



Munich Personal RePEc Archive

**A statistical evaluation of GAP's
forecasting performance for the Albanian
economy**

Papavangjeli, Meri and Rama, Arlind

Bank of Albania

2018

Online at <https://mpra.ub.uni-muenchen.de/116104/>
MPRA Paper No. 116104, posted 25 Jan 2023 14:22 UTC

A STATISTICAL EVALUATION OF GAP'S FORECASTING PERFORMANCE FOR THE ALBANIAN ECONOMY

Meri Papavangjeli¹, Arlind Rama

ABSTRACT

This paper aims at offering a statistical evaluation methodology on the forecasting performance of the GAP model, a semi-structural economic model used to support monetary policy decisions at the Bank of Albania since 2011. In this paper we evaluate the forecasts produced purely by the model, and not those used by the Monetary Policy Department, which also include the expert judgment and are not made public. The analytical approach used in the discussion material combines a statistical diagnostic look-up consisting in statistical measurements as RMSE and BIAS important to understand the forecasting performance of the model as an instrument in one, two and three years ahead time horizons. A VAR model is constructed resembling the economic relations represented in the GAP model as a most commonly used tool to obtain economic projections based solely on the information that the data series provide. Comparing the forecasting performance of the two models on a common statistical diagnostic metrics helps us to create a broader understanding of the forecasting abilities of the GAP model and draw discussion issues for potential improvements of the model that would potentially lead to an improved representation of the Albanian economy and increased accuracy in its forecasting performance.

Key words: forecasting performance, GAP model, statistical metrics

¹ The views expressed herein are of the authors and do not necessarily represent the views of the Bank of Albania. This paper has benefited from the invaluable comments of Dr Altin Tanku, Head of Research Department, Mr Erald Themeli, Head of Monetary Policy Department and Mr Ilir Vika, Chief of Economic Modelling Sector (at the Bank of Albania), as well as from the comments received by the participants in the 10th Research Workshop organized by Bank of Albania in Tirana, on 1-2 December 2016. email: mpapavangjeli@bankofalbania.org

1. INTRODUCTION

The forecasting accuracy of economic projections is always a matter of primary importance in central banks when it comes to models used for policy-making purposes. Ability of models to read and predict eventual economic conditions deterioration becomes even more important in a time of unprecedented complex nature of crises and when unconventional monetary policy has become a viable option. Hence, having in place reliable economic forecasting models is of a paramount importance to properly orient the policy-making debate around crucial issues needed to be tackled appropriately by the decision-makers. In this light, the prepared discussion material tends to actively contribute in understanding the accuracy of the GAP model forecasting performance aiming critically consider the present stance of the model and think potential perspectives to consider potential future improvements and model expansion in order to properly resemble the Albanian economy.

In this working paper we aim at introducing a statistical methodology to evaluate the forecasting performance of the GAP model, an instrument utilized by the Bank of Albania to draw mid-term projections on potential macroeconomic developments in the Albanian economy. Although we focus only on the pure model forecasts and not on the judgemental ones used by Monetary Policy Department, it is worth to mention that through this contribution, the paper intends to add value to the portfolio of tools in disposal of forecasting at the Bank.

GAP is one of the two core economic models used in the Bank of Albania to make projections on future developments of the Albanian economy, the later being taken into account during policy-making debates in the Bank of Albania. A small New Keynesian structural model, introduced first by Dushku and Kota (2011), the GAP is constructed in four building blocks representing aggregate demand, inflation, exchange rate and the interest rate policy rule. The model works with quarterly data flow variables and has exogenously defined steady-states. The aim of our paper is to evaluate whether this model constructed to replicate the developments of the Albanian economy, by also capturing turning points and having reasonable

impulse response functions to given shocks, performs accordingly with economic expectations in the present macroeconomic framework, without adding the expert judgment. We are going to statistically analyze the performance of the pure model forecasts and comparing them to the forecasts obtained by a VAR, as a commonly used and accepted model to understand the data causal relationship and made projections, built to match the Albanian economy.

The paper is structured as follows. Section 2 describes the data used for model estimation and variables transformations. Section 3 continues with the methodology utilized to deliver the statistical diagnostic analysis, including the construction of the VAR model. Section 4 discusses the empirical results obtained by the comparative analysis and the potential paths to be explored when thinking in eventual improvements of the GAP model. Final remarks are presented in Section 5.

2. LITERATURE REVIEW

Evaluation of the forecasting performance of structural econometric models used to provide macroeconomic projections in central banks is a continuous process, useful to understand and measure accuracy as well as biasedness of these econometric tools that contribute to the policymaking process. In this perspective there are a number of studies that try to explore more on evaluation techniques of present macroeconomic models and shed light on important issues worth to consider when certain statistical analysis are performed.

Fawcett, Körber, Masolo and Waldron (2015) of the Bank of England investigate the real-time performance of Bank's main DSGE model, COMPASS, before, during and after the financial crisis with reference to statistical and judgmental benchmarks. Specifically, the forecast performance of a relatively 'judgment-free' version of COMPASS is evaluated against the performance of the judgmental forecasts made by the Monetary Policy Committee (MPC) and published in the Inflation Report, as well as against the performance of a statistical benchmark forecast from the Bank's

Suite of Statistical Models. The authors find that, at shorter horizons, the MPC's Inflation Report projections (both point forecasts and complete probability density forecasts) are more accurate for both GDP growth and inflation than either the COMPASS forecasts or the forecasts produced by the Statistical Suite. At longer horizons (horizons of more than one year), COMPASS has the more accurate inflation point forecasts, and the Statistical Suite the more accurate GDP point forecasts. The authors note that not all these differences are statistically significant, and forecast accuracy in itself is not the only metric by which models should be assessed. An in-depth evaluation of Bank forecasts for GDP growth and CPI inflation is provided by Hackworth, Radia and Roberts (2013). The authors provide a detailed comparison of economic outturns over the 2010-'13 period with MPC central expectations made in August 2010. Hackworth et al provide statistical diagnostics similar to those detailed in this paper on the performance of MPC mean projections for GDP growth and inflation over 1997-2013; results are presented for one quarter ahead and one year ahead projections. For GDP growth, mean projections were found to be unbiased at the 5% level at both horizons, although there was some evidence of bias at the 10% significance level for one year ahead GDP growth (MPC mean projections too high), as well as some evidence of strong inefficiency in one quarter ahead projections. For inflation, mean projections were found to be unbiased at the 5% level at both horizons, although there was some evidence of bias at the 10% significance level for one year ahead inflation (MPC mean projections too low), as well as some statistically significant evidence of weak inefficiency at both horizons. Groen, Kapetanios and Price (2009) evaluate the performance of central estimates for inflation and growth contained in the Bank's Inflation Report for the 1997 Q3 to 2006 Q2 period. Specifically, Groen et al compares RMSEs of Bank forecasts with those of pure statistical models, using the Diebold-Mariano-West test of significance. Groen et al conclude that while GDP forecasts produced by statistical models perform as well or better than Bank forecasts at all horizons, the reverse was typically true of Bank inflation forecasts. For inflation, the Inflation Report forecasts are clearly dominant, often significantly so.

Other studies have compared forecasts from DSGE and BVAR models. Iversen et al. (2014) investigate the case of the Sveriges Riksbank and explicitly contrast DSGE and BVAR real-time forecasts since 2007. They find that the BVAR model forecasts for inflation performed well both in absolute terms and relative to the DSGE model forecasts and the Riksbank's published forecasts. Another study, by Christoffel et al. (2011), examines the forecasting performance of NAWM, the ECB's DSGE model, against Bayesian VAR benchmarks. They assess NAWM against four BVARs which vary in size and type of prior and the models are re-estimated annually. They also find that the DSGE model is outperformed by a BVAR benchmark, both in terms of point and density forecasts.

The study of Hong and Tan (2014) evaluates and compares the forecasting performance of GDP growth rate forecasts at global, regional and individual country levels between three international organizations: the United Nations, the International Monetary Fund and the World Bank. The authors use the following statistical indicators for forecast evaluation: RMSE (Root Mean Squared Errors), MAE (Mean Absolute Errors) and MAPE (Mean Absolute Percentage Errors).

3. DATA AND METHOD OF ANALYSIS

3.1 DATA DESCRIPTION

In estimating the GAP model, we use quarterly data on real Gross Domestic Product (GDP), Consumer Price Index (CPI), nominal exchange rate ALL/EUR and monetary policy rate for the period 2002Q1 to 2015Q4. For updating the database with the latest available data, Institute of Statistics and Bank of Albania sources are utilized and the data are transformed accordingly, respectively data on GDP and CPI are taken from INSTAT, and those on interest rate and exchange rate from Bank of Albania. We have chosen 2002 as starting point of the sample period, because there are some structural breaks of the series in the previous years. Data on foreign variables are taken from the Eurostat database and the European

Central Bank. Prior to estimation, real GDP, CPI and nominal exchange rate are transformed into quarter-on-quarter and year-on-year growth rates, approximated by the first difference of their logarithm. An extensive discussion of the empirical implementation of the GAP model is beyond the scope of this article, and the reader is thus referred to Dushku and Kota (2011) for details on the calibration of the model's steady state and the distribution of model parameters.

This section describes the methodology used in this article to evaluate the forecasting performance of GAP model. First, the accuracy and the biasedness of model forecasts were assessed using some statistical indicators and then the GAP model's forecasting properties were evaluated against a less theoretical oriented forecasting tool such as Vector Autoregressive (VAR) model.

3.2 ACCURACY

The accuracy of forecasts was measured using root mean squared errors (RMSE), defined as: s

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (1)$$

where e_i is a forecast error, defined as outturn less forecast.

The larger forecast errors, the larger are the RMSEs, but the relationship is not linear. RMSEs will be disproportionately large (small) if errors are very large (small). RMSEs are a standard loss function used in the forecast evaluation literature. More importantly, a quadratic loss function is appropriate for our purposes because policymakers will care more about big forecast errors that could lead to big policy mistakes and have a damaging impact on the economy, than small errors, which may not impact much on policy. However, RMSEs represent one particular form of loss function (a quadratic loss function) and alternative loss functions could lead to different results (Timmermann, 2006).

To allow comparability of RMSEs across variables, RMSEs are scaled by the standard deviation of the data outturns (over the same period as the RMSE was calculated). This is a simple way to account for data volatility, since greater volatility makes a given variable inherently more difficult to forecast. It is important to note that the value of the scaled RMSE is not itself informative (i.e. whether it should be above or below a certain value), and is only useful to compare with scaled RMSEs of other variables or of the same variable in other periods of time.

3.3 BIASEDNESS

Biasedness of forecasts was assessed using ordinary least squares (OLS) regression. Forecast errors were regressed on a constant with a null hypothesis that the constant was zero, which would be the case if the forecasts were unbiased. Otherwise, the forecasts could have been made more accurate by adding a constant amount to them. As in equation (1), forecast errors are defined as:

$$e_t^{t-h} = y_t - y_t^{t-h} \quad (2)$$

where y_t is the outturn of variable y in period t and y_t^{t-h} is the forecast for variable y in period t made in period $t-h$.

To test for biasedness, the following regression is estimated:

$$e_t^{t-h} = \beta_0 - u_t \quad (3)$$

where u_t is a zero-mean error term. Under the null hypothesis of unbiasedness $\beta_0 = 0$. If $\beta_0 > 0$, forecasts have been systematically too low. If $\beta_0 < 0$, forecasts have been too high. We estimated the regression using OLS with Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors. Forecasts made in consecutive quarters will cover mostly the same forecast period and the associated forecast errors are therefore likely to be auto-correlated.

Using HAC standard errors should account for this and other potential autocorrelation and heteroscedasticity issues (Andrews, 1991).

3.4 BENCHMARKING OF GAP FORECASTS

Comparing GAP' dynamic properties and forecasting accuracy with those of more data-driven benchmarks such as VARs can be helpful because DSGE models place a great number of restrictions on the time-series behaviour of the variables they seek to explain and forecast. Their size poses challenges to both estimation and specification analysis, which entails risks for the reliability of their forecasts. VARs often provide a reasonably good fit to macroeconomic time series data (Domit et. al., 2016).

In this article, we introduce an unrestricted VAR with 4 lags with the same macroeconomic variables as GAP. As in GAP model, we treat Albania as a small open economy and model the rest of the world as exogenous. We then assess the relative performance of both models in forecasting inflation, GDP growth and exchange rate growth. All the variables in the VAR are expressed as annual growth rates, which in addition avoids the need for seasonal adjustment for those who have a seasonal behavior during the year. Several diagnostic tests are performed to check for variables stationarity, model stability, autocorrelation, heteroscedasticity and normality of residual distribution. Test results suggest that the VAR model satisfies all the necessary assumptions of an OLS estimation procedure.

To reflect the information available at the time the forecasts would have been produced, both models are re-estimated between 2002 and 2015 using real-time data. The real-time estimation approach means that each forecast is produced only with information that would have been available at each forecast round (Iversen et al., 2014). The forecast performance of GAP is assessed using an extending procedure. The initial estimation period is 2002 Q1-2009 Q4 and it is gradually extended by four quarters. First, we estimate the parameters up to the fourth quarter of 2009 and then we compute out-of-sample forecasts for one year, two years

and three years ahead. We extend the estimation sample by four quarters and then compute again the forecasts for one year, two years and three years ahead. We repeat this process several times until the end of the sample.

We then used a Diebold-Mariano test (Diebold, Mariano, 1995) to assess whether differences in accuracy (measured by squared forecast errors) between the GAP and VAR forecasts were statistically significant. To conduct this test, a difference in squared forecast errors for the forecasts of the two models is defined as:

$$d_t = e_{t,1}^2 - e_{t,2}^2 \quad (4)$$

where $e_{t,1}^2$ and $e_{t,2}^2$ are the squared forecast errors at time t respectively for the first and the second forecasting model.

The following equation is then estimated using Ordinary Least Squares (OLS) with HAC standard errors:

$$d_t = \beta_0 + u_t \quad (5)$$

where u_t is a zero-mean error term. The null hypothesis is that there is no difference in accuracy between the two forecasts, i.e. $\beta_0 = 0$. If $\beta_0 > 0$, the second model forecast has tended to be more accurate than the first one, and vice versa, if $\beta_0 < 0$.

For forecast accuracy comparison between GAP and VAR model, we use RMSEs. Firstly, we compared scaled RMSEs over different horizons and used Diebold-Mariano tests to assess whether differences in accuracy between GAP and VAR models were statistically significant. RMSEs are compared for each variable, scaled by the standard deviation of data outturns over the same period we calculated the RMSEs. And again, Diebold-Mariano tests are used to assess whether differences in accuracy between GAP and VAR models were statistically significant.

4. ESTIMATED RESULTS

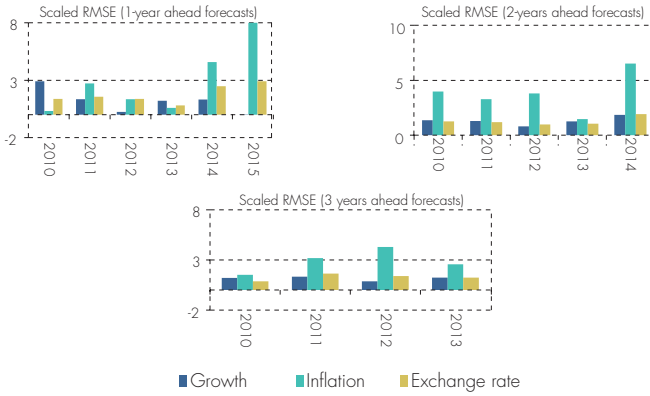
In this section we set out our main empirical findings. This section starts with the accuracy and biasedness of pure model forecasts of GAP model across the three macroeconomic variables: inflation, real growth and exchange rate. We then compare the accuracy of GAP forecasts with forecasts from a VAR model. It is important to note that the relatively small sample size means that we need to interpret the results below with caution.

4.1 ACCURACY OF GAP FORECASTS

As described in the previous section, we measure forecast accuracy over a given period using RMSEs scaled by the standard deviation of data outturns over that same period. Scaled RMSEs for (quarterly) GAP forecasts at all three horizons are shown in Figure 1. As we noted before, the value of a scaled RMSE is not in itself informative, but we can compare the degree of accuracy across variables and across time periods. A higher scaled RMSE indicates that forecasts of a particular variable have tended to be less accurate, relative to the volatility of the data outturns.

For inflation, forecast accuracy tends to decrease as the forecast horizon expands, except for 2010. This result is reasonable as with the increase of forecast horizon, it is more difficult to predict the likely path of a given variable. For growth and exchange rate, the behavior of forecasts accuracy is more irregular: sometimes it increases, sometimes it decreases with the increase of forecast horizon, but scaled RMSE values remain low. Growth and exchange rate forecasts have tended to be the most accurate at all the forecast horizons, while inflation forecasts have tended to be the least accurate.

Figure 1 Forecast accuracy for different time horizons.



Note: The year in the horizontal axis means that the forecast sample starts from that year and the estimated sample is until the last quarter of the previous year.

4.2 BIASEDNESS

Consistent with our measure of forecast accuracy, and in order to allow comparability across variables and time periods, the estimated degree of bias in forecast errors (given by β_0 in equation (3)) are scaled by the standard deviation of the data outturns (over the same period as the bias was calculated). Inflation forecasts tend to be substantially biased for the years 2014 and 2015, which explains also the low scaled RMSE values for the same years. As scaled RMSE, forecast biasedness of inflation forecasts for the same estimated sample tends to decrease as the forecast horizon increase, except 2010. For growth and exchange rate the values of scaled bias are quite low.

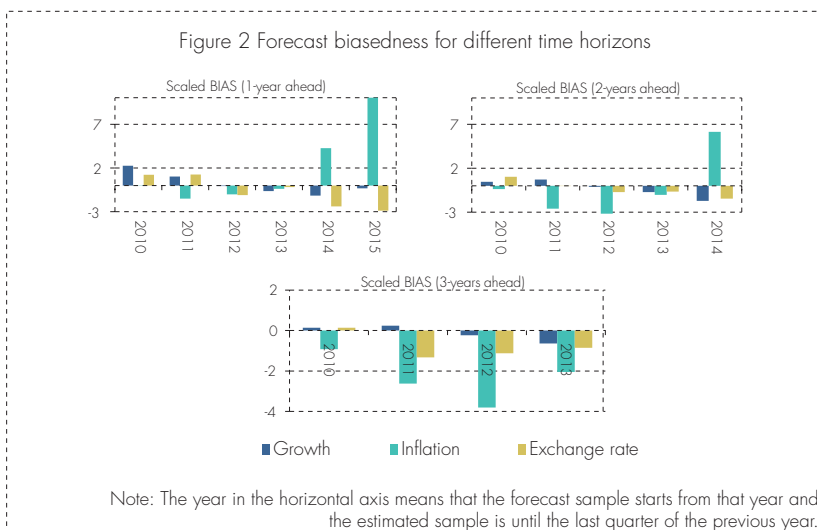


Table 1 presents the estimated bias coefficients for GAP forecasts over different time horizons and their statistical significance. As it can be seen, there is statistically significant evidence of bias only for inflation forecasts at 10% level for the 3-years ahead horizon. Relative to data outturns, growth forecasts have tended to be high during all the forecast horizons. Exchange rate forecasts have been lower for 1-year and 2-years ahead horizons and higher during 3-years ahead horizon. Inflation forecasts have been lower for the 1-year ahead horizon and higher for 2-years and 3-years horizons, but only the coefficient of the latest is statistically significant at 10%. The estimated (scaled) bias coefficient for inflation forecasts was the largest at the three years period, followed by growth bias in the 3-years ahead horizon and exchange rate bias in the 2-years ahead horizon.

Table 1. Statistical significance of bias coefficients.

	Growth	Inflation	Exchange rate
1-year	-0,299	0,199	0,258
2-years	-0,291	-1,057	0,336
3-years	-0,406	-1,347*	-0,228

Note: *significance at 10%.

4.3 COMPARISON BETWEEN GAP AND VAR FORECASTS

4.3.1 ACCURACY OF GAP FORECASTS

As discussed in the previous section, we benchmarked the accuracy and the biasedness of GAP forecasts against a simple VAR model. From Figure 3, it is possible to notice that the GAP model systematically outperforms the VAR model when it comes to forecasting exchange rate trends in one, two and three years ahead forecast horizon. In terms of accuracy of forecast for inflation, the VAR model outperforms the GAP model in the three time horizons of the forecast on which the analyses has been developed. GAP forecasts have lower Scaled RMSEs compared to VAR even for growth.



4.3.2 BIASEDNESS OF GAP FORECASTS

Forecasts obtained for growth and exchange rate with the GAP model are slightly biased, very close to zero for all time horizons, while the same cannot be said for the VAR model. In one year time horizon the GAP results are less biased than those obtained by the VAR.

Persistently, the VAR forecasting results are negatively biased for all the variables. While significantly less biased than the VAR forecasts for Growth and inflation, the GAP forecasts in time horizons four and eight quarters ahead are negatively biased. Results obtained for Inflation by the VAR in eight and twelve quarters ahead are less biased than those obtained by the GAP. In terms of biasness the GAP outperforms the VAR significantly when it comes in predicting Growth in one, two and three years ahead time horizons. GAP forecasting outputs are less biased when run in four quarters ahead horizon, outperforming the VAR in comparative terms.



4.3.3 DIEBOLD-MARIANO TEST

We conducted Diebold-Mariano test according to the steps described earlier, to test whether there are differences in the accuracy of model forecasts and we get the following results:

Table 2. Diebold-Mariano test results.

	GROWTH	INFLATION	EXCHANGE RATE
1-YEAR	-1.442217	1.070123*	-90.2
2-YEARS	-3.139118	4.006239*	-30.78867*
3-YEARS	-3.143100	1.354761	-35.85594*

Note: *significance at 10%.

Among the coefficients above, only those for inflation and for exchange rate in the respective horizons are significant. The results suggest that VAR outperforms GAP model for inflation forecasts for 1-year and 2-years horizons, but it has a worse performance when it comes to the exchange rate for 2 and 3-years horizons. Regarding growth forecasts, there are no statistically significant differences between the two models. Even though, we should be cautious when interpreting the results, because the small sample affects results reliability. An outcome difficult to be predicted may affect substantially the RMSE estimates and forecast accuracy in short horizons, meanwhile in longer horizons its effect is insignificant.

5. CONCLUDING REMARKS

This paper provides a methodology to evaluate the forecasting performance of GAP model in several statistical diagnostic aspects, including benchmarking its performance against a VAR model. The forecasts used in this paper are produced purely by the GAP model, and are not those used by the Monetary Policy Department, which also include the expert judgment. This section tries to draw some concluding remarks based on key results obtained from the forecasting comparative analysis and sets out some notes to take into account particular aspects of forecasting performance.

Different performance measures suggest that the forecast ability of the GAP model, referring to the 2011th version used in our case which was introduced by Dushku and Kota (2011), but excluding the expert judgement used by Monetary Policy Department to generate their forecasts for the policymaking process, is moderate and there is still room for further improvement in its structure. Specifically, GAP model estimates show that the model forecasts for economic growth and the exchange rate are relatively better than medium-term inflation forecasts. Benchmarking with an alternative data-driven model shows that the model can be improved through better identification of the structure and parameters that define inflation in the model. This brings to attention the need for better identification of the inflation determinants in the model, contributing to the accuracy of projections used in the design of monetary policy. Even though, we should be cautious when interpreting the results, because the small sample does not allow us to reach deterministic conclusions on the forecasting performance of the instrument and that is why we see the added value on the evaluation methodology rather than the estimated results.

The need to increase the accuracy of GAP model statistical predictions should serve as an incentive for further and deep research work that focus on improving the projections derived from the model, periodical recalibration of the parameters to reflect changes in the economy, as well as the enrichment with other possible blocks as effective paths toward better inclusion of policy oriented expert judgment in the model.

REFERENCES

Adolfson, M., M. K. Andersson, J. Lindé, M. Villani & A. Vredin, (2007), "Modern forecasting models in action: Improving analyses at central banks", *International Journal of Central Banking*, forthcoming

Andersson, M K, Karlsson, G and Svensson, J (2007), 'The Riksbank's forecasting performance', Working Paper Series 218, Sveriges Riksbank.

Andrews, D W K (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, Vol. 59, No. 3, pages 817–58.

Diebold, F X and Mariano, R S (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics*, Vol. 13, No. 3, pages 253–63.

Domit, S., Francesca Monti and Andrej Sokol (2016), 'A Bayesian VAR benchmark for COMPASS', Staff Working Paper No. 583.

Dushku, E. and Kota, V. (2011), 'A Bayesian Estimation of A Small Structural Model for The Albanian Economy', BoA Working Paper.

Hackworth, C, Radia, A and Roberts, N (2013), 'Understanding the MPC's forecast performance since mid-2010', *Bank of England Quarterly Bulletin*, Vol. 53, No. 4, pages 336–50, available at www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2013/qb130405.pdf

Hong and Tan (2014), "A comparative study of the forecasting performance of three international organizations", DESA Working Paper No. 133 ST/ESA/2014/DWP/133.

Iversen, J., S. Laseen, H. Lundvall and U. Soderstrom (2014), "Real-Time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank," mimeo.

Fawcett, N, Körber, L, Masolo, R M and Waldron, M (2015), 'Evaluating UK point and density forecasts from an estimated DSGE model: the role of off-model information over the financial crisis', Bank

of England Working Paper No. 538, available at www.bankofengland.co.uk/research/Documents/workingpapers/2015/swp538.pdf.

Groen, J J J, Kapetanios, G and Price, S (2009), 'A real time evaluation of Bank of England forecasts of inflation and growth', *International Journal of Forecasting*, Vol. 25, No. 1, pages 74–80.

Sims, C. A. and T. Zha (1998), "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, vol. 39(4), pages 949-68.

Timmermann, A (2006), 'An evaluation of the World Economic Outlook Forecasts', *International Monetary Fund Working Paper WP/06/59*.

CIP Katalogimi në botim BK Tiranë

Papavangjeli, Meri

A statistical evaluation of GAP's forecasting performance for the Albanian economy / Meri

Papavangjeli, Arlind Rama. – Tiranë : Banka e

Shqipërisë, 2018

28f. ; 15x21cm. – (Working paper)

Bibliogr.

ISBN 978-9928-262-02-8

I.Rama, Arlind

1.Ekonomia 2.Ekonometria 3.Shqipëri

330.43(496.5)