\Big(\log|F_t| + \upsilon_t' F_t^{-1}\upsilon_t\Big) \tag{5}$$

where the quantities $\upsilon_t$ and $F_t$ represent the prediction residuals $(y_t - \Lambda a_{t|s})$ and the predicted variance $(\Lambda P_{t|s}\Lambda' + \Omega)$, which are evaluated by the Kalman filter.

### 4.2 Expanding the series using air quality information

In this section we propose a simple methodology for expanding the "Google Mobility Index" obtained in the previous section by using air quality-related information. Specifically, we use the Air Quality Open Data Platform for extracting temperature and fine particulate matter (PM2.5) information per city in each country worldwide. Then, the information is averaged per country such as it can be easily associated with the Google Mobility Index over time.

Let's consider $p_j$ and $q_j$ for $j = t-J,\ldots,t,\ldots,T$ the normalized temperature and PM2.5 information at time $j$, while $f_j^{MLE}$ for $j = t,\ldots,T$ is the Google Mobility Index. Notice that $p_j$ and $q_j$ contain $J$ more data points than $f_j^{MLE}$. Therefore, we recover backwards the information as follows

$$f_{j-1}^{MLE} = \frac{f_j^{MLE}}{1+\rho_1\times p_j + \rho_2\times q_j} \tag{6}$$

for all $j = t - J, \ldots, t$ and $\rho_{1,2}$ the weighted correlation coefficient between the Google Mobility Index and the normalized series.

### 4.3 Nowcasting industrial production

In this section we consider the Mixed Data Sampling (MIDAS) regression of Ghysels et al. (2004) for nowcasting industrial production growth rates (sourced from OECD Main Economic Indicators database), $x_t$. Industrial production is published on a monthly basis. $f_t^{(d),MLE}$ represents the daily "Google Mobility Index", which is observed $d$ days in a particular $M$ month. Specifically, we want to predict the variable $x_t$ onto a history of lagged observations of $f_{t-j}^{(d),MLE}$. The superscript $(d)$ denotes the higher frequency sampling and its exact timing lag is expressed as a fraction of the unit interval between months $M$ and $M-1$. The MIDAS regression is expressed as follows:

$$x_t = \beta_0 + \beta_1 B(L^{1/d}; \Theta) f_t^{d,MLE} + u_t^d \tag{7}$$

for $t = 1, \ldots, T$, and where $B(L^{1/d}; \Theta) = \sum_{k=0}^{K} B(k; \Theta) L^{k/d}$ and $L^{1/d}$ is a lag operator such that $L^{1/d} f_t^{d,MLE} = f_{t-1}^{d,MLE}$, and the lag coefficient in $B(k; \Theta)$ of the corresponding lag operator $L^{k/d}$ are parameterized as a function of a small-dimensional vector of parameters $\Theta$. In order of addressing the parameter proliferation, in a MIDAS regression the coefficients of the polynomial in $L^{1/d}$ are captured by a known function $B(L^{1/d}; \Theta)$ of a few parameters summarized in a vector $\Theta$, typically, polynomial specifications.

## 5 Results

Figure 2 presents the results of the two-step estimator for extracting one common factor of the six Google Mobility indicators, the "Google Mobility Index", for each Latin America and the Caribbean (LAC) country available in the Google database. The gray lines represent the non-smoothed indicators while the bold red line represents the smoothed Kalman filter

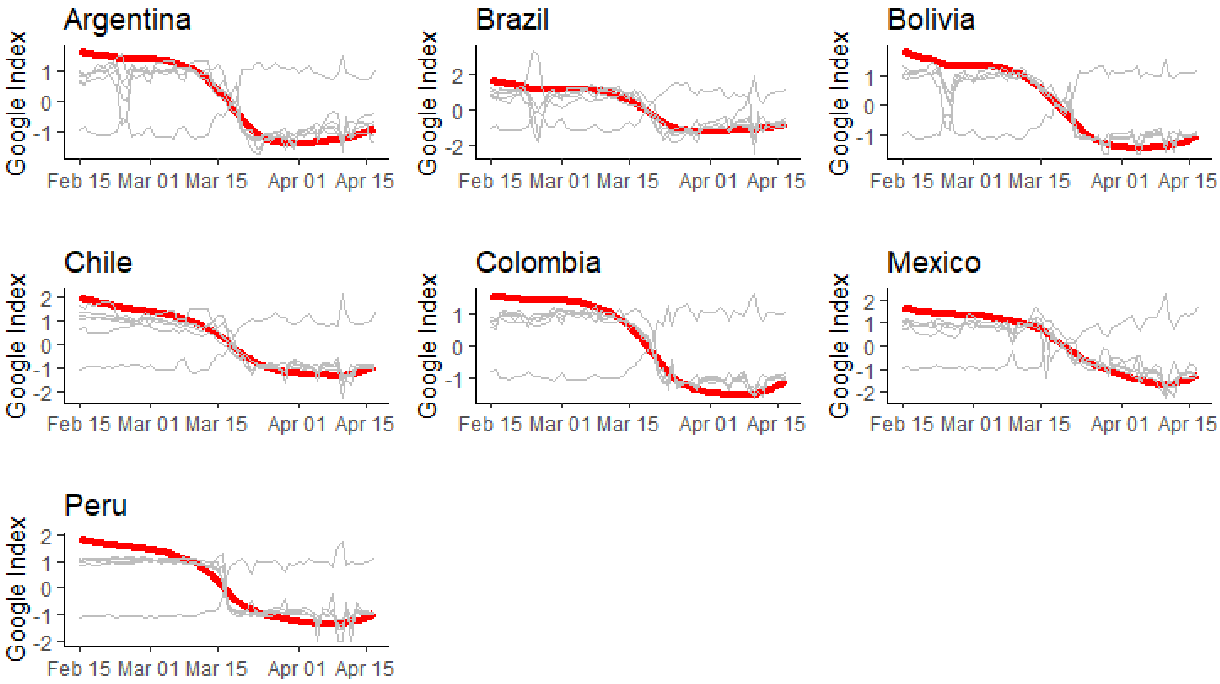


Figure 2: Google leading indicator for Latin American and the Caribbean(LAC) economies

estimate. As expected, in all countries the index declined significantly from mid-February. Interestingly, the bottom of the indicator is in early April with the index starting to recover to its baseline (which is a value reflected in February 2020). Notably, there are economies in which the decline is steeper than in others. As an example, the index suggests a stronger decline in Mexico compared to Brazil or Chile.

Figure 3 presents the correlation coefficients between pollution, measured as the fine particulate matter(PM2.5), and the average temperature per country with the estimated Google Mobility Index. In most cases, the correlation coefficient is greater than 0.2 in absolute terms. We found a negative correlation between pollution and Google Mobility Index in three countries, Brazil, Chile and Mexico. Temperature is positively correlated with the Google Mobility Index in all cases but Mexico. The rationale is as follows: with few people in the street, the average temperature should decline, while high pollution will prevent people from spending longer hours in the street. Obviously, there are caveats to

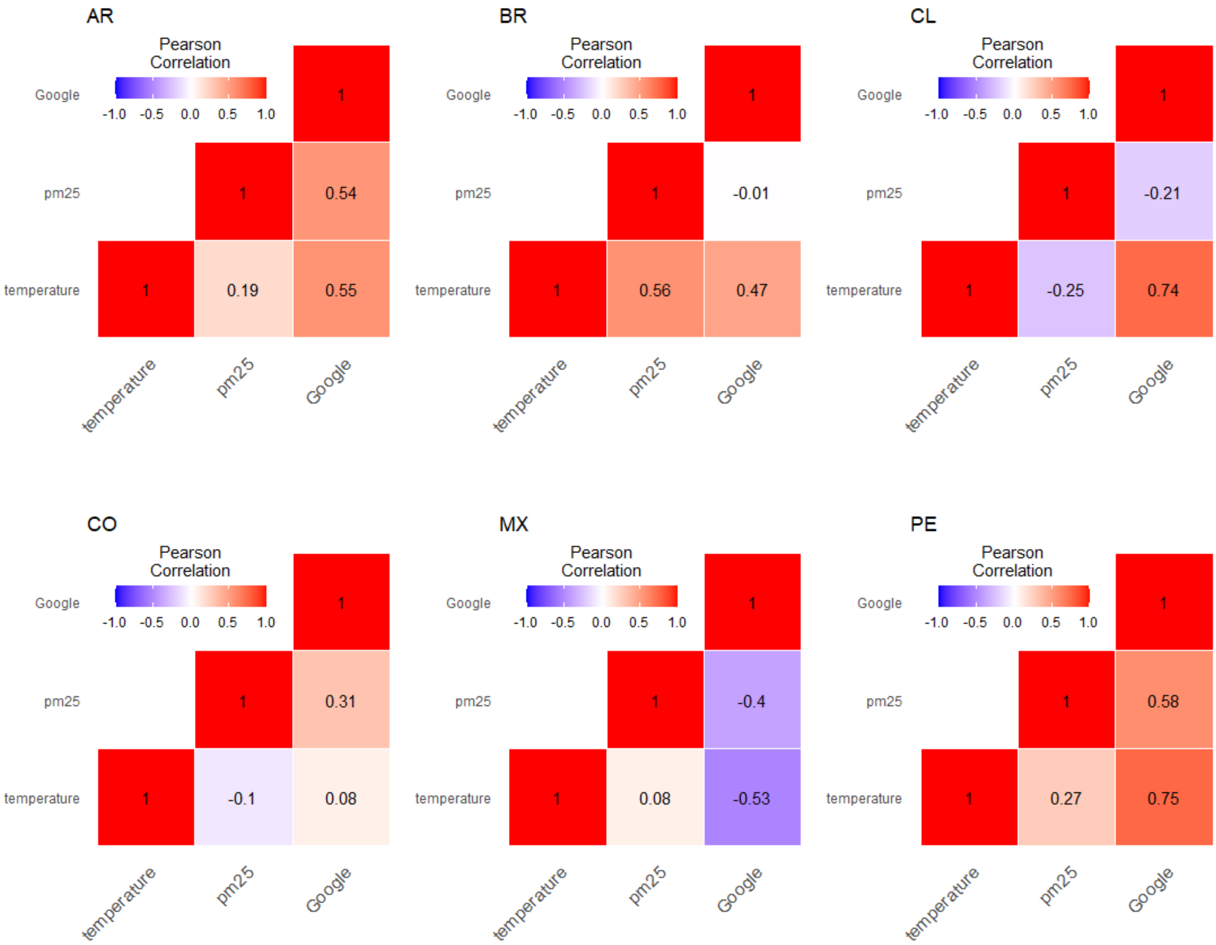


Figure 3: Google index correlations with air quality related information. Note: AR = Argentina, BR = Brazil, CL = Chile, CO = Colombia, MX = Mexico and PE = Peru

this anecdotal explanation - since the baseline matters - e.g. if pollution is persistent.

Figure 4 presents the results of extending the Google Mobility Index (which always lags by one week) with air quality-related information: temperature and fine particulate matter (PM2.5), for countries for which data are available. The information is gathered from the Air Quality Open Data Platform from January 1, 2019 to April 30, 2020 for Argentina, Brazil, Chile, Colombia, Mexico, Peru and El Salvador. The shaded areas represent the information estimated backward and forward using Equation 6. The extended information suggests that the period prior to the COVID-19 crisis was signaling a recovery in Argentina, and a significant decline in Chile and Mexico, although stronger in the former. Meanwhile, the index points to stability in Brazil, Peru and El Salvador. Appending air quality data to the mobility index is warranted on both statistical and economic grounds, with the latter being the main motivation for this analysis. The predicted value of the Google Mobility Index using more up-to-date information confirms that economies may have bottomed-out in April, with Mexico being the exception.

The final set of results, which nowcasts industrial production using the appended Google Mobility Index, is summarized in Table 1 for Brazil, Chile, Colombia and Mexico.[1] The monthly growth rates are gathered from the Economic Indicators database of the OECD. The $R-sq.$ achieves a maximum of 25 percent in Brazil and a minimum of 18 percent in Colombia. In most cases, the results point to a deterioration of March growth rates compared to February, and an even stronger decline in April. Specifically, Mexico is expected to decline by 5 and 6 percent for March and April, respectively, from a 0.7 percent decline in February. Similarly, Brazil is expected to decline by nearly 7 and 3 percent, while Chile is expected to decline by 1 and 2 percent; and Colombia 0.4 and 2 percent, respectively. In all regressions the optimal number of lags is 3, while the polynomial degree varies from 3

[1] In Table 3 in the Appendix section, various consistency checks with different combinations of the Google Index and the air quality data are compared. The differences between the various explanatory variables are insignificant.

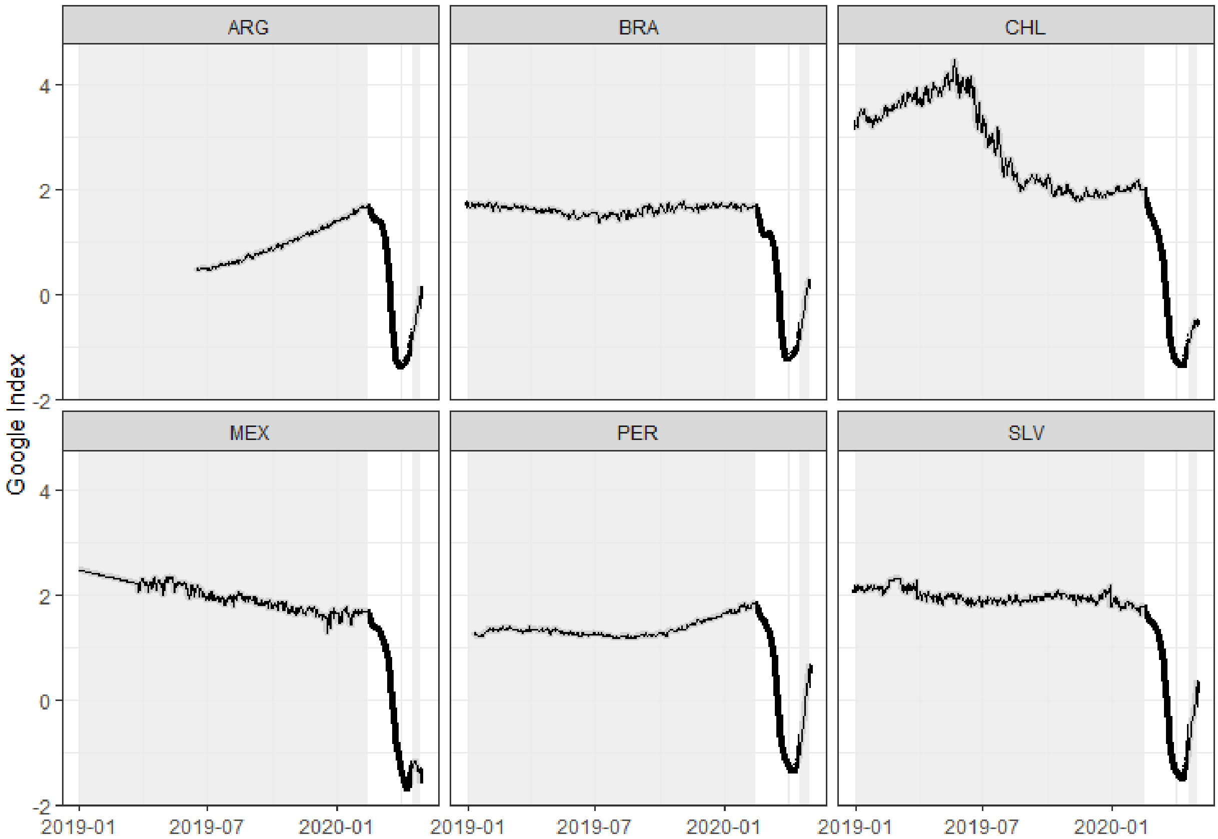


Figure 4: Google index expanded by the air quality related information. Note: The shaded area represents the information estimated backward and forward by using Equation 6

| Country | $R-sq.$ | Lag | Polynomial | Feb. Actual | Mar. proj. | Apr. proj. |
|---|---|---|---|---|---|---|
| Brazil | 0.25 | 3 | 3 | -0.009 | -0.069 | -0.034 |
| Chile | 0.21 | 3 | 3 | -0.018 | -0.010 | -0.020 |
| Colombia | 0.18 | 5 | 3 | -0.004 | -0.004 | -0.018 |
| Mexico | 0.20 | 5 | 3 | -0.007 | -0.047 | -0.059 |

Table 1. MIDAS results for nowcasting industrial production growth rate (m/m). The $R-sq.$ reflects the one-step ahead projection residuals for the period 2019-Feb until 2020-Feb, while Lag and Polynomial represent the optimal number of lags and polynomial degree in Equation 7.

to 5 in Colombia and Mexico. In addition, the one-step-ahead predicted values are plotted for each economy in Figures 5 to 8

## 5.1 Comparison with other methods

The predictive test contrasts the MIDAS approach to an autoregressive method of first and second orders in Table 2. The results reveal a significant forecast improvement when incorporating high-frequency information from Google indicators in all cases except Colombia. We use the root-mean-square error ($RMSE$) for model comparison. The $RMSE$ represents the quadratic mean of the differences between the one-step ahead predicted values and the observed data. Therefore, the lower the $RMSE$ the better model performance. Table 2 shows that $RMSE$ is almost three times higher when compared to $AR$ specifications in Brazil, 26 percent higher in the case of Chile, while 14 percent higher in case of Mexico. Overall, the results provide strong evidence in favor of using a MIDAS regression combining high-frequency indicators gathered from Google mobility in comparison to standard methods.

| Country | $AR(1)$ | $AR(2)$ | $MIDAS$ | $MIDAS$ with $AR(1)$ | $MIDAS$ with $AR(2)$ |
|---|---|---|---|---|---|
| Brazil | 0.041 | 0.038 | 0.009 | 0.009 | 0.011 |
| Chile | 0.024 | 0.024 | 0.022 | 0.019 | 0.019 |
| Colombia | 0.010 | 0.011 | 0.011 | 0.011 | 0.011 |
| Mexico | 0.009 | 0.008 | 0.007 | 0.007 | 0.007 |

Table 2. Comparison for predicting the industrial production growth rates (m/m) by using the Root-mean-square error ($RMSE$). Note: The data sample ranges from February 2018 until February 2020 for $AR$ regressions while January 2019 until February 2020 for MIDAS regression.

# 6 Conclusion

A novel database is used to generate high frequency forecasts of economic activity in the wake of COVID-19. The World Bank and IMF have revised growth estimates significantly downward during COVID-19. The health and economic policy responses and subsequent economic outcomes are very uncertain. To reduce some of this uncertainty this paper details the use of daily mobility and air quality data to predict movements in industrial production, which is typically used to assess within-year movement of GDP growth. The database includes Google's Community Mobility Report data, air quality data and OECD industrial production data.

Estimation proceeds in three steps: (i) lagged mobility data are patched with air quality data; (ii) the mobility data are then combined to extract a common Mobility Index via Kalman filtering; and finally (iii) a MIDAS approach nowcasts industrial production from the smoothed Mobility Index. The results can be updated daily. This paper illustrates its use for a set of Latin American countries.

The Mobility Index is compared to a standard auto-regressive forecast model. The results of the exercise suggest that our approach beats the AR models for pseudo out of sample forecasts. The index predicts a strong decline in industrial production monthly growth rates of 7 (5) and 4 (6) percent for March and April, respectively, for Brazil (Mexico). Chile and Colombia follow a similar decline. Finally, the index, while still negative, suggests that the trough in output occurred in April 2020.

## References

Andreou, E., Ghysels, E. and Kourtellos, A.: 2010, Regression Models with Mixed Sampling Frequencies, *Journal of Econometrics* **158**, 256–261.

Armesto, M., Engemann, K. and Owyang, M.: 2010, Forecasting with Mixed Frequencies, *Review* **92**.

Bai, J.: 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71**, 135–171.

Bai, J. and Ng, S.: 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70**, 191–221.

Doz, C., Giannone, D. and Reichlin, L.: 2012, A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models, *Review of Economics and Statistics* **94**, 1014–1024.

Doz, C. and Reichlin, L.: 2011, A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering, *Journal of Econometrics* **164**, 188–205.

Durbin, J. and Koopman, S. J.: 2012, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

Geweke, J. F.: 1989, Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica* **57**, 1317–1339.

Ghahramani, Z. and Hinton, G.: 1996, Parameter Estimation for Linear Dynamical Systems, *University of Toronto* . Technical Report CRG-TR-96-2.

Ghysels, E. and Marcellino, M.: 2016, The Econometric Analysis of Mixed Frequency Data Sampling, *Journal of Econometrics* **193**, 291–293.

Ghysels, E., Santa-Clara, P. and Valkanov, R.: 2004, The Midas Touch: Mixed Data Sampling Regression Models, *CIRANO Working Papers* . Working Paper.

Gorgi, P., Koopman, S. J. and Mengheng, L.: 2018, Forecasting Economic Time Series using Score-driven Dynamic Models with Mixed-data Sampling, *International Journal of Forecasting* **35**, 1735–1747.

Onatski, A.: 2012, Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors, *Journal of Econometrics* **168**, 244–258.

Onatski, A.: 2015, Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models, *Journal of Econometrics* **186**, 388–406.

Stock, J. and Watson, M.: 1989, New Indexes of Coincident and Leading Economic Indicators, *in* O. Blanchard and S. Fischer (eds), *NBER Macroeconomics Annual*, MIT Press, Cambridge.

# Appendices

| Explanatory variables | BRA | CHL | COL | MEX |
|---|---|---|---|---|
| Google Index | -5.96 | -4.40 | -5.69 | -6.47 |
| PM2.5 | -6.18 | -4.58 | -5.64 | -6.57 |
| Temperature | -5.95 | -4.31 | -5.60 | -6.30 |
| PM2.5 and Temperature | -6.22 | -4.99 | -5.48 | -6.48 |
| PM2.5 and Temperature and Google Index | -6.93 | -5.59 | -6.50 | -8.88 |

Table 3. AIC values for different model specifications for nowcasting industrial production growth rate (m/m).

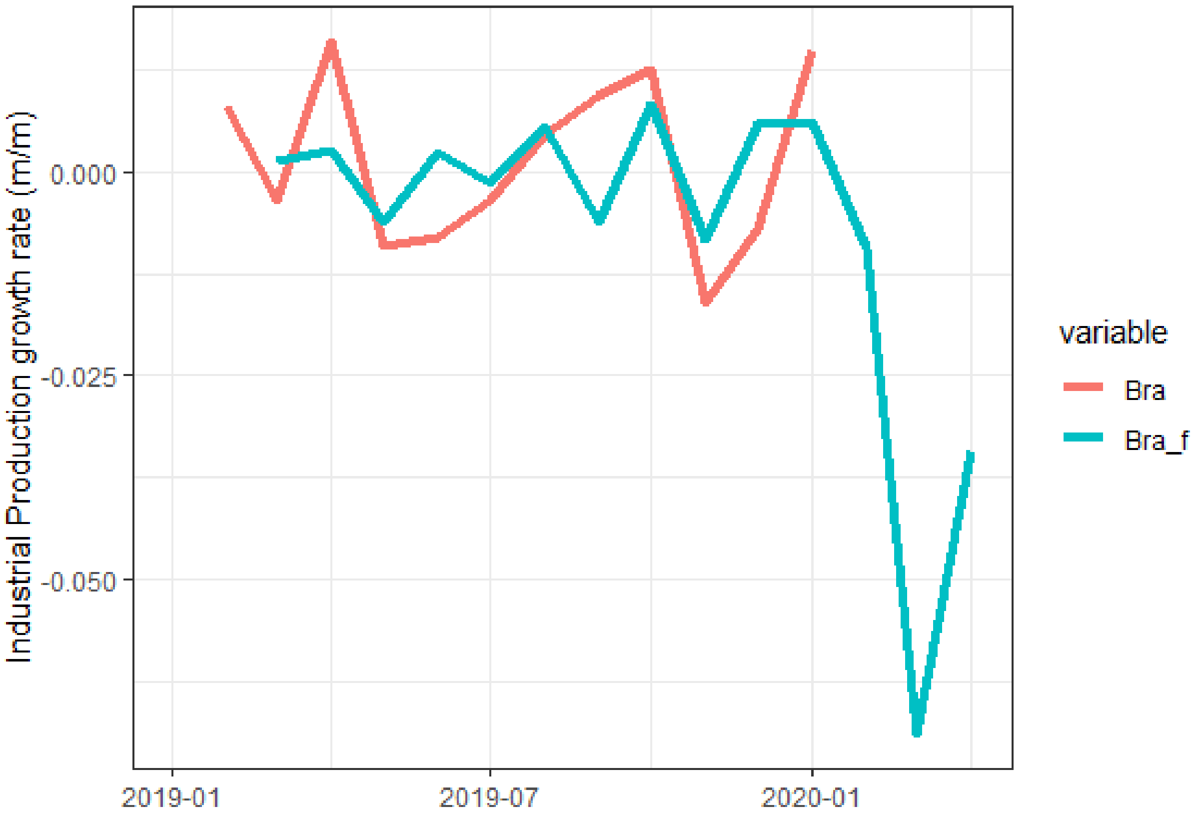


Figure 5: Actual versus one-step ahead predicted industrial production growth rate in Brazil

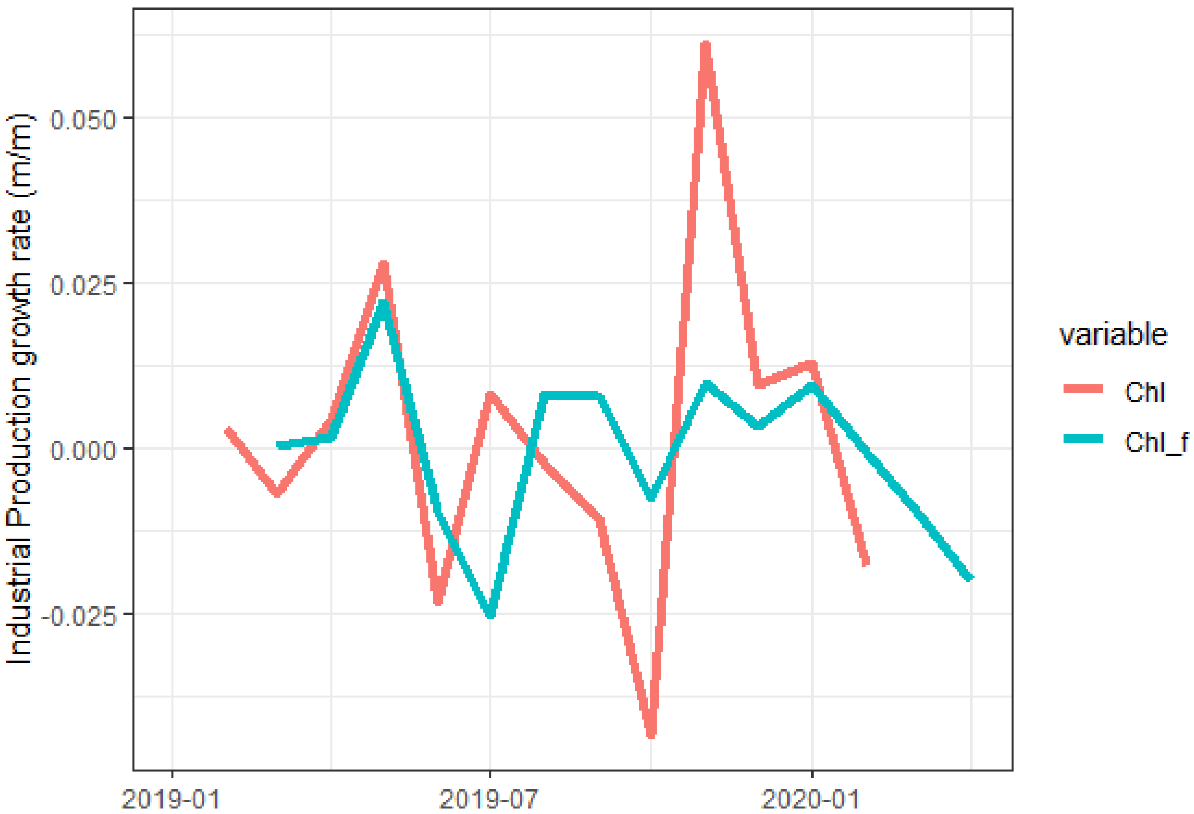


Figure 6: Actual versus one-step ahead predicted industrial production growth rate in Chile

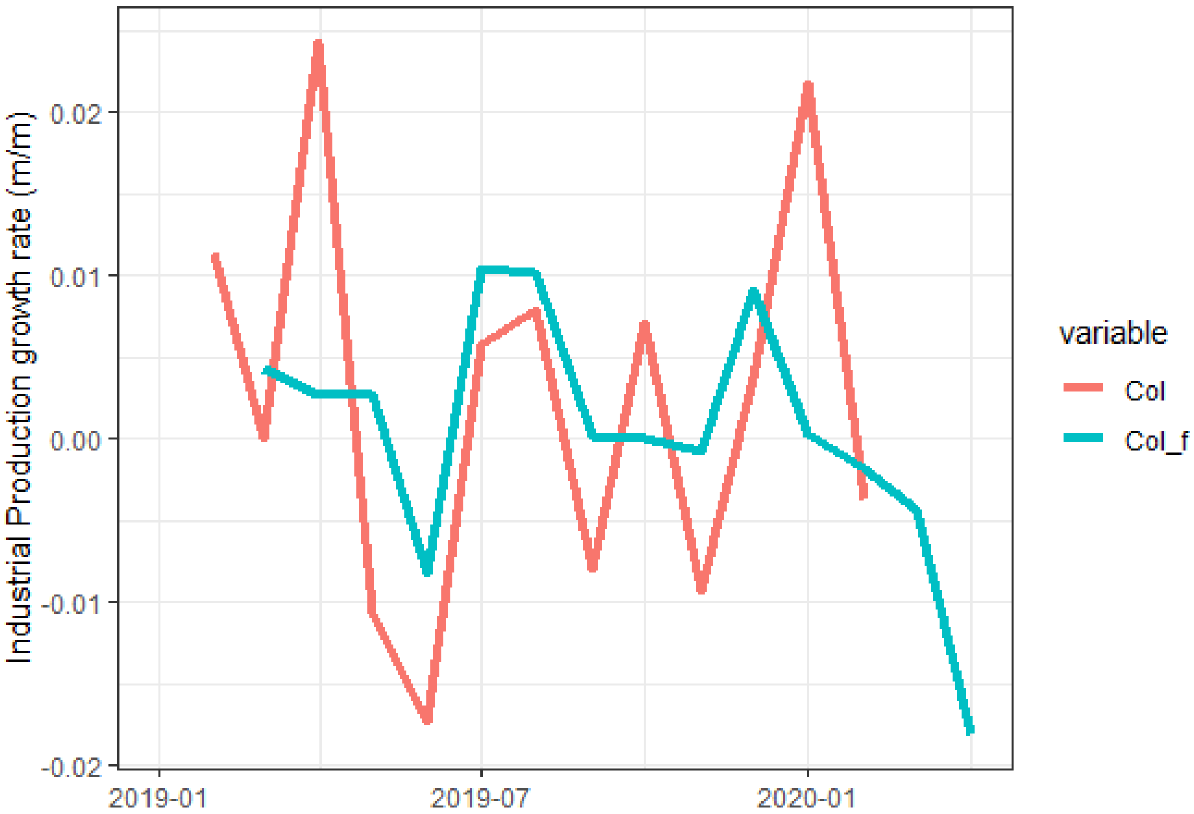


Figure 7: Actual versus one-step ahead predicted industrial production growth rate in Colombia

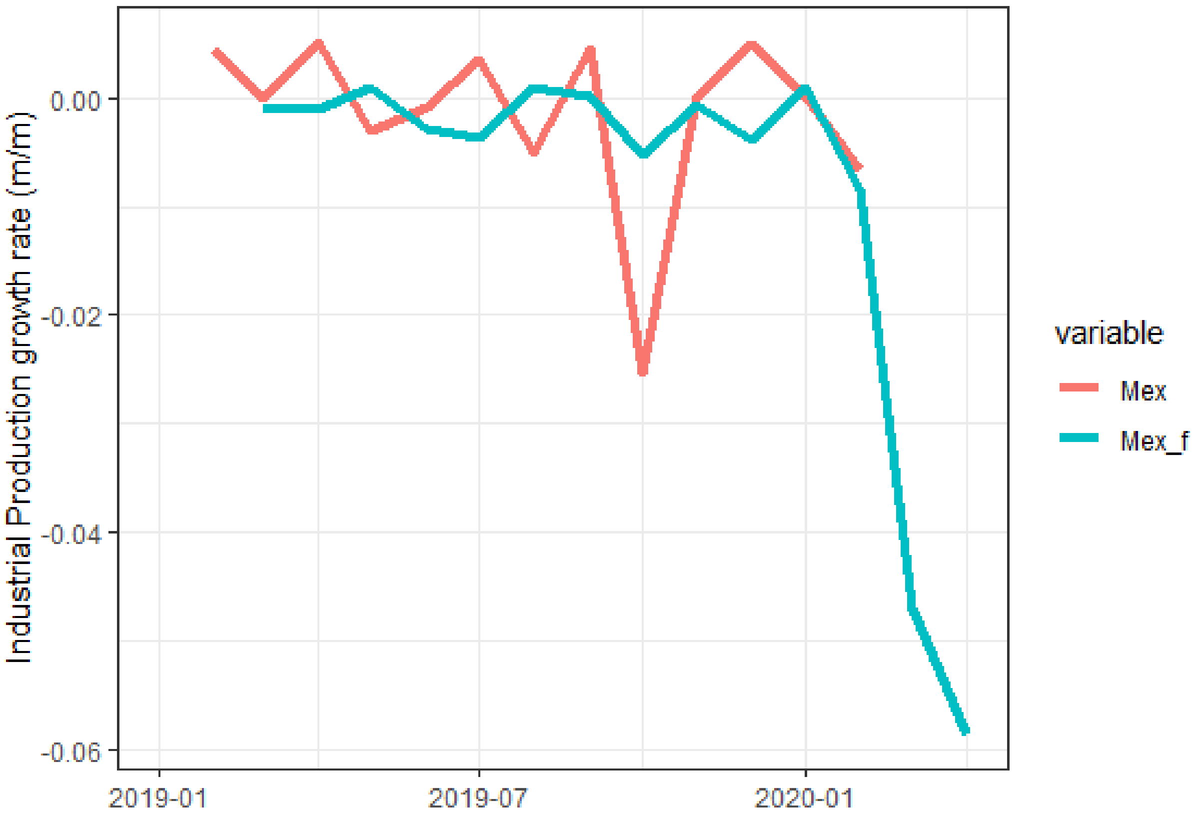


Figure 8: Actual versus one-step ahead predicted industrial production growth rate in Mexico