Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods

Marcelo C. MEDEIROS
Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro 22451-900, Brazil (*mcm@econ.puc-rio.br*)

Gabriel F. R. VASCONCELOS
Department of Economics, University of California, Irvine, 3201 Social Sciences Plaza B, Irvine, CA 92617 (*gabriel.vasconcelos@uci.edu*)

Álvaro VEIGA
Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro 22451-900, Brazil (*alvf@ele.puc-rio.br*)

Eduardo ZILBERMAN
Central Bank of Chile and Department of Economics, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Agustinas 1180, Santiago 867, Chile (*ezilberman@bcentral.cl*)

Inflation forecasting is an important but difficult task. Here, we explore advances in machine learning (ML) methods and the availability of new datasets to forecast U.S. inflation. Despite the skepticism in the previous literature, we show that ML models with a large number of covariates are systematically more accurate than the benchmarks. The ML method that deserves more attention is the random forest model, which dominates all other models. Its good performance is due not only to its specific method of variable selection but also the potential nonlinearities between past key macroeconomic variables and inflation. Supplementary materials for this article are available online.

KEY WORDS: Big data; Inflation forecasting; LASSO; Machine learning; Random forests.

1. INTRODUCTION

It is difficult to overemphasize the importance of forecasting inflation in rational economic decision-making. Many contracts concerning employment, sales, tenancy, and debt are set in nominal terms. Therefore, inflation forecasting is of great value to households, businesses, and policymakers. In addition, central banks rely on inflation forecasts not only to inform monetary policy but also to anchor inflation expectations and thus enhance policy efficacy. Indeed, as part of an effort to improve economic decision-making, many central banks release inflation forecasts on a regular basis.

Despite the benefits of forecasting inflation accurately, improving simple models has proved challenging. As Stock and Watson (2010) emphasized, "it is exceedingly difficult to improve systematically upon simple univariate forecasting models, such as the Atkeson and Ohanian (2001) random walk model [...] or the time-varying unobserved components model in Stock and Watson (2007)." This conclusion is supported by a large literature (Faust and Wright 2013), but this literature has largely ignored the recent machine learning (ML) and "big data" boom in economics.[1] With a few exceptions, previous works either considered a restrictive set of variables or were based on a small set of factors computed from a larger number of predictors known as "diffusion indexes" (Stock and Watson 2002). In addition, most of these works focused on a time period when inflation was very persistent, which favors models that treat inflation as nonstationary.

"Big data" and ML methods are not passing fads, and investigating whether the combination of the two is able to provide more accurate forecasts is of paramount importance. Gu, Kelly, and Xiu (2018), for example, showed that ML methods coupled with hundreds of predictors improve substantially out-of-sample stock return predictions. In a similar spirit, despite the previous skepticism, we argue that these methods lead to more accurate inflation forecasts. We find that the gains of using ML methods can be as large as 30% in terms of mean squared

[1]See Mullainathan and Spiess (2017) for discussions of ML methods and big data in economics. In this article, an ML model is any statistical model that is able to either handle a large set of covariates and/or describe nonlinear mappings nonparametrically. Some of these methods even predate "machines."

Journal of Business & Economic Statistics
January 2021, Vol. 39, No. 1
DOI: 10.1080/07350015.2019.1637745
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

What does machine learning say about the drivers of inflation?

Emanuel Kohlscheen[1,2]

Abstract

This paper examines the drivers of CPI inflation through the lens of a simple, but computationally intensive machine learning technique. More specifically, it predicts inflation across 20 advanced countries between 2000 and 2021, relying on 1,000 regression trees that are constructed based on six key macroeconomic variables. This agnostic, purely data driven method delivers (relatively) good outcome prediction performance. Out of sample root mean square errors (RMSE) systematically beat even the in-sample benchmark econometric models, with a 28% RMSE reduction relative to a naïve AR(1) model and a 8% RMSE reduction relative to OLS. Overall, the results highlight the role of expectations for inflation outcomes in advanced economies, even though their importance appears to have declined somewhat during the last 10 years.

JEL Classification: E27; E30; E31; E37; E52; F41
Keywords: expectations; forecast; inflation; machine learning; oil price; output gap; Phillips curve.

[1] Bank for International Settlements. Centralbahnplatz 2, 4051 Basel, Switzerland. *E-Mail address*: emanuel.kohlscheen@bis.org.
[2] I am grateful for comments from Deniz Igan and Daniel Rees. The views expressed in this paper are those of the author and do not necessarily reflect those of the Bank for International Settlements.

Inflation Prediction model using Machine Learning

Omprakash Yadav[#1] Cynara Gomes[#2] Abhishek Kanojiya[#3] Abhishek Yadav[#4]
Department of Computer Engineering Xavier Institute of Engineering, Mumbai University
[1]omprakash.y@xavierengg.com
[2]c98gomes@gmail.com
[3]abkanojiya8998@gmail.com
[4]yadavabhishek3008@gmail.com

***Abstract* — Inflation can be defined as the loss of purchasing power of a fiat currency over a period of time. It plays a key role in designing the macro economic policy of central banks across the world. During periods of high growth rate, inflation rises and during periods of low or negative growth rates, deflation takes place. We undertook a study to find the correlation between inflation and CPI(Consumer Price Index). We have implemented various machine learning algorithms to find out pattern of Inflation over the time. This has helped us predicting the Inflation. Finally by using Karl Pearson's coefficient we have een successful in finding out top 5 factors which has affected inflation. Inflation is also a state when purchasing power of a currency i s falling,our model also predicts the value of money.To derive the best model for our dataset we have compared various algoritms like Linear Regression, Ridge Regression, Lasso Regression, XGboost Regression, Random Forest Regression.**

Keywords - Consumer Price Index, Linear Regression, Ridge Regression, Lasso Regression, XGboost Regression, Random Forest Regression.

I. INTRODUCTION

Severe inflation can cause a country's economic downturn. Therefore, inflation needs to be controlled. One of inflation control conducted by the government is to calculate and predict inflation every month using CPI indicators. The Consumer Price Index (CPI) is one of the most commonly used indicators to measure the inflation rate. Prediction with monthly frequency, could be too late, because inflation has been a few days and it is not known quickly. With the development of internet technology today, various data sources related to inflation are easily obtained in real-time. This data can be used for daily CPI prediction variables. Daily predictions will allow policy-makers to make better policies.

The daily inflation data can be obtained online or real time by using web scrapping techniques, data crawling techniques (twitter) and data API features. By using web scrapping techniques, we do not have to wait for web providers to generate API data. Scrapping the web will take the primary commodity price of many e-commerce and put it into CPI calculations and predictions. For example daily data that can be associated with inflation are commodities data, many hypermarkets publishing their prices on their sites. Social media data such Twitter and facebook data, many people discuss several topics including economics in social media. Twitter data and commodity prices in the e-commerce market can be used as predictive variables.

Many central banks use forecasting models based on machine learning methodologies for estimating various macroeconomic indicators, like inflation, GDP Growth and currency in circulation etc. Particularly those central banks that run an inflation targeting regime urgently require high quality inflation forecasts. A sound base for monetary policy decisions requires deep insights into the process that generates future inflation in general and the transmission mechanism from short-term interest rates to long term interest rates, exchange rates, real economic activity and inflation in particular. Inflation forecasting plays a major role in monetary policy and daily life. Based on different diagnostic and evaluation criteria, the best forecasting model for predicting inflation is identified. The results will enable policy makers and businesses to track the performance and stability of key macroeconomic indicators using the forecasted inflation.

In this report we have attempted monthly inflation of consumer price index (CPI) for India by using conventional time series forecasting based machine learning algorithms on the basis of monthly data between January 2013 to April 2018. The focus of

How is Machine Learning Useful for Macroeconomic Forecasting?* 60 citations

Philippe Goulet Coulombe[1†] Maxime Leroux[2] Dalibor Stevanovic[2‡]
Stéphane Surprenant[2]

[1]University of Pennsylvania
[2]Université du Québec à Montréal

First version: October 2019
This version: August 31, 2020

Abstract

We move beyond *Is Machine Learning Useful for Macroeconomic Forecasting?* by adding the *how*. The current forecasting literature has focused on matching specific variables and horizons with a particularly successful algorithm. To the contrary, we study the usefulness of the underlying features driving ML gains over standard macroeconometric methods. We distinguish four so-called features (nonlinearities, regularization, cross-validation and alternative loss function) and study their behavior in both the data-rich and data-poor environments. To do so, we design experiments that allow to identify the "treatment" effects of interest. We conclude that **(i)** nonlinearity is the true game changer for macroeconomic prediction, **(ii)** the standard factor model remains the best regularization, **(iii)** K-fold cross-validation is the best practice and **(iv)** the L_2 is preferred to the $\bar{\epsilon}$-insensitive in-sample loss. The forecasting gains of nonlinear techniques are associated with high macroeconomic uncertainty, financial stress and housing bubble bursts. This suggests that Machine Learning is useful for macroeconomic forecasting by mostly capturing important nonlinearities that arise in the context of uncertainty and financial frictions.

JEL Classification: C53, C55, E37
Keywords: Machine Learning, Big Data, Forecasting.

*The third author acknowledges financial support from the Fonds de recherche sur la société et la culture (Québec) and the Social Sciences and Humanities Research Council.
†Corresponding Author: gouletc@sas.upenn.edu. Department of Economics, UPenn.
‡Corresponding Author: dstevanovic.econ@gmail.com. Département des sciences économiques, UQAM.

Inflation Forecasting Using Machine Learning Methods

Ivan Baybuza, *Ludwig Maximilian University of Munich**
ibaybuza@nes.ru

Inflation forecasting is an important practical problem. This paper proposes a solution to this problem for Russia using several basic machine learning methods: LASSO, Ridge, Elastic Net, Random Forest, and Boosting. Despite the fact that these methods already existed in the early 2000s, for a long time they remained almost unnoticed in the professional literature related to the forecasting of inflation in general, and Russian inflation in particular. This paper is one of the first attempts to apply machine learning methods to the forecasting of inflation in Russia. The present empirical study demonstrates that the Random Forest model and the Boosting model are at least as good at inflation forecasting as more traditional models, such as Random Walk and autoregression. The main result of this paper is the confirmation of the possibility of more accurate forecasting of inflation in Russia using machine learning methods.

Keywords: *inflation forecast, machine learning, boosting, random forest*
JEL Codes: *C53, E37*

Citation: Baybuza, I. (2018). Inflation Forecasting Using Machine Learning Methods. *Russian Journal of Money and Finance*, 77(4), pp. 42–59.
doi: 10.31477/rjmf.201804.42

1. Introduction

It is difficult to overestimate the importance of inflation forecasting for rationally thinking and acting economic agents: numerous economic obligations, including wages and interest rates, are usually expressed in nominal prices. In practice, central banks implement monetary policy guided mainly by their expectations of how inflation will behave in the short or medium term, rather than by its current values, since the rate of inflation does not react immediately to the tightening or easing of monetary policy, but rather with a certain lag. Therefore, price forecasting is important both for households and businesses, and for official authorities.

* The author is a graduate of the HSE/NES Joint Bachelor's Program in Economics (BAE 2018) and a student on the Master's Programme in Economics of Ludwig Maximilian University of Munich, Germany.