Evaluation of Inflation Forecasting Models in Guatemala

Juan Carlos Castañeda-Fuentes
Carlos Eduardo Castillo-Maldonado
Héctor Augusto Valle-Samayoa
Douglas Napoleón Galindo-Gonzáles
Juan Carlos Catalán-Herrera
Guisela Hurtarte-Aguilar
Juan Carlos Arriaza-Herrera
Edson Roger Ortiz-Cardona
Mariano José Gutiérrez-Morales

Abstract

Inflation forecasts are a relevant input for monetary policy decision in central banks, particularly for those operating under inflation targeting. Therefore, central banks must continuously evaluate the forecasting accuracy of the models used to produce inflation forecasts. In this study, we present the results of an exhaustive evaluation of historical forecast performance for the most important models used to forecast inflation at Banco de Guatemala. We find evidence supporting the claim that time series models perform better for short time horizons while conditional DSGE models (both structural and semi-structural) do better in medium and long time horizons.

Keywords: economic forecasting, forecasting accuracy, forecasting efficiency.

JEL classification: C530.

J. C. Castañeda-Fuentes <jccf@banguat.gob.gt>, Director, Economic Research Department, Banco de Guatemala. Corresponding Author. This research project was developed at Banco de Guatemala's Economic Research Department, within the framework of the CEMLA's Joint Research Program 2017, coordinated by the Banco de la República, Colombia. The authors thank counseling and technical advice provided by the Financial Stability and Development (FSD), Group of the Inter-American Development Bank in the process of writing this document. The opinions expressed herein are those of the authors and do not necessarily reflect the views of CEMLA, the FSD group, the Inter-American Development Bank or Banco de Guatemala.

1. INTRODUCTION

Banco de Guatemala adopted a monetary policy framework based on inflation targeting (IT) in 2005. Because of the forward-looking nature of that regime, central bank authorities should base their policy decisions on reliable inflation forecasts. In fact, Banco de Guatemala employs an array of models to forecast inflation, which include OLS, ARIMA, structural and semi-structural DSGE type of models, as well as forecast combinations of all, or some of these approaches. Since each of these models provides different information about the future path of inflation, a rigorous evaluation of their performance is required in order to determine their reliability, so that the central bank staff could give more weight to more reliable models, and improve the less reliable ones or get rid of them.

This document presents the results of a thorough evaluation of the most frequently used models by Banco de Guatemala to forecast inflation. Our evaluation is divided according to the type of model employed to produce a forecast. First, we evaluate models that produce unconditional forecasts, based on four different approaches: *1)* forecasting accuracy and bias; *2)* ability to predict a change of trend; *3)* prediction similarity; and *4)* forecast efficiency. Second, we assess the performance of models that produce conditional forecasts, by generating in-sample projections for different scenarios of exogenous and endogenous variables. Our main findings indicate that time series models perform better for short time horizons, while the DSGE models are more efficient forecasting longer time horizons.

The remaining of this document is organized as follows. Section 2 presents a description of all unconditional and conditional models employed by Banco de Guatemala to generate inflation forecasts. Section 3 describes the data and methodology employed for evaluation purposes. Section 4 shows the results obtained. Finally, Section 5 concludes.

2. FORECASTING EVALUATION AT THE CENTRAL BANK OF GUATEMALA

The prediction of the inflation rate is very important in the case of an inflation targeting regime, because it allows the central bank to take the monetary policy actions to keep inflation on target and keep the credibility of the regime. Therefore, Banco de Guatemala uses an array of models to forecast the inflation rate. The main forecast models are divided between those that produce unconditional forecasts and those producing conditional forecasts.

2.1 Unconditional-forecasts Models

In this section, we describe the main models used in this paper to evaluate unconditional inflation forecasts. We start by explaining the three main models used to explain the inflation rate. The first one is the indicator variable (IV), which is the inflation forecast employed at Banco de Guatemala as the main short-term forecast in the conduction of its monetary policy, and it is estimated by the Department of Macroeconomic Analysis and Forecasts. The forecast is based on a set of time series models plus the expert knowledge that the economic analysts have about the inflation series. In particular, they complement the inflation forecasts generated by the models with considerations about trend, seasonality, and temporary shocks, in addition to the overall domestic and foreign economic conditions. The second one is the forecast combination through individual time-varying efficient weights (EFP). This model is based on assessing past forecast performance efficiency at each of eight quarters ahead, according to an algorithm called the efficient forecast path (EFP), described in Castillo y Ortiz (2017). The model is explained in detail in Annex 3, which is delivered upon request. The third one is the average macroeconomic models (AMM), used by the Economic Research Department (DIE[1]). The DIE uses two macroeconomic models to make forecasts: the semi-structural macroeconomic model 4.0.1 (MMS) and the macroeconomic structural model (MME).

Furthermore, we evaluate inflation expectations with two measures available at Banco de Guatemala. Both are measured monthly. The first one is from an Economic Expert Panel (EEP). Banco de Guatemala surveys an independent panel of experts from the private sector every month on economics, finance, and business in Guatemala. The objective of the survey is to assess their perception of the future trend of inflation, economic activity, and public confidence in the economy. The second one is from the DIE, which also carries out an inflation expectations survey among its staff.

2.2. Conditional-forecasts Models

In this section, we evaluate the performance of three conditional models to predict the inflation rate. The first model is the MMS 4.0.1 which is a reduced form model, characterized by a difference-equations system, representing the transmission mechanisms of monetary policy for quarterly data. The current version (MMS 4.0.1) is part of the set of non-micro funded general equilibrium macroeconomic models used at Banco de Guatemala that have evolved from the first version launched in 2006. It was built on the basis proposed by Berg, et al. (2006a and 2006b), who provided a practical guide to non-micro funded DSGE models and their implementations for central banks. In this regard, the MMS 4.0.1 is a semi-structural model (non-micro funded) for a small, open economy, where monetary authorities operate policy within an inflation-targeting framework and implement monetary policy through a Taylor-type rule. All variables in the model are specified in annual growth rates. The MMS 4.0.1 has 40 equations (and 40 variables), of which 28 (70%) are endogenous and 12 (30%) are exogenous variables. The model delivers forecasts for both core inflation and headline inflation, and it is currently used for producing inflation and monetary policy interest rate forecasts that are inputs for Banco de Guatemala's monetary policymaking process. Those variables that display high volatility are transformed through a moving sum (or average) scheme in order to reduce that volatility and avoid possible outliers. At that respect, we get smoothed series.

The second model is a macroeconomic model of inflation forecast for Guatemala (PIGU). It is also a semi-structural macroeconomic model, very similar to the MMS 4.0.1. Variables in PIGU are also expressed as annual rates of change. There are three main differences between PIGU and MMS 4.0.1: the set of exogenous variables, the exogenous variables' volatility, and the type of inflation. First, the set of exogenous variables: Even though some exogenous variables are common to both models, others are not. For example, foreign inflation in MMS 4.0.1 is the US core-PCE inflation, while in PIGU is US Headline CPI inflation. Second, the exogenous variables' volatility: many MMS 4.0.1's exogenous variables are smoothed (four-quarter averages), while PIGU uses quarterly variables. Finally, the type of inflation: MMS 4.0.1 forecasts both core and headline inflation, while PIGU forecasts headline inflation only. The model is currently available to all the central bank's staff, through a custom-made interface.

The third model is the macroeconomic structural model (MME), which is a medium scale DSGE model, built within the new-Keynesian framework. It features a financial accelerator à la Bernanke, Gertler and Gilchrist (1999) and other frictions relevant for emerging or developing economies, such as deviations from the law of one price and the UIP. It is a model of heterogeneous agents; households supply labor services to entrepreneurs. They consume domestic and foreign goods, constitute deposits in domestic currency, take foreign debt and collect remittances from abroad. Firms, operating in a perfectly competitive market, assemble differentiated varieties to produce the home (or domestic) homogeneous final good. There are other firms producing the intermediate good, operating in a monopolistic competitive market; they buy a homogeneous wholesale good from entrepreneurs to differentiate it and produce a particular variety. When these firms decide to change their prices, they face adjustment costs, à la Rotenberg (1982), introducing nominal price rigidities into the model. Entrepreneurs use three inputs to produce the wholesale good: capital, labor, and imported raw materials. They buy capital from capital producing firms using their own wealth and loans granted by banks since they are not able to self-finance their entire capital purchases. The financial sector is comprised of private banks divided into two activities: narrow banks that carry out passive operations gathering deposits from households and retail banks using those deposits to grant loans to entrepreneurs. There is also a central bank setting the short-term interest rate–the policy rate–according to a Taylor-type rule and a central government carrying out unproductive spending.

3. DATA AND FORECAST EVALUATION METHODOLOGY

In this section, we describe the data and explain the methodology chosen in order to examine the forecasting accuracy of both the unconditional and conditional models. In the case of the forecast evaluation of unconditional models, the statistical tests are not included in this paper; however, they can deliver upon request (see Annex 3).

3.1 Data

First, we begin describing the dataset used for the unconditional models. First, we use quarterly data to evaluate the forecasting accuracy of the unconditional models. Each quarter, the IV and the AMM model forecast inflation for the next eight quarters, starting at 2011Q1 and finishing at 2017Q2. The EFP model starts forecasting inflation every quarter for the next eight quarters only from 2014Q2 to 2017Q2. Then, we classify the forecasts of each quantitative model into different time-horizons (one, two, three, four, and eight quarter) to evaluate the forecasting performance of each time horizon, in order to find which model is best to forecast the inflation patterns in every one of them. The evaluation sample is rather short, especially in the case of the EFP's forecasts, for which there are only 13 quarters. Also, we evaluate how well the quantitative models predict the inflation rate in December the current and the next year. Second, we use the monthly data on inflation expectations from both an economic experts' panel (EEP) and the DIE to examine the accuracy of the inflation expectations in prediction the inflation of December over a one and two-year horizon. The sample of forecasting errors is from 2015M07 to 2017M06 in the case of the one-year horizon and from 2016M07 to 2017M06 in the case of two-year horizon predictions.

Second, we describe the data used in the case of the conditional models. For each of the three evaluated models, we generate quarterly headline inflation forecasts with a sample from 2011Q1 to 2017Q2.[2] In addition, we consider five forecasting horizons: One quarter, two quarters, four quarters, six quarters, and eight quarters.

3.2. Forecast Evaluation Methodology

First, we explain the methodology to evaluate the forecasting accuracy of the unconditional models. We evaluate the key properties of the forecasting errors; i.e., we perform precision, accuracy, directional

[2] A first evaluation was conducted considering a wider sample (2006Q1-2017Q2), but results from this exercise were not as expected, in particular for headline inflation forecasts. This could be due to some periods of high volatility in headline inflation. For example, inflation went from 14.16% in the third quarter of 2008 towards a negative value (–0.73%) one year later (in August 2009). Therefore, in order to get robust results, we began our evaluation from 2011Q1.

change, and efficiency tests to evaluate which model is best to predict the future path of inflation. We start examining the residuals distribution of the forecast, checking for normality and skewness. Then, we compare the root mean square error (RMSE) values to find which model predicts the inflation rate best. After that, we use the Diebold-Mariano (DM) test to examine if the difference between the MSE of the two competing models is statistically significant at least at the 10% level. Also, we use the Giacomini-Rossi fluctuation (GR) test to examine the forecasting accuracy between the two competing models over forecasting horizons with rolling windows of four. With this test, we examine if the forecasts of one model are better than another in every rolling window or if there is a change (fluctuation) in the accuracy. In addition, we use the Pesaran-Timmerman (PT) test to determine if the forecasts of the models can correctly predict the directional change of inflation. Finally, we test the efficiency of the forecasts by calculating the weak and strong efficiency tests.

Second, we explain the methodology to evaluate the performance of the conditional models to predict the inflation rate. The quality of any variable's conditional forecasts depends on two elements: The performance of the forecasting model (as such) and the quality of the forecasting model's inputs on which the forecasts are conditioned (e.g., the quality of the exogenous variables' forecasts). We evaluate the forecasting model's performance by generating in-sample forecasts in hindsight for different scenarios for the exogenous variables and for some endogenous variables as well. Some of these scenarios involve historically observed values for the exogenous and some endogenous variables, to evaluate forecasts as if we had the best possible forecast for these variables and thus, eliminate one source of error. In the case of the semi-structural models (MMS and PIGU), we plug, for each forecasted period, the historically observed values of exogenous and some endogenous variables. In the case of the structural model (MME), exogenous variables are represented by stochastic processes, typically of autoregressive nature. Therefore, alternative scenarios are only conditioned by historically observed values of two endogenous variables: inflation and output.

First, the MMS 4.0.1 considers the scenarios: free, anchor 1, anchor 2, and anchor 3. In the free scenario, the exogenous variables' forecasts are generated by the model's laws of motion and all endogenous' forecasts are generated by the model. In the anchor 1 scenario, the exogenous variables' forecasts are generated by the corresponding historically observed data, and some endogenous variables' forecasts generated by the corresponding historically observed data: monetary aggregates and economic output. The anchor 2 scenario considers that the inflation forecast for the first quarter in the forecasting horizon is anchored by the corresponding historically observed data, besides the characteristics of the anchor 1 scenario. The last scenario (anchor 3) considers that the monetary policy interest rate is anchored by the corresponding historically observed data, as well as the characteristics of the anchor 2 scenario.

Second, PIGU considers the scenarios: free, anchor 1, anchor 2, and anchor 3. The free scenario contains the same characteristics than in the case of the MMS 4.0.1. In the anchor 1 scenario, the exogenous variables' forecasts are generated by the corresponding historically observed data, and all endogenous variables' forecasts are generated by the model. In the anchor 2 scenario, the exogenous variables' forecasts are generated by the corresponding historically observed data, the inflation forecasts for the first two quarters in the forecasting horizon are anchored by the corresponding historically observed data, while all other endogenous forecasts are generated by the model. In the anchor 3 scenario, the exogenous variables' forecasts are generated by the corresponding historically observed data, the inflation forecasts for the first two quarters in the forecasting horizon are anchored by the corresponding historically observed data, and all other endogenous variables' forecasts are anchored by the corresponding historically observed data.

Third, the MME considers two scenarios: free and anchor 1. In the free scenario, the exogenous variables forecasts are generated by the model's law of motion. In the anchor 1 scenario, the exogenous variables are generated by the model's laws of motion; and the inflation and output forecasts for the first quarter in the forecasting horizon are anchored by the corresponding historically observed data.[3]

For each model's horizon-scenario combination, we compute the mean error and the root mean squared error. The quantitative results allow us to compare the models' forecasting performances (provided that they are fed with the best possible inputs; i.e., they

[3] Anchored values of inflation are slightly different from the corresponding observed values because the inflation series generated by the model has a quarterly frequency; hence, its annualized inflation rate is the sum of four quarterly values rather than a 12-month variation rate.

are fed with historically observed data for the relevant variables) and to assess the informative contribution of exogenous and endogenous variables for forecasting headline inflation.

4. RESULTS

In this section, we present the main results of the forecasting accuracy of both the unconditional and the conditional models. Most of the tables and figures are presented in Annex 5, which do not appear in this paper. However, they are delivered upon request.

4.1. Unconditional Forecast Evaluation

We compare the forecasting performance to predict the inflation patterns between the AMM, the IV, and the EFP model. Also, we evaluate the forecasting performance of the inflation expectations generated by both the EEP and the DIE. First, we compare the performance of the forecasts of the models to predict inflation one, two, three, four, and eight quarters ahead. Second, we analyze the accuracy of the forecasts to predict the inflation rate in December in either the current or the following year. The December inflation forecast is a monetary policy indicator variable at Banco de Guatemala; hence, its evaluation is very important.

4.1.1. Skewness and Normality

We start by evaluating the key properties of the forecasting error distribution: normality and bias. To examine normality, we use the JB test developed by Jarque and Bera (1980). The tables are included in Annex 5, which is delivered upon request. First, we evaluate the properties of the forecasts through different forecasting horizons. The forecast errors of the three models follow a normal distribution according to the Jarque-Bera test, at the conventional levels of significance. Also, the IV's forecast shows a negative skewness while the AMM's and EFP's forecasts show a positive skewness. However, the skewness is low in all cases. Also, the forecast errors of the inflation expectation predictions (both the EEP and the DIE) also follow a normal distribution. There is a positive bias in the inflation expectations predictions in the case of the DIE in both one- and the two-year horizons.

Second, we evaluate the properties of the forecasts in the case of December evaluation. The forecast errors of the three models also follow a normal distribution in all forecasting horizons. In addition, there is a positive bias in the EFP's forecast in the first three quarters while there is no skewness in the remaining ones. IV's and AMM's forecasts both tend to have a negative bias.

4.1.2. RMSE and MPE

We compute the RMSE and MPE to determine which forecasting model performs best, in the case of both the quantitative and the inflation expectations. The tables are included in Annex 5, which is delivered upon request. In the case of the quantitative models, the forecasts of the IV model are better in the short run–one and two quarters–based on the RMSE. In the middle run, the forecasts of the AMM model are more accurate. However, in the long run–eight quarters–, the forecasts of the EFP model outperform the others. Also, we also analyze the inflation expectations predictions. Based on the RMSE, the EEP's inflation expectations are more accurate than those of the DIE's in both the one- and two-years horizons

Second, we proceed to analyze the forecasting accuracy of the quantitative models in their ability to predict the inflation rate in December for the current and the following year, based on the RMSE. We observe that the forecasts of the AMM model are better than the others in the first five forecasting horizons, while the IV's forecasts are best for the last three horizons.

4.1.3. Diebold-Mariano Test

First, we use the DM test developed by Diebold and Mariano (1999) to compare the predictive accuracy between two competing models, of both the quantitative and the inflation expectations predictions. The null hypothesis is that the two models have equal accuracy. The results of the DM test in the case of the quantitative models are presented in Table 1 (the p-values of the test are shown in parenthesis). In Column 2, it is shown the test between the AMM and the IV model. Only in the case of four- and eight-quarter forward forecasting horizons, the DM-statistic is negative and statistically significant at 5% level; therefore, we reject the null hypothesis, and conclude that the forecasting accuracy of the AMM model is best for both the intermediate and long time horizons. Then, the DM-statist between the EFP and the IV model is presented in Column 3. The statistic is positive and statistically significant at 5% level in all forecasting horizons, which means that all MSE of the IV model are lower than those of the EFP model; therefore, the forecasting accuracy of the IV model is best to predict inflation.

After that, the DM-statistic between the EFP and the AMM model is presented in Column 4. The statistic is only positive and statistically significant for the one-, two-, and three-quarter forward forecasting horizons. This means that for those horizons, the MSE of the AMM model are lower than those of the EFP model; therefore, the AMM's forecasts are best to predict inflation in the short run. Also, we evaluate the predictive performance of the inflation expectations of both the EEP and the DIE. The DM-statistic is only statistically significant for the two-year horizon with a sample of 12 months. This means that the MSE of the EEP is lower than the MSE of the DIE. Thus, we conclude that the inflation expectation predictions of the EEP are more accurate than those of the DIE, only on this horizon.

Second, we compare the forecasting accuracy of the quantitative models to predict the December inflation rate, from different horizons. The results of the DM test are presented in Table 2. In Column 2, it is shown the test between the AMM and the IV model. The DM-statistics are negative and statistically significant starting from three-quarter forward forecasting horizon, so the MSE of the AMM model are lower than those of the IV model. Therefore, the forecasts of the AMM model are best to predict inflation.

Table 1

DM TEST, QUANTITATIVE MODELS

Forecasting horizons in quarters	*dm statistic (AMM-IV)*	*dm Statistic (EFP-IV)*	*dm statistic (EFP-AMM)*
1	1.44 (0.15)	1.71 (0.087)	1.65 (0.09)
2	1.30 (0.19)	1.97 (0.049)	2.03 (0.04)
3	0.21 (0.84)	1.79 (0.074)	1.70 (0.09)
4	–2.95 (0.00)	1.76 (0.079)	1.61 (0.11)
8	–3.35 (0.02)	2.91 (0.004)	–0.87 (0.38)

Sources: author's elaboration, central bank's forecasts.

Table 2

DM TEST, QUANTITATIVE MODELS

Forecasting horizons in quarters	*dm statistic (AMM-IV)*	*dm statistic (EFP-IV)*	*dm statistic (EFP-AMM)*
1	1.44 (0.15)	2.10 (0.036)	1.65 (0.10)
2	−0.95 (0.34)	2.70 (0.007)	2.55 (0.01)
3	−4.60 (0.00)	2.62 (0.009)	2.58 (0.00)
4	−2.33 (0.01)	4.75 (0.000)	7.32 (0.00)
5	−3.20 (0.00)	2.09 (0.036)	3.16 (0.00)
6	−2.93 (0.00)	−5.61 (0.000)	−22.50 (0.00)
7	−2.98 (0.00)	−62.39 (0.000)	2.58 (0.01)
8	−1.95 (0.05)	–	–

Sources: author´s elaboration, central bank's forecasts.

Then, the DM-statistic between the EFP and the IV model is presented in Column 3. The statistic is statistically significant in all forecasting horizons, which means that the MSEs of the IV model are lower than those of the EFP model. Hence, we reject the null hypothesis of equal accuracy. Also, the statistic is positive for the one- to five-quarter horizons, which means that the MSEs of the IV model are lower than those of the EFP. Hence, the IV model is more accurate in its prediction of December inflation rate in the short and intermediate time horizons. On the other hand, the statistic is negative from six to seven quarters ahead; therefore, the EFP model is best in the long run to predict the inflation rate. After that, the DM-statistic between the EFP and the AMM model is presented in Column 4. This is statistically significant in all forecasting horizons, which means that we reject the null hypothesis of equal accuracy. Also, in almost all forecasting horizons, the MSEs of the AMM model are lower than those of the EFP model. Therefore, the AMM model is best to predict inflation rate in December.

4.1.4. Pesaran-Timmerman Test

We use the PT test developed by Pesaran and Timmerman (1992) to evaluate the directional forecasting of both the quantitative models and the inflation expectations predictions. The critical values to reject the null hypothesis of independence are ± 1.645 for 10% level of significance. First, we examine the directional forecasting accuracy in the case of the IV model (see Annex 1, Table A.1.1). The S_n statistic is only higher than its critical value in the case of one-, two- and three-quarter horizons, so we can reject the null hypothesis of independence and conclude that the forecasts of the IV model can predict successfully the direction of inflation in the short run. Now, we evaluate the directional accuracy in the case of the AMM model (see Column 3). We observe that the S_n statistic is higher than its critical value only in the case of one- and two-quarter horizons, so we can reject the null hypothesis of independence only for those two horizons and conclude that the model can successfully predict the direction of the inflation in the short run. We proceed to analyze the directional accuracy of the forecast in the case of the EFP mode (see Column 4). The S_n statistic is higher than the critical value only in the case of one-quarter horizon; therefore, we can only reject the null hypothesis of independence for this horizon and conclude that the forecast of the EFP model can predict successfully the direction of the inflation in the case of that particular horizon. Also, we analyze the directional forecasting accuracy of the inflation expectations predictions of both the EEP and the DIE. We reject the null hypothesis of independence only in the case of the EEP's forecasts in the case of a two-year horizon. Hence, we can conclude that the panel can predict successfully the direction of inflation.

Second, we examine the directional forecasting accuracy of the inflation rate for December (see Table Annex 2, A.2, which is delivered upon request) only for the case of the IV and AMM models, since we do not have enough data for the case of the EFP model. We start with the IV model (see the first column). We can reject the null hypothesis of independence in the case of one-, three-, four-, five-, and six-quarter horizons, so the model can predict successfully the directional change of inflation in the short and middle run. Then, evaluate the performance of the AMM model (see the second column). We can reject the null hypothesis of independence in the case of one-, two-, three-, six-, seven-, and eight-quarter horizons, which implies that the model can predict successfully the directional change of inflation in both the short and the long run.

4.1.5. *Giacomini-Rossi Fluctuation Test*

We use the Giacomini and Rossi fluctuation test developed by Giacomini and Rossi (2010) to examine the performance of two competing models in the presence of possible instabilities. We use the IV model as the benchmark model in the case of the quantitative model, and the inflation expectations' predictions of the EEP in the case of expectations' forecasts. The test is only used in some of the forecasting horizons due to data availability. We set the rolling windows equal to four quarters to make the forecasting analysis. Also, we use graphical analysis to examine the performance of the forecasts of the two competing models in the different rolling windows to see whether there is a fluctuation in the forecasting accuracy. This is available in Annex 4, which is delivered upon request.

First, we start with the forecasting accuracy evaluation of the quantitative models (see Annex 1, Table A1.3). We define the loss function between the AMM and the IV model in Equation 1. If the loss function turns out to be negative, we conclude that the forecasts of the AMM model are more accurate than those of the IV model. On the other hand, if the loss function turns out to be positive, the forecasts of the IV model are better at predicting inflation than those of the AMM model. We observe that we reject the null hypothesis of equal forecasting accuracy over every forecasting horizon since the GR-statistic is higher than its critical value (see Table A1.3, Column 2). This means that one model displays better predictive ability to forecast inflation in at least one period of time. Also, the graphical analysis reveals that the forecasts of the IV model are more accurate than those of the AMM one step ahead. However, it seems that the forecasts of the AMM model predict better the inflation patterns in four- and eight-quarter horizons.

$$L_t\left(\hat{\theta}_{j-h,R}, \hat{\gamma}_{j-h,R}\right) = MSE_{AMM,t} - MSE_{IV,t} \tag{1}$$

Then, we compare the forecasting accuracy between the EFP and the IV model with the use of the GR test (see Column 3). The loss function between the two models is defined by Equation 2. In this case, the null hypothesis of equal accuracy is rejected in every forecasting horizon since the GR-statistic is higher than the critical value. This means that, at least in one period, one model generates more accurate forecasts of inflation. The graphical analysis shows that the forecasts of the IV model are more accurate in almost all the evaluation sample in each forecasting horizon. Therefore, the forecasts of the IV model seem to be more accurate than the EFP model in all forecasting horizons (see Annex 5, delivered upon request).

$$L_t\left(\hat{\theta}_{j-h,R},\hat{\gamma}_{j-h,R}\right)=MSE_{EFP,t}-MSE_{IV,t} \tag{2}$$

Second, we use the GR test to examine the performance of the inflation expectations predictions from the DIE and the EEP. We consider the EEP data as a benchmark model. The loss function is set up in Equation 3. The graphical analysis shows that there is a fluctuation of the forecasting accuracy of the inflation expectations between the two models in the case of one-year horizon. However, the inflation expectations of the EEE predict better the inflation patterns in the case of the two-year horizon (see Annex 4, delivered upon request).

$$L_t\left(\hat{\theta}_{j-h,R},\hat{\gamma}_{j-h,R}\right)=MSE_{EEP,t}-MSE_{DIE,t} \tag{3}$$

4.1.6. Weak Efficiency Test

We examine the efficiency of the unconditional forecasts of both the quantitative and the qualitative models with a variant of the weak efficiency test developed by Mincer and Zarnowitz (1969). First, we start with the quantitative models (see Annex 1, Table A1.4). From the second column, we observe that AMM's forecasts satisfy the weak efficiency hypothesis only in the case of one quarter ahead. From the third column, we analyze the weak efficiency of the IV forecasts (see the third column) We observe that forecasts of the model satisfy the weak efficiency only in the case of one and two forecasting horizons. From the fourth column, we evaluate the weak efficiency of the EFP forecasts (see the fourth column). We observe that the forecasts of the model satisfy the weak efficiency in almost all forecasting horizons with the exception of four quarters ahead. In sum, the forecast of the EFP is more efficient than those of the other models based on the results of the weak efficiency test. Also, the forecast of the AMM and the IV are weakly efficient in the short run. In addition, the inflation expectations predictions of both the EEP and the DIE model do not satisfy de weak efficiency test at 5% level in all forecasting horizons.

Second, we test for the weak efficiency only in the case of the AMM and the IV models, in the prediction of the inflation rate of December, because of data availability (see Annex 5, which is delivered upon request). In the case of the AMM's forecasts, we cannot reject the null hypothesis of weak efficiency only in the case of two and three quarters ahead. Also, the forecasts of the IV model satisfy the weak efficiency tests in five out of eight forecasting horizons. In sum, the forecasts of the IV model are more efficient than those of the AMM model in evaluating the December predictability of inflation.

4.1.7. Strong Efficiency Test

We perform the strong efficiency test for the two econometric models: IV and EFP. The null hypothesis establishes that a new variable (which is not included in the econometric models) does not explain the forecasting error. Therefore, the rejection of the null hypothesis means that the errors are strongly efficient. Otherwise, if the null hypothesis is not rejected, then the inclusion of a new variable can add information to improve the forecasts. We consider five variables in logs of the structural model of the Banco de Guatemala to make the test: consumption, index of raw materials, investment, government spending, and credit.

First, we start with the IV model; the tests are shown in Annex 5, Table A5.7, which is delivered upon request. In the second column, we list the coefficient of consumption. We cannot reject the null hypothesis at the 5% level of significance in the case of one and two quarters ahead. Therefore, the forecasts are strongly efficient for those horizons. However, for three to eight quarters ahead, consumption does explain the forecasting error, which means that they are not strongly efficient for these horizons. Similarly, in the third column, the null hypothesis is not rejected at the 5% significance level. Therefore, the forecasts are strongly efficient in those horizons. However, from three to eight quarters ahead, the inclusion of the raw material index can improve the forecasts, which mean that they are not strongly efficient. Then, in the fourth column, we observe that the null hypothesis is not rejected in one, two and three quarters ahead, which means that the forecasts are strongly efficient in those horizons. However, from four to eight quarters ahead, investment explains the forecasting errors, therefore; the forecasts are not strongly efficient. After that, in the fifth column, we observe that the null hypothesis is not rejected in all forecasting horizons, which means that the forecasts are strongly efficient, and the inclusion of government spending will not improve them. Finally, in the sixth column, we observe that the forecasts are strongly efficient from one to three quarters ahead. However, from four to eight quarters ahead, the inclusion of credit can improve the forecasts, which implies that they are not strongly efficient in those horizons.

We continue with the EFP model; the tests are shown in Annex 5, Table A5.8, which is delivered upon request. We observe that we reject the null hypothesis for one-quarter predictions for the five variables, which means that the forecasts of the IV model are not strongly efficient and the inclusion of the consumption, raw material index, investment, government spending, and credit can improve the forecasts for this forecasting horizon. However, the forecasts are strongly efficient in the case of the remaining forecasting horizons for the five variables, because we cannot reject the null hypothesis.

Second, we perform the strong efficiency tests in the case of the evaluation of December, only for the IV model due to data availability (see Annex 5, Table A5.9). We observe that we cannot reject the null hypothesis for all forecasting horizons in the case of the raw material index, investment, government spending, and credit, at the 5% level of significance, which means that the forecast are strongly efficient. However, in the case of consumption, we cannot reject the null hypothesis in all forecasting horizons except for the three quarters ahead, which means that the forecast is strongly efficient for most horizons.

4.2 Conditional Forecast Evaluation

We make a headline inflation forecasting exercise in hindsight for the three models. Also, we consider four scenarios for both the MMS 4.01.1 and PIGU and two scenarios for MME. The forecasting horizon begins on 2011Q1. First, we show the inflation patterns and the forecasts of each model (see Annex 5, Figures A5.1, A5.2, and A5.3, which are delivered upon request). Second, we calculate the ME and the RMSE (see Annex 1, Tables A1.6, A1.7 y A1.8).

In the case of the MMS 4.0.1, the model generates core inflation forecasts, and therefore headline inflation is constructed based on those projections. This explains that, in the case of anchor 2 and anchor 3, we have values different from zero in 1 and 2 quarters ahead for the ME and RMSE (see Annex 1, Table A1.6). PIGU model minimizes the RMSE in the fourth scenario (anchoring exogenous variables, all other endogenous variables and two quarters of inflation) for all forecasting horizons (see Annex 1, Table A1.7). In this case, the model's forecasts are negatively biased for all relevant horizons (the first two horizons are trivially unbiased since the historically observed inflation values are imposed as the model's forecasts). In order to compare the two models' forecasting performances, we pick the best scenario for each model. In particular, we compare the MMS 4.0.1's performance in the third scenario with the PIGU's performance in the fourth scenario. We focus on the last three forecasting horizons since PIGU's RMSE for the first two horizons is trivially equal to zero. The results show that PIGU's RMSE for the three relevant horizons are less than the corresponding values for MMS 4.0.1 and, hence, PIGU is preferred in this evaluation exercise, even though its forecasts tend to underestimate inflation (i.e., its forecasts are negatively biased). See Table 3.

For the MME, the ME suggests that there is a positive inflation bias (see Annex 1, Table A1.8). Results also suggest that forecasts generated by the model can benefit from anchoring inflation and output one quarter ahead since doing so reduces the RMSE (or its mean across different forecasting horizons). This improvement will require that better short-term projections (from outside the model) are available.

Table 3

COMPARISON OF THE BEST SCENARIOS BETWEEN MMS 4.0.1 AND PIGU

Forecasting horizons in years	*mms 4.0.1, anchor 2*	*PIGU, anchoring exogenous and endogenous variables, plus two periods of inflation*
4	1.37	0.61
6	1.36	0.62
8	1.57	0.65

Source: author's elaboration, central bank's forecasts.

5. CONCLUSIONS

In this paper, we evaluated Banco de Guatemala's most important models used to forecast inflation. Forecast accuracy for unconditional models (i.e., IV, AMM, and forecast combinations of OLS and time series models) was evaluated for end of the year forecasts, and for a two-year forecast horizon, using a variety of measurements and tests (i.e., normality, RMSE, DM, PT, GR, and weak and strong efficiency tests). In the case of a conditional forecast, we evaluated the forecasting accuracy of three models: MMS 4.0.1, PIGU, and MME.

We found empirical evidence supporting a higher degree of accuracy for time series models for the short forecast-horizons, and better performance for models generating conditional-forecasts in longer forecast-horizons. The main purpose of this study was to assess the accuracy and precision of the main inflation forecasts generated at Banco de Guatemala. The next step is to take advantage of the obtained results in order to improve the quality of the inflation forecasting models in use at the central bank. In particular, we should continuously reevaluate model specifications, the quality of the data sets, and the variable-transformation procedures. In addition, we should perform a complete evaluation of the inflation forecasts at least once a year, as some central banks already do.

ANNEX

Annex 1. Tables of the Unconditional Forecast Evaluation

Table A1.1

PT TEST, QUANTITATIVE MODELS

Forecasting horizons in quarters	*S_n statistic (IV)*	*S_n Statistic (AMM)*	*S_n statistic (EFP)*
1	4.28	3.98	2.41
2	3.77	2.93	1.62
3	2.57	1.54	0.73
4	0.00	0.88	−1.01
8	0.00	−1.49	−1.53

Sources: author's elaboration, central bank's forecasts.

Table A1.2

PT TEST, QUANTITATIVE MODELS, DECEMBER EVALUATION

Forecasting horizons in quarters	*S_n statistic (IV)*	*S_n statistic (AMM)*
1	1.67	1.67
2	1.02	1.67
3	−1.67	1.67
4	1.67	−1.46
5	−2.31	−1.33
6	−2.31	−2.31
7	–	−2.31
8	–	−2.31

Sources: author's elaboration, central bank's forecasts.

Table A1.3

GR TEST, QUANTITATIVE MODELS

Forecasting horizons in quarters	*GR statistic (AMM-IV)*	*GR statistic (EFP-IV)*
1	4.77	5.68
2	15.28	5.93
3	9.93	11.29
4	9.07	7.39
8	11.28	–

Sources: author's elaboration, central bank's forecasts.

Table A1.4

WEAK EFFICIENCY TEST, QUANTITATIVE MODELS

Forecasting horizons in quarters	*Weak efficiency test (AMM)*	*Weak efficiency test (IV)*	*Weak efficiency test (EFP)*
1	0.11 (0.89)	6.33 (0.24)	3.58 (0.08)
2	4.41 (0.02)	3.29 (0.12)	0.22 (0.89)
3	6.18 (0.00)	11.57 (0.01)	12.08 (0.97)
4	5.39 (0.01)	21.81 (0.00)	–
8	104.62 (0.00)	62.16 (0.00)	0.20 (0.83)

Sources: author's elaboration, central bank's forecasts.

Table A1.5		
WEAK EFFICIENCY TEST, QUANTITATIVE MODELS, DECEMBER EVALUATION		
Forecasting horizons in quarters	*Weak efficiency test (AMM)*	*Weak efficiency test (IV)*
1	83.48 (0.00)	1.17E+12 (0.00)
2	1.36 (0.35)	1.5268 (0.32)
3	1.45 (0.34)	1.2242 (0.38)
4	8.87 (0.034)	9.5156 (0.03)
5	1.71E+11 (0.00)	14.1267 (0.03)
6	1.85E+11 (0.00)	1.6197 (0.33)
7	2.03E+10 (0.00)	0.9950 (0.47)
8	1.66E+10 (0.00)	1.8451 (0.30)

Sources: author's elaboration, central bank's forecasts.

Tables of the Conditional Forecast Evaluation

Table A1.6								
ME AND RMSE, MMS 4.01, 2011Q1-2017Q2								
Forecasting horizons in quarters	*Free model*		*Anchor 1*		*Anchor 2*		*Anchor 3*	
	ME	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*
1	–0.11	0.73	–0.03	0.71	0.01	0.33	0.01	0.33
2	–0.13	1.21	–0.02	1.27	0.01	0.87	0.01	0.87
4	0.22	1.43	0.29	1.58	0.29	1.37	0.29	1.37
6	0.55	1.47	0.54	1.4	0.52	1.36	0.52	1.36
8	0.27	1.72	0.55	1.63	0.54	1.57	0.54	1.57
Mean	0.16	1.31	0.26	1.32	0.27	1.1	0.27	1.1

Sources: author's elaboration, central bank's forecasts.

Table A1.7

ME AND RMSE, PIGU, 2011Q1-2017Q2

Forecasting horizons in quarters	*Free model*		*Anchoring exogenous variables*		*Anchoring exogenous variables and two periods of inflation*		*Anchoring exogenous and endogenous variables, plus two periods of inflation*	
	ME	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*
1	−0.22	0.83	−0.25	0.72	0	0	0	0
2	−0.3	1.26	−0.38	0.9	0	0	0	0
4	0	1.44	−0.47	0.88	−0.39	0.82	−0.27	0.61
6	0.34	1.11	−0.58	1.12	−0.56	1.13	−0.32	0.62
8	0.41	0.89	−0.79	1.29	−0.79	1.29	−0.38	0.65
Mean	0.05	1.11	−0.49	0.98	−0.35	0.65	−0.19	0.38

Sources: author's elaboration, central bank's forecasts.

Table A1.8

ME AND RMSE, MME, 2011Q1-2017Q2

Forecasting horizons in quarters	*Free model*		*Anchor 1*	
	ME	*RMSE*	*ME*	*RMSE*
1	0.3	0.62	−0.09	0.1
2	0.89	1.28	0.36	0.61
4	2.37	2.72	1.81	2.09
6	2.82	2.98	2.87	3.04
8	2.82	2.93	2.86	2.96
Mean	1.84	2.11	1.56	1.76

Sources: author's elaboration, central bank's forecasts.

References

Berg, Andrew, Philippe D. Karam, and Douglas Laxton (2006a), *A Practical Model-based Approach to Monetary Policy Analysis-Overview.*

Berg, Andrew, Philippe D. Karam, and Douglas Laxton (2006b), *Practical Model-based Monetary Policy Analysis: A How-to Guide,* No. 6-81, International Monetary Fund.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist (1999), "The Financial Accelerator in a Quantitative Business Cycle Framework," *Handbook of Macroeconomics 1,* pp. 1341-1393.

Castillo-Maldonado, Carlos, and Edson R. Ortiz-Cardona (2018), "Evaluación de combinaciones de inflación en Nicaragua (NICA): Un método eficiente para combinar pronósticos," *Revista de Economía y Finanzas,* Vol. 5, Banco Central de Nicaragua, octubre, pp. 1-34.

Diebold, F. X., and R. S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* Vol. 13, No. 3, July, pp. 253-263.

Giacomini, Raffaella, and Barbara Rossi (2010), "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics,* Vol. 25, No. 4, pp. 595-620.

Jarque, Carlos M., and Anil K. Bera (1980), "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economics Letters,* Vol. 6, No. 3, pp. 255-259.

Mincer, Jacob A., and Victor Zarnowitz (1969), "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance,* NBER, pp. 3-46.

Pesaran, M. Hashem, and Allan Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance," *Journal of Business & Economic Statistics,* Vol. 10, No. 4, pp. 461-465.

Rotemberg, Julio J. (1982), "Sticky Prices in the United States," *Journal of Political Economy,* Vol. 90, No. 6, pp. 1187-1211.