LATIN AMERICAN ECONOMIC REVIEW

2026 | Volume 37 | Article 5

Evaluating the Predictive Power of Mexico's Timely Economic Activity Indicator: Real and Pseudo Real-Time Performance

Francisco Corona[*,1], Jesús López-Pérez[2],
Edgar René Benavidez-Maruri[1] and Francisco J. López-Ballesteros[1]

Submission received: [April 19, 2025]
Final version received: [February 19, 2026]
Accepted: [February 24, 2026]
Published: [March 24, 2026]

Abstract

This paper evaluates the performance of Mexico's Timely Economic Activity Indicator (IOAE), published by the National Institute of Statistics and Geography, as a nowcasting tool for short-term economic activity. Using both real-time evidence (October 2020– December 2024) and a pseudo-real-time comparison with alternative nowcasting approaches over 2018–2024, we assess the IOAE against individual indicators, econometric models, and machine-learning methods. Nowcast accuracy is evaluated using Diebold–Mariano tests with Heteroskedasticity and Autocorrelation Consistent correction, the Superior Predictive Ability test, and the Model Confidence Set procedure. The results indicate that the IOAE consistently belongs to the set of best-performing models and tends to exhibit relatively stronger predictive accuracy, particularly at the two-month-ahead horizon. A comprehensive robustness analysis—covering alternative window schemes, numbers of factors, loss functions, and sample definitions, including evaluations conducted both with and without the COVID-19 period—supports the stability of these findings. Overall, the evidence supports the view that the IOAE provides timely and informative nowcasts of economic activity, reinforcing its usefulness as a complementary tool for short-term economic monitoring in Mexico.

Keywords: Forecast evaluation, IOAE, Machine learning; Nowcasting; Robustness analysis.

[1]Instituto Nacional de Estadística y Geografía
[2]City University of New York
[*]Corresponding author: franciscoj.corona@inegi.org.mx

1. Introduction

The Mexico's Timely Economic Activity Indicator (IOAE)[1] was introduced by the National Institute of Statistics and Geography (INEGI) in October 2020 as an experimental dataset. Its objective is to provide timely estimates of Mexico's Global Indicator of Economic Activity (IGAE),[2] a national aggregate measure of economic activity (rather than an indicator of the global or world economy), as well as its secondary and tertiary sectors, one and two months ahead of the latest official figures.

The IGAE can be understood as a proxy for monthly Gross Domestic Product (GDP) and is therefore one of the most important variables for monitoring the economy in the short run. Consequently, having timely estimates is crucial for decision-making. Given the IGAE's role as a proxy for GDP, the IOAE can also serve as a timely estimate of quarterly GDP—available up to 15 days before the official preliminary GDP release (PIBO).[3] In other words, the IOAE functions as a timely estimator for both the IGAE and GDP.

The IOAE follows the methodology proposed by Corona et al. (2022), which combines macro-financial indicators and Google Trends data within a Dynamic Factor Model (DFM) framework. That approach involves: (i) selecting macro-financial time series with high contemporaneous correlation with the IGAE; (ii) applying transformations that maximize such correlations; (iii) identifying relevant Google Trends topics via Least Absolute Shrinkage and Selection Operator (LASSO) regression; (iv) estimating dynamic factors following Doz et al. (2011); and (v) generating nowcasts through factor-based regressions with Autoregressive Moving Average (ARMA) error structures.

Initial evaluations reported in Corona et al. (2022), using pseudo-real-time data for the period November 2017–October 2020, suggest that the IOAE outperforms simple benchmarks such as autoregressive models, Neural Networks (NNs) without covariates, and DFMs excluding Google Trends data. The IOAE was also shown to yield smaller forecast errors than INEGI's Monthly Manufacturing Indicator and to compare favorably with GDP projections produced by analysts at the Bank of Mexico.

Despite these encouraging results, several limitations of the initial IOAE evaluation remain. Most importantly, the original assessment was necessarily restricted to a pseudo-real-time setting, as the IOAE had not yet been released as an official experimental statistic. As a result, its performance could not be evaluated under genuine real-time conditions, which are crucial for assessing the practical usefulness of timely indicators in official statistics.

In addition, the benchmark set considered in Corona et al. (2022) was relatively narrow, relying mainly on simple time-series models and basic NNs. The analysis did not assess the IOAE against a broader range of econometric and Machine-Learning (ML) approaches commonly used in the nowcasting literature. Moreover, although LASSO was employed as a variable-selection tool, the contribution of individual predictors or subsets of information—such as the Purchasing Managers' Index (PMI), a well-known leading indicator of economic activity Lahiri and Monokroussos (2013)—was not examined in detail. Taken together, these limitations leave open the question of how the IOAE performs in real time and how its predictive accuracy compares with alternative nowcasting strategies under comparable evaluation frameworks.

The objective of this paper is to provide a comprehensive and rigorous reassessment of the IOAE's predictive performance. Specifically, we evaluate the IOAE both in real time (from its official release in October 2020 onward) and in a pseudo-real-time framework starting in October 2018, and compare its performance against a broader set of competing methods. These include individual-indicator regression models, LASSO implemented as a nowcasting method, Factor-Augmented Vector Autoregressive (FAVAR) models following

[1] https://www.inegi.org.mx/investigacion/ioae/
[2] https://www.inegi.org.mx/temas/igae/
[3] https://www.inegi.org.mx/temas/pibo/

Bernanke et al. (2005), Multilayer Perceptron (MLP) neural networks using dynamic factors as covariates, and Autoregressive Integrated Moving Average (ARIMA) benchmark specifications. Forecast accuracy is assessed using formal statistical tests and multiple-comparison procedures to ensure robust inference.

In addition, given the IOAE's role as an official experimental statistic, we complement the real-time evaluation with a brief descriptive sentiment analysis based on social-media content. This analysis is not intended to improve the nowcasting model itself; rather, it aims to document public engagement with and perceptions of the IOAE as a timely indicator of economic activity. Such evidence is relevant for communication, transparency, and the potential adoption of experimental indicators within official statistics.

Against this backdrop, the paper contributes to the growing literature on nowcasting, which was initially introduced into macroeconomic analysis by Giannone et al. (2008) through DFMs. Since then, the nowcasting literature has developed along multiple methodological lines, including DFMs, bridge equations, and, more recently, ML techniques that exploit high-frequency and alternative data sources for short-term economic monitoring. Recent studies document gains from using ML methods alongside traditional econometric approaches, either as competing benchmarks or in combination with them (Fornaro and Luomaranta, 2020; Maehashi and Shintani, 2020; Marcellino and Sivec, 2021; Medeiros et al., 2021). In the Latin American context, prior work shows that nowcasting models can substantially improve timely GDP estimates by integrating conventional indicators with high-frequency or non-traditional data sources Gálvez-Soriano (2020); Bantis et al. (2023); Bravo-Higuera et al. (2024); Bolivar (2024), among many others.

The remainder of the paper is structured as follows. Section 2 describes the methods employed in this study, including the econometric and ML models as well as the statistical criteria used for forecast evaluation. Section 3 presents the data. Section 4 evaluates the real-time performance of the IOAE, including forecast accuracy and sentiment analysis. Section 5 reports the results of the pseudo-real-time comparison with alternative nowcasting approaches. Section 6 discusses the implications for policymakers. Finally, Section 7 concludes.

2. Methods and techniques

This section outlines the notation and methods used to perform nowcasts, including the IOAE approach and the alternative models considered for comparison. It also describes the statistical criteria employed to evaluate and compare forecast performance across methods.

2.1. Nowcasting

We define y_t for $t = 1\ldots,T$ as the objective time series to be estimated in a timely manner. Let $\mathbf{X}_t = (X_{1t},\cdots,X_{Nt})'$ represent the set of time series closely related to y_t. Given this, the nowcasts are formulated using the following expression:

$$E(y_{T+H}) = E\left[y_{T+H} \mid \Theta(f(\mathbf{X}_{T+H}), y_{T+H-p})\right]. \tag{1}$$

That is, the nowcast of y for H steps forward is conditional to the estimation of the parameters Θ obtained as a function of covariates $\mathbf{X}_{T+H}$ and of their possible p lags of y. Having defined the elements of nowcasting in this manner, the subsequent subsections will detail the most crucial components of the models employed in this study.

2.2. Transformations

We suppose that $\mathbf{Y}_i = (Y_{i1}, \ldots Y_{iT})$, is a vector of macroeconomic or financial time series, the IOAE proposes that for each $i = 1, \ldots, M$ the optimal transformation satisfies the following condition

$$\mathbf{X}_i^* := \left[f(\mathbf{Y}_i) \mid \max_{\rho} f(\mathbf{Y}_i), g(\mathbf{y}) \right], \tag{2}$$

where $\mathbf{y} = (y_1, \ldots, y_T)$ and ρ is the correlation coeficient. Consequently, $\mathbf{X}i^*$ tends to covary contemporaneously with the target time series. In the context of the IOAE, the methodology considers three alternative representations of each series:

- the level of the series (index, base 2018 = 100),
- monthly growth rates (MV), defined as month-over-month percentage changes, and
- annual growth rates (AV), defined as year-over-year percentage changes.

These transformations are applied symmetrically to both the predictors and the target variable and define the functional forms of $f(\cdot)$ and $g(\cdot)$. In other words, $\mathbf{X}i^*$ is constructed by selecting, for each series, the transformation that maximizes its contemporaneous correlation with the corresponding transformation of the IGAE.

2.3. Nowcasting methods

This subsection describes the nowcasting approaches considered in the empirical analysis. LASSO regression is included both as a variable-selection tool and as a standalone nowcasting method. The IOAE serves as the baseline method and is based on a DFM that aggregates information from a broad set of timely indicators.In addition, we consider a FAVAR model and a MLP as representative econometric and machine learning alternatives, respectively. Finally, an automatic ARIMA model selected using information criteria is used as a naïve benchmark that relies solely on the past dynamics of the target series.

2.3.1 LASSO: Selecting relevant time series and regression method

LASSO regression was originally introduced by Tibshirani (1996). Its objective is to predict y by selecting variables that minimize the estimation error, which may include either the entire set or a subset of the time series $\mathbf{G}$ associated with the target series. For this purpose, the following penalized Sum of Squared Residuals (SSR) is minimized:

$$(\mathbf{y} - \boldsymbol{\psi}\mathbf{G})'(\mathbf{y} - \boldsymbol{\psi}\mathbf{G}) + \lambda \sum_{i=1}^{n} \mid \psi_i \mid . \tag{3}$$

As is well known in the literature, this penalized minimization problem can also be posed as minimizing the SSR subject to $\sum_{i=1}^{n} \mid \psi_i \mid < \lambda$, where λ is the regularization parameter. Expression (3) has no closed-form solution. In practical terms, when $\lambda \to 0$, all variables tend to contribute to reducing the estimation error of y, whereas when $\lambda \to \infty$, the coefficients $\psi \to 0$, and no variable retains predictive power. Parameter estimation is carried out using cross-validation adapted to a time-series setting.

In this sense, LASSO serves as a variable-selection method by identifying $\mathbf{G}_t^* = (G_{1t}, \ldots, G_{mt})$, where $m \leq n$, as the subset of relevant time series for the timely estimation of y_t. In the context of this work, $\mathbf{G}$ represents the matrix of Google Trends topics.

Moreover, assuming that

$$\mathbf{X}_t = (\mathbf{X}_t^*, \mathbf{G}_t^*), \tag{4}$$

that is, fixed macroeconomic and financial time series augmented with the relevant Google Trends topics selected via LASSO, with $N = M + m$, expression (1) can be employed to generate nowcasts. In addition, LASSO is also implemented directly as a nowcasting method by relying on $\mathbf{X}_t$ rather than solely on $\mathbf{G}_t$.

Given the real-time and pseudo-real-time nature of the exercises conducted in this paper, it is important to clarify how the LASSO regression is implemented in practice. Specifically, LASSO is used both as a variable-selection device and as a nowcasting method within real-time evaluation schemes based on either expanding or rolling estimation windows, depending on the design of the forecasting exercise. At each nowcast origin, the model is re-estimated using only the information available up to that date, thereby replicating the information set that would have been available in real time.

When LASSO is implemented as a nowcasting device, all predictors are standardized to have zero mean and unit variance within the estimation window. This transformation ensures scale-invariant penalization given the heterogeneous measurement units of macroeconomic indicators (see, for instance, (Tibshirani, 1996; Friedman et al., 2010).) Importantly, this standardization is performed separately at each estimation date, using only information available up to that point in time.

By contrast, when LASSO is used for Google Trends topic selection, predictors are not standardized. The underlying variables are derived from normalized indices expressed on a common 0–100 scale, rendering additional rescaling unnecessary.

At each nowcast origin, the regularization parameter λ is selected in a data-driven manner via cross-validation. In particular, λ is searched over a fixed grid of 100 values ranging from 0.0001 to 0.1. Cross-validation is carried out using a rolling time-series scheme based exclusively on historical data, so that no future information enters the tuning or estimation process.

This fixed λ grid is applied consistently across all real-time and pseudo-real-time vintages in order to preserve comparability over time and to avoid look-ahead bias when updating the model. Accordingly, we employ either expanding or rolling estimation windows, depending on whether the specification corresponds to a real-time or a pseudo-real-time vintage. Given the standardization of regressors, the chosen range is economically meaningful and sufficiently wide to span from lightly penalized specifications to highly regularized models that retain only the most informative predictors. In practice, this approach is equivalent to searching over a fraction of the maximum admissible penalty $\lambda_{\max}$—defined as the smallest value of λ for which all coefficients are shrunk to zero—while ensuring numerical stability and computational tractability in a rolling real-time setting Friedman et al. (2010).

The LASSO algorithm is estimated using the `glmnet` package, while time-series cross-validation is performed using the `caret` library.

2.3.2 Timely Indicator of Economic Activity

While the original IOAE approach is based on the estimation of a dynamic factor—typically implemented within the framework of Doz et al. (2011)—we do not impose a single-factor restriction in this paper. Instead, we adopt a general factor structure in which the covariates $\mathbf{X}_t$ are driven by $r < N$ latent factors, yielding the following representation:

$$\mathbf{X}_t = \mathbf{P}\mathbf{F}_t + \boldsymbol{\varepsilon}_t, \tag{5}$$

that is, the movements of the observed variables are explained as a function of common factors that follows an Vector Autoregressive (VAR) process, $\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \cdots + \mathbf{A}_p\mathbf{F}_{t-p} + \boldsymbol{\eta}_t$, where $\mathbf{P}$ is a vector that indicates the contribution of the factors to the observations whose dimension is $N \times r$. The errors, $\boldsymbol{\eta}_t$ are assumed to be white noise. In addition, we assume that the idiosyncratic component, $\boldsymbol{\varepsilon}_t$, follows a stationary process.

Assuming that the common factors are either $I(1)$ or $I(0)$, as is typical in macroeconomic applications, the factor process $\mathbf{F}_t$ may be non-stationary. In that case, its first difference satisfies $\Delta\mathbf{F}_t \sim I(0)$. Consequently, in line with Bai (2004), we assume that the idiosyncratic component $\boldsymbol{\varepsilon}_t$ is stationary. This assumption is required to ensure consistent estimation of the model under the two-step procedure of Doz et al. (2011). Empirical support for this consistency in a related application is provided by Corona et al. (2020). The estimation proceeds as follows:

1. **Transformations.** The optimal transformation is estimated using seasonally adjusted $\mathbf{Y}_t$, while the relevant Google Topics time series are selected according to expressions (2) and (3) respectively, which jointly determine $\mathbf{X}_t$.

2. **Asymptotic Principal Components (PC).** Given the identifiability condition $\mathbf{P}'\mathbf{P}/N = \mathbf{I}_r$, it follows that $\hat{\mathbf{P}}$ is $\sqrt{N}\mathbf{V}$ where $\mathbf{V}$ is the first r eigenvectors of the matrix $\mathbf{X}'\mathbf{X}$, and $\mathbf{X} = (\mathbf{X}_1, ..., \mathbf{X}_T)'$. By Regressing $\mathbf{X}$ on $\hat{\mathbf{P}}$ and using the identifiability restrictions, the Ordinary Least Squares (OLS) common factor estimator can be expressed as $\tilde{\mathbf{F}} = \mathbf{X}\hat{\mathbf{P}}/N$.

3. **Kalman filter for real-time nowcasts.** For reasons of computational convenience and tractability, we assume that $\tilde{\mathbf{F}}_t$ follows a VAR(1) process (see, for example, Bai and Ng, 2007 for similar representations). The coefficient matrix $\hat{\mathbf{A}}$ is estimated by OLS, along with $\hat{\mathbf{P}}$, $\hat{\boldsymbol{\Sigma}}_\varepsilon = \text{diag}(\hat{\sigma}^2_{\varepsilon_i})$, $\hat{\boldsymbol{\Sigma}}_\eta = \text{diag}(\hat{\sigma}^2_{\eta_j})$, and the diagonal matrix $\hat{\boldsymbol{\Sigma}}_{F_{0j}}$, which contains the unconditional variances of the an AR(1) process for each common factor, $j = 1, \ldots, r$. When the common factors are non-stationary, they are modeled as random walks with $\hat{\sigma}^2_{\eta_j} = 10^7$ (see Corona et al., 2020 for details).

   These estimates allow us to specify the state-space representation, from which the Kalman filter for real-time nowcasts is derived:

   $$\hat{\mathbf{F}} = E(\mathbf{F} \mid \Xi).$$

   This formulation refines the factors initially extracted by PC. In other words, the final factor estimates are obtained by explicitly modeling the dynamics of the PC-based factors and conditioning on all available information from the DFM series up to $T + H$, as summarized in Ξ and collected in $\mathbf{X}_t$, while excluding y_t, whose values remain unknown.[4] Consequently, the time dimension of $\hat{\mathbf{F}}$ extends to $T + H$.

4. **Stationary errors.** Note that stochastic nature of $\mathbf{X}_t$ depends on $\mathbf{F}_t$ and $\boldsymbol{\varepsilon}_t$. In any case, if $\mathbf{F}_t \sim I(1)$ and $\boldsymbol{\varepsilon}_t \sim I(0)$, $\mathbf{F}_t$ are the common trends of $\mathbf{X}_t$. On the other hand, if $\mathbf{F}_t \sim I(0)$ and $\boldsymbol{\varepsilon}_t \sim I(1)$, $\mathbf{X}_t \sim I(1)$ but they are not cointegrated, consequently, the estimation of $\mathbf{F}_t$ is not consistent (Bai, 2004). Therefore, we apply Panel Analysis of Nonstationarity in Idiosyncratic and Common components (PANIC) tests (Bai and Ng, 2004) to $\hat{\mathbf{F}}_t$ and $\hat{\boldsymbol{\varepsilon}}_t$ in order to conclude stationarity or cointegration in $\mathbf{X}_t$.

5. **Regression with possible ARMA errors.** Once the factor is estimated, the following regression model is used $y_t = \alpha + \boldsymbol{\beta}'\hat{\mathbf{F}}_t + e_t$, where the error structure may be $\boldsymbol{\phi}(L^p)e_t = \boldsymbol{\theta}(L^q)u_t$. Using expression (1) we obtain the nowcasts for y_{T+H}. Note that alternatively, we can use each element of $\mathbf{X}_t$ instead of the $\hat{\mathbf{F}}_t$ so that we can have nowcasts using each of the variables instead of the common factor.

   As a remark, note that whereas the target y_t incorporates information only up to T, $\hat{\mathbf{F}}_t$ can be computed up to $T + H$, using the timely availability of explanatory variables. This additional information is therefore exploited to produce a timely estimate of the IGAE. Consequently, the nowcasts are constructed under the assumption that the values of y_t for $T + 1, \ldots, T + H$ are unknown.

[4] As an alternative representation, missing values are imputed using a random walk for each time series, extending $\mathbf{X}_t$ to $T + H$ and avoiding Kalman-filter-based imputation.

The IOAE approach is implemented using the `nowcasting` package together with the `Arima` function. The analysis also makes use of several custom-built functions developed by the authors.

2.3.3 Factor Augmented Vector Autoregressive Model

Following, for example, Bernanke et al. (2005) and Stock and Watson (2005), we can use the predictive capacity of the factor with respect to the variable of interest, that is, use the past of $\mathbf{F}_t$ to forecast y_t. Furthermore, in this work we exploit the contemporaneity of the factor to fit y_t such that the FAVAR model is expressed as follows

$$\begin{pmatrix} y_t \\ \mathbf{F}_t \end{pmatrix} = \begin{pmatrix} \mu \\ \boldsymbol{\mu}_r \end{pmatrix} + \sum_{i=1}^{p} \begin{pmatrix} \pi_i & \boldsymbol{\Pi}_{y,i} \\ \mathbf{0} & \boldsymbol{\Pi}_{F,i} \end{pmatrix} \begin{pmatrix} y_{t-i} \\ \mathbf{F}_{t-i} \end{pmatrix} + \begin{pmatrix} \varphi_1 \\ 0 \end{pmatrix} \mathbf{F}_t + \boldsymbol{\epsilon}_t. \tag{6}$$

The estimation of the expression (6) is carried out by OLS in such a way that it is inefficient, however, for predictive purposes, it is considered valid. Consequently, using the functional form (1), it is possible to obtain the nowcasts or forecasts of y_t and even the predictions $\mathbf{F}_t$ the latter for the purposes of this work are irrelevant.

This method incorporates not only the contemporaneous values of $\mathbf{F}_t$ if not their lags, consequently, we expect that this method will be competitive.

The FAVAR model is estimated using the `vars` package, with the lag length selected according to the Schwarz information criterion. The estimated VAR is then restricted to the FAVAR representation via the `restrict` function.

2.3.4 Multilayer perceptron for time series forecasts

Following Kourentzes et al. (2014) NNs are very attractive for forecasting time series because they are highly flexible since they do not require distributional assumptions as traditional time series models frequently require. Thus, the MLP for time series forecasts used in Corona et al. (2022) generates competitive estimation errors with respect to the IOAE, however, they only use information from the past of the time series of interest to make the appropriate estimates. In this work, the predictive capacity of $\hat{\mathbf{F}}_t$ is used as a predictive element of $\mathbf{y}$. The MLP performs a series of sequential transformations to the input variables, thus defining a series of "hidden" variables (neurons), organized in blocks called "hidden layers", where the term "hidden" is since such variables are not directly observable. The particular form of the MLP is defined according to a certain architecture, where there is an initial layer that contains the original variables, generally standardized, and an output layer, which provides us with the response variables according to the particular task being addressed: either classification or regression. The MLP can have many hidden layers, leading to deep and complex networks, however, for a large number of applications, including ours, an MLP with one hidden layer is enough. Thus, the MLP used to predict H steps forward to the variable of interest is written as follows

$$\hat{y}_{T+H} = \bar{\omega}_0 + \sum_{j=1}^{J} \bar{\omega}_j s \left(\sum_{i=1}^{p^*+1} \gamma_{ji} \mathbf{Z}_i \right), \tag{7}$$

where $\bar{\omega}$ and γ are the weights attributable to the network, the first and the second $\bar{\omega}_0$ and γ_{0i} respectively are the biases of each neuron that play a similar role as the intercept in a regression, J is the total number of hidden nodes in the network and $s(\cdot)$ is non-linear function. The matrix $\mathbf{Z}$ contains the p^* lags of y_t and the factor $\hat{\mathbf{F}}_t$. NNs can model interactions between inputs if there are any. The outputs of the hidden nodes are connected to an output node that produces the appropriate estimate.

It is worth noting that the essence of this nowcasting approach lies in replacing the regression with ARMA errors used in the IOAE procedure with the MLP method.

The MLP is implemented using the `nnfor` package, where the estimated common factors are included as regressors in the network architecture.

2.3.5 ARIMA benchmark model

As a univariate benchmark, we consider an ARIMA model estimated using the automatic model selection procedure proposed by Hyndman and Khandakar (2008) and implemented in the `auto.arima` function of the `forecast` package in R.

Specifically, for each nowcast origin, the ARIMA model is selected in a data-driven manner by minimizing an information criterion, subject to standard parsimony and stationarity constraints. The procedure jointly determines the orders (p, d, q) of the ARIMA(p, d, q) specification, where d is selected via unit-root testing and (p, q) are chosen by optimizing the corrected Akaike Information Criterion. Empirically, the target time series is seasonally adjusted prior to estimation. As a result, seasonal patterns are largely removed from the data, and the automatic identification procedure implemented in `auto.arima` typically selects non-seasonal ARIMA specifications.

Importantly, the ARIMA benchmark is estimated using only past realizations of the target variable and does not incorporate any exogenous regressors or latent factors. As such, it provides a naïve yet widely used reference model that captures purely autoregressive dynamics, thereby constituting a natural lower bound for forecast performance against which more information-rich approaches—such as the IOAE, LASSO-based regressions, DFMs, and ML methods—can be meaningfully evaluated.

To ensure generality and comparability across vintages, the benchmark is implemented within the broader ARIMA framework, which allows for a flexible treatment of the stochastic properties of the target series. When the series is found to be stationary, the differencing order is set to zero, so that the specification reduces to a standard ARMA model nested within the ARIMA class.

Finally, the ARIMA model is re-estimated at each real-time and pseudo-real-time forecast origin using only the information that would have been available at that date, thereby ensuring that no look-ahead bias is introduced into the evaluation exercise.

2.4. Forecast evaluation and decision criteria

This subsection describes the statistical criteria used to evaluate and compare the nowcasting performance of the IOAE and its competing models in a pseudo-real-time setting. Given the multiplicity of nowcasting approaches and evaluation objectives, we rely on complementary decision rules that address different dimensions of forecast comparison. Specifically, we combine pairwise tests of predictive accuracy with multiple-comparison procedures that explicitly account for data-snooping effects. This layered evaluation strategy allows us to assess both bilateral performance differences and the relative standing of each model within a broader set of alternatives.

2.4.1 Diebold–Mariano test with heteroskedasticity and autocorrelation consistent standard errors

First, we evaluate forecast accuracy using pairwise comparisons based on the Diebold and Mariano (1995) test with Heteroskedasticity and Autocorrelation Consistent standard errors (DM–HAC). The IOAE is treated as the benchmark model and is compared against each competing nowcasting approach at horizons up to $T+2$. Let $e_{0,s}$ denote the forecast error of the IOAE and $e_{j,s}$ the forecast error of an alternative model j, for $s = 1, \ldots, H_t$. Forecast performance is evaluated using a loss function $\mathcal{L}(\cdot)$. The loss differential is defined as

$$\ell_{j,s} = \mathcal{L}(e_{0,s}) - \mathcal{L}(e_{j,s}), \tag{8}$$

and its sample mean as

$$\bar{d}_j = \frac{1}{H_t} \sum_{s=1}^{H_t} \ell_{j,s}. \tag{9}$$

The DM–HAC statistic is given by

$$DM_{HAC,j} = \frac{\bar{d}_j}{\sqrt{\widehat{\mathrm{Var}}(\bar{d}_j)}}, \tag{10}$$

where the long-run variance estimator is

$$\widehat{\mathrm{Var}}(\bar{d}_j) = \frac{1}{H_t}\hat{\gamma}_0 + \frac{2}{H_t} \sum_{l=1}^{h-1} \omega_l \hat{\gamma}_l. \tag{11}$$

Here, $\hat{\gamma}_l$ denotes the sample autocovariance of $\ell_{j,s}$ at lag l, h is the forecast horizon, and $\omega_l = 1 - \frac{l}{h}$ are Bartlett weights, which ensure a positive semi-definite variance estimator.

The null hypothesis of the DM–HAC test is

$$H_0 : \ \mathbb{E}[\ell_{j,s}] = 0,$$

which corresponds to equal predictive accuracy between the IOAE and model j. The alternative hypothesis is

$$H_1 : \ \mathbb{E}[\ell_{j,s}] < 0,$$

indicating that the IOAE provides superior nowcasts, in the sense of lower expected loss, relative to the competing model. Under the null hypothesis, $DM_{HAC,j}$ is asymptotically standard normal. However, given the relatively limited forecast evaluation sample, statistical inference is conducted using HAC variance estimators together with the small-sample correction proposed by Harvey et al. (1997), so that reported p-values are obtained from the corresponding finite-sample t distribution.

2.4.2 Superior Predictive Ability test

To complement the pairwise comparisons, we apply the Superior Predictive Ability (SPA) test proposed by Hansen (2005). Unlike the DM–HAC test, which evaluates forecast accuracy on a bilateral basis, the SPA test provides a joint assessment of whether any competing model outperforms a prespecified benchmark.

In our application, the IOAE is treated as the benchmark model. Let $\ell_{j,s}$ denote the loss differential between the IOAE and an alternative model j, defined as in equation (8). Hence, the SPA test evaluates the null hypothesis

$$H_0 : \ \max_{j=1,\ldots,K} \mathbb{E}[\ell_{j,s}] \leq 0,$$

which states that none of the competing models provides superior predictive accuracy relative to the IOAE. The alternative hypothesis is

$$H_1 : \exists j \text{ such that } \mathbb{E}[\ell_{j,s}] > 0,$$

indicating that at least one competing model delivers lower expected loss than the benchmark.

The SPA test statistic is defined as the maximum, across competing models, of a one-sided studentized mean loss differential. Its null distribution is approximated via a circular block bootstrap applied to centered loss differentials using positive-part centering. This procedure accommodates potential serial dependence in the evaluation sample and controls for data snooping arising from multiple model comparisons. In the empirical analysis, the SPA test is implemented using the same evaluation sample and loss function as in the DM–HAC comparisons.

2.4.3 Model Confidence Set procedure

To further assess forecast performance in a multiple-comparison framework, we apply the Model Confidence Set (MCS) procedure of Hansen et al. (2011), which extends the SPA test of Hansen (2005). The MCS is a sequential testing algorithm designed to evaluate the null hypothesis of equal predictive ability across a finite set of competing models using loss differentials. At each iteration, the model with the worst relative performance is eliminated until a subset of models remains for which the null hypothesis cannot be rejected at a given confidence level. This approach explicitly controls for data-snooping effects and provides a formal framework for identifying a set of models with statistically indistinguishable predictive accuracy.

Let $\ell_{i,s}$ denote the loss associated with model i at forecast origin s, with $i, j = 1, \ldots, K$ and $s = 1, \ldots, H_t$. The pairwise loss differential between models i and j is defined as

$$d_{ij,s} = \ell_{i,s} - \ell_{j,s}, \tag{12}$$

and its sample average is given by

$$\bar{d}_{ij} = \frac{1}{H_t} \sum_{s=1}^{H_t} d_{ij,s}. \tag{13}$$

For each model i, the average loss differential relative to all other models in the candidate set is defined as

$$d_{i\cdot} = \frac{1}{K-1} \sum_{j \neq i} \bar{d}_{ij}. \tag{14}$$

The null hypothesis tested by the MCS procedure is

$$H_0 : \mathbb{E}(\ell_{i,s}) = \mathbb{E}(\ell_{j,s}) \quad \forall\, i, j \in \mathcal{M},$$

which states that all models in the candidate set $\mathcal{M}$ have equal predictive ability. The alternative hypothesis implies that at least one model exhibits inferior predictive performance and should therefore be excluded from the confidence set.

In this study, we adopt the range statistic proposed by Hansen et al. (2011), defined as

$$T_{R,i} = \max_{j \neq i} \frac{\bar{d}_{ij}}{\widehat{\sigma}_{ij}}, \tag{15}$$

where $\widehat{\sigma}_{ij}$ denotes a heteroskedasticity and autocorrelation consistent estimator of the standard deviation of $\bar{d}_{ij}$. The global test statistic is then given by

$$T_R = \max_i T_{R,i}. \tag{16}$$

The MCS procedure iteratively eliminates the worst-performing model according to this statistic until a superior set of models remains. The range statistic yields a stringent and global elimination rule, ensuring that any model that performs significantly worse than at least one alternative is excluded from the confidence set.

The use of the T_R statistic is standard in the applied forecasting literature; see, for instance, Bernardi and Catania (2015). This choice enhances the comparability of our results with previous studies. The reported $Rank_R$ is obtained by ordering models according to their relative performance within this iterative procedure: models with lower values of the test statistic exhibit better relative predictive accuracy, whereas models with larger values are more likely to be excluded from the confidence set. Statistical inference is conducted using a bootstrap procedure that approximates the distribution of the test statistics under the null hypothesis of equal predictive ability.

3. Data

This section describes the variables employed in the previously described nowcasting models and provides remarks on the empirical specifications used for estimating these models.

The variables used in this work are consistent with those included in the IOAE released by INEGI on February 14, 2025, and incorporate Google Trends topics (see Appendix A.1). In this case, the transformation function $g(\cdot)$ corresponds to the MV, so that the target time series is defined as $y_t = (IGAE_t/IGAE_{t-1}) \times 100 - 100$.

The time series are of monthly frequency, starting from January 2004, with some available through January 2025. All variables have been obtained in seasonally adjusted form; if not, they have been adjusted using the X-13ARIMA-SEATS method. This seasonal adjustment relies on the automatic modeling procedure implemented in the `seas` function from R's `seasonal` package. Table 1 provides a summary of the dataset, including information on each variable's availability up to T, $T+1$, and/or $T+2$ (as variables available only up to time T do not contain observed information to directly inform the nowcasting exercise beyond the contemporaneous period), as well as the optimal transformation applied. For notational convenience, we use interchangeably the calendar-time notation ($T+1$, $T+2$) and the forecast-horizon notation ($h=1$, $h=2$) throughout the manuscript.

Table 1: *IOAE macroeconomic and financial time series*

Variable	Description	Transformation	Correlation
IMAI*	Industrial production	MV	0.923
Sales	Retail sales	MV	0.827
U*	Unemployment	MV	-0.132
M*	Imports	MV	0.567
Conf Constr**	Construction confidence index	MV	0.257
Conf Manuf**	Manufacturing confidence index	MV	0.262
Conf Trade**	Trade confidence index	MV	0.546
Conf Serv**	Services confidence index	MV	0.529
Trend L Manuf*	Manufacturing employment trend	MV	0.403
IPC**	Mexican Stock Exchange Index	AV	0.146
E**	Exchange rate	MV	-0.141
IR 28**	Short-term interest rate	AV	-0.052
SP 500**	S&P 500 Index	AV	0.153
Industrial USA**	U.S. industrial production	MV	0.713
ANTAD*	ANTAD retail sales	MV	0.559
Automotive**	Vehicle production	MV	0.493
Hotel*	Hotel occupancy	Level	0.026
Gas*	Gasoline sales	MV	0.790
IMSS**	IMSS formal employment	MV	0.472
Remmitances*	Remittances inflows	MV	0.113
X*	Exports	MV	0.650
PMI**	Manufacturing orders	MV	0.547
Cards**	Credit and debit card transactions	MV	0.567
SPEI**	Interbank Electronic Payment System	MV	0.086
U USA**	U.S. unemployment rate	MV	-0.710
Manuf USA**	U.S. manufacturing activity	MV	0.713
PIBO*	Timely GDP estimate	MV	1.000

* Variables available up to $T + 1$. ** Variables available up to $T + 2$. All variables are seasonally adjusted using X-13ARIMA-SEATS methodology. Correlation refers to the linear correlation with the MV of IGAE.

Note that, given the definition of the target time series, most transformations are also monthly. The linear correlations are generally strong, with some exceptions such as *IR 28*, *Hotel*, *SPEI*, *U*, and *Remittances*, among others. Nonetheless, these variables are retained in the analysis to ensure consistency in the IOAE estimates.

The *PIBO* variable is available on a quarterly basis and is therefore particularly useful for the timely estimation of March, June, September, and December.[5]

The correlation with the IGAE is equal to one because, in months other than those mentioned above, the observed values of the IGAE are used directly. Officially, the IOAE is updated based on U.S. industrial activity.

To illustrate how data become available over time—allowing us to generate IOAE nowcasts with varying shares of updated information—Figure 1 provides a visual representation of the IOAE estimates for October 2024 through January 2025.

[5]We define a new variable as $x_t = y_t$, and let Y_{t^*} represent the annual GDP growth for $t^* \in (5, \cdots, T/3)$, where $T/3$ is the number of quarters in the sample given the total number of months T. The relationship between quarterly and monthly observations is then given by:

$$x_{3t^*} = \left(1 + \frac{Y_{t^*}}{100}\right)\left(\sum_{i=1}^{3} x_{3t^*-i-11}\right) - \left(\sum_{i=1}^{2} x_{3t^*-i}\right).$$

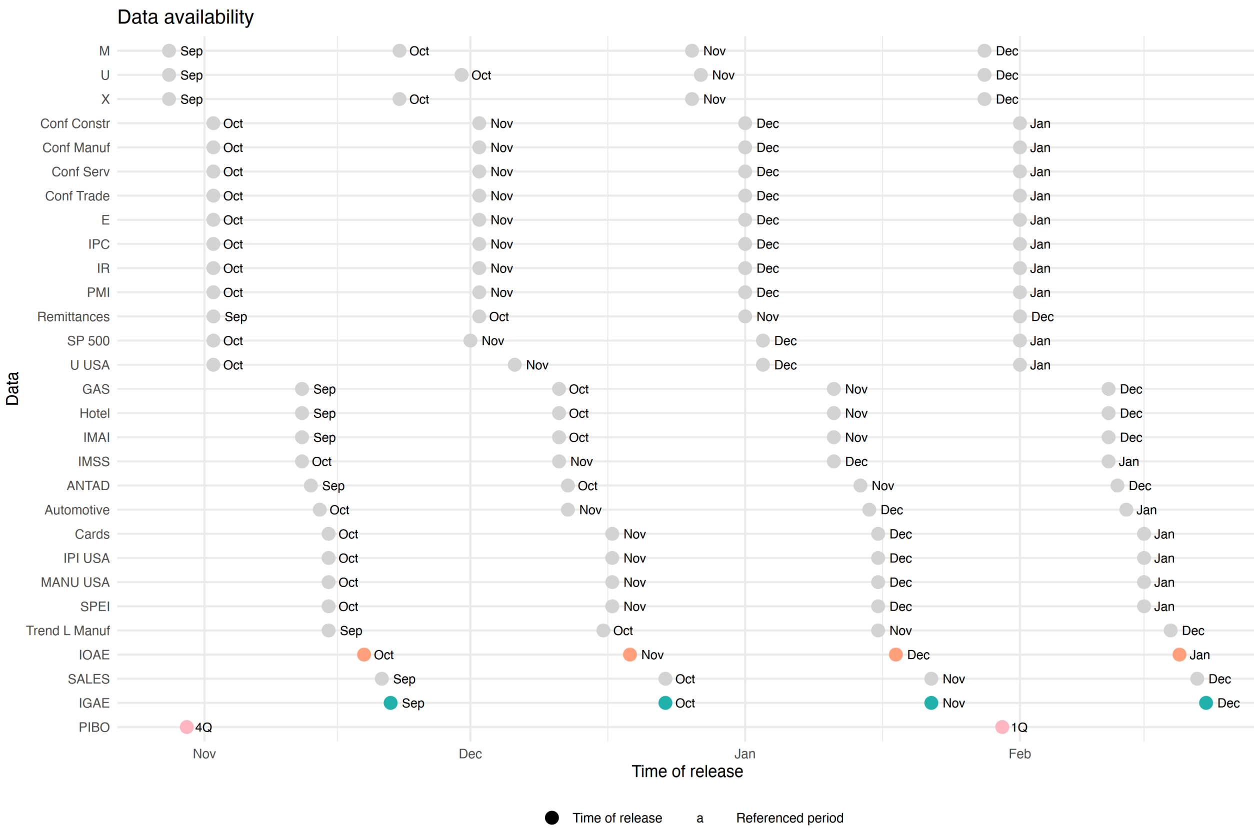

Figure 1: Schedule of data releases used for constructing the IOAE nowcasts. Solid circles indicate the timing at which each variable becomes available relative to the reference month. The staggered release structure explains how the IOAE can generate nowcasts at both horizons, with the T+2 estimates providing up to a 45-day lead over the official publication of the IGAE.

For example, if the objective is to generate nowcasts for November and December 2024, all variables are updated through November except for *Sales*. Some variables, however, are only available through December, including U.S. time series, *IMSS*, financial variables, among others. Consequently, the nowcast for $T + 1$ (November) incorporates nearly all available time series, whereas the nowcast for $T + 2$ (December) relies on ten series that remain outdated.

It is important to note that *PIBO* data are available for $T + 1$ when producing an estimate for December, as they correspond to the fourth quarter. Moreover, the *IOAE* is more timely than the *IGAE*, as it is released approximately one month before the official publication.

4. Real time performance of the IOAE

Having described the IOAE approach and the data employed, we now present the historical performance of the IOAE as a predictor of the IGAE and, consequently, of GDP. In addition, we include a sentiment analysis to examine how it is perceived on the social platform X (formerly Twitter).

4.1. IOAE vs IGAE

The IOAE was officially published by INEGI in October 2020,[6] and since then, until today (comparison with information released up to March 2025), it has produced estimates of the IGAE and its secondary and tertiary activities. Reviewing the historical estimates of the IOAE and focusing on the timely estimation of the IGAE. Figure 2 illustrates the behavior of the estimates in levels of the index, MV, and AV in relation to the observed IGAE data.[7]

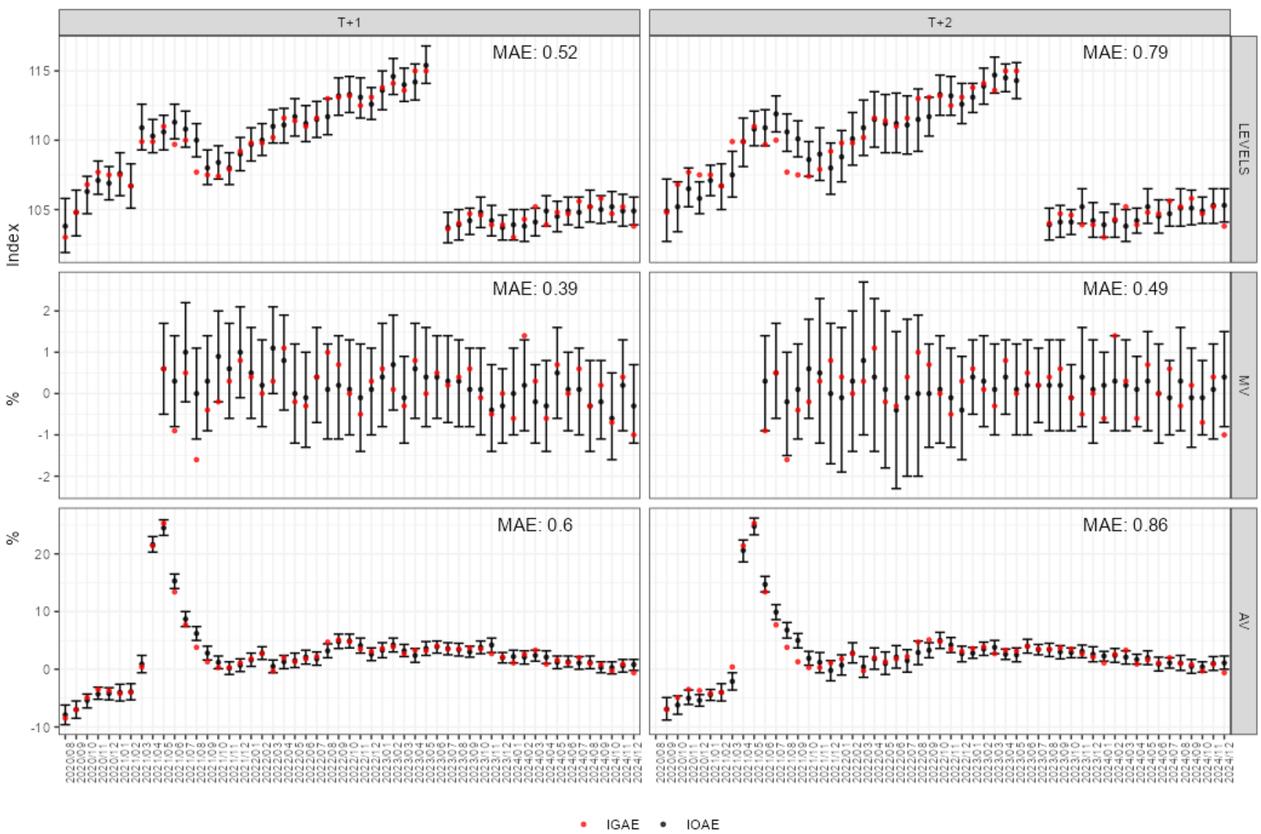

Figure 2: The real-time performance of the IOAE is shown by black points, together with its 95% confidence intervals (black lines), and is compared with the observed values (red points). The top panel displays the level of the index, the middle panel shows the MV, and the bottom panel depicts the AV.

We can observe that the estimation error, as expected, increases for $T+2$. Specifically, for $T+1$, the Mean Absolute Error (MAE) is 0.52 for levels, 0.39 for MV, and 0.6 for AV. These errors increase for T+2 to 0.79, 0.49, and 0.86, respectively. It is important to note that INEGI has been releasing nowcast results for MV since April 2021.

It is worth noting that, in terms of timeliness, $T+1$ provides estimates for the reference month approximately 45 days after its conclusion, while $T+2$ does so in just 15 days, whereas the IGAE is published around 55

[6]https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2020/IOAE/IOAE2020_10.pdf

[7]It is important to emphasize that Figure 2 is constructed using the original nowcasts and compared with the official IGAE values available on INEGI's website (https://www.inegi.org.mx/investigacion/ioae/#tabulados). Moreover, the confidence intervals are not necessarily symmetric and may vary over time due to methodological refinements, as specified in the IOAE methodology (https://www.inegi.org.mx/investigacion/ioae/#documentacion).

days later. Consequently, the time gain is 10 days for $T + 1$ and 40 days for $T + 2$.

Due to INEGI's base year revision from 2013 to 2018, the June 2023 level estimate is not available. Accordingly, data from August 2020 to May 2023 are expressed with the 2013 base year, while subsequent observations are reported with the 2018 base year.

The trade-off between timeliness and accuracy in nowcasting has been widely documented across countries. In Norway, Aastveit (2010) report a 45-day publication lag with relative Mean Squared Error (MSE) close to 0.60, while in Bosnia and Herzegovina, Alihodzic (2020) find longer delays of around 60 days and higher errors with Mean Absolute Percentage Error between 1.05 and 1.49. By contrast, Spain achieves shorter lags of about 30 days with low Root Mean Squared Error (RMSE) near 0.20 (Camacho, 2021), and for the euro area, Banbura et al. (2013) show that Bayesian VARs and DFMs reduce RMSE relative to autoregressive benchmarks. In Italy, mixed-frequency models lower the RMSE of industrial production forecasts by roughly 10% compared with bridge regressions (Bragoli et al., 2015). Similarly, Statistics Canada (MAE $\approx$ 0.02) and the U.S. GDPNow (MAE $\approx$ 1.26) deliver ultra-rapid estimates, though at the cost of higher dispersion. Mexico's PIBO demonstrates that competitive results (MAE as low as 0.08–0.14) can be obtained within 30 days, while the Timely Private Consumption Indicator broadens coverage to consumption but with larger errors (MAE between 0.50 and 1.06). Less timely estimates (60–90 days), such as those for the U.K. (RMSE $\approx$ 0.14), tend to benefit from more complete datasets and therefore higher precision. Overall, these experiences highlight that while ML can provide very rapid signals, factor and econometric models are generally more effective at balancing speed and reliability.

Consequently, the performance of the IOAE is expected to decline when estimates are released with shorter publication lags.

4.2. IOAE as timely estimates of Quarterly GDP

As we have commented, the IGAE serves as the monthly proxy variable for GDP. In this sense, the IOAE can even be used as an initial nowcast for the timely GDP which in this study is denoted as PIBO and, consequently, for GDP itself. The retrospective exercise is straightforward if we first consider the levels of the IGAE and IOAE, then aggregate them to a quarterly frequency by averaging the monthly observations, and finally compute the quarterly and annual variations that allow for a comparison with GDP values.

In this context, the PIBO provides information about the reference quarter one month after its conclusion. The IOAE at $T + 1$ offers an advance of half a month compared to PIBO, meaning it is slightly less timely. However, at $T + 2$, it provides a lead of half a month relative to PIBO. It is important to note that the IGAE is contemporaneous with GDP.

Figure 3 summarizes the role of the IGAE as a predictor of GDP and PIBO, as well as the behavior of the IOAE at $T + 1$ and $T + 2$.

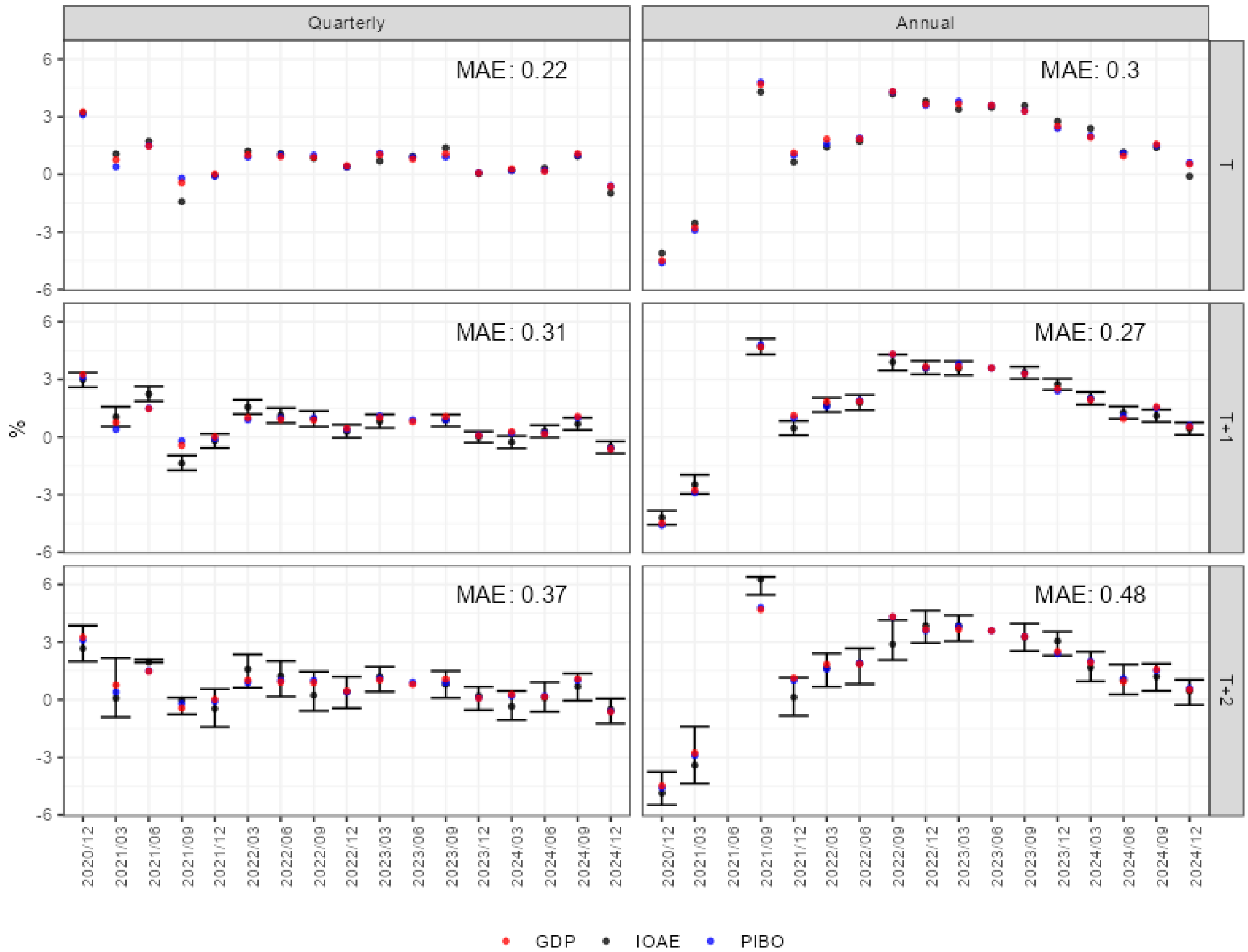

Figure 3: The real-time performance of the IOAE is shown by black points, together with its 95% confidence intervals (black lines), and is compared with observed GDP (red points) and PIBO estimates (blue points). The MAE is computed with respect to the observed GDP.

It is interesting to observe that the IGAE, in terms of annual variations, is less accurate than the IOAE at $T + 1$. Moreover, the errors at $T + 2$ are well below the 1.06 threshold and even below the 0.61 reference for $T + 1$. In all cases, the estimates follow the trend of the observed values, with correlations at $T + 1$ of 0.99 for both quarterly and annual variations, and 0.88 and 0.99, respectively, for $T + 2$.

Consequently, given that the PIBO serves as a flash estimate of GDP within the framework of official statistics, the IOAE can also function as a nowcast of GDP, providing relatively accurate estimates that offer timely signals about GDP dynamics.

As a remark, for the second quarter of 2023, the IOAE approach does not generate GDP estimates due to the change in the base year, which makes the results not comparable with the levels of the index.

4.3. Sentiment analysis

Social media platforms such as X provide timely information on how economic indicators are discussed and perceived by analysts and the general public. While the previous sections focus on the predictive performance of the IOAE, it is also relevant—particularly from the perspective of official statistics—to document how this experimental indicator is received in public discourse.

Accordingly, this subsection presents a brief descriptive sentiment analysis of social-media posts related to the IOAE. Importantly, this analysis is not incorporated into the nowcasting framework and does not affect any of the forecast evaluation results reported in this paper. Its sole purpose is to assess public engagement with and perceptions of the IOAE as a timely indicator of economic activity, which is relevant for communication, transparency, and the potential adoption of experimental indicators within official statistics.

The analysis is based on 2,290 posts collected from X that explicitly reference the IOAE. Sentiment classification is performed using BETO, a Spanish-language model based on BERT, which assigns each post to one of three categories: positive, neutral, or negative. Standard text-cleaning procedures were applied prior to classification to ensure consistency and linguistic accuracy; full technical details and code are provided in the online replication repository.[8]

Figure 4 summarizes the distribution of sentiment across posts.

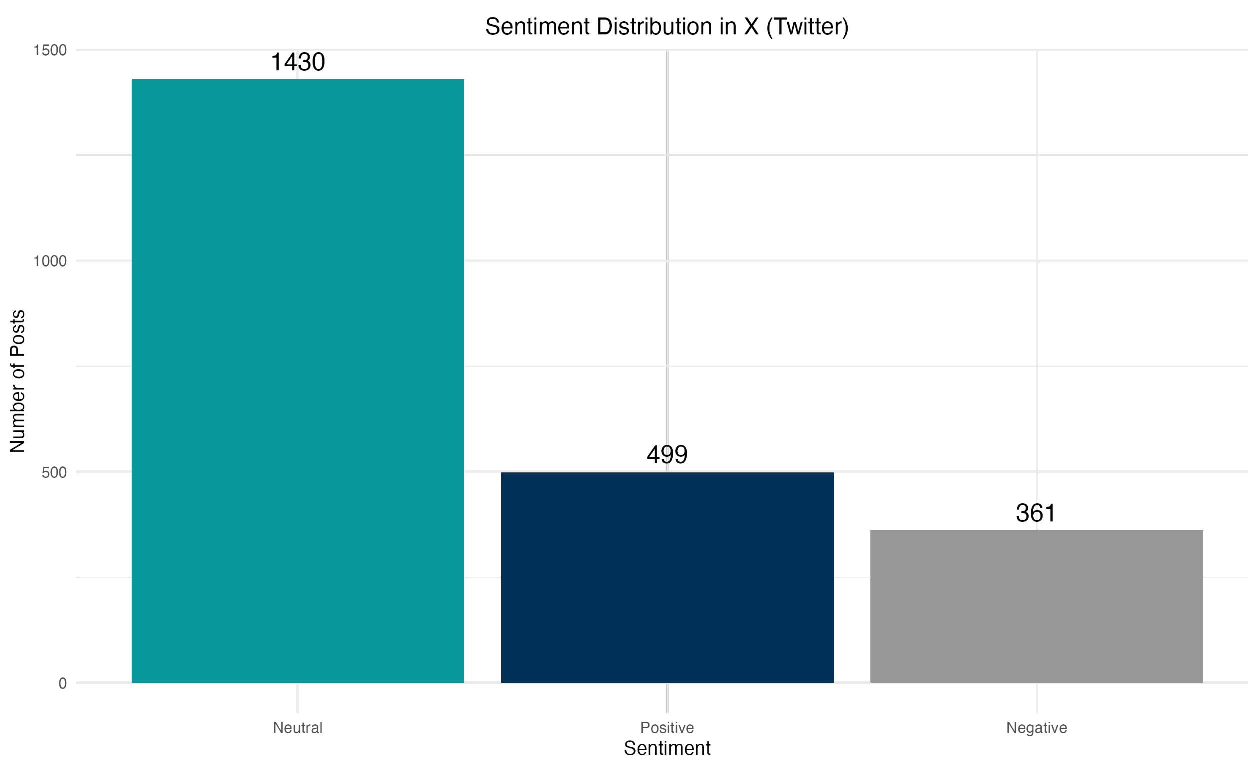

Figure 4: Distribution of sentiment (positive, neutral, and negative) in X posts related to the IOAE.

Neutral sentiment predominates, accounting for approximately 62% of the observations, followed by positive (22%) and negative (16%) sentiment. This pattern suggests that discussions of the IOAE on social media are largely informational and data-oriented rather than strongly opinionated or emotionally charged.

[8]https://github.com/ReneMaruri/ioae-sentiment-analysis

Overall, this descriptive evidence indicates that the IOAE has attracted attention among economic agents and is primarily discussed in a neutral and analytical manner. While not intended to enhance forecasting performance, this sentiment analysis complements the real-time evaluation by illustrating the IOAE's relevance and visibility in the Mexican context as an experimental official statistic.

5. Empirical analysis

In this section, we present the results of the components of the DFM using the IOAE procedure. Additionally, we conduct a pseudo real-time analysis to compare the results with those obtained from other nowcasting techniques, and complement the analysis with a robustness assessment.

5.1. Components of the DFM

As previously noted, the IOAE approach assumes $r = 1$. Applying the three information criteria proposed by Bai and Ng (2002), the edge distribution procedure of Onatski (2010), and the eigenvalue ratio approach of Ahn and Horenstein (2013) to determine the number of factors by scaling $\mathbf{X}_t$, we obtain $\hat{r} = 1$ under the second and third methodologies, whereas Bai and Ng (2002) yields $\hat{r} = 0$ in all three cases. Several conclusions can be drawn from this outcome. In practice, Onatski (2010) and Ahn and Horenstein (2013) are robust to weak factors or to large variance in the idiosyncratic errors relative to the common factors, while Bai and Ng (2002) tends to perform better when the common component is stronger relative to the idiosyncratic errors. Thus, one possible explanation for this result is that the DFM exhibits serial correlation in the idiosyncratic errors, leading Bai and Ng (2002) to estimate $\hat{r} = 0$, whereas the other approaches are able to detect the correct number of factors. Repeating this analysis over the last $H_t = 73$ months (November 2018–November 2024), the results remain unchanged, showing consistency under the pseudo-real-time sample. Consequently, there are statistical reasons to support the argument that the number of factors in the IOAE procedure is $\hat{r} = 1$.

Hence, the components of the DFM, obtained using the methods described in Section 3 and the data presented in Section 4, are shown below. Figure 5 displays the estimated factor loadings.

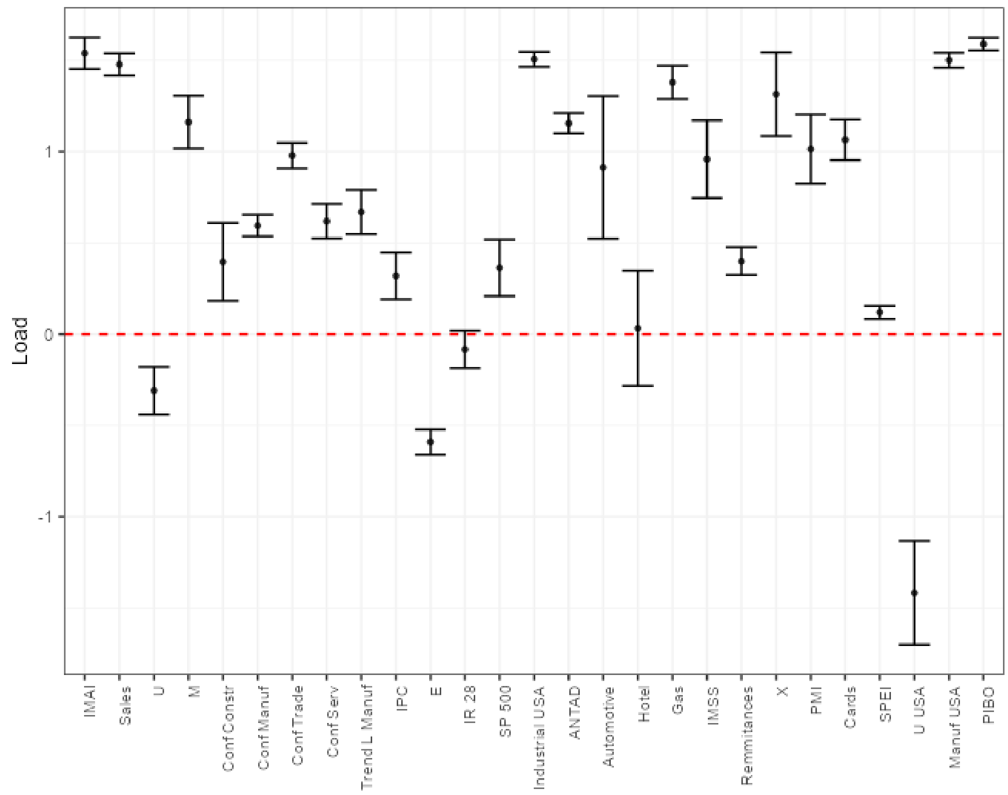

Figure 5: Estimated factor loadings for each variable in the IOAE, with 95% confidence intervals.

We observe that most of the time series are relevant, with the exception of *IR 28* and *Hotel*, which are not statistically significant at the 95% confidence level. This result is consistent with the linear correlations presented in Table 1. Additionally, the signs of the estimated loadings are economically meaningful; for instance, the coefficients for the Mexican and U.S. unemployment rates, as well as the exchange rate, are negative. In contrast, all other series exhibit a positive relationship, with *IMAI*, *Manuf USA*, and *PIBO* standing out as particularly influential.

Note that in this case, no Google Trends topic was selected by the LASSO procedure. However, this does not imply that some of them were not selected in previous periods, for instance, in the pseudo real-time analysis.

Figure 6 shows the behavior of the target time series and the corresponding factor estimates.

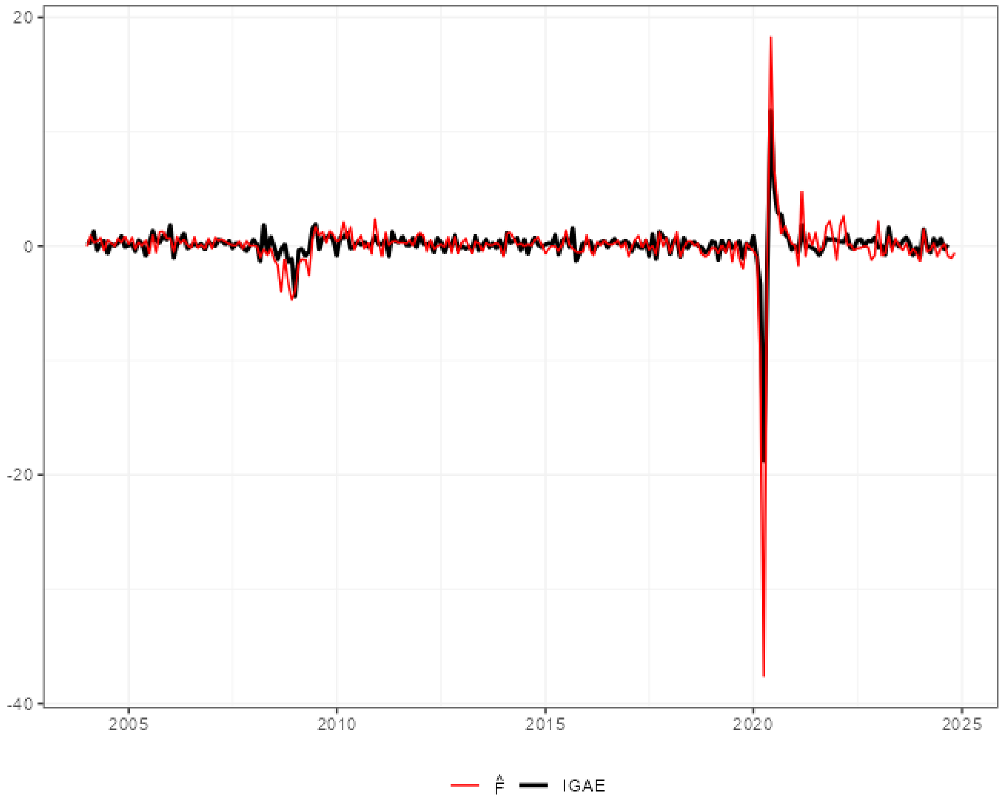

Figure 6: Common factor (red line) and target time series (black line) on a normalized scale, January 2004–January 2025.

We can observe that both time series are extremely close. In fact, the linear correlation between them is 0.96, with the advantage that the factor provides two additional observations, allowing for current nowcasts to be produced. This timely information is used in the pseudo real-time analysis to evaluate predictive performance during an out-of-sample period. In each nowcast iteration, we exploit the timeliness of the common factor to generate a prompt estimate of the IGAE.

The target time series, as well as the transformed variables collected in X_t, tend to exhibit stationarity. In this context, the DFM is assumed to be stationary, and according to PANIC tests, both the dynamic factor and the idiosyncratic components are stationary. When $r = 1$, testing the stationarity of the factors is equivalent to implementing the Augmented Dickey–Fuller (ADF) test (Bai and Ng, 2004), while the stochastic properties of the idiosyncratic components are examined using the pooled test:

$$S = \frac{-2\sum_{i=1}^{N} \log s_i - 2N}{\sqrt{4N}},$$

where s_i denotes the p-value corresponding to the Dickey–Fuller test of the i-th idiosyncratic residual.

Specifically, the ADF test applied to the estimated common factor, with a constant and three lags selected according to the Schwarz information criterion, yields a statistic of -9.40, rejecting the null of a unit root. By contrast, the pooled stationarity test for the idiosyncratic components yields a statistic of 17.91, indicating strong evidence of stationarity.

5.2. Pseudo real-time analysis

The pseudo-real-time analysis aims to provide a timely estimation of the MV of the IGAE for the period October 2018–October 2024 when $h = 1$, and for the period November 2018– November 2024 when $h = 2$ ($H_t = 73$). Hence, the two-step-ahead nowcasts are generated using an expanding window. This procedure relies on the information available at each point in time, as detailed in Table 1, in order to replicate the real-time conditions under which the IOAE produces its nowcasts. It is referred to as a pseudo-real-time analysis because the dataset includes revised values available as of February 14, 2025.

Hence, in each iteration the true IGAE values at $T + 1$ and $T + 2$ are treated as unknown, while $\mathbf{X}_t$ is updated according to Table 1 and illustrated in Figure 1. This ensures that it contains only the information available at that point in time, thereby replicating past real-time conditions.

For illustration, during the most critical months of the pandemic—April and May 2020—nowcasts are constructed using the database available as of February 14, 2025, while restricting the information set to mimic real-time data availability at horizons $T + 1$ and $T + 2$. Specifically, at horizon $T + 1$ the *Sales* series is treated as unavailable, whereas at horizon $T + 2$ additional late-reporting indicators—such as *IMAI*, exports, and imports—are also excluded from the information set.

In this way, although the underlying database may incorporate subsequent revisions, certain observations are deliberately treated as unknown in order to replicate the informational constraints prevailing at the forecast origin. This strategy allows us to emulate the real-time conditions under which the nowcasts would have been produced, thereby preventing look-ahead bias in the forecast evaluation exercise.

The same information-set restrictions are applied consistently across all competing approaches. In the case of LASSO-based nowcasts, since the regression does not rely on factors estimated up to $T + 2$, some predictors remain unobserved at the forecast origin; in such cases, the most recently available observation is used. For regressions based on individual indicators, only information updated up to $T + 2$ is incorporated.

In this paper, IOAE-based nowcasts are obtained by estimating the regression equation

$$\hat{y}_t = \hat{\alpha} + \hat{\boldsymbol{\beta}}' \hat{\mathbf{F}}_t,$$

without incorporating an additional ARMA structure in the error term. This specification is adopted deliberately in order to isolate and exploit the predictive content of the dynamic factor itself, yielding a parsimonious implementation of the original approach. The implications of augmenting this baseline specification with an ARMA component in the disturbance term—thereby combining factor-based and serial-dependence dynamics—are left for future research and discussed separately. Therefore, we follow the IOAE specification proposed by Giannone et al. (2008), i.e., we refrain from incorporating an ARMA component in the regression error term.

Once the MV of the IGAE are estimated, we reconstruct the implied levels and, subsequently, the AV rates. Figure 7 summarizes the pseudo-real-time performance of the IOAE over 2018:10–2024:11 for the three transformations (levels, MV, and AV), reporting observed values and 95% confidence intervals.

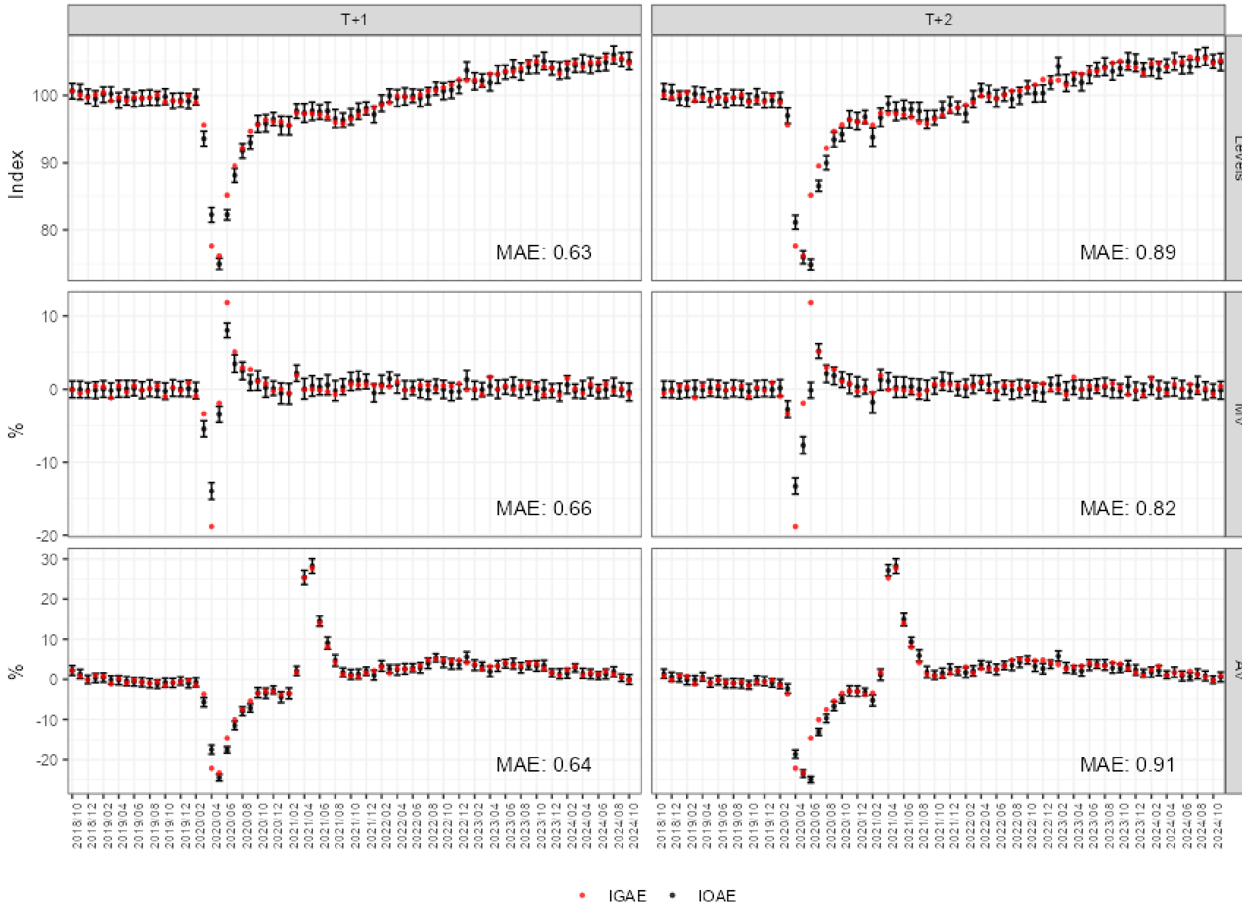

Figure 7: IOAE performance under pseudo-real-time conditions, 2018:10–2024:11. The figure reports results in levels, MV, and AV. Shaded bands denote 95% confidence intervals, and the red line shows observed values.

Overall, nowcast accuracy deteriorates as the horizon increases: for levels, the MAE is 0.63 at $T+1$ and 0.89 at $T+2$. These values are broadly comparable to those obtained in the real-time evaluation (August 2020–December 2024), despite the longer pseudo-real-time window and the inclusion of the COVID-19 shock. During April–June 2020, observed values fall outside the 95% confidence bands, highlighting the exceptional nature of that episode; outside the pandemic window, the IOAE tracks realized IGAE dynamics closely.

For MV and AV growth rates, the same qualitative message holds: errors tend to increase from $T+1$ to $T+2$, and the pandemic period remains the most challenging segment of the sample. Relative to Corona et al. (2022), the MAE values obtained here (0.64 and 0.91 for $T+1$ and $T+2$, respectively) are comparable and slightly lower than the 0.80 and 0.86 reported for November 2017–October 2020.

5.3. Comparison of IOAE to other estimation techniques

In this subsection, we compare the IOAE nowcasts obtained in the previous section using the set of forecast evaluation tests described in Subsection 2.4. Throughout the analysis, forecast accuracy is assessed using the absolute error as the loss function.

Table 2 reports the results of the DM–HAC test, where the IOAE is treated as the benchmark model and compared against individual indicators as well as alternative econometric and ML approaches. The test is conducted under the null hypothesis of equal predictive accuracy against the one-sided alternative that the IOAE yields lower expected loss.

Table 2: DM-HAC tests based on absolute error loss: IOAE versus alternative predictors and models.

	$h = 1$		$h = 2$	
Variable / Model	DM_{HAC}	p-value	DM_{HAC}	p-value
Individual indicators				
Construction confidence index	-1.0692^{*}	0.0681	-1.0466	0.1051
Manufacturing confidence index	-1.8738^{***}	0.0051	-1.2624^{*}	0.0658
Trade confidence index	-2.5681^{***}	0.0003	-1.7542^{**}	0.0188
Services confidence index	-3.2370^{***}	0.0000	-2.7592^{***}	0.0007
Mexican Stock Exchange Index	-1.9319^{***}	0.0041	-1.3862^{**}	0.0492
Exchange rate	-2.2556^{***}	0.0011	-1.6393^{**}	0.0258
Interest rate	-1.9705^{***}	0.0035	-1.3987^{**}	0.0477
S&P 500 Index	-2.0565^{***}	0.0025	-1.4811^{**}	0.0389
U.S. industrial production	-2.5087^{***}	0.0004	-1.0301	0.1087
Vehicle production	-1.1879^{**}	0.0492	-1.1687^{*}	0.0812
IMSS formal employment	-2.5587^{***}	0.0003	-1.9664^{**}	0.0101
Manufacturing orders	-2.5591^{***}	0.0003	-1.5098^{**}	0.0362
Credit and debit card transactions	-2.0248^{***}	0.0028	-1.1902^{*}	0.0774
SPEI electronic transfers	-1.8582^{***}	0.0054	-1.2715^{*}	0.0645
U.S. unemployment rate	-1.7441^{***}	0.0082	-0.1187	0.4432
U.S. manufacturing activity	-2.4811^{***}	0.0004	-0.8566	0.1522
Econometric and ML models				
LASSO	0.3839	0.7049	-1.5962^{**}	0.0289
FAVAR	-0.9231^{*}	0.0987	-1.8416^{**}	0.0146
MLP	-2.0571^{***}	0.0025	-1.9866^{***}	0.0095
ARIMA	-2.2617^{***}	0.0011	-1.9708^{**}	0.0100

Notes: The DM–HAC test is computed using absolute error loss. Negative statistics indicate lower average loss for the IOAE relative to the competing model. Statistical inference is based on HAC variance estimators. Reported p-values correspond to the one-sided alternative of superior predictive accuracy of the IOAE. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Several patterns emerge from Table 2. First, for horizon $h = 1$, the majority of DM–HAC statistics are negative and statistically significant at the 1% or 5% levels, leading to a rejection of the null hypothesis of equal predictive accuracy in favor of the IOAE. This result holds across most individual indicators and econometric benchmarks, indicating that the IOAE delivers systematically lower absolute forecast errors in the very short run.

Notable exceptions at $h = 1$ include the construction confidence index and LASSO. In the first case, the null hypothesis cannot be rejected at conventional significance levels, pointing to statistically similar predictive accuracy relative to the IOAE. In the case of LASSO, although the null hypothesis of equal predictive accuracy is also not rejected, the DM–HAC statistic is positive, suggesting weak evidence in favor of LASSO at the one-month-ahead horizon. This finding highlights that, at very short horizons, alternative regularization-based approaches may exploit overlapping information sets and occasionally match or marginally outperform the

IOAE. At this horizon, the null hypothesis of equal predictive accuracy between the IOAE and the FAVAR model is rejected at the 10% significance level, providing weak statistical evidence in favor of the IOAE under the one-sided alternative.

Second, at horizon $h = 2$, while all DM–HAC statistics remain negative—indicating that the average loss differential continues to favor the IOAE—the statistical evidence becomes more heterogeneous. In particular, the null hypothesis of equal predictive accuracy cannot be rejected for the construction confidence index and several U.S. indicators, including industrial production, manufacturing activity, and the unemployment rate. These results suggest that, at the longer horizon, external indicators capturing U.S. economic conditions contain information that is broadly comparable to that embedded in the IOAE. It is worth noting, however, that this outcome may be partly influenced by the larger variance of nowcasting errors at longer horizons, which directly affects the variability of the loss differential and, consequently, the power of the DM–HAC statistic.

Finally, comparisons with alternative econometric and ML models at $h = 2$ indicate that the IOAE delivers statistically superior nowcasts relative to LASSO, FAVAR, MLP, and ARIMA. The uniformly negative and statistically significant DM–HAC statistics for these models provide strong evidence in favor of the IOAE at the two-month-ahead horizon, reinforcing the role of the IOAE as a robust and competitive benchmark for short-term economic monitoring.

Overall, these results indicate that the informational aggregation embedded in the IOAE becomes more relevant at $h = 2$, supporting more stable predictive performance at longer nowcast horizons.

It is worth highlighting that, at horizon $h = 1$, the comparison with the LASSO model yields a positive but statistically insignificant DM–HAC statistic, pointing to comparable predictive accuracy in the very short run, as discussed above. This result is consistent with the idea that, at the $T+1$ horizon, the LASSO procedure may temporarily benefit from selecting highly timely indicators (such as industrial production, exports, imports, and, when available, PIBO) that exhibit strong contemporaneous correlation with economic activity. This short-run advantage reflects the ability of LASSO to exploit very recent information rather than evidence of systematic dominance over the IOAE. At horizon $h = 2$, this effect vanishes, and the broader information set and dynamic factor structure of the IOAE become dominant, leading to superior predictive performance.

To complement the pairwise DM–HAC comparisons, we report the results of the SPA test, using the IOAE as the benchmark, as described in Subsection 2.4. At horizon $h = 1$, the SPA statistic equals 0.3839 with a p-value of 0.4955, while at horizon $h = 2$ the test yields a statistic of 0.0000 with a p-value of 1.0000. In both cases, the null hypothesis that no competing model outperforms the IOAE cannot be rejected, providing joint evidence in favor of the IOAE once multiple comparisons are taken into account.

Overall, the DM–HAC results provide strong bilateral evidence in favor of the IOAE—particularly beyond the contemporaneous month—while the SPA test confirms that these gains remain robust in a multiple-comparison setting. These findings naturally motivate the use of the MCS procedure to further assess the relative standing of the IOAE within a broader class of nowcasting models. Intuitively, the MCS framework evaluates the joint predictive performance of all competing models and identifies a subset whose forecast accuracy is statistically indistinguishable at a given confidence level, thereby accounting for multiple comparisons. In addition, the procedure provides a relative ranking within this set, which helps to gauge how close each model is to the best-performing benchmark without imposing a strict winner–loser dichotomy.

Table 3 reports the results of the MCS procedure based on the range statistic T_R at the 10% significance level.

Table 3: MCS results based on absolute error loss.

	$h = 1$			$h = 2$		
Variable / Model	$T_{R,i}$	Rank$_R$	MCS	$T_{R,i}$	Rank$_R$	MCS
Individual indicators						
IOAE	0.3127	2	✓	-0.1650	1	✓
Construction confidence index	1.0727	5	✓	1.0843	4	✓
Manufacturing confidence index	1.2009	8	✓	1.1638	6	✓
Trade confidence index	1.7355	16	✓	1.9481	18	✓
Services confidence index	2.7494	21	✓			×
Mexican Stock Exchange Index	1.1892	7	✓	1.1885	7	✓
Exchange rate	2.1168	19	✓	2.1595	19	✓
28-day interbank interest rate	1.4930	13	✓	1.5474	14	✓
S&P 500 Index	1.4974	14	✓	1.3205	10	✓
U.S. industrial production	1.9178	18	✓	1.7781	16	✓
Vehicle production	1.0698	4	✓	1.0762	3	✓
IMSS formal employment	1.5312	15	✓	1.8227	17	✓
Manufacturing PMI	2.5906	20	✓	2.6877	20	✓
Credit and debit card transactions	1.2795	9	✓	1.2312	8	✓
SPEI electronic transfers	1.1719	6	✓	1.1241	5	✓
U.S. unemployment rate	1.4111	10	✓	0.1650	2	✓
U.S. manufacturing activity	1.7503	17	✓	1.4267	13	✓
Econometric and ML models						
LASSO	-0.3127	1	✓	1.3321	11	✓
FAVAR	0.9749	3	✓	1.3531	12	✓
MLP	1.4354	11	✓	1.5549	15	✓
ARIMA	1.4789	12	✓	1.3033	9	✓
Global p-value		0.1386			0.1488	

Notes: The statistic $T_{R,i}$ is computed using absolute error loss. Lower values indicate better relative predictive performance. Rank$_R$ denotes the relative ranking across indicators and models. A check mark (✓) indicates that the predictor is retained in the MCS for the corresponding forecast horizon, while a cross (×) denotes that the indicator is eliminated from the set.

At horizon $h = 1$, the global p-value (0.1386) exceeds the chosen significance level, implying that the null hypothesis of equal predictive ability cannot be rejected. Accordingly, no model is eliminated and all candidates belong to the MCS. In this case, the reported $T_{R,i}$ values should be interpreted as descriptive indicators of relative performance rather than evidence of statistical dominance. In particular, the LASSO model attains a low $T_{R,i}$ value and ranks among the top-performing alternatives, indicating that it delivers predictive accuracy that is statistically indistinguishable from the IOAE in the very short run.

At horizon $h = 2$, the MCS procedure excludes the services confidence index, reflecting its inferior predictive performance. All remaining models form the confidence set, with the IOAE attaining the lowest $T_{R,i}$ value and ranking first among the surviving alternatives. Although the global p-value is larger at horizon $h = 2$ (0.1488), the procedure identifies a clearly underperforming model and removes it from the confidence set. By contrast, at horizon $h = 1$, no model exhibits sufficiently poor relative performance to warrant elimination.

Overall, the MCS results indicate that several nowcasting approaches exhibit statistically indistinguishable predictive accuracy at short horizons. Nevertheless, under absolute error loss, the IOAE consistently belongs to the confidence set and tends to rank favorably as the forecast horizon lengthens. These findings complement the DM–HAC and SPA tests by highlighting the robustness of the IOAE within a multiple-comparison framework when forecast performance is evaluated using absolute errors.

To assess robustness, we replicate the DM–HAC, SPA tests, and the MCS procedure using squared error loss, $\mathcal{L}(\cdot)$, so that forecast evaluation is interpreted in terms of MSE/RMSE rather than MAE. Relative to the absolute-error case, the DM–HAC evidence becomes weaker under squared loss for both $h = 1$ and $h = 2$, indicating a reduction in statistical power in favor of the IOAE. While most DM–HAC statistics remain negative—suggesting lower average squared forecast errors for the IOAE—fewer comparisons reach conventional significance levels.

In particular, for squared loss the DM–HAC statistic associated with the U.S. unemployment rate remains negative at both horizons, indicating relative evidence in favor of the IOAE, although the null hypothesis of equal predictive accuracy cannot be rejected. This pattern illustrates that, under squared loss, differences in forecast performance become less pronounced in statistical terms, even when point estimates continue to favor the IOAE.

Consistent with this interpretation, the SPA test continues to support the benchmark role of the IOAE: for both horizons, the null hypothesis that no competing model outperforms the IOAE cannot be rejected, confirming that none of the alternative predictors delivers statistically superior performance once multiple comparisons are taken into account.

Turning to the MCS results based on squared error loss, the IOAE remains within the confidence set but its relative position changes, ranking second at $T+1$ and third at $T+2$. At the two-month-ahead horizon, U.S. unemployment and U.S. manufacturing indicators emerge as particularly competitive within the MCS, highlighting the relevance of external economic conditions for short-term activity forecasting in Mexico.

Taken together, these robustness checks indicate that although squared loss attenuates the apparent advantage of the IOAE in pairwise DM–HAC comparisons and affects its relative ranking within the MCS, the IOAE remains one of the most stable and consistently well-performing approaches across horizons under the squared error loss criterion. Importantly, the SPA tests provide complementary inferential evidence within this same loss framework, pointing to a relative predictive advantage of the IOAE despite the more demanding penalty imposed by squared errors. Additional details based on squared error loss are provided in Appendices A.2, A.3, and A.4.

It is important to emphasize that the MAE provides the most direct measure of typical forecast accuracy and is therefore adopted as our preferred loss function. By contrast, the MSE assigns disproportionate weight to large errors and is used only as a robustness check. Furthermore, when applying the Mean Absolute Scaled Error (MASE), the relative ranking of competing models remains identical to that obtained with the MAE, since the MASE is simply the MAE of each model divided by the MAE of a naïve benchmark. The main advantage of the MASE is that it standardizes forecast errors, allowing interpretation relative to the benchmark: values below one indicate superior performance. In this sense, the results obtained with the MASE are equivalent to those derived from the MAE.

As emphasized by Patton (2011) and Gneiting (2011), the choice of loss function is crucial when comparing forecast performance. The MAE is associated with the conditional median and is therefore more robust to outliers, while the RMSE corresponds to the conditional mean and penalizes large errors disproportionately. This difference implies that the ranking of models may vary depending on whether typical forecast accuracy (captured by the MAE) or sensitivity to large deviations (captured by the RMSE) is prioritized. In practice, this justifies analyzing results under both loss functions to ensure that conclusions are not driven by the particular properties of a single metric. Nevertheless, in this study we place greater emphasis on the MAE, given its robustness to outliers and its more direct interpretation of nowcast accuracy.

An explanation for the relative performance of the IOAE across the different forecast evaluation exercises can be summarized as follows:

- The IOAE aggregates predictive information by capturing the contemporaneous comovement of a broad set of individual time series. Unlike traditional models that rely primarily on lagged relationships, the IOAE exploits the real-time availability of factor estimates up to $T+2$, allowing it to incorporate timely signals that are not yet fully reflected in individual series.
- The LASSO regression may exhibit competitive performance at the $T+1$ horizon due to the availability of highly informative contemporaneous indicators, such as industrial production, exports, and imports, which account for substantial shares of the IGAE. However, at the $T+2$ horizon, when several of these series are not yet updated, the performance of the LASSO deteriorates, highlighting its sensitivity to data availability in real time.
- The relatively weaker performance of the MLP and FAVAR models, which rely on lagged representations of the common factor, suggests that for nowcasting applications the inclusion of additional temporal or nonlinear structure may introduce noise rather than signal. In such settings, increased model complexity—whether linear or nonlinear—does not necessarily translate into improved predictive accuracy.

In the following subsection, we assess the robustness of the IOAE, based on DFMs, relative to alternative procedures by varying the number of factors, evaluating nowcast errors during the non-COVID-19 period, and implementing rolling windows in the pseudo-real-time analysis.

5.4. Robustness analysis

This subsection assesses the robustness of the IOAE nowcasting results under alternative and economically plausible evaluation designs. The objective is to verify that the main conclusions regarding the IOAE's predictive performance do not hinge on a particular choice of sample period, window scheme, loss function, or model specification.

All robustness exercises are conducted within a pseudo real-time framework. At each forecast origin, model estimation, factor extraction, and parameter selection rely exclusively on information available up to that date, thereby emulating real-world nowcasting conditions. While revised data are used for evaluation purposes—as is standard in the nowcasting literature—no future information is employed in model estimation or tuning. This design ensures that the robustness results reflect feasible real-time implementation rather than ex post optimization.

To this end, we examine the sensitivity of the results along four dimensions:

- **Expanding and rolling windows:** The IOAE is estimated using both expanding and rolling window schemes. Under expanding windows, the estimation sample initially spans $t = 1, \dots, T$ and is updated sequentially as new information becomes available. Under rolling windows, the sample shifts at each forecast origin, taking the form $t = s, \dots, T + s - 1$, for $s = 1, \dots, H_t$. This comparison assesses whether results are driven by long-run information accumulation or by more local dynamics.
- **Number of factors:** While the baseline IOAE specification uses $\hat{r} = 1$ factor, we also consider $\hat{r} = 2$ and $\hat{r} = 3$ to evaluate the impact of including additional common components in the nowcasting exercise.

- **Sample periods:** We evaluate performance using the full sample and a subsample that excludes the months most severely affected by the COVID-19 pandemic in Mexico (March–June 2020).[9]
- **Loss functions:** Forecast accuracy is evaluated using both the MAE and the RMSE, the latter placing greater weight on large forecast errors.

The results of these robustness checks are summarized in Table 4.

Table 4: MAE and RMSE of the IOAE across window schemes, number of factors, and sample periods.

	Expanding window						Rolling window					
	$r = 1$		$r = 2$		$r = 3$		$r = 1$		$r = 2$		$r = 3$	
Sample	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$	$h = 1$	$h = 2$
Full sample												
MAE	0.659	0.824†	0.638	1.024	0.582†	1.103	0.671	0.867	0.649	0.864	0.611	0.953
RMSE	0.994	1.796	0.830	2.500	0.704†	2.691	1.024	1.868	0.880	1.536†	0.799	2.047
Excluding COVID-19												
MAE	0.520†	0.526†	0.545	0.570	0.541	0.594	0.523	0.559	0.556	0.616	0.548	0.607
RMSE	0.634†	0.634†	0.636	0.695	0.634†	0.727	0.657	0.666	0.668	0.774	0.649	0.748

Notes: The symbol † denotes the lowest forecast error within each horizon ($h = 1$ or $h = 2$), loss function (MAE or RMSE), and window scheme (expanding or rolling), across alternative factor specifications. Comparisons are conditional on the evaluation design and are intended to highlight relative performance rather than a unique globally optimal specification.

Several clear patterns emerge. First, differences between expanding and rolling windows are generally modest. Across both horizons, the relative performance of the IOAE remains stable, indicating that its nowcasting accuracy is not driven by a specific window scheme. While rolling windows occasionally yield slightly higher errors, the overall ranking of the IOAE relative to alternative specifications does not change materially.

Second, the number of factors mainly affects forecast accuracy at horizon $h = 1$, particularly in the full sample that includes the COVID-19 period. Multi-factor specifications can improve short-horizon performance in the presence of large and abrupt shocks. However, at horizon $h = 2$, additional factors tend to increase forecast errors, and the baseline one-factor IOAE specification consistently performs best under expanding windows. This result suggests that the gains from additional factors are context-specific rather than structural.

Third, excluding the months most severely affected by the COVID-19 pandemic substantially alters the error dynamics across horizons. When these months are removed, the increase in forecast errors from $h = 1$ to $h = 2$ becomes negligible. This finding indicates that the pronounced deterioration in nowcasting accuracy at longer horizons is largely driven by an extreme and unprecedented shock, rather than by an intrinsic limitation of the IOAE framework. It also highlights an important avenue for future research, namely the development of auxiliary strategies to improve nowcasting performance during rare but severe disruptions.

A fourth robustness insight concerns the role of the loss function used to evaluate nowcasting performance. As expected, RMSE values are systematically larger than MAE values, particularly when the full sample is considered. This difference is driven by the higher sensitivity of RMSE to large forecast errors, which are concentrated during extreme episodes such as the COVID-19 shock. When the most disruptive pandemic

[9]In Mexico, the strongest economic disruption associated with the COVID-19 pandemic took place between March and June 2020, coinciding with the *Jornada Nacional de Sana Distancia* and the suspension of non-essential activities. Although subsequent waves followed, this initial period concentrated the sharpest declines in mobility, employment, and production.

months are excluded, the gap between MAE and RMSE narrows substantially, and RMSE increases only marginally from $h = 1$ to $h = 2$.

This result indicates that the relative deterioration of RMSE-based performance at longer horizons is not a pervasive feature of the IOAE, but rather a reflection of a small number of extreme observations. Importantly, the main ranking and qualitative conclusions regarding the IOAE remain unchanged across loss functions, reinforcing the view that its strong performance is not an artifact of a particular evaluation metric but holds under alternative and economically meaningful loss criteria.

To complement the full-sample perspective, we examine the temporal evolution of forecast errors through cumulative loss profiles reported in Appendices A.5 and A.6 for MAE-based evaluations, and in Appendices A.7 and A.8 for RMSE-based evaluations. In this exercise, econometric and ML methods are evaluated relative to a benchmark defined by a DFM based on a single common factor, estimated using an expanding-window scheme—i.e., the IOAE approach. While formal tests of forecast instability across subsamples—such as those proposed by Giacomini and Rossi (2010)—are not implemented in this study, these cumulative metrics provide an expanding-window diagnostic view akin to fluctuation-type analyses. This framework allows us to assess the sensitivity of forecast performance to major macroeconomic disruptions and to track the relative stability of competing models over time.

The evidence indicates that the IOAE exhibits comparatively lower error sensitivity, particularly during the COVID-19 shock, reinforcing the robustness of its predictive performance even under periods of pronounced macroeconomic stress. This pattern is consistently observed under both MAE and RMSE loss functions, suggesting that the indicator's relative advantage is not driven by a specific evaluation metric but reflects stable gains in nowcasting accuracy.

At the same time, alternative approaches such as LASSO and FAVAR display competitive performance under certain configurations—especially at the one-month-ahead horizon—highlighting that regularization-based and factor-augmented structures can effectively exploit short-run informational content embedded in the data. Nevertheless, across expanding windows, loss functions, and sample partitions, the IOAE remains systematically among the best-performing models.

Taken together, these robustness exercises provide reassurance that the main conclusions of the paper are not driven by a particular evaluation design, window choice, or loss function. In particular, the superior performance of the IOAE at horizon $T+2$ remains intact across expanding and rolling windows, alternative factor specifications, and both MAE- and RMSE-based evaluations. Moreover, the comparison between MAE and RMSE reveals that differences in forecast accuracy are largely attributable to a limited number of extreme observations concentrated during the COVID-19 episode, while outside this exceptional period error dynamics and model rankings remain broadly stable.

Finally, our evaluation strategy deliberately prioritizes full-sample inferential procedures. In this regard, the DM–HAC, SPA, and MCS frameworks provide complementary and conclusive evidence on relative predictive performance by jointly assessing forecast accuracy over the entire evaluation period and controlling for multiple-model comparisons.

6. Discussion and policy implications

This paper builds directly on the IOAE framework originally proposed in Corona et al. (2022), which introduced a DFM–based approach for the timely monitoring of Mexico's economic activity. Rather than re-framing the analysis as a direct competition between alternative IOAE specifications, the objective of the present study is to extend and strengthen the empirical foundations of that framework by providing a comprehensive and transparent evaluation of its predictive performance.

A first and central contribution of this paper relative to the original framework is the systematic comparison of the IOAE against a broad set of alternative nowcasting approaches. While the original study documented the operational feasibility and descriptive accuracy of the IOAE, it did not formally assess its performance relative to competing econometric and machine learning models. By incorporating pairwise tests of predictive accuracy (DM–HAC), joint tests that account for multiple comparisons (SPA), and the MCS procedure, this paper establishes that the IOAE is statistically competitive with, and often superior to, widely used alternatives. This result substantially strengthens the empirical credibility of the IOAE as a real-time monitoring tool within the system of official statistics.

Second, the extended evaluation conducted in this paper reveals an important interpretative implication that was not emphasized in the baseline approach. The strong and robust performance of the IOAE at horizon $h = 2$ suggests that it can serve not only as a contemporaneous indicator of the IGAE, but also as an informative short-term predictor of quarterly GDP. In this sense, the IOAE provides timely signals about aggregate economic conditions before the release of official quarterly GDP estimates and, in particular, before the availability of PIBO. This finding broadens the potential policy relevance of the IOAE, especially for institutions that require early signals about turning points in economic activity.

Third, the results of this study suggest that a relatively parsimonious implementation of the IOAE—augmented with PIBO as an additional predictor in the factor estimation stage—may be sufficient for practical nowcasting purposes. The inclusion of PIBO constitutes a substantive difference with respect to the foundational approach, where this variable was not incorporated into the information set. Our results indicate that exploiting this quarterly signal within a dynamic factor framework can enhance short-horizon nowcasting performance, particularly at $h = 1$, without compromising stability at longer horizons.

In contrast to the initial methodological framework, which modeled the idiosyncratic error component using ARMA structures, we adopt the original regression-on-factors representation of Giannone et al. (2008). The pseudo real-time evidence presented in this paper suggests that this simpler specification delivers competitive and robust nowcasting results. While a formal statistical test of whether modeling serial correlation in the idiosyncratic errors yields systematic improvements lies beyond the scope of this study, the comparison of pseudo real-time performance across vintages provides indicative evidence that the gains from ARMA-based error modeling may be limited in this setting.

From a policy and operational perspective, this finding is relevant. It implies that the IOAE can be implemented in a more transparent and computationally efficient manner, while maintaining predictive accuracy comparable to more complex specifications. Nonetheless, a deeper investigation of the conditions under which modeling serial correlation in the measurement errors improves nowcasting performance remains an important avenue for future research.

Beyond these methodological considerations, the IOAE provides accurate and timely signals of Mexico's economic activity that can be directly incorporated into policy analysis and decision-making. In particular, combining the IOAE with Mexico's System of Cyclical Indicators (SIC)[10] can enhance the real-time assessment of the business cycle, given that the SIC uses the IGAE as its reference variable.

Like any nowcasting tool, the IOAE is subject to forecast uncertainty. Sudden policy announcements, financial shocks, or extraordinary events may not be fully captured by the available information set, and the IGAE itself is subject to subsequent revisions. A key methodological strength of the IOAE framework, however, is its ability to provide not only point nowcasts but also confidence intervals. Unlike many ML approaches that focus exclusively on point prediction, the factor-based structure of the IOAE allows for formal uncertainty quantification, which is essential for informed policy decisions, particularly during periods of

[10]https://www.inegi.org.mx/app/reloj/tablero.html

heightened volatility such as the COVID-19 pandemic.

Overall, the evidence presented in this paper indicates that the IOAE should be viewed not as a replacement for official statistics, but as a complementary tool that enhances the timeliness, interpretability, and robustness of short-term economic monitoring in Mexico. In this sense, the IOAE also provides a preliminary assessment of the PIBO, offering early signals of aggregate economic activity while remaining fully consistent with the official statistical framework.

7. Conclusions and further research

This study presented a comprehensive analysis of the performance of the IOAE, published monthly by INEGI since October 2020. The evaluation was conducted in two ways: first, in real time, by comparing the initial nowcasts with the subsequently observed IGAE values published each month; and second, through a pseudo real-time exercise, in which estimates obtained via an expanding window were compared against those generated by alternative nowcasting methods.

The results lead to several important conclusions. The IOAE is highly competitive when compared to international practices, particularly at horizon $T + 2$, and can be reliably used as a nowcasting tool for GDP, providing a timelier alternative to the final GDP estimate. Furthermore, when compared to other estimation techniques—such as individual regressions, LASSO, FAVAR, MLP, or ARIMA models—the IOAE generally demonstrates superior performance, especially at $T + 2$. Notably, however, LASSO exhibits strong predictive power at $T + 1$.

In this context, several avenues for future research are proposed. It would be relevant to compare the IOAE's performance with additional techniques such as elastic net, Random Forest, Extreme Gradient Boosting, Support Vector Machines, and Partial Least Squares. Moreover, combining LASSO with DFMs could yield improvements in nowcasting accuracy. Further research will also focus on evaluating the IOAE's performance in relation to timely estimates of the primary, secondary, and tertiary sectors. Finally, related with the modeling, combining results from various estimation methods is a promising line of inquiry.

Finally, an avenue for future research is to extend the evaluation of the IOAE through a more granular analysis of forecast instability across horizons and subperiods using rolling-window schemes in the spirit of Giacomini and Rossi (2010), where instability is evaluated through local loss differentials with HAC correction. While our robustness analysis already shows that the main conclusions are stable across alternative window specifications, factor structures, loss functions, and sample periods—including exercises that exclude the COVID-19 shock—and is further supported by expanding cumulative forecast error profiles, a rolling instability framework could provide additional insights into the temporal evolution of nowcasting performance across forecast horizons and economic regimes. Such an approach would be particularly informative during periods of heightened macroeconomic stress (see, e.g., González-Rivera et al., 2024).

A second promising direction for future research is the use of nowcast ensembles. By combining forecasts from econometric and ML models, ensemble methods may further improve accuracy and robustness, particularly during periods characterized by heightened uncertainty or structural change.

In addition, while the evidence presented in this paper suggests that a parsimonious IOAE specification delivers competitive performance, a more formal and systematic evaluation of the role of ARMA-type dynamics in the regression errors remains an interesting extension. Future work could explicitly assess whether modeling serial correlation in the measurement errors yields statistically and economically meaningful gains relative to simpler implementations, and under which conditions such gains might arise.

Exploring these avenues could further enhance the reliability and operational appeal of the IOAE framework for short-term economic monitoring and policy analysis.

Declarations

- **Funding**
  The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
- **Competing interests**
  The authors have no relevant financial or non-financial interests to disclose.

Data and Program Availability

All data and programs used in this study are publicly available. The complete set of datasets and replication codes can be accessed at the following GitHub repository: https://github.com/FranciscoL-B/ioae_assesment. This repository contains all materials necessary to reproduce the empirical results presented in the paper.

Disclaimer

The views and opinions expressed in this paper are those of the authors and do not reflect the official position of the INEGI. The authors are solely responsible for any errors or omissions.

References

Aastveit, Knut Are (2010), "Forecasting GDP with a dynamic factor model for Norway." *Empirical Economics*, 39, 95–114.

Ahn, Seung C. and Alex R. Horenstein (2013), "Eigenvalue Ratio Test for the Number of Factors." *Econometrica*, 81, 1203–1227.

Alihodzic, Adnan (2020), "Nowcasting GDP of Bosnia and Herzegovina: A Comparative Approach." *Croatian Review of Economic, Business and Social Statistics*, 6, 1–15.

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

Bai, J. and S. Ng (2004), "A PANIC attack on unit roots and cointegration." *Econometrica*, 72, 1127–1177.

Bai, Jushan (2004), "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data." *Journal of Econometrics*, 122, 137–183.

Bai, Jushan and Serena Ng (2007), "Determining the number of primitive shocks in factor models." *Journal of Business & Economic Statistics*, 25, 52–60.

Banbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2013), "Nowcasting." In *Handbook of Economic Forecasting* (G. Elliott and A. Timmermann, eds.), volume 2A, 193–224, Elsevier.

Bantis, Evripidis, Michael P Clements, and Andrew Urquhart (2023), "Forecasting GDP growth rates in the United States and Brazil using Google Trends." *International Journal of Forecasting*, 39, 1909–1924.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz (2005), "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics*, 120, 387–422.

Bernardi, Mauro and Leopoldo Catania (2015), "Model Confidence Set and Forecast Combination." *International Journal of Forecasting*, 31, 517–527.

Bolivar, Osmar (2024), "GDP nowcasting: A machine learning and remote sensing data-based approach for Bolivia." *Latin American Journal of Central Banking*, 5, 100126.

Bragoli, Daniela, Luca Metelli, and Michele Modugno (2015), "Nowcasting the Italian economy: A comparison of bridge and mixed-frequency models." *International Journal of Forecasting*, 31, 735–749.

Bravo-Higuera, Diego Fernando, León-Darío Parra.Bernal, Milenka-Linneth Argote-Cusi, and Grace Andrea Torres-Pineda (2024), "Colombian Agricultural Sector's Early Estimator of Gross Domestic Production Using Nowcasting and Big Data Methods." *Journal of technology management & innovation*, 19, 54–66.

Camacho, Maximo (2021), "A new economic activity indicator for Spain." *Economics Letters*, 200, 109739.

Corona, Francisco, Graciela González-Farías, and Jesús López-Pérez (2022), "Timely Estimates of the Monthly Mexican Economic Activity." *Journal of Official Statistics*, 38, 733–765.

Corona, Francisco, Pilar Poncela, and Esther Ruiz (2020), "Estimating Non-stationary Common Factors: Implications for Risk Sharing." *Computational Economics*, 55, 37–60.

Diebold, Francis X. and Roberto S. Mariano (1995), "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics*, 13, 253–263.

Doz, C., D. Giannone, and L. Reichlin (2011), "A two-step estimator for large approximate dynamic factor models based on Kalman filtering." *Journal of Econometrics*, 164, 188–205.

Fornaro, Paolo and Henri Luomaranta (2020), "Nowcasting Finnish real economic activity: a machine learning approach." *Empirical Economics*, 58, 55–71.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010), "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, 33, 1–22.

Gálvez-Soriano, O. (2020), "Nowcasting Mexico's quarterly GDP using factor models and bridge equations." *Estudios Económicos*, 35, 213–265.

Giacomini, Raffaella and Barbara Rossi (2010), "Forecast comparisons in unstable environments." *Journal of Applied Econometrics*, 25, 595–620.

Giannone, Domenico, Lucrezia Reichlin, and David Small (2008), "Nowcasting: The real-time informational content of macroeconomic data." *Journal of Monetary Economics*, 55, 665–676.

Gneiting, Tilmann (2011), "Making and evaluating point forecasts." *Journal of the American Statistical Association*, 106, 746–762.

González-Rivera, Gloria, Carlos V. Rodríguez-Caballero, and Esther Ruiz (2024), "Expecting the unexpected: Stressed scenarios for economic growth." *Journal of Applied Econometrics*, 39, 926–942.

Hansen, Peter R., Asger Lunde, and James M. Nason (2011), "The Model Confidence Set." *Econometrica*, 79, 453–497.

Hansen, Peter Reinhard (2005), "A test for superior predictive ability." *Journal of Business & Economic Statistics*, 23, 365–380.

Harvey, David, Stephen Leybourne, and Paul Newbold (1997), "Testing the equality of prediction mean squared errors." *International Journal of Forecasting*, 13, 281–291.

Hyndman, Rob J. and Yeasmin Khandakar (2008), "Automatic time series forecasting: The forecast package for R." *Journal of Statistical Software*, 27, 1–22.

Kourentzes, Nikolaos, Devon K Barrow, and Sven F Crone (2014), "Neural network ensemble operators for time series forecasting." *Expert Systems with Applications*, 41, 4235–4244.

Lahiri, Kajal and George Monokroussos (2013), "Nowcasting US GDP: The role of ISM business surveys." *International Journal of Forecasting*, 29, 644–658.

Maehashi, Kohei and Mototsugu Shintani (2020), "Macroeconomic forecasting using factor models and machine learning: an application to Japan." *Journal of the Japanese and International Economies*, 58, 101104.

Marcellino, Massimiliano and Vasja Sivec (2021), "Nowcasting GDP growth in a small open economy." *National Institute Economic Review*, 256, 127–161.

Medeiros, Marcelo C, Gabriel FR Vasconcelos, Álvaro Veiga, and Eduardo Zilberman (2021), "Forecasting inflation in a data-rich environment: the benefits of machine learning methods." *Journal of Business & Economic Statistics*, 39, 98–119.

Onatski, Alexei (2010), "Determining the Number of Factors from Empirical Distribution of Eigenvalues." *The Review of Economics and Statistics*, 92, 1004–1016.

Patton, Andrew J. (2011), "Volatility forecast comparison using imperfect volatility proxies." *Journal of Econometrics*, 160, 246–256.

Stock, J. H. and M. W. Watson (2005), "Implications of dynamic factor models for VAR analysis." *NBER*, Working Paper 11467.

Tibshirani, Robert (1996), "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.

Appendices

Table A.1: Google Trends topics

Topics	
AH1N1	Economic measures
AMLO	Migration
Ayotzinapa	Migrants
Calderon	Morena
Cartel	Dead
White House	Wall
Chapo	Omicron
China	Outsourcing
Coronavirus	Pact
Corruption	PAN
Economic crisis	Pandemic
Health crisis	PEMEX
Quarantine	Peso
Face mask	Oil
Unemployment	PRI
Dollar	Resurgence
Elections	Recession
EPN	Reforms
Gas	Salary
Homicides	Earthquake
Huachicol	Exchange rate
Inflation	Trump
Insecurity	Vaccine
N95 Mask	Violence

Table A.2: DM-HAC tests based on squared error loss: IOAE versus alternative predictors and models.

	$h = 1$		$h = 2$	
Variable / Model	DM_{HAC}	p-value	DM_{HAC}	p-value
Individual indicators				
Construction confidence index	-1.0053^{*}	0.0804	-1.0117	0.1129
Manufacturing confidence index	-1.3875^{**}	0.0272	-1.0196	0.1111
Trade confidence index	-1.4337^{**}	0.0235	-0.9407	0.1298
Services confidence index	-1.6090^{**}	0.0132	-1.2314^{*}	0.0706
Mexican Stock Exchange Index	-1.4578^{**}	0.0218	-1.1316^{*}	0.0880
Exchange rate	-1.4308^{**}	0.0237	-1.0323	0.1082
Interest rate	-1.4297^{**}	0.0238	-1.0940^{*}	0.0953
S&P 500 Index	-1.4378^{**}	0.0232	-1.0934^{*}	0.0954
U.S. industrial production	-1.3269^{**}	0.0328	-0.3284	0.3464
Vehicle production	-1.0389^{*}	0.0738	-1.0448	0.1055
IMSS formal employment	-1.7061^{***}	0.0094	-1.4323^{**}	0.0440
Manufacturing orders	-1.2334^{**}	0.0432	-0.6686	0.2110
Credit and debit card transactions	-1.4473^{**}	0.0225	-0.9127	0.1370
SPEI electronic transfers	-1.4222^{**}	0.0244	-1.0814^{*}	0.0978
U.S. unemployment rate	-1.1903^{**}	0.0489	0.9034	0.8606
U.S. manufacturing activity	-1.5625^{**}	0.0154	-0.1862	0.4114
Econometric and ML models				
LASSO	0.6533	0.8199	-1.1638^{*}	0.0820
FAVAR	-0.4342	0.2712	-1.3500^{*}	0.0537
MLP	-1.4116^{**}	0.0252	-1.2782^{*}	0.0635
ARIMA	-1.7840^{***}	0.0071	-1.7260^{**}	0.0203

Notes: The DM–HAC test is computed using squared error loss. Negative statistics indicate lower average loss for the IOAE relative to the competing model. Statistical inference is based on HAC variance estimators with the Harvey et al. (1997) small-sample correction. Reported p-values correspond to the one-sided alternative of superior predictive accuracy of the IOAE. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A.3: SPA test results based on alternative loss functions for forecast horizons $h = 1$ and $h = 2$, comparing the IOAE against competing predictors and models.

	$h = 1$		$h = 2$	
Loss function	Statistic	p-value	Statistic	p-value
Absolute error	0.3839	0.4955	0.0000	1.0000
Squared error	0.6533	0.3665	1.0062	0.2225

Table A.4: MCS results based on squared error loss.

	$h = 1$			$h = 2$		
Variable / Model	$T_{R,i}$	Rank_R	MCS	$T_{R,i}$	Rank_R	MCS
Individual indicators						
IOAE	0.5506	2	✓	0.9237	3	✓
Construction confidence index	1.0760	4	✓	1.0744	6	✓
Manufacturing confidence index	1.1595	6	✓	1.0960	8	✓
Trade confidence index	1.2471	14	✓	1.5313	20	✓
Services confidence index	2.1340	21	✓	2.0385	21	✓
Mexican Stock Exchange Index	1.1757	9	✓	1.1366	9	✓
Exchange rate	1.2306	13	✓	1.2656	16	✓
28-day interbank interest rate	1.2942	18	✓	1.2960	19	✓
S&P 500 Index	1.3600	20	✓	1.2794	17	✓
U.S. industrial production	1.2649	16	✓	1.0029	4	✓
Vehicle production	1.0794	5	✓	1.0802	7	✓
IMSS formal employment	1.2285	12	✓	1.2549	15	✓
Manufacturing PMI	1.2646	15	✓	1.2802	18	✓
Credit and debit card transactions	1.1708	7	✓	1.0713	5	✓
SPEI electronic transfers	1.2720	17	✓	1.2115	12	✓
U.S. unemployment rate	1.1725	8	✓	-0.7863	1	✓
U.S. manufacturing activity	1.3118	19	✓	0.7863	2	✓
Econometric and ML models						
LASSO	-0.5506	1	✓	1.1426	10	✓
FAVAR	0.8340	3	✓	1.2345	13	✓
MLP	1.1815	10	✓	1.1866	11	✓
ARIMA	1.2203	11	✓	1.2425	14	✓
Global p-value	0.2150			0.2316		

Notes: The statistic $T_{R,i}$ is computed using squared error loss. Lower values indicate better relative predictive performance. Rank_R denotes the relative ranking across indicators and models. A check mark (✓) indicates that the predictor is retained in the MCS for the corresponding forecast horizon, while a cross (×) denotes that the indicator is eliminated from the set.

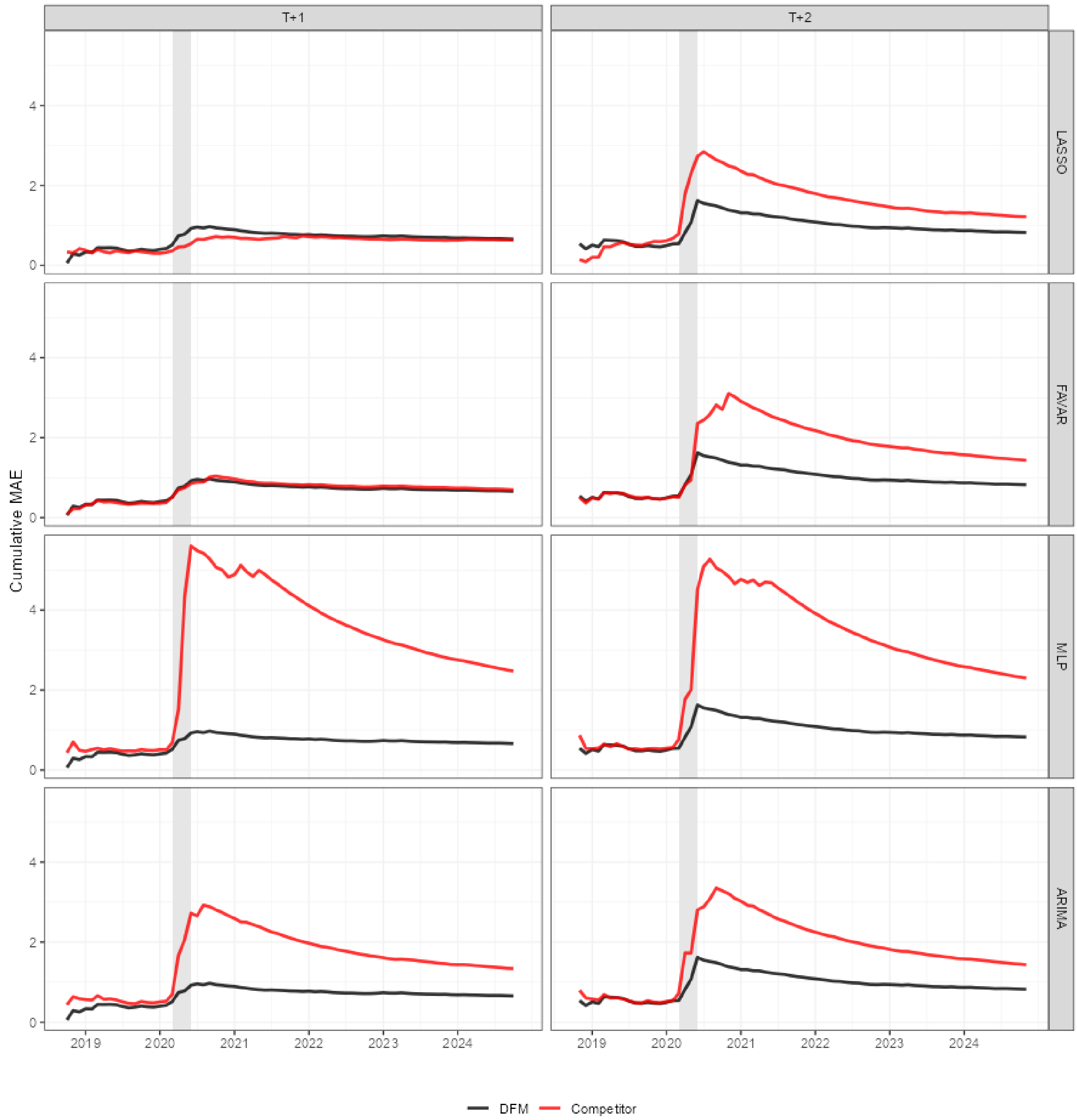

Figure A.5: Cumulative MAE for each nowcasting method. The black line depicts the IOAE approach, while the red line corresponds to the respective competitor, evaluated across the entire sample period. The shaded area corresponds to the COVID-19 period.

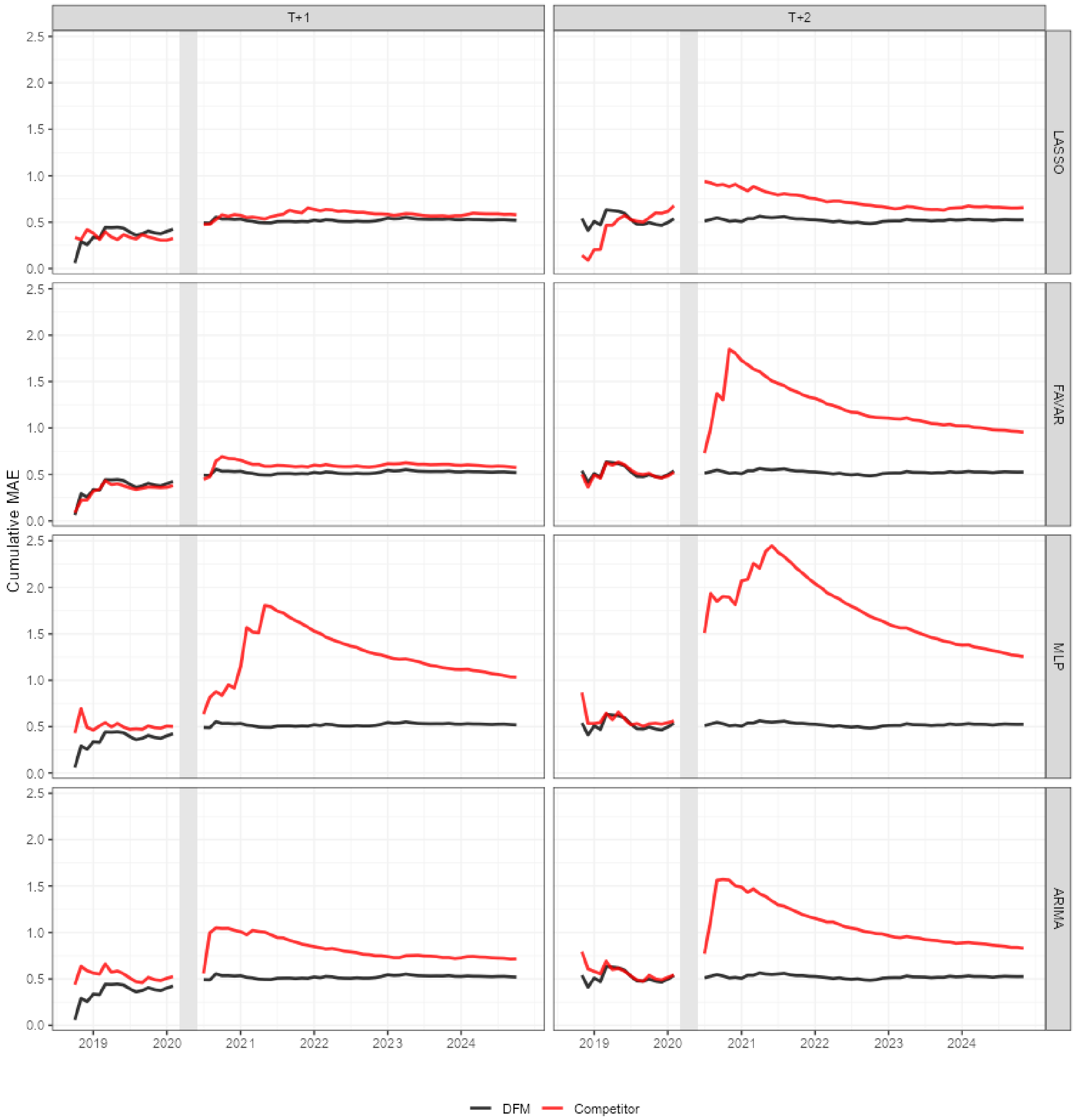

Figure A.6: Cumulative MAE for each nowcasting method. The black line depicts the IOAE approach, while the red line corresponds to the respective competitor, evaluated over the sample excluding observations from March to June 2020.The shaded area corresponds to the COVID-19 period.

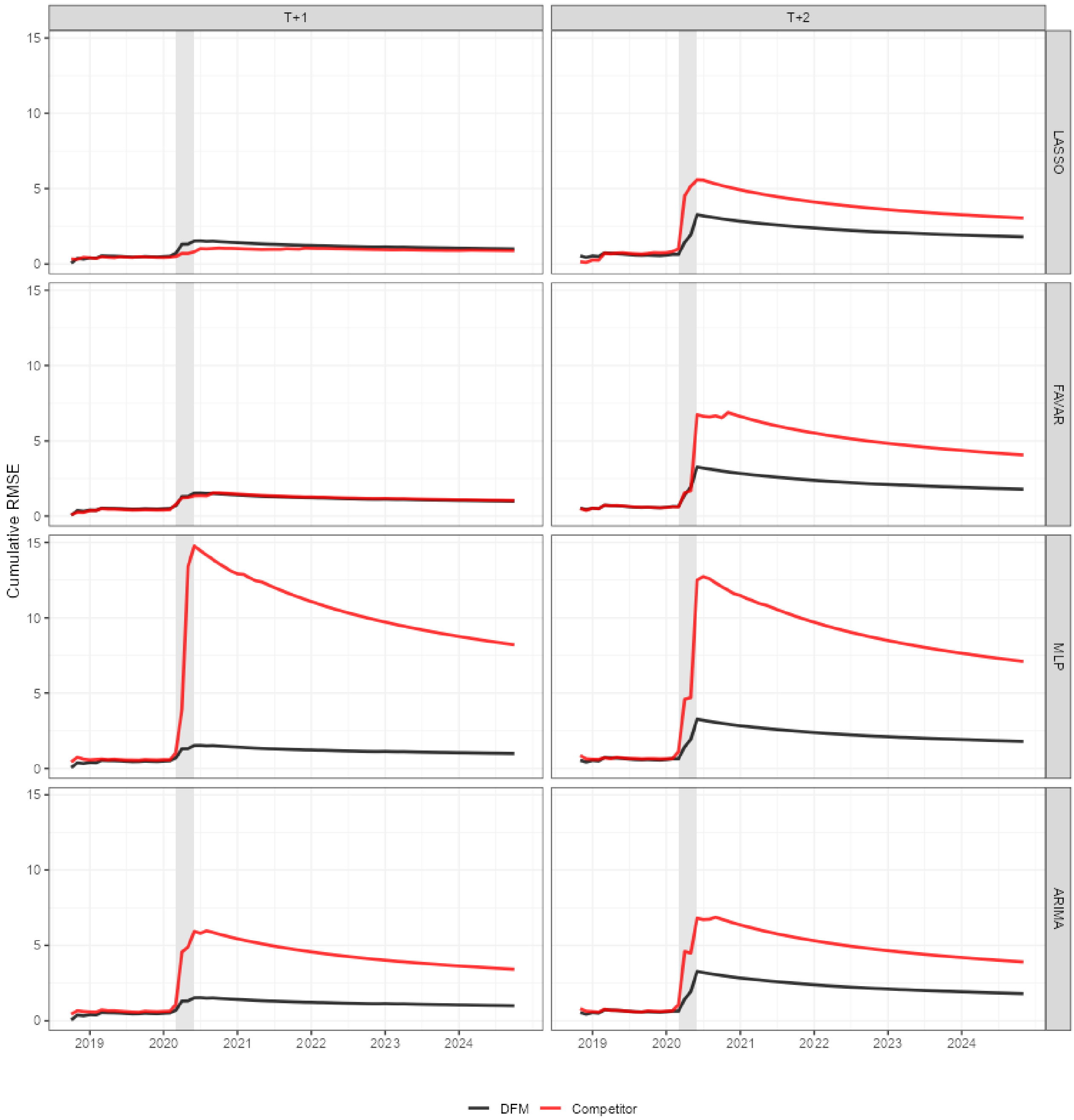

Figure A.7: Cumulative RMSE for each nowcasting method. The black line depicts the IOAE approach, while the red line corresponds to the respective competitor, evaluated across the entire sample period. The shaded area corresponds to the COVID-19 period.

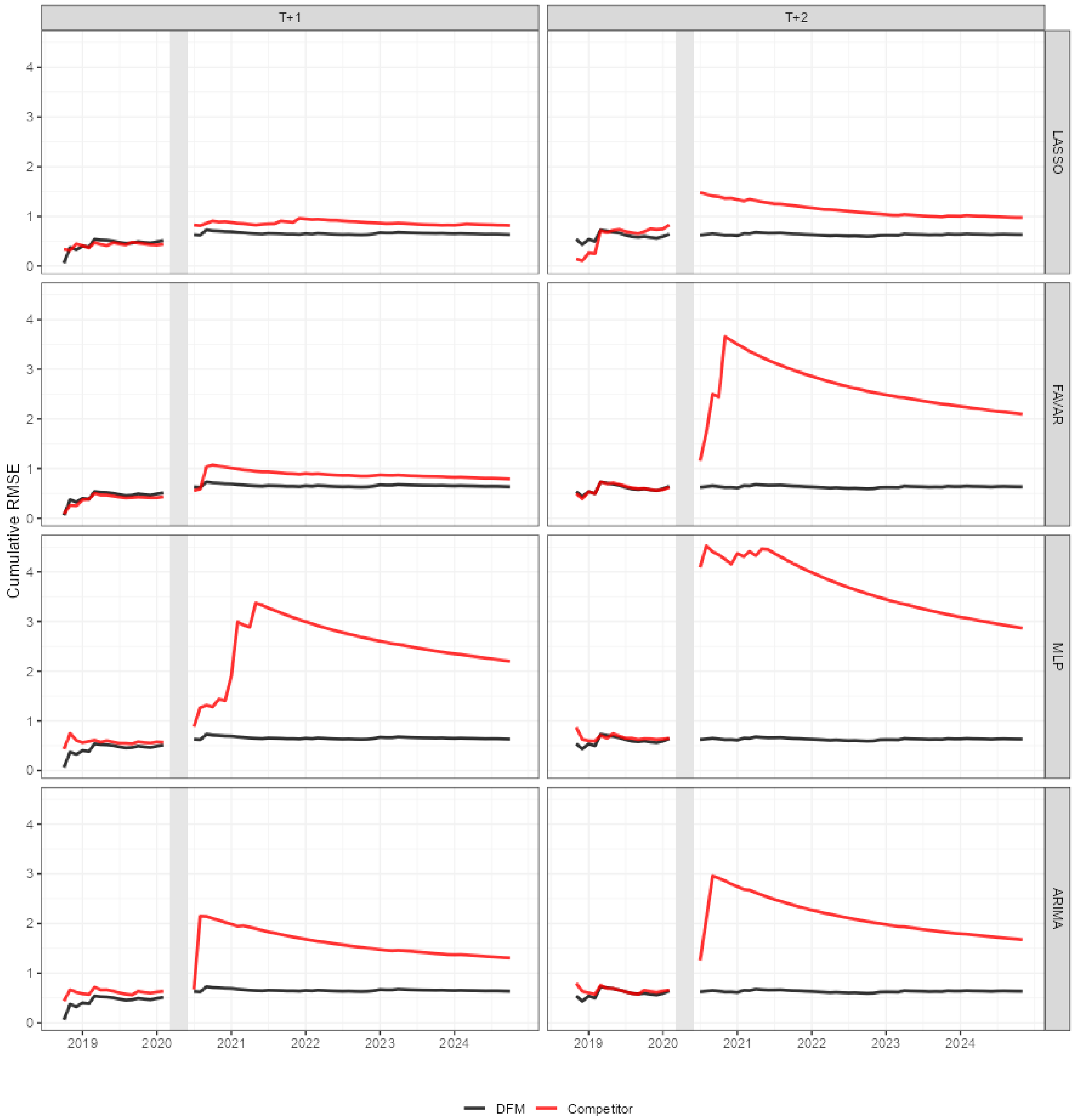

Figure A.8: Cumulative RMSE for each nowcasting method. The black line depicts the IOAE approach, while the red line corresponds to the respective competitor, evaluated over the sample excluding observations from March to June 2020.The shaded area corresponds to the COVID-19 period.