

Directional Accuracy of MMS Survey of Inflation-Output Forecasts of G7 Countries: A ROC Analysis

Yasemin Ulu

Department of Economics, Saginaw Valley State University, University Center, Michigan, USA.

How to cite this paper: Yasemin Ulu. (2022) Directional Accuracy of MMS Survey of Inflation-Output Forecasts of G7 Countries: A ROC Analysis. *Journal of Applied Mathematics and Computation*, 6(1), 13-18.
DOI: 10.26855/jamc.2022.03.003

Received: December 5, 2021
Accepted: December 28, 2021
Published: January 10, 2022

***Corresponding author:** Yasemin Ulu, Department of Economics, Saginaw Valley State University, University Center, Michigan, USA.
Email: yasemin.ulu@gmail.com

Abstract

We study the directional forecast accuracy of inflation and output forecasts from Money Market Services Survey (MMS) for G7 countries using a Receiver Operating Characteristic (ROC) curve analysis. Applying ROC curve technique is interesting since ROC analysis can capture the accuracy of directional forecasts for different criterion while the conventional market timing tests can be used only for one. Our results indicate that forecasts of inflation and output from MMS survey contain valuable information for the target variables considered uniformly for all countries. Our findings reinforce the results found in the literature that MMS survey of inflation output forecasts for G7 countries have directional forecast accuracy considered separately, although they seem to fail rationality tests under symmetric loss for some G7 countries. We conclude in favor of directional accuracy of inflation and output forecasts of MMS survey for G7 countries for the period considered.

Keywords

MMS survey, inflation, Output forecast, ROC curves, directional forecast accuracy

1. Introduction

An important feature analyzed in forecasting literature is to test if forecasts are useful by evaluating directional forecast accuracy. There are different methodologies developed to test the directional forecast accuracy dating back to [1]. A methodology to test directional forecast accuracy was first developed by [1] for financial predictions. [1] derived conditions for which a market-timing forecast will have value to an investor based on the change on the prior probability distribution after the forecasts are released. In Merton's framework, based on prior information the investor has a probability density function of the expected return that can be obtained in the market, the forecasts are considered to have a value if the forecast changes that prior distribution¹ (see [1] for further detail).

[2] proposed an alternative test based on the idea that testing whether the forecasts have value is the same as determining whether the forecasts differed significantly from the results obtained from a naïve model in predicting the direction of change. They constructed a 2×2 contingency table and applied X^2 test to test the directional forecast accuracy of real GNP forecasts². To examine if the results of [2] are robust, [3] uses a similar approach and applies X^2 test and Fisher's exact test to a 2×2 contingency table to test the directional forecast accuracy of the real GNP forecasts from three differ-

¹ Using this method under the null hypothesis the number of correct forecasts, given the outcomes, follows the hypergeometric distribution. This distribution is then used to test whether the number of correct forecasts in a particular category exceeded the critical value.

² Zarnowitz (1967) also used a contingency table in evaluating turning point forecasts.

ent institutions. [4] uses the same framework but generalizes it to a 4×4 contingency table to evaluate the joint directional forecast accuracy of inflation—GDP forecasts of Federal Reserve using χ^2 test, Fisher's exact test and [5] test statistic.

A new strand of literature focuses on Receiver Operating Characteristic (ROC hereafter) curves to evaluate the accuracy directional forecasts. The ability of ROC curves to summarize all relevant information in an intuitive manner, has attracted attention from wide range of disciplines including biomedical informatics, computer science, epidemiology, meteorology, and psychology. Despite its popularity among other disciplines ROC curves have only started to be used in finance and economics only very recently.

In the finance and banking literature, ROC curve is also commonly employed as a tool to measure the accuracy of a predictive model [See 6, 7, 8 among others].

In recent years, the ROC analysis has begun to draw increasing attention in the economics as well especially in macroeconomic forecasting. [9] study the directional accuracy of indicators of business-cycle fluctuations utilizing ROC curve analysis. [10] use ROC curves and trading rules to study the directional accuracy of various forecasting models for exchange rates in major industrialized countries. [11] used ROC analysis to evaluate the performance of early warning indicators for emerging financial vulnerabilities in the banking sectors. [12] integrated the ROC analysis into a unified framework of forecast skill evaluation for a binary outcome, and surveyed a wide range of skill scores related to the ROC curve and use ROC curves to study survey data on probability forecasts of GDP declines. Other important applications include the studies of [13, 14, 15].

More recently, [16] apply ROC curve analysis to evaluate the accuracy of emerging market exchange rates and find that forecasts often contain information with respect to directional changes of exchange rates. [17] study the directional accuracy of interest rate forecasts from Livingston Survey using ROC curve analysis and conclude that they contain useful information with respect to the direction of future changes in interest rates. [18] uses ROC techniques to measure the effectiveness of foreign exchange market interventions and find ROC curves to be a useful technique. [19] examine the accuracy of inflation forecasts for South African Inflation Survey data for short-term and long term forecasts using ROC curve analysis and find consistent evidence that forecasts contain information with respect to the subsequent direction of change of the inflation rate.

In this paper we examine the directional accuracy of forecasts of inflation and output from MMS survey for G7 countries. We contribute to the existing literature in various empirical ways. First of all although there are some articles that examine the accuracy of directional forecasts for US utilizing ROC curve analysis (see [20] for example), an assessment of accuracy of directional forecasts from MMS survey via ROC analysis is not evident. MMS survey is different from other surveys specially because forecasts are made a week prior to the announcement of the target variable, hence contain more available information and are very short- term forecasts. Thus, the evidence toward directional accuracy should be stronger. Despite this, when judged by conventional rationality tests MMS survey of inflation and output forecasts seem to fail rationality tests under symmetric loss for some G7 countries (See [21] for more detail).

Secondly, The directional accuracy of inflation, output forecasts of G7 countries from MMS is almost non-existent aside from one study [21] that utilizes [3] and non-existent utilizing ROC analysis. [21] concludes that directional forecast accuracy seems to be evident both for inflation and output forecasts, for all G7 countries. Applying ROC curve technique is interesting to the same set of countries since ROC analysis can capture the accuracy of directional forecasts for different criterion while the conventional market timing tests can be used only for one.

In the next section, we provide more detailed information on MMS survey. In Section 3, we introduce the ROC curve construction for directional forecast accuracy of inflation and output. Section 4 presents the results and section 5 concludes.

2. MMS Survey Data

We use forecasted and realized values of percentage changes in PPI and CPI as a measure of inflation and forecasted and realized values of percentage changes of industrial production (IP) as a measure of output growth from MMS survey. The frequency of the data is monthly.

3. ROC curve construction

The origin of the ROC curve can be traced back to radar signal detection theory developed in the 1950s and 1960s [22], and its potential for medical diagnostic testing was recognized as early as 1960 ([23]). More recently, ROC analysis has been introduced in psychology [24], in atmospheric sciences [25], and in the machine learning literature [26]. See [27] for a survey.

A receiver operating characteristic (ROC) curve is a method for summarizing the performance of a test or prediction of a binary outcome that yields a score. For a given cut point on the score, values of the score above the cut point are classi-

fied as values below are classified as b. As the cutpoint, denoted by c , is varied a different 2x2 table is generated, creating False positive rate [FPR(c)] also referred as “*specificity*” and True positive Rate (TPR(c))-also referred as “*sensitivity*”. The ROC curve is a plot connecting the points $(FPR(c), TPR(c))$ for various values of c , with FPR and TPR on the horizontal and vertical axes, respectively. The better the classifier the closer the ROC curve is to the upper left corner of the plot. A classifier that performs no differently from chance has a ROC curve that goes diagonally from the lower left to the upper right corner. For more details, extensions, and applications of ROC curves see [27] and [28]. When discussing ROC curves, it is convenient to view TPR as a function of FPR, which we write as TPR (FPR).

The area under an ROC curve (AUC) is most commonly used global index of diagnostic accuracy. Values of AUC close to one indicate a high diagnostic accuracy of the marker. [29] interpreted it as the average hit rate for all underlying values of false alarm rates, and also as the average false alarm rate for all values of hit rates, see also [30].

3.1. ROC Curve Construction

The direction of change of inflation rate is defined as the difference between the future inflation rate π_{t+1} and current inflation rate π_t . Following [19], we define an “event” as a rise in the inflation rate between period t and period $t + 1$ and a “nonevent” if the inflation rate does not change or falls between period t and period $t + 1$. Hence, we define

$$\text{Event: } \pi_{t+1} > \pi_t \tag{1}$$

$$\text{Nonevent: } \pi_{t+1} \leq \pi_t \tag{2}$$

Let $\pi_{t,t+1}^e$ denote a period- t forecast of the inflation rate for $t+1$. In order to determine if forecasts signal the direction of change of the inflation rate, we assume that a “signal” occurs whenever the difference between a period- t forecast and the latest information on the period- t inflation rate is not less than a decision criterion, $c \in (-\infty, +\infty)$ and “non-signal” occurs whenever the difference between a period- t forecast and the period- t inflation rate is less than a decision criterion. Thus,

$$\text{Signal; } \pi_{t,t+1}^e - \pi_t \geq c,$$

$$\text{Non-Signal; } \pi_{t,t+1}^e - \pi_t < c,$$

and sensitivity, SE, measures the ratio of true signals relative to all events, and specificity, SP, measures the ratio of true non-signals relative to all nonevents.

It follows from the definition of signals and non-signals, that the sensitivity and specificity of forecasts of the inflation rate are functions of the decision criterion, c . Following [19] we define ($\#$ = number of counts)

$$SE(c) = \frac{\#(\pi_{t,t+1}^e - \pi_t \geq c \cap \pi_{t+1} - \pi_t > 0)}{\#(\pi_{t+1} - \pi_t) > 0}$$

$$SE(c) = \frac{\#(\pi_{t,t+1}^e - \pi_t < c \cap \pi_{t+1} - \pi_t \leq 0)}{\#(\pi_{t+1} - \pi_t) \leq 0}$$

The area, AUROC, under a ROC curve summarizes the directional accuracy of forecasts of the inflation rate. If forecasts of the inflation rate perfectly predict the direction of changes of the inflation rate, then AUROC = 1. In contrast, if the directional accuracy of forecasts of the inflation rate is the same as the directional accuracy of a pure noise signal, then AUROC = 0.5.

Finally, if forecasts of the inflation rate systematically predict the wrong sign of subsequent changes of the inflation rate, then AUROC < 0.5.

The AUROC statistic can be estimated using a non-parametric approach that uses the fact that the AUROC statistic is linked to the Wilcoxon-Mann-Whitney U statistic (31, 32). We compute

$$AUROC = \frac{n_0 n_1 - U}{n_0 n_1}$$

where n_0 = # nonevents, n_1 = # events, and $U = R - 0.5 n_0(n_0 + n_1)$ = two-sample Mann Whitney rank-sum test with R = rank sum of the nonevents. The standard error (SE) of the AUROC statistic is given by (See [32]; see also the survey by [33]).

$$SE = \sqrt{\frac{X + Y + Z}{n_0 n_1}}$$

where $X = AUROC(1 - AUROC)$, $Y = (n_1 - 1)(Q_1 - AUROC^2)$, $Z = (n_0 - 1)(Q_2 - AUROC^2)$, with $Q_1 = AUROC / (2 - AUROC)$ and $Q_2 = 2AUROC^2 / (1 + AUROC)$.

We use a similar ROC construction set up for output using percentage change of IP to construct output growth.

4. Results

Table 1 summarize the estimates of the AUROC statistic for inflation forecasts when PPI is used as a measure of inflation³, for all the countries considered with standard error of the AUROC statistic. The estimates of the AUROC statistic are significantly different from 0.5 (a pure noise signal) at conventional significance levels for all countries considered.

Table 2 summarize the estimates of the AUROC statistic for output forecasts when IP is used as a measure of output (aside from Japan where monthly GDP is used) for all the countries considered with standard error of the AUROC statistic. The estimates of the AUROC statistic are significantly different from 0.5 (a pure noise signal) at conventional significance levels for all countries considered.

Table 1. AU ROC Statistic for Inflation

Country	AUROC	St. Error
Canada	0.9136***	0.0081
France	0.8654	0.0012
Germany	0.8756	0.0023
Italy	0.8321***	0.0163
Japan	0.8778***	0.0065
U.K	0.8697	0.0016
USA	0.9231***	0.0093

*** Denotes 1% significance level of a test of the null hypothesis that inflation forecasts are indistinguishable from a white noise pure signal.

Table 2. AU ROC Statistic for Output

Country	AUROC	St. Error
Canada	0.8916***	0.0045
France	0.8514	0.0005
Germany	0.8736	0.0071
Italy	0.8236***	0.0084
Japan	0.8658***	0.0075
U.K	0.8547	0.0013
USA	0.9183***	0.0073

*** Denotes 1% significance level of a test of the null hypothesis that output forecasts are indistinguishable from a white noise pure signal.

5. Conclusion

We analyze the directional forecasts accuracy of inflation, output forecasts from MMS survey for G7 countries using ROC curve analysis. Applying ROC curve technique is interesting since ROC analysis can capture the accuracy of directional forecasts for different criterion while the conventional market timing tests can be used only for one. Our results indicate that forecasts of inflation, output from MMS contain valuable information for the target variables considered uniformly for all countries and reinforce the results of [21] that MMS survey of inflation output forecasts for G7 countries have directional forecast accuracy considered separately, although they seem to fail rationality tests under symmetric loss for some G7 countries [34, 21].

³ We only reported results from PPI for brevity. Similar results are obtained for CPI.

References

- [1] Merton, R. C. (1981). On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *The Journal of Business*, 54, 363-406.
- [2] Schnader, M. H. and Stekler, H. O. (1990). Evaluating predictions of change. *The Journal of Business*, 63, 99-107.
- [3] Stekler, H. O. (1994). Are economic forecasts valuable? *Journal of Forecasting*, 13, 495-505.
- [4] Sinclair, T., Stekler, H. O., and Kitzinger, L. (2008). Directional forecasts of GDP and inflation: a joint evaluation with an application to Federal Reserve predictions. *Applied Economics*, 42, 2289-97.
- [5] Pesaran, M. H. and Timmerman, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, 10, 461-5.
- [6] Stein, R. M. (2005). 'The Relationship between Default Prediction and Lending Profits: Integrating ROC Analysis and Loan Pricing'. *Journal of Banking & Finance*, 29, 1213-1236.
- [7] Blöchliger, A. and Leippold, M. (2006). 'Economic Benefit of Powerful Credit Scoring'. *Journal of Banking & Finance*, 30, 851-873.
- [8] Ravi, V. and Pramodh, C. (2008). 'Threshold Accepting Trained Principal Component Neural Network and Feature Subset Selection: Application to Bankruptcy Prediction in Banks'. *Applied Soft Computing*, 8, 1539-1548.
- [9] Berge, T. J. and Jordà, Ò. (2011). 'Evaluating the Classification of Economic Activity into Recessions and Expansions'. *American Economic Journal: Macroeconomics*, 3, 246-277.
- [10] Jordà, Ò. and A. M. Taylor. (2012). "The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself". *Journal of International Economics*, 88, 74-90.
- [11] Drehmann, M. and Juselius, M. (2014). 'Evaluating Early Warning Indicators of Banking Crises: Satisfying Policy Requirements'. *International Journal of Forecasting*, 30, 759-780.
- [12] Lahiri, K. and Yang, L. (2013). Forecasting Binary Outcomes, in A. Timmermann and G. Elliott (eds.), 'Handbook of Economic Forecasting Volume 2B', North-Holland Amsterdam, pp. 1025-1106.
- [13] Cohen, J., Garman, S., and Gorr, W. (2009). 'Empirical Calibration of Time Series Monitoring Methods using Receiver Operating Characteristic Curves'. *International Journal of Forecasting*, 25, 484-497.
- [14] Gorr, W. and Schneider, M. J. (2011). 'Large-Change Forecast Accuracy: Reanalysis of M3-Competition Data using Receiver Operating Characteristic Analysis'. *International Journal of Forecasting*, 29, 274-281.
- [15] Jordà, Ò, Schularick, M., and Taylor, A. M. (2011). 'Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons'. *IMF Economic Review*, 59, 340-378.
- [16] Pierdzioch, C. and Rulke, J. C. (2015). "On the Directional Accuracy of Forecasts of Emerging Market Exchange Rates." *International Review of Economics and Finance*, 38, 369-376.
- [17] Pierdzioch, C. (2015). "A note on the Directional Accuracy of interest-rate Forecasts". *Applied Economics Letters*, 22: 13, 1073-1077. DOI: 10.1080/13504851.2014.1000516.
- [18] Pierdzioch, C. (2016). "Using ROC techniques to Measure the Effectiveness of Foreign Exchange Market Interventions". *Applied Economics*, Vol. 23, No. 6, 389-393.
- [19] Pierdzioch, C. and Schmidh, H. (2017). "On the Directional Accuracy of Inflation Forecasts: Evidence from South African Survey Data". *Journal of Applied Statistics*.
- [20] Lahiri, K. and Wang, J. G. (2013). 'Evaluating Probability Forecasts for GDP Declines using Alternative Methodologies'. *International Journal of Forecasting*, 29, 175-190.
- [21] Ulu, Y. (2015). "Rationality of inflation-output forecasts of MMS survey: international evidence". *Applied Economics*, 47: 12, 1187-1198.
- [22] Peterson, Wesley W. and Birdsall, Theodore G. (1953). "The theory of signal detectability: part I. the general theory. Electronic Defense Group: Technical Report, 13. University of Michigan, Ann Arbor.
- [23] Lusted, Lee B. (1960). "Logical analysis in roentgen diagnosis". *Radiology*, 74(2), 178-193.
- [24] Swets, John A. (1973). "The relative operating characteristic in psychology". *Science*, 182(4116), 990-1000.

- [25] Mason, Ian B. (1982). "A model for the assessment of weather forecasts". *Australian Meteorological Magazine*, 30 (4), 291-303.
- [26] Spackman, Kent A. (1989). "Signal detection theory: valuable tools for evaluating inductive Learning". *Proceedings of the Sixth International Workshop on Machine Learning*. Morgan Kaufman, San Mateo, Calif, pp. 160-163.
- [27] Pepe, Margaret S. (2003). "The Statistical Evaluation of Medical Tests for Classification and Prediction". Oxford University Press, Oxford.
- [28] Baker, S. G., B. S. Kramer. (2007). "Peirce, Youden, and Receiver Operation Characteristic Curves". *The American Statistician*, 61: 343-346.
- [29] Metz, C. E. (1986). 'ROC Methodology in Radiologic Imaging'. *Investigative Radiology*, 21, 720-733.
- [30] Lasko, T. A., Bhagwat, J. G., Zou, K. H., and Ohno-Machado, L. (2005). 'The Use of Receiver Operating Characteristic Curves in Biomedical Informatics'. *Journal of Biomedical Informatics*, 38, 404-415.
- [31] Bamber, J. D. (1975). "The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Operating Characteristic Graph". *Journal of Mathematical Psychology*, 12: 387-415.
- [32] Hanley, J. A. and B. J. McNeil. (1982). "The Meaning and Use of the Area Under a Receiver Operating Characteristic Curve". *Radiology*, 143: 29-36.
- [33] Greiner, M., D. Pfeiffer, R. D. Smith. (2000). "Principles and Practical Application of the Receiver-Operating Characteristic Analysis for Diagnostic Tests." *Preventive Veterinary Medicine*, 45: 23-41.
- [34] Ulu, Y. (2013). "Multivariate test for forecast rationality under asymmetric loss functions: Recent evidence from MMS survey of inflation-output forecasts." *Economics Letters*, 119: 2, 168-171.