IDB WORKING PAPER SERIES Nº IDB-WP-882

Evaluation of Inflation Forecasting Models in Guatemala

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

Inter-American Development Bank
Department of Research and Chief Economist

July 2018

Evaluation of Inflation Forecasting Models in Guatemala

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

Banco de Guatemala

July 2018

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Evaluation of inflation forecasting models in Guatemala / Juan Carlos Castañeda-Fuentes, Héctor Augusto Valle-Samayoa, Juan Carlos Catalán-Herrera, Juan Carlos Arriaza-Herrera, Mariano José Gutiérrez-Morales, Carlos Eduardo Castillo-Maldonado, Douglas Napoleón Galindo-Gonzáles, Guisela Hurtarte-Aguilar, Edson Roger Ortiz-Cardona.

p. cm. — (IDB Working Paper Series ; 882)

Includes bibliographic references.

1. Inflation (Finance)-Guatemala-Forecasting. 2. Inflation (Finance)-Guatemala-Econometric models. I. Castañeda-Fuentes, Juan Carlos. II. Valle-Samayoa, Héctor Augusto. III. Catalán-Herrera, Juan Carlos. IV. Arriaza-Herrera, Juan Carlos. V. Gutiérrez-Morales, Mariano José. VI. Castillo-Maldonado, Carlos Eduardo. VII. Galindo-Gonzáles, Douglas Napoleón. VIII. Hurtarte-Aguilar, Guisela. IX. Ortiz-Cardona, Edson Roger. X. Inter-American Development Bank. Department of Research and Chief Economist. XI. Series.

IDB-WP-882

http://www.iadb.org

Abstract[1]

Forecasting the inflation path is an important task for central banks in an inflation targeting regime. Therefore, central banks must continuously evaluate the forecasting accuracy of the models used to generate inflation forecasts. This paper evaluates the performance of most of the models that produce either unconditional or conditional inflation forecasts at Banco de Guatemala. The paper is divided in two main parts. The first part evaluates the forecasting accuracy and efficiency of the models that produce unconditional forecasts, applying different measures as normality test, root mean square error (RMSE), mean percentage errors (MPA), and tests as Diebold-Mariano, Pesaran-Timmerman, Giacomini-Rossi, as well as both weak and efficiency tests. The second part evaluates the conditional forecasting performance of the central bank's main macroeconomic models by generating sample forecasts in hindsight for different scenarios for exogenous and some endogenous variables. We find evidence supporting the claim that the time series models perform better in forecasting inflation for short time horizons while the structural macroeconomic models perform better in medium and long time horizons.

JEL classification: C53
Keywords: Economic forecasting, Forecasting accuracy, Forecasting efficiency

[1] This research project was developed at Banco de Guatemala's Economic Research Department, within the framework of the CEMLA's Joint Research Program 2017 that was coordinated by the Central Bank of 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

The purpose of this research is to evaluate the main inflation forecasting models in use by the Banco de Guatemala. Banco de Guatemala adopted an inflation targeting regime to conduct its monetary policy in 2005. To have accurate and unbiased forecasts of inflation is a matter of paramount importance in this monetary policy framework. Currently Banco de Guatemala employs an array of models to this end. This set of models encompasses time series econometrics models, macroeconomic models and forecast combination technics. Each model provides different information on the future evolution of inflation according to its nature. The forecasts are a key element in the conduct of monetary policy.

This paper is divided into two main parts. In the first part we evaluate the forecasting performance of all the models, both time-series and structural macroeconomic models, in terms of accuracy and bias. This is done based on the typical measures of evaluation, i.e., the root mean square error, the mean percentage error and normality tests. In addition, we assess the capability of each model to forecast a change in the direction of inflation using the Pesaran-Timmerman test. Also, we use the Diebold and Mariano test to compare the predictive accuracy of the forecasts between two competing models, and the Giacomini-Rossi tests to examine the performance of two competing models in the presence of possible instabilities over different evaluation rolling windows. Finally, we test the efficiency of the models with both weak and strong efficiency tests.

In the second part, we assess the performance of the macroeconomic models in producing conditional forecasts. We evaluate the forecasting model's performance by generating in-sample forecasts in hindsight for different scenarios for exogenous and some endogenous variables. Some of those scenarios involve historically observed values for the exogenous and some endogenous variables. In general, we have found that the time series models perform better for short time horizons, while the structural macroeconomic models perform better for longer time horizons, as expected.

The paper is organized as follows. In Section 2, we focus on the unconditional forecast evaluation. In Section 3, we look at the conditional forecast evaluation. Finally, in Section 4, we present the main conclusions of the research paper.

2. Unconditional Forecasts Evaluation

2.1 The Models

2.1.1 Indicator Variable (IV)

This is the inflation forecast employed at Banco de Guatemala as the main short-term forecast in the conduct of its monetary policy. It is estimated by the Department of Macroeconomic Analysis and Forecasts. The forecast is based on a set of time series models plus economic analysts' expert knowledge about the inflation series. In particular, economists complement the inflation forecasts generated by the models with considerations of trend, seasonality and temporary shocks in addition to the overall domestic and foreign economic conditions.

2.1.2 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 and Ortiz (2017). It consists of a five-step method to construct a weighting scheme for model combination that includes an ex ante endogenous procedure for model trimming. The model is discussed in Appendix 1.

2.1.3 Average Macroeconomic Models (AMM)

The Economic Research Department (DIE[2]) uses two macroeconomic models to make forecasts: The Semi-Structural Macroeconomic Model 4.0.1 (MMS) and the Macroeconomic Structural Model (MME).

The MMS 4.0.1 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 in Banco de Guatemala that have evolved from the first version launched in 2006.

The MME is a medium-scale DSGE model, built within the New Keynesian framework. It features a financial accelerator à la Bernanke et al. (1999) and other frictions relevant for emerging or developing economies, such as deviations from the law of one price and the uncovered interest parity.

[2] DIE is the Spanish acronym of the Economic Research Department.

2.1.4 Inflation Expectations

There are two measures of inflation expectations available at Banco de Guatemala. Both are updated monthly, and they are described below.

Panel of Economic Experts

Banco de Guatemala conducts a monthly survey with an independent panel of private-sector experts on economics, finance and business in Guatemala. The objective is to gather information on their perceptions of the future trend of inflation, economic activity and confidence in the economy.

Economic Research Department

The Economic Research Department carries also an inflation expectations survey among its members.

2.2 Evaluation Methodology

In this section, we describe the methodology chosen to evaluate the forecasting accuracy of the unconditional forecasts models.

We consider three competing models: Indicator Variable, Average Macroeconomic Models, and the Efficient Forecasting Path model. Also, we evaluate the inflation expectations of both the Panel of Economic Experts and Economic Research Department of the Central Bank.

2.2.1 Evaluation Sample

First, we use quarterly data to evaluate the accuracy of the forecasts of the quantitative models. Each quarter, the competing models forecast inflation for the next eight quarters, starting at 2011q1 and finishing at 2017q2 in the case of both the Indicator Variable (IV) and the Average Macroeconomic Models (AMM). However, the Efficient Forecasting Path model forecasts inflation for the next eight quarters from 2014q2 to 2017q2.

Then, we classify the forecasts of each quantitative model into different time-horizons (one, two, three, four, and eight quarters) to evaluate the forecasting performance at each time horizon and determine which model is best for forecasting the inflation patterns in every one of them. The evaluation sample is rather short, though, especially in the case of the EFP'S forecasts, for which there are only 13 quarters. We additionally evaluate how well the models predict the December inflation rate for both the current and the next year.

Second, we use the inflation expectations monthly data from both the economic experts' panel and the Economic Research Department to examine the accuracy of the inflation expectations to predict December inflation at one and two-year horizons. The sample of forecasting errors is from 2015q7 to 2017q6 in the case of the one-year horizon and from 2016q7 to 2017q6 in the case of the two-year horizon.

2.2.2 Forecasting Evaluation

In this study, we evaluate the key properties of the forecasting errors; i.e., we perform precision, accuracy, directional change, and efficiency tests to evaluate which model is best to predict the inflation path. We start by examining the residuals distribution of the forecast to check for normality and skewness.

We then use the Root Mean Square Error (RMSE) measure to find which model best predicts the inflation rate, and the Diebold–Mariano test to examine if the difference between the MSE of the two competing models is statistically significant at least at the 10% level of significance. Also, we use the Giacomini–Rossi Fluctuation test in order to examine the forecasting accuracy between two competing models over the forecasting horizons with rolling windows of four. With this test, we examine whether the forecasts of one model are better than the other's in every rolling window or whether there is a change (fluctuation) in the accuracy.

In addition, we use the Pesaran–Timmerman test to determine whether the forecasts of the models can correctly predict the directional change of inflation, and Finally, we test the efficiency of the forecasts by examining bias, autocorrelation, and (weak and strong) efficiency tests.

The tests are discussed in detail in Appendix 2.

3. Main Results

This section compares forecasting performance to predict the inflation patterns of the Average Macroeconomic Models (AMM), the Indicator Variable model (IV), and the Efficient Forecasting Path framework (EFP). Also, we evaluate the forecasting performance of the respective inflation expectations generated by the Economic Experts Panel (EEP) and the Economic Research Department (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 forecast to predict the December inflation rate 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.

3.1 Skewness and Normality

We start evaluating the key properties of the forecasting error distribution: normality and bias. To examine normality, we use the Jarque–Bera test.

3.1.1 Forecasting Horizons

We begin analyzing the forecasting error distribution of the quantitative models (see Table A3.1, Appendix 3). The p-values of the Jarque-Bera test are shown in parenthesis.

Table A3.1 of Appendix 3 shows that the forecast errors follow a normal distribution according to the Jarque–Bera test in all forecasting horizons for the three models at least at 10 percent level of significance. In addition, we observe that the IV forecast shows a negative skewness, while the AMM and EFP forecasts show a positive skewness. In all cases, the skewness is low.

We then analyze the error distribution of the inflation expectations (both EEP and DIE; see Table A3.2, Appendix 3). The forecast error follows a normal distribution. In this case, there is a positive bias in the DIE's inflation expectations at both the one- and the two-year horizons.

In sum, the models' forecast error distributions show good statistical properties.

3.1.2 December Evaluation

Here, we evaluate the forecasting accuracy to predict the December inflation rate (for the current and the following year). Also, the p-values are shown in parenthesis in the case of the Jarque-Bera test.

From Table A3.3, Appendix 3, we observe that the forecast error follows a normal distribution according with the Jarque – Bera test in all forecasting horizons for the three models. In addition, there is a positive bias in the case of the EFP'S forecast in the first three quarters, while there is no skewness in the remaining ones. In the case of the IV's forecasts, they tend to have a negative bias, and so do the AMM's forecasts as well. All in all, we observe good statistical properties of these forecast errors too.

3.2 Forecasting Accuracy

The main goal in this section is to evaluate the forecasting accuracy of both the quantitative models and the inflation expectations. In the first place, we compute the root mean square error (RMSE) and the mean percentage error (MPE) to determine which forecasting model performs best. Secondly, we use the Diebold–Mariano test to compare the predictive accuracy between two competing models. Thirdly, we use the Pesaran-Timmerman test to analyze the accuracy of directional forecasting of the models. Finally, we use the Giacomini-Rossi test to examine the performance of two competing models in the presence of possible instabilities

3.2.1 Root Mean Square Error and Mean Percentage Error

Forecasting Horizons

In this section, we evaluate the forecasting accuracy with two measures: the root mean square error (RMSE) and the mean percentage error (MPE) for the models' forecasts and inflation expectations.

We assess the forecasts of the quantitative models in Table A3.4, Appendix 3. 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.

We also analyze inflation expectations (see Table A3.5, Appendix 3). Based on the RMSE, the EEP's inflation expectations are more accurate than those of the DIE at both the one and two-year horizons.

December Evaluation

We proceed to analyze the forecasting accuracy of the quantitative models to predict the December inflation rate for the current and the following year, based on the RMSE (see Table A3.6, Appendix 3). We observe that the forecasts of the AMM model are better than those of the other models in the first five forecasting horizons, while the IV's forecasts are best for the last three horizons.

3.2.2 Diebold–Mariano Test

In this section, we use the Diebold–Mariano test to compare the predictive accuracy of the forecasts between two competing models. We describe the test in detail in Appendix 2.

Forecasting Horizons

- Quantitative Models

We use the Diebold–Mariano test to compare the accuracy of the two models' forecasts. We start with the AMM and IV models (see Table A3.7, Appendix 3). The forecasting evaluation sample is from 2011q1 to 2017q2. The p-values of the DM statistic are shown in parenthesis.

In the one-, two- and three-quarter forecasting horizons, the mean square error (MSE) of the IV model is lower than that of the AMM model. However, we cannot reject the null hypothesis that the difference between the MSEs is equal to zero; therefore, in this sense, the accuracy of the forecasts of the two models is the same.

In the case of four- and eight-quarter forecasting horizons, the mean square error of the AMM model is lower than the IV model, and we reject the null hypothesis of equal accuracy, so we conclude that the accuracy of the forecasts of the AMM model is better for both intermediate and long time horizons.

Then, we compare the forecasting accuracy between the EFP and the IV models (see Table A3.8, Appendix 3). For this purpose, we use a forecasting evaluation period spanning from 2014q2 to 2017q2. The p-values of the DM Statistic are shown in parenthesis.

We observe that the MSE of the IV model is lower than the EFP model in all forecasting horizons. Furthermore, we also reject the null hypothesis of equal forecasting accuracy in all forecasting horizons. Therefore, the forecasts of the IV model are more accurate than those of the EFP model.

Third, we compare the prediction performance between the forecast of the EFP model and the AMM model with a forecasting evaluation sample from 2014q2 to 2017q2 (see Table A3.9, Appendix 3). The p-values of the DM statistic are shown in parenthesis.

We observe that the MSE of the AMM model are lower than those of the MSE of the EFP model for the one-, two-, and three-quarters forecasting horizons. Also, the DM statistics for those forecasting horizons are statistically significant at 10 percent, which means that we can reject the null hypothesis and conclude that the forecasts of the AMM model are more accurate. It seems that the forecasts of the AMM model are better in the short run than those of the EFP model.

Similarly, the MSE of the AMM model are lower than those of the EFP model in the case of the four-quarters horizon. However, the DM statistic is not statistically significant; therefore, we cannot reject the null hypothesis of equal accuracy between the forecasts.

Also, we cannot reject the null hypothesis of equal accuracy between the forecast of the AMM and the EFP in the case of the eight-quarters horizon.

- Inflation Expectations

We also evaluate the predictive performance of the inflation expectations of both the EEP and the DIE (see Table A3.10, Appendix 3).

First, we compare the forecasting accuracy for the 1-year horizon and a sample of 24 months. We observe that the MSE of EEP is lower than the MSE of the Economic Research Department (DIE). However, the DM statistic is not statistically significant; therefore, we cannot reject the null hypothesis of equal accuracy.

Then, we evaluate the predictive performance for the two-year horizon and a sample of 12 months. We observe that the MSE of the EEP is lower than the MSE of the DIE, and we can reject the null hypothesis of equal accuracy; therefore, we conclude that the EEP's inflation expectations are more accurate than those of the DIE in the case of the two-year horizon.

December Evaluation

We compare the forecasting accuracy of the AMM and the IV models to predict the December inflation rate, for different time horizons, with the DM test. The p-values are shown in parenthesis.

From Table A3.11, Appendix 3, we observe that the DM statistic is not statistically significant in the one- and two-quarter horizons, so we cannot reject the null hypothesis of equal accuracy.

However, the DM statistic is statistically significant for the three- to eight-quarter horizons; therefore, we reject the null hypothesis of equal accuracy and conclude that the AMM's forecasts are more accurate than the IV model's, since the former's MSE is lower than the latter's.

We now compare the forecasting accuracy for the EFP and the IV models (see Table A3.12, Appendix 3). The DM statistic is statistically significant in every forecasting horizon, so we can reject the null hypothesis of equal accuracy. Therefore, we can compare the MSE of both models to determine which one is better.

The MSE of the IV model is lower than the EFP for the one- to five-quarter horizons, so we can conclude that the IV model is more accurate at predicting the December inflation rate for short and intermediate time horizons. On the other hand, from six to seven quarters ahead, the EFP model is more accurate; therefore, it is more accurate for long time horizons.

Finally, we compare the forecasting accuracy for the EFP and the AMM models to predict the December inflation rate (see table A3.13, Appendix 3). We observe that the DM statistic is not statistically significant in the case of the one-quarter horizon, so we cannot reject the null hypothesis.

However, the DM statistic is statistically significant for two- to seven-quarter horizons, so we can reject the null hypothesis and examine the MSEs to determine which model predicts the December inflation rate best.

The AMM model predicts the December inflation rate best in the cases of two-, three-, four-, and five-quarter time horizons, while the EFP model outperforms the AMM one in the cases of six- and seven-quarter time horizons.

Therefore, we can conclude that the AMM model predicts the December inflation rate best in the short and middle run while the EFP model does so in the long run.

3.2.3 Pesaran–Timmermann (PT) Test

In this section, we use the Pesaran–Timmerman test to examine the ability of the forecast to predict the direction of change for the inflation rate. We describe the test in detail in Appendix 2.

Forecasting Horizons

- *Quantitative Models*

We start with the directional forecasting evaluation of the quantitative models. The critical values to reject the null hypothesis of independence are ± 1.645 at 10 percent level of significance. The Pesaran–Timmermann values of the test for each model are shown in Table A3.14, Appendix 3.

First, we examine the directional accuracy of the forecast in the case of the IV model. From table A3.14, the S_n statistic is 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 successfully predict the direction of inflation in the short run.

However, the S_n statistic is not statistically significant in the cases of four- and eight-quarter horizons, so we cannot reject the null hypothesis of independence for those horizons.

Now, we evaluate the directional accuracy in the case of the AMM model. 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 model. 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 and conclude that the forecast of the EFP model can predict successfully the direction of the inflation in the case of that particular horizon.

In sum, the best model to predict the direction of the inflation is the IV model in the case of the quantitative models.

- Inflation Expectations

Finally, we analyze the directional forecasting ability of the inflation expectations (see table A3.15, Appendix 3). In the case of the one-year horizon, we cannot reject the null hypothesis of independence because in both models the S_n statistic is lower than its critical value. However, in the case of the two-year horizon, we reject the null hypothesis of independence only in the case of the forecast of the economic experts' panel; hence, in this case we can conclude that the panel can predict successfully the direction of inflation.

December Forecast Evaluation

We evaluate the directional change accuracy of the inflation forecast for December only for the case of the IV and AMM models, since we do not have enough data for the case of the EFP model.

First, we start with the IV model. We observe from table A3.16, Appendix 3, that 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.

Second, we evaluate the performance of the AMM model. We observe from table A3.16, Appendix 3, that we can reject the null hypothesis of independence in the case of the one-, two-, three-, six-, seven-, and eight-quarter horizons; therefore, the model can predict successfully the directional change of inflation in both the short and the long run.

3.2.4 Giacomini–Rossi Fluctuation Test

In this section, we use the Giacomini and Rossi fluctuation test 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 forecasts of the economic experts panel 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 graphic 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. We explain in the test in detail in Appendix 2.

Forecasting Horizons

- *Quantitative Models*

First, we start with the comparison of the forecasting performance of the AMM and the IV models with the use of the GR statistic. We define the loss function for the two models as follows:

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

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.

From Table A3.17, Appendix 3, 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. It means that one model displays better predictive ability to forecast inflation in at least one period of time.

The fluctuation test allows us to determine whether there is a change in predictability between the forecasts of the IV and the AMM models. Unlike the Diebold–Mariano test, the GR test compares the performance of both models in different sub-samples through rolling windows.

In Table A3.18, Appendix 3, we deliver a summary of the Giacomini–Rossi fluctuation test for each forecasting horizon. We observe that the forecasts of the IV model are more accurate than those of the AMM model one step ahead. However, it seems that the forecasts of the AMM model better predict the inflation patterns at four- and eight-quarter horizons. In addition, in Appendix 4 we explain in further detail the results of the test with a graphic analysis.

Second, we compare the forecasting accuracy between the EFP and the IV model with the use of the Giacomini-Rossi fluctuation test. The loss function between the two models is as follows:

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

In the same way than the previous analysis, if the loss function is negative then the forecasts of the EFP model are more accurate than those of the IV model. On the other hand, if the loss function is positive, the forecasts of the IV model are better at predicting inflation.

From Table A3.19, Appendix 3, we see that the null hypothesis of equal accuracy is rejected in every forecasting horizon since the GR–statistic is higher than the critical value. It means that, at least in one period, one model generates more accurate forecasts of inflation.

In Table A3.20, Appendix 3, we show a summary of the Giacomini–Rossi test for each forecasting horizon. We observe 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 at all forecasting horizons. In Appendix 4, we explain in more detail the results of the test with a graphic analysis.

- Inflation Expectations

We use the Giacomini-Rossi Fluctuation test to examine the performance of the inflation expectations from the Economic Research Department (DIE) and the Economic Experts Panel (EEP), where we assume the latter as a benchmark model. We estimate the loss function of the inflation expectations of the DIE and the EEP.

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

In the qualitative forecasts, we have two scenarios: one and two-year forecasting horizons. Also, if the loss function is negative then the inflation expectations of the EEE are more accurate than those of the DIE. On the contrary, if the loss function is positive, the inflation expectations of the DIE are better at predicting inflation.

From Table A3.21, Appendix 3, the GR statistic is higher than its critical values at both one and two-year horizons. Therefore, we reject the null hypothesis of equal accuracy. It means that one model outperforms the other in at least one period of the time.

In Table A3.22, Appendix 3, we show a summary of the Giacomini-Rossi test. From the table, there is a fluctuation of the forecasting accuracy of the inflation expectations between the two models in the case of the one-year horizon. However, the inflation expectations of the EEE better predict the inflation patterns in the case of the two-year horizon.

3.2.5 Forecasting Efficiency

In this section, we test the efficiency of the unconditional forecasts of both the quantitative and the qualitative models. These tests will complement the previous tests of accuracy and precision in order to evaluate the overall forecast performance of the models.

Before evaluating the efficiency of the forecasts, we start examining the absence of bias and serial autocorrelation. Then, we use the weak and strong efficiency test to evaluate the efficiency of the forecasts. We explain the tests in more detail in Appendix 2.

Unbiasedness Test

- Forecasting Horizons
  - *Quantitative Models*

We start with the quantitative models: AMM, IV, and EFP.

From Table A3.23, Appendix 3, we observe the unbiasedness tests (the p-values are shown in the parenthesis) for the quantitative models. If the p-value is lower than 0.05 then we reject the null hypothesis and the model is biased in the forecasting horizon.

The forecasts of the AMM model are unbiased from one to third forecast horizons because we cannot reject the null hypothesis that the constant term is equal to zero. However, in the case of four and eight quarters ahead, the forecasts are biased.

We observe a similar pattern in the case of the forecasts of the IV model. In this case, the forecast are unbiased only for one and two quarters ahead. After that, the forecasts are biased because we reject the null hypothesis of unbiasedness.

Finally, the forecasts of the EFP model are only biased at the one quarter ahead horizon. After that, they are unbiased.

In sum, the forecasts of the AMM and the IV become biased in the middle and long run.

- *Inflation Expectations*

Table A3.24, Appendix 3 shows shown the unbiasedness tests for the inflation expectations of both the panel of economic experts (EEP) and the economic research department (DIE). Again, if the p-value is lower than 0.05, the forecasts are biased.

We observe that both the EEP and DIE models are unbiased in the case of the one-year horizon because we cannot reject the null hypothesis at the 5 percent level of confidence. However, the models are biased in the case of the two-year horizon.

Autocorrelation Test

- Forecasting Horizons
- *Quantitative Models*

We start with the quantitative models: AMM, IV, and EFP. The null hypothesis of the test is that there is no autocorrelation. If the p-value is lower than 0.05, we reject the null hypothesis of no autocorrelation. Both the unbiasedness and the autocorrelation test represent a preliminary weak efficiency test.

From table A3.25, Appendix 3, we show the autocorrelation tests where the p-values are shown in parenthesis. In the case of the forecasting errors of the AMM model, we observe autocorrelation in all forecasting horizons because we reject the null hypothesis at the 5 percent level of significance.

Similarly, the forecasting errors of the IV model show serial correlation in almost all forecasting horizons with the exception of one step ahead. In the case of the EFP model, there is evidence of autocorrelation only in the case of one quarter ahead, after that; there is no serial correlation in the remaining forecasting horizons o *Inflation Expectations*

We continue with the forecasts of the inflation expectations. From A3.26, Appendix 3, we observe that both the inflation expectations of the DIE and the EEP, shows autocorrelations in all forecasting horizons.

Weak Efficiency Tests

- Forecasting Horizons

o *Quantitative Models*

We start with the quantitative models: AMM, IV, and EFP. Table A3.27, Appendix 3 presents appears the F-statistic of the weak efficiency test, where the p-values appear in parenthesis.

From the second column, we observe that the forecast of the AMM model satisfies the weak efficiency hypothesis only in the case of one quarter ahead, because we cannot reject the null hypothesis (see Appendix 2). In the case of the remaining forecasting horizons, the model does not satisfy the weak efficiency hypothesis.

Then, we analyze the weak efficiency of the IV forecasts. From the third column, we observe that forecasts of the model satisfy the weak efficiency only in the case of one and two forecasting horizons.

Finally, we evaluate the weak efficiency of the EFP forecasts. From 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 are more efficient than those of the other models based in the results of the weak efficiency test. Also, the forecast of the AMM and the IV are weakly efficient in the short run.

o *Inflation Expectations*

From Table A3.28, Appendix 3, we observe the weak efficiency tests of the inflation expectations. In this case, the forecasts of both the EEP and DIE model do not satisfy the weak efficiency test at the 5 percent level of significance in all forecasting horizons. Therefore, the forecasts are not weakly efficient.

o *December Evaluation*

We test for weak efficiency only in the case of the AMM and the IV models because of data availability (see table A3.29, Appendix 3). In the case of the forecasts of the AMM model, we cannot reject the null hypothesis of weak efficiency only in the case of two and three quarters ahead. The null hypothesis is rejected in the remaining forecasting horizons. 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.

Strong Efficiency Tests

- Forecasting Horizons

We perform the strong efficiency test for the two econometrics models: IV and EFP in this sub-section. 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 and it may be included in them.

We consider five variables in logs of the structural model of the Central Bank of Guatemala to make the test: Consumption, index of raw materials, investment, government spending, and credit.

We start with the IV model; the tests are shown in table A3.30, Appendix 3. In the second column, it appears the coefficient of consumption. We observe that we cannot reject the null hypothesis at the 5 percent level of significance in the case of one and two quarters ahead. Therefore, the forecasts are strongly efficient at those horizons. However, from three to eight quarters ahead, consumption does explain the forecasting error, which means that they are not strongly efficient in those horizons.

Similarly, in the third column, the null hypothesis is not rejected at 5% level of significance. 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 mean 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 mean that they are not strongly efficient in those horizons.

We continue with the EFP model; the tests are shown in Table A3.31, Appendix 3. We observe that we reject the null hypothesis in the case of one quarter ahead for the five variables, which mean 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 in 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.

In sum, the forecasts of the IV model are strongly efficient in the short run but lose the efficiency in the long run. Also, the forecasts of the EFP model are strongly efficient from two quarters ahead in all structural variables.

- December Evaluation

In the case of the evaluation of December, we perform the test only for the IV model due to data availability.

Table A3.32, Appendix 3 shows the tests for the eight forecasting horizons. 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 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 with the exception of three quarters ahead, which means that the forecasts are strongly efficient in most of the forecasting horizons.

4. Conditional Forecast Evaluation

4.1 The Models

4.1.1 The Semi-Structural Macroeconomic Model 4.01 (MMS 4.0.1)

The MMS 4.0.1 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 (Banguat) that have evolved from the first version launched in 2006.

The MMS 4.0.1 was built on the basis proposed by Berg, Karam and Laxton (2006a and 2006b), who provided a practical guide to non-micro funded DSGE models and its implementation 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 percent) are endogenous and 12 (30 percent) are exogenous variables. The MMS 4.0.1 delivers forecasts for both core inflation and headline inflation. The model is currently used for producing inflation and monetary policy interest rate forecasts that are inputs for Banco de Guatemala's monetary policymaking process.

Variables that display higher volatility are transformed through a moving sum (average) scheme in order to reduce that volatility and avoid possible outliers. We thus obtain smoothed series.

4.1.2 Macroeconomic Model of Inflation Forecast for Guatemala (PIGU)

The second model (PIGU) 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. are the following: the set of exogenous variables, the exogenous variables' volatility, and the type of inflation.

First, the set of exogenous variables: 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.

PIGU is currently available to all the central bank's staff, through a custom-made interface. The interface is a tool built on MATLAB that allows the user to become familiar with the equations of the PIGU model, to create different forecasting scenarios based on alternative paths for any particular variable (endogenous or exogenous), and to modify the calibration of the model and visualize the corresponding effects on the conditional forecasts. It also shows the dynamic effects of different shocks (impulse-response functions). PIGU's interface is designed to be a user-friendly technological tool, which does not require programming or modeling skills from the user.

4.1.3. Macroeconomic Structural Model (MME)

The structural model is a medium scale DSGE model, built within the New-Keynesian framework. It features a financial accelerator à la Bernanke et al. (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 by 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.

Some parameters of the model are calibrated, and a subset is estimated using Bayesian technics with quarterly data. Bayesian estimation was performed using Dynare. The first sample runs from 2001Q1 to 2010Q4. Then we add one quarter at a time, until we reach 2017Q1, the last observed data used to forecast corresponds to 2017Q2. Bayesian estimation deals with outliers by assigning a small probability (in the tails of the posterior distribution) to the parameter values that could generate such extreme values. Observable variables were introduced (for estimation) as stationary series. They were transformed, in accordance with the model, as the first difference of the natural logarithm, of seasonal adjusted data. In the case of the consumer price index (from which inflation rate is derived), there are no structural breaks.

4.2 Evaluation Methodology

The quality of any variable's conditional forecasts depends on two elements: i) the performance of the forecasting model (as such) and ii) the quality of the forecasting model 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 those 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 in, for each forecasted period, the historically observed values of exogenous and endogenous variables. In the case of the structural model (MME), exogenous variables are represented by stochastic processes, typically of an autoregressive nature. Therefore, alternative scenarios are only conditioned by historically observed values of two endogenous variables: inflation and output.

4.2.1 Evaluation Sample[3]

For each of the three evaluated models, we generate quarterly headline inflation forecasts with a sample from 2011Q1 to 2017Q2. In addition, we consider five forecasting horizons: one quarter, two quarters, four quarters, six quarters, and eight quarters.

4.2.2 Forecasting Scenarios

For each of the three evaluated models, we consider different scenarios.

MMS 4.0.1 Scenarios

Four scenarios are considered: 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 are generated by the corresponding historically observed data on 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.

PIGU Scenarios

Four scenarios are considered: 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.

[3] 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 percent in the third quarter of 2008 to the negative value of -0.73 percent one year later (in August 2009). Therefore, in order to get robust results, we began our evaluation from 2011Q1.

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, and all other endogenous variables' 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.

MME Scenarios

It 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.[4]

4.2.3 Forecasting Evaluation

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 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.

5. Main Results

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.

[4] 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.

First, we show the inflation patterns and the forecasts of each model (see Tables A5.1, A5.2, and A5.3, Appendix 5)

Second, we calculate the mean error (see Tables A5.1, A5.2, and A5.3, Appendix 5) and the Root Square Mean Error (see Tables A5.4, A5.5, and A5.6, Appendix 5).

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 Mean Error (ME) and Root Mean Squared Error (RMSE).

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 Table A5.4, Appendix 5). 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 below.

Table 1. 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 2 periods of inflation**
4	1.37	0.61
6	1.36	0.62
8	1.57	0.65

Source: Authors' compilation, Central Bank forecasts.

For the MME, the Mean Error suggests that there is a positive inflation bias (see Table A5.3, Appendix 5). 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) become available.

6. Conclusions

In this paper, we evaluate the accuracy and precision of both the unconditional and conditional inflation forecasts of Banco de Guatemala.

In the case of the unconditional forecasts, we evaluate the forecasting accuracy of three quantitative models: Indicator Variable (IV), Average Macroeconomic Models (AMM), and the Efficient Forecasting Path (EFP) in two scenarios: forecasting horizons, and the prediction of the inflation for the month of December. In each case, we evaluate the precision, accuracy, and efficiency to predict inflation with different measures and tests: normality, RMSE, Diebold-Mariano (DM), Pesaran-Timmerman (S_n), Giacomini-Rossi, and the weak and strong efficiency tests. The sample evaluation is rather short, especially in the case of the EFP model where we have only 13 quarters. Also, the sample period to evaluate the forecasting accuracy to predict the inflation in the month of December is rather short. We have at most six years of data.

First, we evaluate the forecasting accuracy of the models from one to eight steps ahead. We found empirical evidence that the forecasts of the AMM model are more accurate than those of the IV model in the middle and long term according to the Diebold–Mariano and Giacomini–Rossi Fluctuation test. Similarly, the forecasts of the IV model are more accurate than those of the EFP model in all forecasting horizons.

Also, the forecasts of the AMM and the IV model are weakly efficient in the short run (one and two steps ahead), while the forecasts of the EFP are weakly efficient from two quarters ahead

In addition, the forecasts of the IV model can predict better the directional change of inflation than the others according to the Pesaran–Timmerman test.

Second, we evaluate the forecasting accuracy of the models to predict the inflation in the month of December with both the Diebold–Mariano and Giacomini–Rossi tests. We found empirical evidence that the forecast of the AMM model are more accurate than those of the IV model in the prediction of inflation in the month of December from three to eight quarters horizons. Also, the forecasts of the IV model are better at predicting inflation in December than those of the EFP model for all forecasting horizons.

In addition, the forecasts of the IV model are more weakly efficient than those of the AMM model.

Third, we compare the predictive ability of the inflation expectations of both the Economic Expert Panel (EEE) and the Economic Research Department (DIE). We found empirical evidence that the inflation expectations of the EEE are more accurate than those of the DIE in the case of the two-year horizon, according to the Diebold–Mariano and the Giacomini–Rossi Fluctuation tests.

Furthermore, we found empirical evidence that the inflation expectations of the economic expert's panel can predict the directional change of inflation in the two-year forecasting horizon.

In the case of the conditional forecast, we evaluated the forecasting accuracy to predict headline inflation with three models: MMS 4.0.1, PIGU, and MME. Both the MMS 4.0.1 and PIGU are evaluated in four scenarios (free, anchor 1, anchor 2, and anchor 3), while the MME considers only two scenarios (free and anchor 1), depending on which (if any) exogenous and/or endogenous variables are anchored by their historically observed values.

First, across different scenarios of the same model, the MMS 4.01 model minimizes the RMSE in both the anchor 2 (the inflation forecast for the first quarter and all the exogenous variables in the forecasting horizon are anchored by the corresponding historically observed data) and anchor 3 (in addition, the monetary policy interest rate is also anchored by the historically observed data) scenarios for all forecasting horizons.

Also, PIGU model minimizes the RMSE in the anchor 3 scenario (anchoring exogenous variables, two quarters of inflation, and all other endogenous variables) for all forecasting horizons.

In addition, in the case of the structural model (MME), anchoring one quarter of inflation and output reduces both the mean and root mean square errors. This reduction lasts only for the first three quarters of the forecast horizon.

Secondly, across different models, the model that minimizes the mean error average (across different forecasting horizons) is PIGU without any anchoring. In the case of the RMSE, again PIGU model renders the lowest average across the different forecasting horizons considered, when anchoring exogenous and endogenous variables, plus two periods of inflation.

In addition, when forecasting eight quarters ahead, the lowest RMSE is observed in PIGU and the highest is observed for the MME.

Finally, the mean error shows a positive forecast bias for all relevant quarters in both MMS 4.0.1 and MME. Only for PIGU shows a negative bias for all relevant forecasting horizons.

The main purpose of this study was to learn about 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 variables transformation procedures. In addition, we should perform a complete evaluation of the inflation forecasts at least once a year, as some central banks already have.

Appendix 1. Efficient Forecasting Path (EFP)

The method described in this appendix was first developed in Castillo and Ortiz (2017). It is designed to generate an efficient out-of-sample forecast path (EFP) for a finite time series, based on a variety of forecasting models. The following sections summarize its methodology, the models and data employed to generate inflation forecasts.

1. Methodology

Let us assume that $\boldsymbol{y}_t$is an N-size vector containing a finite time series, which we are interested in forecasting for *h* periods ahead of the sample size. To perform this task we can use *k* models of the following types: ARIMA, OLS, SWLS, VAR, and VEC. In addition, let $\boldsymbol{y}_t^w$ be an *s*-size vector containing an ordered subsample of the original series, such that there could be a total of *w* continuous-size-*s* windows that can be derived from the original sample. Based on this information, the EFP algorithm can be computed through the following five steps.

Step 1: Rolling regressions

The first step consists of estimating rolling regressions of sample size *s* for each model *k*, so as to generate in-sample forecasts for *h* periods ahead.

Step 2: Absolute forecast errors and model weights

A different weight, $q_{s+h,k}^w$ is assigned to each model *k* according to its performance (or forecast accuracy) to project each period *s+h*. Such a weight is equivalent to the inverse absolute forecast error of each model expressed as a ratio of the summation overall absolute forecast error inverse values generated by every model *k* when forecasting each s+*h* period at the specific window *w:* $\hat{q}_{s+h,k}^w = \frac{\widehat{fi}_{s+h,k}^w}{\widehat{FI}_{s+h}^w}$. Notice that each element of $\hat{q}_{t,k}^w$ has a value between zero and one. The higher its value, or the closer is it to one, the more accurate its forecast of *s+h*, and the more important would be model *k* when forecasting outside the sample size.

Step 3: Average weight for each model

For each model *k*, it is computed a simple average of weights, overall rolling regression windows. This term is denoted as $\hat{q}_{s+h,k}$, and its value fluctuates between zero and one.

Step 4: Average weight distributions and endogenous trimming

The last step involves getting rid of the models whose forecast efficiency, denoted by its average weight $\hat{q}_{S+h,k}$, is too low relatively to the rest of models. This is done by computing the mean and standard deviation for all weights that forecast the same period $s+h$, so as to obtain an average weight distribution, which allows us to select models based on above-average performance. Through this endogenous trimming method we could get rid of (set to zero) those average weight values at the lower extreme of each distribution. The proportion of models, whose average weights were set to zero, represented those who were below two standard deviations above the mean, so we ended up with the top 5 percent performers. Hence, we also renormalized each weight, $\hat{Q}_{s+h,k}$, so that the overall summation is still add up to one.

Step 5: Efficient Forecast Path

The final step in the algorithm is to generate out-of-sample forecasts for *h* periods ahead for each model *k*, and to weight each forecast by each model's normalized weight $\hat{Q}_{s+h,k}$. Therefore, each weighted forecast, say $\hat{Q}_{s+h,k}\hat{y}_{N+h,k}$, can be interpreted as the contribution of model k to y_t's forecast of period $N+h$. Hence, the summation overall *k* models is equivalent to a single forecast for y_t, and it is defined as the Efficient Forecast Path (EFP) of y_t.

2. Forecasting Models

In order to proceed with the forecast combination we created a set of 1,380 models to forecast Guatemalan inflation. In particular, we employed five types of models: ARMA (*p*, *q*), Ordinary Least Squares (OLS), Stepwise Least Squares (SWLS), Vector Autoregression (VAR), and Vector Error Correction (VEC) models. Different lag and variable combinations were performed for each kind of model. All variables were transformed to their logarithmic form, and rolling estimation through moving windows was performed for each model specification.

2.1 Autoregressive Moving Average (ARMA) Models

ARMA(*p*,*q*) models are univariate representations that express a variable to be modeled (y_t) as a function of its own lags (p), as well as lags (q) of the error term (ε_t). Let $\Omega_{ARMA} = ARMA(1,0), ARMA(2,0), \ldots ARMA(p,0), ARMA(1,1), \ldots ARMA(p,q)$, be the set of models estimated per country. Therefore, the total number of models contained in such a set is equal to $\Omega_{ARMA} = 2p(1+q)$, where p and q varied per country according to data availability. We considered values for p and q of 12 and 11, respectively. Hence, Ω_{ARMA} is composed of 288 models

2.2 Ordinary Least Squares (OLS) Models

The OLS models that we employed are classical econometric representations of variables as a function of their past values, and one or more independent variables and their lags. Let us assume that Ω_{OLS} is the set of all OLS models estimated, which contains all possible combinations of multivariable models, along with all possible combinations of lags. In this case, the total number of OLS models contained in Ω_{OLS} equals to $5p_0 + 4p_1 + 6p_2 + 4p_3 + p_4$. The sub-index in the lag expression denotes the number of inflation fundamentals included in each estimation. By letting p_0, p_1, p_2, p_3 and p_4 to be equal to 31, 21, 15, 12, and 10, respectively, it would imply that Ω_{OLS} contains 387 (=155+84+90+48+10) models.

2.3 Stepwise Least Squares (SWLS) Models

SWLS is an iterative algorithm proposed by Efroymson (1960) to automatically obtain the best fit of OLS regressions.[5] Each model representation is identical to those of the previous section, but the final results differ, since the SWLS algorithm was established to select just those regressors whose *p-value* was lower or equal to 0.05. The set of all SWLS models estimated per country is equal to the OLS set ($\Omega_{SWLS} = \Omega_{OLS}$).

2.4 Vector Autoregression (VAR) Models

Unrestricted VAR models are systems of equations that express each variable as a function of its own past values, and lags of the remaining variables in the system. We tried all combinations from two to five variables (domestic IPC and its four fundamentals), and different lag specifications. The set of all VAR models estimated per country, Ω_{VAR}, contained $4P_2 + 6P_3 + 4P_4 + P_5$ models. As before, the sub-index in the lag expression denotes the number of inflation fundamentals included in each VAR model. By letting P_2, P_3, P_4 and P_5 to be equal to 21, 15, 12, and 10, respectively, it would imply that Ω_{VAR} is composed of 232 (=84+90+48+10) models.

[5] Derksen and Keselman (1992), and Burnham and Anderson (1998) provide a description of the algorithm and describe some advantages and disadvantages of this methodology.

2.5 Vector Error Correction (VEC) Models

We estimated equilibrium VEC models for one cointegrating relationship through the Johansen procedure. As with the previous case, combinations for all possible variable and lag specifications were performed. Although the procedure to compute the number of models resembles the VAR case, the number of lags considered to estimate each VEC model specification was lower: 8, 6, 4, and 2, respectively. Hence, Ω_{VEC} is composed of 86 (=32+36+16+2) models.

3. Data

Guatemalan inflation was estimated and forecast based on the models described above, using quarterly data from 1995Q1 to 2017Q1. The data set also included information for four well-identified inflation fundamentals (Castillo, 2014): i) US Inflation; ii) Nominal Exchange Rate (Q/US$); iii) Real Money Supply; and iv) Real Banking Credit. As the United States is Guatemala's main trading partner, US price fluctuations, along with Q/US$ variations are transferred quickly and almost fullly to domestic prices. Moreover, changes in domestic money supply and banking credit affect domestic inflation indirectly through their effect on domestic GDP.[6] Following Clements and Hendry (1999), data were neither deseasonalized nor detrended in order to avoid missing important forecasting information. Finally, for empirical purposes all variables were transformed to their log form. The results obtained are described in the following section.

Appendix 2. Forecasting Evaluation Tests

In this appendix, we explain in detail the different tests considered to evaluate the forecasting performance of the unconditional models.

1. Tests of the Residuals

We examine the key statistical properties of the forecasting errors: Normality and Skewness.

First, the normality of the forecasting errors is evaluated with the Jarque–Bera test defined as

$$JB = \left(\frac{n-k-1}{6}\right)\left(S^2 + \frac{1}{4}(C-3)^2\right) \qquad \text{(A2.1)}$$

[6] A quarterly series for Guatemalan GDP is not available for the whole period under consideration.

where n is the number of observations, k is the number of repressors, S is the sample skewness, and C is the sample kurtosis

The null and the alternative hypotheses of the test are the following:

H_0: The forecast errors follow a standard normal distribution.

H_1: The forecast errors do not follow a standard normal distribution.

Second, the skewness allows us to know whether the forecasts overestimate or underestimate the inflation patterns.

2. Root Mean Square Error and Mean Percentage Error

We consider the Root Mean Square Error (RMSE) and the Mean Percentage Error (MPE) to test the precision of the forecasts of each model. The model with the lowest RMSE is considered the best.

The RMSE is defined as follows:

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{t=1}^{n}(y_i - \hat{y}_i)^2} \tag{A2.2}$$

The MPE is defined as follows:

$$MPE = \frac{100}{n}\sum_{t=1}^{n}\frac{y_i - \hat{y}_i}{y_i} \tag{A2.3}$$

where y_i is the value of inflation, $\hat{y}_i$ is the inflation forecast, and n is the number of forecasting evaluation periods.

The RMSE gives us an initial evaluation about the precision of the models to forecast inflation. After that, we use the Diebold-Mariano test to examine whether the difference between the MSE is statistically significant or not to determine which model has better forecasting accuracy.

3. Diebold-Mariano Test

We use the Diebold-Mariano (DM) test developed by Diebold and Mariano (1995) to compare the predictive accuracy of two competing models in both quantitative models and inflation expectations.

The null and the alternative hypotheses are the following:

$$H_0: \; g(e_{1t}) - g(e_{2t}) = 0$$
$$H_1: \; g(e_{1t}) - g(e_{2t}) \neq 0$$

where $g(e_{it})$ is the loss function associated with the forecasting error term of each model. In our paper, we assume the mean square error as the loss function. The null hypothesis establishes that the forecasts have the same forecasting accuracy. If we reject the null hypothesis, the model with the lower mean square error is the best at forecasting inflation.

The Diebold Mariano statistic (DM) is defined as

$$DM_{12,t} = \frac{d_{12}}{\sigma_{d_{12}}} \;\rightarrow\; \mathcal{N}(0,1) \qquad \text{(A2.4)}$$

where $d_{12} = (e_{1t}) - g(e_{2t})$ and $\sigma_{d_{12}}$ is a consistent estimate of the standard deviation of the DM statistic. The statistics follows a standard normal distribution with mean zero and standard deviation equal to 1.

4. Pesaran-Timmerman Test

We also use the Pesaran – Timmerman (PT) test developed by Pesaran and Timmerman (1993) to examine the ability of the forecasts to predict the direction of the change of inflation in the case of the quantitative models as well as those of the inflation expectations.

The null and alternative hypotheses of the test are the following:

H_0: The forecasts are not able to predict the change of inflation.

H_1: The forecasts are able to predict the change of inflation.

The Pesaran – Timmerman Statistic is defined as:

$$S_n = \frac{\hat{P} - \hat{P}_*}{[\hat{V}(\hat{P}) - \hat{V}(\hat{P}_*)]^{1/2}} \;\rightarrow\; \mathcal{N}(0,1) \qquad \text{(A2.5)}$$

where $\hat{P} = \frac{1}{n}\sum_{t=1}^{n} I(y_t, x_t)$, in which y_t represents the change of the consumer price index, x_t represents the change of inflation forecasts of every model, and $\widehat{\mathrm{P}}$ represents the proportion of times where both the change of inflation and the change of the forecast of inflation go in the same direction either increasing or decreasing.

We additionally have the following definitions:

$$\hat{P}_* = \hat{P}_y \hat{P}_x + \left(1 - \hat{P}_y\right)\left(1 - \hat{\mathrm{P}}_{\mathrm{x}}\right) \quad \text{(A2.6)}$$

$$\hat{\mathrm{P}}_{\mathrm{y}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{t}=1}^{\mathrm{n}} \mathrm{I}(\mathrm{y}_{\mathrm{t}}) \quad \text{(A2.7)}$$

$$\hat{\mathrm{P}}_{\mathrm{x}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{t}=1}^{\mathrm{n}} \mathrm{I}(\mathrm{x}_{\mathrm{t}}) \quad \text{(A2.8)}$$

where $\hat{\mathrm{P}}_{\mathrm{y}}$ and $\hat{\mathrm{P}}_{\mathrm{x}}$ represents the proportion of the positive changes of inflation and the forecasts of every model.;

$$\mathrm{I}(\cdot) = \begin{cases} 1 \text{ if } (\cdot) > 0 \\ 0 \text{ otherwise} \end{cases} \quad \text{(A2.9)}$$

where $\mathrm{I}(\cdot)$ is an indicator variable which takes the value of 1 if the change is positive and 0 otherwise; and

$$\hat{V}\left(\hat{P}\right) = \frac{1}{n}\left(\hat{P}_*\left(1 - \hat{P}_*\right)\right) \quad \text{(A2.10)}$$

$$\hat{V}\left(\hat{P}_*\right) = \frac{1}{n}\left(2\hat{P}_y - 1\right)^2 \hat{P}_x\left(1 - \hat{P}_x\right) + \frac{1}{n}\left(2\hat{P}_x - 1\right)^2 \hat{P}_y\left(1 - \hat{P}_y\right) + \frac{4}{n^2}\hat{P}_y\hat{P}_x\left(1 - \hat{P}_y\right)\left(1 - \hat{P}_x\right)$$

where $\hat{\mathrm{V}}\left(\hat{\mathrm{P}}\right)$ and $\hat{\mathrm{V}}\left(\hat{\mathrm{P}}\right)$ are the variance of $\hat{\mathrm{P}}$ and $\hat{\mathrm{P}}_*$ respectively.

5. Giacomini – Rossi Fluctuation Test

In addition, we use the Giacomini-Rossi Fluctuation test developed by Giacomini and Rossi (2010) to examine the performance of two competing models in the presence of instabilities. In this test, we can analyze the predictive accuracy of the models in different subsamples by using rolling windows to determine if the behavior of the forecasts is the same across the complete sample.

The null and alternative hypotheses of the test are:

$\mathrm{Ho} : \mathrm{E}\left[\Delta \mathrm{L}_{\mathrm{t}}(\hat{\theta}_{\mathrm{j-h,R}}, \hat{\gamma}_{\mathrm{j-h,R}})\right] = 0$ for all $\mathrm{t} = \mathrm{R} + \mathrm{h}, \dots \dots \dots \mathrm{T}$

$H1\text{:}\ E\left[\Delta L_t(\hat{\theta}_{j-h,R}, \hat{\gamma}_{j-h,R})\right] \neq 0$ When $max_t\left|F_{t,m}^{\infty S}\right| > k_\alpha$

The loss function is defined as the difference between the mean square errors of the two competing models as follows:

$$\Delta \mathrm{L}_{\mathrm{t}}\left(\hat{\theta}_{\mathrm{j-h,R}}, \hat{\gamma}_{\mathrm{j-h,R}}\right) = \mathrm{MSE}_{1,\mathrm{t}} - \mathrm{MSE}_{2,\mathrm{t}} = 0 \quad \text{(A2.11)}$$

The GR – Statistics is defined as

$$F_{t,m}^{\infty S} = \hat{\sigma}^{-1} m^{1/2} \sum_{j=t-m/2}^{t+m/2-1} \Delta L_j\left(\hat{\theta}_{j-h,R}, \hat{\gamma}_{j-h,R}\right) \quad \text{(A2.12)}$$

with $t = R + h + m/2 + 1$

Giacomini and Rossi (2010) found that the rolling window, *m*, has to be an even and small number to preserve the power of the test. In this paper, *m* is set as two.

6. Efficiency Tests

We examine the efficiency of the forecasts with two tests: strong and weak efficiency. Before making the efficiency tests, we also check for the presence of bias and autocorrelation in the forecasting errors.

The efficiency tests complement the previous tests. They aim to verify if there are other variables not included in the models that could improve the accuracy of the forecasts.

6.1 Absence of Bias

This is important to know whether the forecasts are overly optimistic or over pessimistic. To test the absence of bias, we run the following regression

$$\hat{e}_t = \alpha + \varepsilon_t \qquad \text{(A2.13)}$$

where $\hat{e}_t$ is the forecast error, ε_t is a normal distributed error term with zero mean and α is a constant which measures the degree of bias of the forecasts.

Under the null and alternative hypothesis,

$$H_0: \alpha = 0$$
$$H_1: \alpha \neq 0$$

If we reject the null hypothesis, then the forecasts are biased. We test the null hypothesis with a 5 percent level of significance.

6.2 Autocorrelation Tests

Another important test is to examine the possible autocorrelation between the forecasting errors of the models. We estimate the following regression

$$\hat{e}_t = \alpha + \beta\hat{e}_{t-1} + \varepsilon_t \qquad \text{(A2.14)}$$

where $\hat{e}_t$ is the forecast error, ε_t is a normal distributed error term with zero mean, and α is a constant which measures the degree of bias of the forecasts.

Under the joint null hypothesis

$$H_0: \propto = 0, \beta = 0$$

If we reject the null hypothesis, then the forecasts are biased. We test the null hypothesis with a 5 percent level of significance.

6.3 Weak Efficiency

We test if the forecasts contain all the information available included in the models. We estimate the forecast error with an MA structure

$$\hat{e}_t = \alpha + w_t + \beta w_{t-1} \tag{A2.15}$$

where $\hat{e}_t$ is the forecast error, $\propto$ is a constant,
and β is a coefficient of the moving average process

Under the joint null hypothesis

$$H_0: \propto = 0, \beta = 0$$

If we do not reject the joint null hypothesis, then the forecasts are weakly efficient.

6.4 Strong Efficiency

In this test, we examine if the addition of new variables different from those of the models can explain the forecasting errors. If they can, then the forecasts of the models are not strongly efficient. We estimate the following regression

$$\hat{e}_t = \alpha + \beta z_t + \varepsilon_t \tag{A2.16}$$

where $\hat{e}_t$ is the forecast error , ε_t is a normal distributed error term with zero mean, $\propto$ is a constant, and β is the coefficient of the new variables.

Under the joint null hypothesis

$$H_0: \propto = 0, \beta = 0$$

We consider five variables of the structural model to perform the strong efficiency tests in the IV and the EFP model, which are the econometrics models.

Appendix 3. Tables of the Unconditional Forecast Evaluation

In this appendix, we include the tables of the unconditional forecasting evaluation.

1. *Statistical Properties of the Forecasting Error*

Table A3.1. Statistical Properties of the Forecasting Error, Quantitative Models

Forecasting Horizons in quarters	Skewness			Normality (Jarque-Bera Test)		
	IV	AMM	EFP	IV	AMM	EFP
1	-0.15	0.21	0.32	0.90 (**0.64**)	0.70 (**0.70)**	0.805 (**0.67**)
2	0.09	-0.07	0.18	1.28 (**0.52**)	0.05 (**0.97**)	0.200 (**0.91**)
3	-0.03	0.07	0.13	3.13 (**0.21**)	0.31 (**0.87**)	2.000 (**0.78**)
4	-0.41	0.06	0.29	2.45(**0.29**)	3.15 (**0.21**)	0.690 (**0.71**)
8	-0.55	0.51	-0.31	4.50 (**0.11**)	4.60 (**0.10**)	0.345 (**0.84**)

Source: Author's compilation, Central Bank forecasts.

Table A3.2: Statistical Properties of the Forecasting Error, Inflation Expectations

Forecasting Horizons in Years	Skewness		Normality (Jarque-Bera Test)	
	Economic Research Department	Economic Experts' Panel	Economic Research Department	Economic Experts
1	0.80	0.56	3.02 (**0.22**)	1.97 (**0.37**)
2	1.11	-0.32	2.88 (**0.24**)	0.39 (**0.82**)

Source: Author's compilation, Central Bank forecasts.

Table A3.3. Statistical Properties of the Forecasting Error, Quantitative Model, December Evaluation

Forecasting Horizons in quarters	Skewness			Normality (Jarque-Bera Test)		
	IV	**AMM**	**EFP**	**IV**	**AMM**	**EFP**
1	-0.39	0.53	0.71	0.42 (**0.81**)	0.69 (**0.70**)	0.53 (**0.77**)
2	-0.20	-0.19	0.36	1.08 (**0.58**)	0.82 (**0.66**)	0.35 (**0.84**)
3	0.40	-0.36	0.68	1.52 (**0.76**)	2.06 (**0.36**)	0.52 (**0.77**)
4	0.59	0.10	0.00	2.48 (**0.29**)	1.40 (**0.50**)	0.33 (**0.84**)
5	-0.28	0.09	0.00	0.86 **(0.65**)	1.04 (**0.59**)	0.33 (**0.85**)
6	-0.40	-0.25	0.00	0.95 (**0.62**)	1.48 (**0.48**)	0.33 (**0.85**)
7	-0.80	-0.28	0.00	1.48 (**0.48**)	1.85 (**0.40**)	0.33 (**0.85**)
8	-1.08	-0.10	-------	2.93 (**0.23**)	1.45 **(0.47**)	---------------

Source: Author's compilation, Central Bank forecasts.

2. RMSE and MPE

Table A3.4. RMSE and MPE, Quantitative Models

Forecasting Horizons in quarters	Root Mean Square Error (RMSE)			Mean Percentage Error (MPE)		
	IV	AMM	EFP	IV	AMM	EFP
1	0.45	0.50	0.74	0.02	0.02	-0.16
2	0.80	0.83	1.19	0.09	0.08	-0.10
3	1.18	1.12	1.68	0.02	0.15	0.00
4	1.53	1.34	1.98	0.31	0.21	0.10
8	1.80	1.26	0.63	0.44	0.21	-0.03

Source: Author's compilation, Central Bank forecasts.

Table A3.5. RMSE and MPE, Inflation Expectations

Forecasting Horizons in Years	Root Mean Square Error (RMSE)		Mean Percentage Error (MPE)	
	Economic Research Department	Economic Experts' Panel	Economic Research Department	Economic Experts' Panel
1	1.34	1.28	-0.01	0.086
2	0.56	0.41	-0.12	-0.068

Source: Author's compilation, Central Bank forecasts.

Table A3.6. RMSE and MPE, Quantitative Models, December Evaluation

Forecasting Horizons in quarters	Root Mean Square Error (RMSE)			Mean Percentage Error (MPE)		
	IV	AMM	EFP	IV	AMM	EFP
1	0.57	0.56	0.88	0.06	0.04	0.14
2	0.84	0.80	1.14	0.14	0.11	0.30
3	0.96	0.58	1.37	0.15	0.08	0.36
4	1.22	0.93	1.29	0.21	0.10	0.36
5	1.65	1.09	1.12	0.37	0.26	0.24
6	1.83	1.26	0.55	0.41	0.26	0.10
7	1.85	1.23	0.88	0.43	0.23	0.18
8	1.80	1.20	0.56	0.44	0.16	-0.13

Source: Author's compilation, Central Bank forecasts.

3. Diebold – Mariano Tests

Table A3.7. Diebold – Mariano Test between the Forecasts of the AMM and the IV Model

Forecasting Horizons in Quarters	DM Statistic	MSE (AMM)	MSE (IV)
1	1.44 **(0.15)**	0.11	0.07
2	1.30 **(0.19)**	0.56	0.48
3	0.21 **(0.84)**	1.21	1.18
4	-2.95 **(0.00)**	1.68	2.15
8	-3.35 **(0.02)**	1.67	3.33

Source: Author's compilation, Central Bank forecasts.

Table A3.8. Diebold – Mariano Test between the Forecasts of the EFP and the IV Model

Forecasting Horizons in Quarters	DM Statistic	MSE (EFP)	MSE (IV)
1	1.71 **(0.087)**	0.55	0.08
2	1.97 **(0.049)**	1.42	0.08
3	1.79 **(0.074)**	2.81	0.08
4	1.76 **(0.079)**	3.91	0.08
8	2.91 **(0.004)**	0.40	0.10

Source: Author's compilation, Central Bank forecasts.

Table A3.9. Diebold – Mariano Test between the Forecasts of EFP and AMM Model

Forecasting Horizons in Quarters	DM Statistic	MSE (EFP)	MSE (AMM)
1	1.65 **(0.09)**	0.55	0.11
2	2.03 **(0.04)**	1.42	0.26
3	1.70 **(0.09)**	2.81	0.53
4	1.61 **(0.11)**	3.91	1.10
8	-0.87 **(0.38)**	0.40	0.53

Source: Author's compilation, Central Bank forecasts.

Table A3.10. Diebold – Mariano Test between the Inflations Expectations of the EEP and the DIE

Forecasting Horizons in Years	DM Statistic	MSE (Economic Experts' Panel)	MSE (Economic Research Department)
1	-0.64 **(0.523)**	1.63	1.79
2	-5.25 **(0.000)**	0.17	0.32

Source: Author's compilation, Central Bank forecasts.

Table A3.11. Diebold – Mariano Test between the Forecasts of the AMM and the IV Model, December Evaluation

Forecasting Horizons in Quarters	DM Statistic	MSE (AMM)	MSE (IV)
1	1.44 **(0.15)**	0.11	0.07
2	-0.95 **(0.34)**	0.87	0.90
3	-4.60 **(0.00)**	0.24	0.54
4	- 2.33 **(0.01)**	0.76	1.41
5	-3.20 **(0.00)**	1.02	2.75
6	-2.93 **(0.00)**	1.46	3.38
7	-2.98 **(0.00)**	1.36	3.46
8	- 1.95 **(0.05)**	1.64	3.48

Source: Author's compilation, Central Bank forecasts.

Table A3.12. Diebold – Mariano Test between the Forecasts of the EFP and the IV Model, December Evaluation

Forecasting Horizons in Quarters	DM Statistic	MSE (EFP)	MSE (IV)
1	2.10 **(0.036)**	0.77	0.06
2	2.70 **(0.007)**	1.30	0.52
3	2.62 **(0.009)**	1.88	0.51
4	4.75 **(0.000)**	1.67	0.08
5	2.09 **(0.036)**	1.24	1.17
6	-5.61 **(0.000)**	0.31	1.45
7	-62.39 **(0.000)**	0.78	1.32

Source: Author's compilation, Central Bank forecasts.

Table A3.13. Diebold – Mariano Test between the Forecasts of the EFP and the AMM Model, December Evaluation

Forecasting Horizons in Quarters	DM Statistic	MSE (EFP)	MSE (AMM)
1	1.65 **(0.10)**	0.77	0.19
2	2.55 **(0.01)**	1.29	0.41
3	2.58 **(0.00)**	1.88	0.16
4	7.32 **(0.00)**	1.67	0.64
5	3.16 **(0.00)**	1.24	0.19
6	-22.50 **(0.00)**	0.31	0.82
7	2.58 **(0.01)**	0.78	0.50

Source: Author's compilation, Central Bank forecasts.

4. Pesaran – Timmerman Tests

Table A3.14. Pesaran – Timmerman 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

Source: Author's compilation, Central Bank forecasts.

Table A3.15. Pesaran-Timmerman Test, Inflation Expectations

Forecasting Horizons in Years	S_n Statistic (Economic Research Department)	S_n Statistic (Economic Experts' Panel)
1	-0.67	-1.37
2	0.35	-2.62

Source: Author's compilation, Central Bank forecasts.

Table A3.16. Pesaran Timmerman 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

Source: Author's compilation, Central Bank forecasts.

5. Giacomini – Rossi Tests

Table A3.17. GR Test between the Forecasts of the AMM and the IV Model

Forecasting Horizons in Quarters	GR Statistic	Critical Values
1	4.77	3.18
2	15.28	3.18
3	9.93	3.18
4	9.07	3.18
8	11.28	3.01

Source: Author's compilation, Central Bank forecasts.

Table A3.18: Summary Results of the GR test between the Forecasts of the AMM and the IV Model

Forecasting Horizons in Quarters	Fluctuation test summary
1	The forecasts of the IV model are more accurate in three different sub- periods of the total sample than those of the AMM model.
2	There is a fluctuation in the accuracy of the forecasts between the two models. In the beginning of the sample, the IV model is the best to predict inflation. However, in the first and second quarter, the forecasts of the AMM model are more accurate.
3	Similarly, in the beginning of the sample, the forecasts of the IV model are more accurate. However, in the first and second quarter of 2012, the forecast of the AMM model predicts better the inflation patterns.
4	The forecasts of the AMM model are more accurate in two different subs – periods of the total sample than those of the IV model.
8	The forecasts of the AMM model are more accurate in almost all the evaluation sample than the IV model.

Source: Author's compilation, Central Bank forecasts.

Table A3.19. GR Test between the Forecasts of the EFP and the IV Model

Forecasting Horizons in Quarters	GR Statistic	Critical Values
1	5.68	2.89
2	5.93	2.89
3	11.29	2.89
4	7.39	2.89

Source: Author's compilation, Central Bank forecasts.

Table A3.20. Summary Results of the GR Test between the Forecasts of the EFP and the IV Model

Forecasting Horizons in Quarters	Fluctuation test summary
1	Over 2015, the forecasts of the IV model are more accuracy than those of the AMM model. After that, the GR statistic is not statistically significant.
2	The forecast of the IV model are more accuracy in almost all the evaluation sample.
3	The forecast of the IV model are more accuracy in almost all the evaluation sample.
4	The forecast of the IV model are more accuracy in two different sub samples.

Source: Author's compilation, Central Bank forecasts.

Table A3.21. GR Test between the Forecasts of the EEP and the DIE

Forecasting Horizons in Years	GR Statistic	Critical Values
1	21.58	3.18
2	11.86	2.89

Source: Author's compilation, Central Bank forecasts.

Table A3.22. Summary Results of the GR test between the Inflation Expectations of the EEP and the DIE

Forecasting Horizons in Quarters	Fluctuation test summary
1	The inflation expectations of the DIE are more accuracy in the beginning of the sample evaluation. However, from January 2016, the inflation expectations of the EEE model predictive better the inflation patterns.
2	The inflation expectations of the EEE are more accuracy in the entire evaluation sample than those of the DIE.

Source: Author's compilation, Central Bank forecasts.

6. *Efficiency Tests*

6.1 Unbiasedness Tests

Table A3.23. Unbiasedness Test, Quantitative Models

Forecasting Horizons in Quarters	Unbiasedness test (AMM)	Unbiasedness test (IV)	Unbiasedness test (EFP)
1	-0.03 (0.63)	-0.04 (0.47)	-0.45 (0.02)
2	-0.24 (0.12)	-0.22 (0.12)	-0.04 (0.69)
3	-0.42 (0.06)	- 0.50 (0.02)	-0.01 (0.92)
4	-0.58 (0.03)	-0.87 (0.00)	0.02 (0.80)
8	-0.35 (0.24)	-1.40 (0.01)	0.04 (0.79)

Source: Author's compilation, Central Bank forecasts.

Table A3.24. Unbiasedness Test, Inflation Expectations

Forecasting Horizons in Years	Unbiasedness test (EEP)	Unbiasedness test (DIE)
1	0.10 (0.71)	0.45 (0.10)
2	0.31 (0.00)	0.53 (0.00)

Source: Author's compilation, Central Bank forecasts.

6.1 Autocorrelation Tests

Table A3.25. Autocorrelation Test, Quantitative Models

Forecasting Horizons in Quarters	Autocorrelation Test (AMM)	Autocorrelation Test (IV)	Autocorrelation Test (EFP)
1	6.67 (0.01)	0.56 (0.58)	6.68 (0.01)
2	3.81 (0.04)	4.89 (0.02)	0.07 (0.93)
3	5.87 (0.01)	13.04 (0.00)	0.03 (0.97)
4	9.34 (0.00)	19.87 (0.00)	0.05(0.96)
8	13.04 (0.00)	70.13 (0.00)	0.02 (0.98)

Source: Author's compilation, Central Bank forecasts.

Table A3.26. Autocorrelation Test, Inflation Expectations

Forecasting Horizons in Years	Autocorrelation Test (DIE)	Autocorrelation Test (EEE)
1	77.28 (0.00)	66.81 (0.00)
2	42.54 (0.00)	9.79 (0.01)

Source: Author's compilation, Central Bank forecasts.

C. Weak Efficiency Tests.

Table A3.27. 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)

Source: Author's compilation, Central Bank forecasts.

Table A3.28. Weak Efficiency Test, Inflations Expectations

Forecasting Horizons in Years	Weak Efficiency test (EEP)	Weak Efficiency test (DIE)
1	1.67E+11 (0.00)	184.40 (0.00)
2	6.51 (0.02)	62.77 (0.00)

Source: Author's compilation, Central Bank forecasts.

Table A3.29. 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)

Source: Author's compilation, Central Bank forecasts.

6.4 Strong Efficiency Tests.

Table A3.30. Strong Efficiency Test, IV Model

Forecasting Horizons in Quarters	Strong Efficiency (consumption)	Strong Efficiency (raw material index)	Strong Efficiency (Investment)	Strong Efficiency (Government Spending)	Strong Efficiency (Credit)
1	0.47 (0.49)	-0.32 (0.43)	0.20 (0.79)	0.33 (0.69)	0.26 (0.51)
2	2.83 (0.09)	-1.60 (0.10)	2.51 (0.19)	0.25 (0.91)	0.71 (0.49)
3	6.17 (0.01)	-2.89 (0.04)	4.86 (0.10)	-0.01 (0.99)	1.64 (0.31)
4	9.36 (0.00)	-4.90 (0.01)	7.47 (0.04)	1.35 (0.78)	4.42 (0.04)
8	11.91 (0.00)	-5.02 (0.01)	10.79 (0.01)	-9.11 (0.29)	5.40 (0.10)

Source: Author's compilation, Central Bank forecasts.

Table A3.31. Strong Efficiency Test, EFP Model

Forecasting Horizons in Quarters	Strong Efficiency (consumption)	Strong Efficiency (raw material index)	Strong Efficiency (Investment)	Strong Efficiency (Government Spending)	Strong Efficiency (Credit)
1	8.49 (0.01)	-3.3091 (0.05)	9.10 (0.02)	-11.93 (0.01)	6.76 (0.09)
2	1.73 (0.37)	-1.1865 (0.24)	2.89 (0.29)	-3.76 (0.17)	3.54 (0.17)
3	1.16 (0.59)	------------------	------------------	-3.27 (0.27)	3.64 (0.27)
4	-0.08 (0.98)	0.0919 (0.10)	0.97 (0.79)	-2.70 (0.37)	------------------
8	------------------	0.7945 (0.82)	------------------	-5.80 (0.33)	------------------

Source: Author's compilation, Central Bank forecasts.

Table A3.32: Strong Efficiency Test, IV Model, December Evaluation

Forecasting Horizons in Quarters	Strong Efficiency (consumption)	Strong Efficiency (raw material index)	Strong Efficiency (investment)	Strong Efficiency (Government Spending)	Strong Efficiency (Credit)
1	1.13 (0.31)	-0.83 (0.23)	0.94 (0.49)	0.94 (0.95)	0.7864 (0.34)
2	7.88 (0.06)	-4.46 (0.11)	6.32 (0.27)	8.55 (0.26)	4.9864 (0.13)
3	5.83 (0.04)	-3.60 (0.05)	4.88 (0.23)	2.66 (0.65)	3.5045 (0.14)
4	5.90 (0.34)	-3.25 (0.42)	2.13 (0.79)	-6.29 (0.54)	1.7364 (0.72)
5	9.54 (0.21)	-4.28 (0.41)	5.47 (0.60)	2.96 (0.92)	4.4500 (0.53)
6	11.90 (0.17)	-6.16 (0.29)	8.99 (0.45)	7.23 (0.83)	7.1250 (0.37)
7	13.00 (0.13)	-6.81 (0.24)	10.49 (0.37)	11.69 (0.73)	8.0250 (0.32)
8	11.59 (0.16)	-6.40 (0.24)	10.78 (0.33)	15.01 (0.63)	7.9000 (0.28)

Source: Author's compilation, Central Bank forecasts.

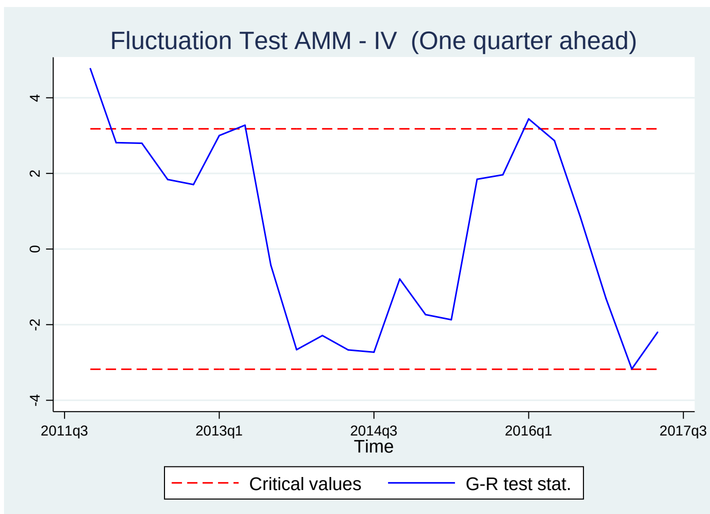
Appendix 4. Graphic Analysis of Giacomini-Rossi Fluctuation Tests

In this appendix, we graphically illustrate the Giacomini-Rossi fluctuation tests between two competing models. The IV model is the benchmark model in this study. If the GR statistics line is outside the critical bands, it means that the null hypothesis of equal accuracy is rejected.

1. GR - Test between the AMM and the IV Model

In Figure A4.1, we observe that the GR statistic line is outside of the critical band, which means that we reject the null hypothesis of equal accuracy. Also, over 2011 the forecasts of the IV model are more accurate than those of the AMM model. Similarly, the forecasts of the former are better than the latter in June 2013 and March 2016.

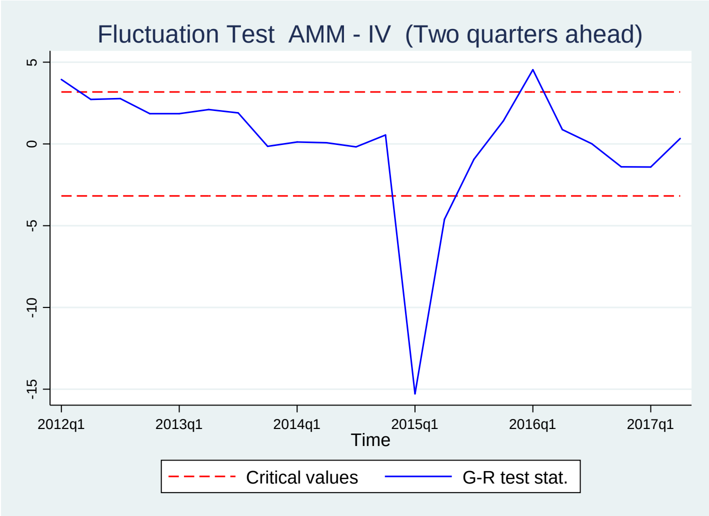
Figure A4.1



Source: Authors' compilation, Central Bank Dataset.

From Figure A4.2, we reject the null hypothesis of equal predictability too. In this case, from the second to the four quarter of 2011, the forecasts of the IV model are more accuracy than the AMM model. However, from the first to the second quarter of 2015, the forecasting accuracy is reverse.

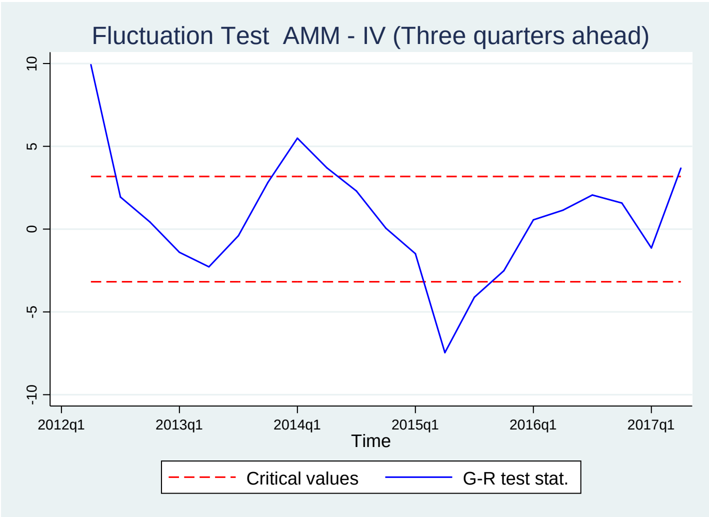
Figure A4.2



Source: Authors' compilation, Central Bank Dataset.

From Figure A4.3, the null hypothesis of equal accuracy is rejected since the GR statistics is outside the critical bands. As in the previous graph, the forecasts of the IV model are more accurate in the beginning, from the first to the second quarter of 2012. After that, the forecasts of the AMM are better from the first to the third quarter of 2015.

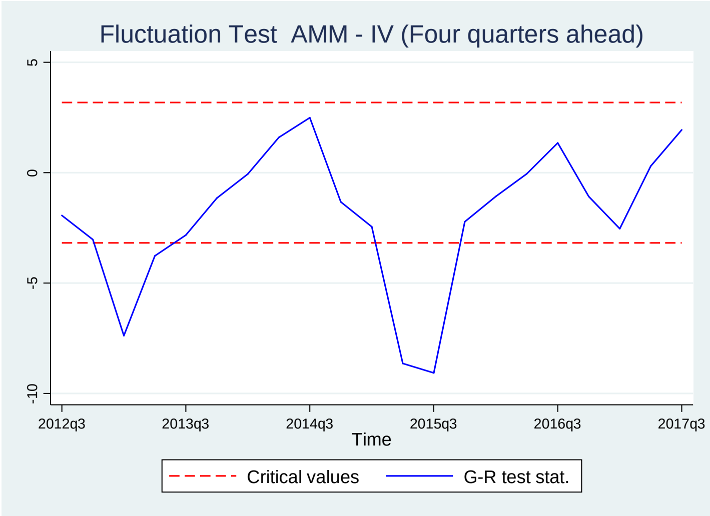
Figure A4.3



Source: Authors' compilation, Central Bank Dataset.

From Table A4.4, the null hypothesis of equal accuracy is rejected since the GR statistic is outside the lower critical bands in two different sub periods. Therefore, the forecasts of the AMM model are more accurate in two different sub-periods of the total sample.

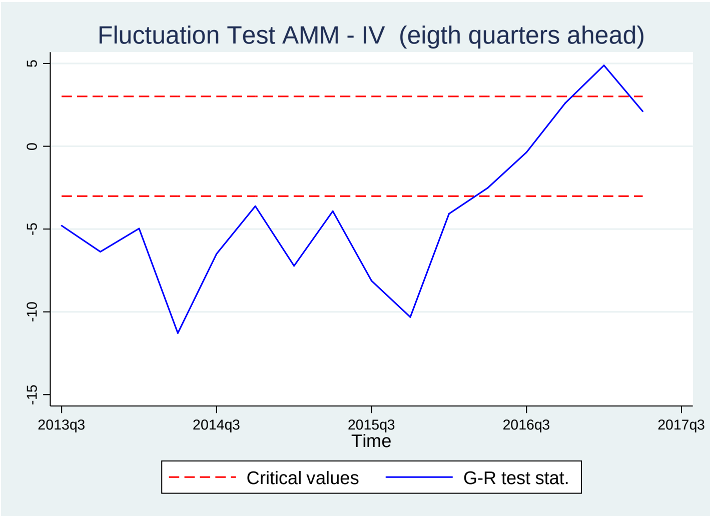
FigureA4.4



Source: Authors' compilation, Central Bank Dataset.

From Figure A4.5, we observe that the GR statistics is below the lower critical band. In this forecasting horizon, the forecasts of the AMM model are more accurate in a significant part of the total sample.

Figure A4.5



Source: Authors' compilation, Central Bank Dataset.

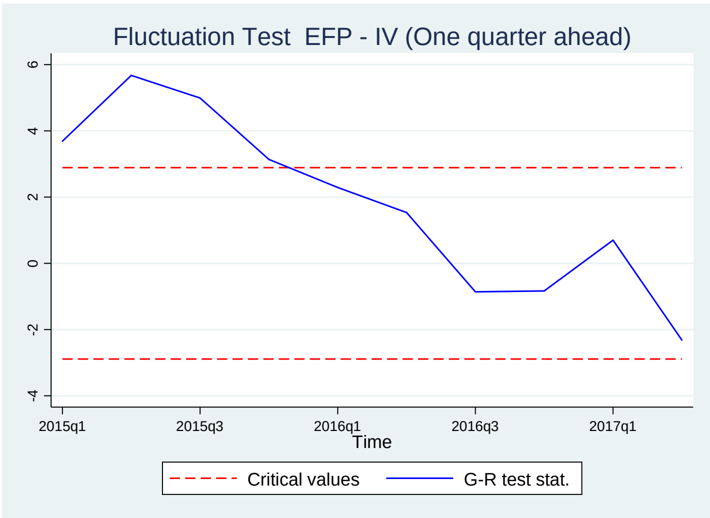
Therefore, there is some evidence that, in the medium and long term, the forecasts of the AMM model are more accurate than those of the IV model.

2. GR Test between the EFP and the IV Model

We use the graphical analysis to examine the behavior of the forecasts in every rolling window as we did previously.

From Figure A4.6, we can observe that in the beginning of the sample, specifically through 2015, the forecasts of the IV model are more accurate than those of the EFP model. After that, there is no statistical difference between them. Therefore, it seems that the IV model generates better forecasts than the EFP model in the case of the one-quarter horizon.

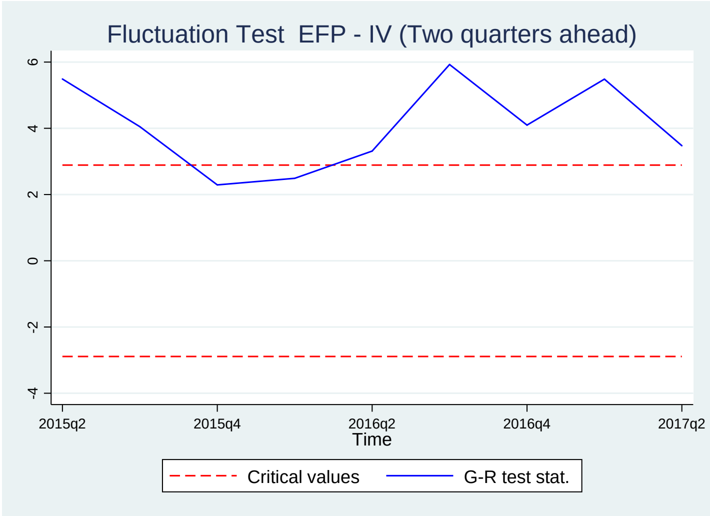
Figure A4.6



Source: Authors' compilation, Central Bank dataset.

From Figure A4.7, we observe that the null hypothesis of equal predictability is rejected because the GR statistic line is outside the upper critical band in most of the sample. The forecasts of the IV model are more accuracy than those of the EFP model. Therefore, the IV model generates better forecast than the EFP model in the case of the two quarters horizon.

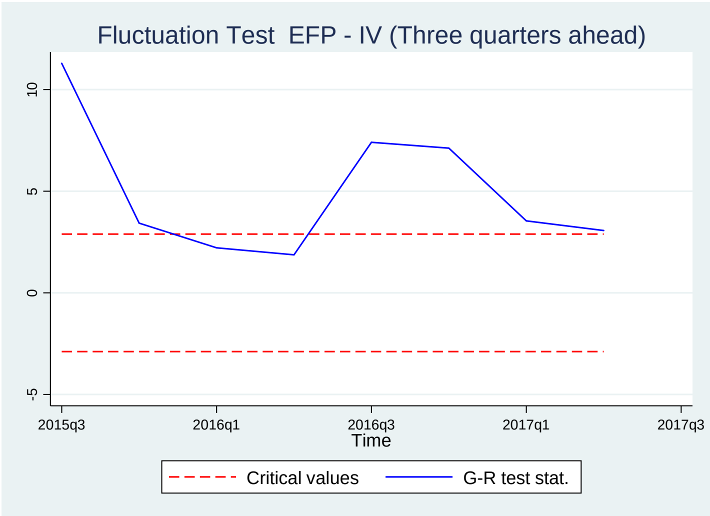
Figure A4.7



Source: Authors' compilation, Central Bank dataset.

From Figure A4.8, we observe that in most of the sample, the GR statistic line is outside the upper critical band, which means that in these subperiods, the forecasts of the IV model are more precise than those of the EFP model. Therefore, the IV model can better predict the patterns of inflation.

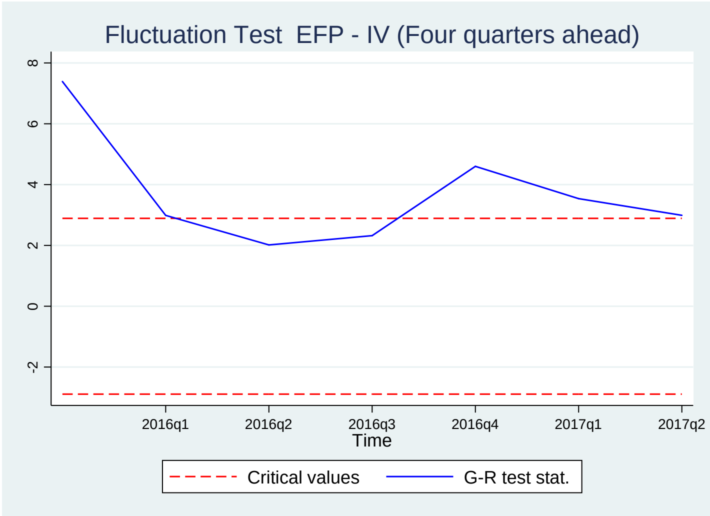
Figure A4.8



Source: Authors' compilation, Central Bank dataset.

From Figure A4.9, we observe the same pattern as before. The GR statistics line is outside the upper critical band, which means that the null hypothesis is rejected. Again, the forecasts of the IV model are more accurate than those of the EFP model in most of the sub-periods of time.

Figure A4.9



Source: Authors' compilation, Central Bank Dataset.

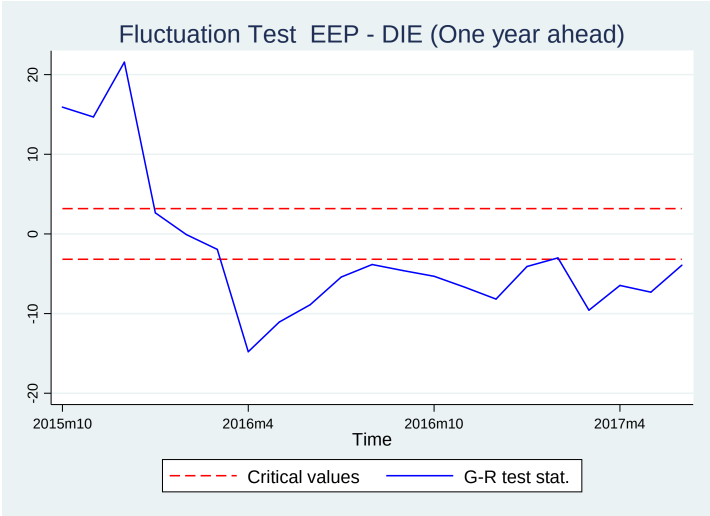
All in all, the IV model better predicts better the inflation pattern than the EFP model in different sub-periods of time.

3. GR Test between the EEP and DIE

We use the graphical analysis to examine the performance of the forecasts in different sub-periods of time. In contrast to the quantitative analysis, we have monthly inflation expectations.

From Figure A4.10, we observe that the GR statistic line is outside the upper and lower critical bands in some sub-periods of time, which means that the null hypothesis of equal accuracy is rejected. From October 2015 to December 2015, the inflation expectations of the DIE are more accurate than those of the EEP. However, the trend is reversed later. From April 2016 to June 2017, the inflation expectations of the EEP are more accurate.

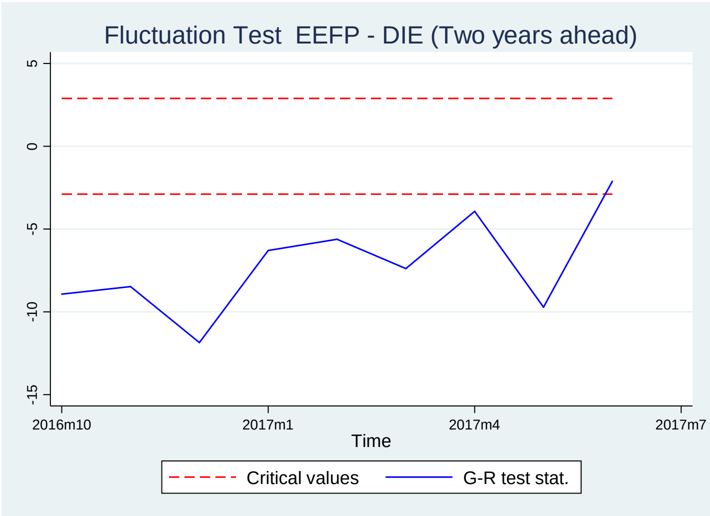
Figure A4.10



Source: Authors' compilation, Central Bank Dataset.

From Figure A4.11, we see that in almost the entire sample the GR statistic line is outside the lower critical band. Therefore, we reject the null hypothesis of equal accuracy between the inflation expectations of the DIE and the EEP. In almost the complete sample, the inflation expectations of the Economic Experts Panel are more accurate at predicting the inflation patterns than those of the Economic Research Department in the case of the two-year horizon.

Figure A4.11



Source: Authors' compilation, Central Bank Dataset.

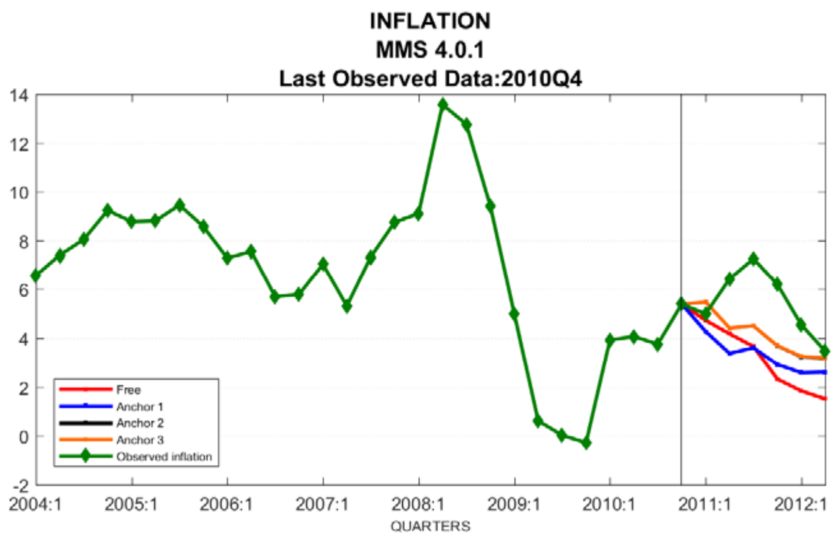
In sum, the inflation expectations of the economic experts panel are more accurate in a two-years horizon in almost the complete sample, while in the case of the one-year horizon, there is some evidence that the inflation expectations of the economic experts panel have been more accurate since 2016m02.

Appendix 5. Graphs and Tables of the Conditional Forecast Evaluation

In this appendix, we include the graphs and tables of the conditional forecast evaluation.

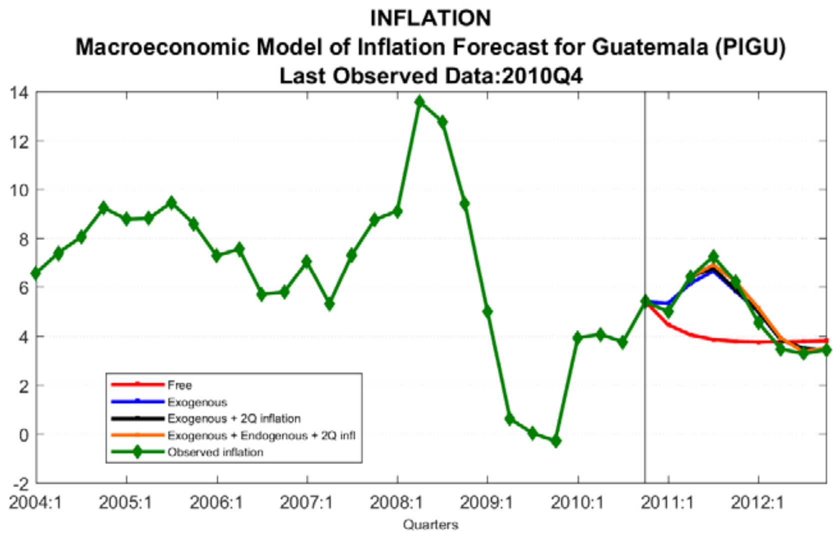
1. Graphs of the Conditional Forecast Evaluation

Figure A5.1



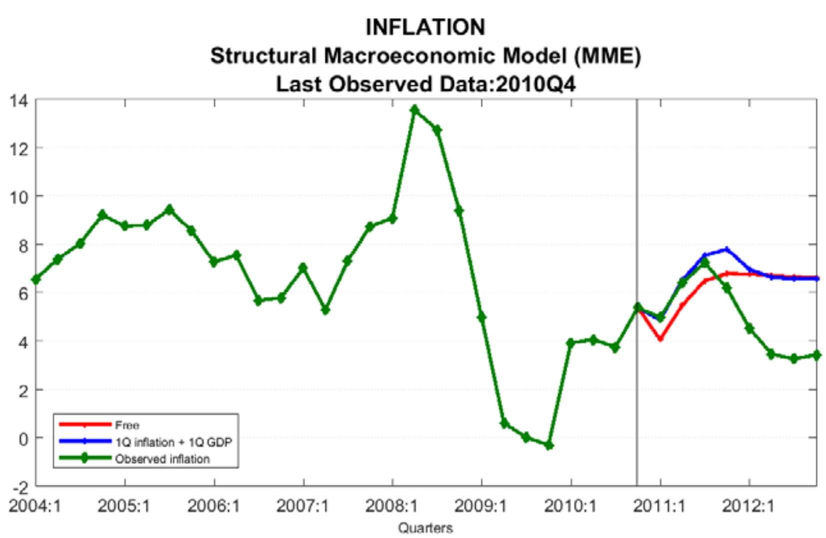
	2011Q1	2011Q2	2011Q3	2011Q4	2012Q1	2012Q2
Free	4.73	4.18	3.66	2.32	1.84	1.51
Anchor 1	4.27	3.38	3.61	2.92	2.58	2.61
Anchor 2	5.48	4.42	4.50	3.69	3.24	3.17
Anchor 3	5.48	4.42	4.51	3.69	3.25	3.19
Observed inflation	5.00	6.42	7.25	6.20	4.55	3.47

Figure A5.2



	2011Q1	2011Q2	2011Q3	2011Q4	2012Q1	2012Q2
Free	4.46	4.05	3.86	3.78	3.76	3.77
Exo	5.33	6.15	6.64	5.82	4.98	3.83
Exo-infla	5.00	6.42	6.76	5.87	5.01	3.84
Exo-infla-endo	5.00	6.42	6.91	6.20	5.14	3.91
Observed inflation	5.00	6.42	7.25	6.20	4.55	3.47

Figure A5.3



	2011Q1	2011Q2	2011Q3	2011Q4	2012Q1	2012Q2	2012Q3	2012Q4
Free	4.08	5.50	6.50	6.81	6.78	6.72	6.67	6.63
1Q inflation + 1Q GDP	4.87	6.54	7.55	7.82	6.97	6.65	6.60	6.60
Observed inflation	5.00	6.42	7.25	6.20	4.55	3.47	3.28	3.45

2. Mean Error

Table A5.1. Mean Error, Semi-Structural Macroeconomic Model 4.0.1, (2011Q1 – 2017Q2)

Forecasting Horizons in Quarters	Free Model	Anchor 1	Anchor 2	Anchor 3
1	-0.11	-0.03	0.01	0.01
2	-0.13	-0.02	0.01	0.01
4	0.22	0.29	0.29	0.29
6	0.55	0.54	0.52	0.52
8	0.27	0.55	0.54	0.54
Mean	0.16	0.26	0.27	0.27

Source: Authors' compilation, Central Bank forecasts.

Table A5.2. Mean Error, Macroeconomic Model of Inflation Forecast for Guatemala (PIGU), 2011Q1 – 2017Q2

Forecasting Horizons in Quarters	Free Model	Anchoring Exogenous variables	Anchoring exogenous variables and 2 periods of inflation	Anchoring exogenous and endogenous variables, plus 2 periods of inflation
1	-0.22	-0.25	0.00	0.00
2	-0.30	-0.38	0.00	0.00
4	0.00	-0.47	-0.39	-0.27
6	0.34	-0.58	-0.56	-0.32
8	0.41	-0.79	-0.79	-0.38
Mean	0.05	-0.49	-0.35	-0.19

Source: Author's compilation, Central Bank forecasts.

Table A5.3. Mean Error, Structural Macroeconomic Model (MME), 2011Q1 – 2017Q2

Forecasting Horizons in Quarters	Free Model	Anchor 1
1	0.30	-0.09
2	0.89	0.36
4	2.37	1.81
6	2.82	2.87
8	2.82	2.86
Mean	1.84	1.56

Source: Author's compilation, Central Bank forecasts.

3. Root Mean Square Error

Table A5.4. Root Mean Square Error, Semi-Structural Macroeconomic Model 4.0.1, (2011Q1 – 2017Q2)

Forecasting Horizons in Quarters	Free Model	Anchor 1	Anchor 2	Anchor 3
1	0.73	0.71	0.33	0.33
2	1.21	1.27	0.87	0.87
4	1.43	1.58	1.37	1.37
6	1.47	1.40	1.36	1.36
8	1.72	1.63	1.57	1.57
Mean	1.31	1.32	1.10	1.10

Source: Author's compilation, Central Bank forecasts.

Table A5.5. Root Mean Square Error, Macroeconomic Model of Inflation Forecast for Guatemala (PIGU), 2011Q1 – 2017Q2).

Forecasting Horizons in Quarters	Free Model	Anchoring Exogenous variables	Anchoring exogenous variables and 2 periods of inflation	Anchoring exogenous and endogenous variables, plus 2 periods of inflation
1	0.83	0.72	0.00	0.00
2	1.26	0.90	0.00	0.00
4	1.44	0.88	0.82	0.61
6	1.11	1.12	1.13	0.62
8	0.89	1.29	1.29	0.65
Mean	1.11	0.98	0.65	0.38

Source: Author's compilation, Central Bank forecasts.

Table A5.6. Root Mean Square Error, Structural Macroeconomic Model (MME), (2011Q1 – 2017Q2)

Forecasting Horizons in Quarters	Free Model	Anchor 1
1	0.62	0.10
2	1.28	0.61
4	2.72	2.09
6	2.98	3.04
8	2.93	2.96
Mean	2.11	1.76

Source: Author's compilation, Central Bank forecasts.

References

Berg, A., P.D. Karam and D. Laxton. 2006. "A Practical Model-Based Approach to Monetary Policy Analysis: Overview." IMF WP/06/80. Washington, DC, United States: International Monetary Fund.

Berg, A., P.D. Karam and D. Laxton. 2006. "Practical Model-Based Monetary Policy Analysis: A How-To Guide." IMF WP/06/81. Washington, DC, United States: International Monetary Fund.

Bernanke, B.S., M. Gertler and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In: J.B. Taylor and M. Woodford, editors. *Handbook of Macroeconomics*. Volume 1. Amsterdam, The Netherlands: Elsevier.

Castillo, C., and E. Ortiz. 2017. "The Nicaraguan Inflation Combination Assessment (NICA): A Forecast Combination System through an Efficient Forecast Path." Forthcoming 2018 in *Revista de Economía y Finanzas*.

Diebold, F.X., and R. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13: 253-263.

Giacomini, R., and B. Rossi. 2010. "Forecast Comparisons in Unstable Environments." *Journal of Applied Econometrics* 25(4): 595-620.

Pesaran, H., and A. Timmerman. 1993. "A Simple Nonparametric Test of Predictive Performance." *Journal of Business and Economic Statistics* 10: 461-465.

Rossi, B., and T. Sekhposyan. 2010. "Have Economic Models' Forecasting Performance for US Output and Inflation Changed over Time, and When?" *International Journal of Forecasting* 26(4): 808-835.

Rotemberg, J.J. 1982. "Sticky Prices in the United States." *Journal of Political Economy* 90(6): 1187-1121.