Economic Modelling xxx (2018) 1–8

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Testing the optimality of inflation forecasts under flexible loss with random forests

Christoph Behrens, Christian Pierdzioch *, Marian Risse

Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

ARTICLE INFO

JEL classification:
C53
E31
E32
E37

Keywords:
Inflation forecasts
German economic research institutes
Forecast optimality
Flexible loss
Random forests

ABSTRACT

We contribute to recent research on the optimality of macroeconomic forecasts. We start from the assumption that forecasters may have a flexible rather than a symmetric (quadratic) loss function assumed in standard tests. This assumption leads to the prediction that variables available to a forecaster when a forecast was formed should have no predictive value for a binary 0/1-indicator that captures the sign of the forecast error. A test of forecast optimality, thus, can be interpreted as a classification problem. We use random forests to model this classification problem. Random forests are a powerful nonparametric modeling instrument originally developed in the machine-learning literature. Unlike conventional linear-probability or logit/probit-models, random forests account in a natural way for potential nonlinear links between the signed forecast error and the variables in a forecaster's information set. Random forests also can handle a situation in which the number of forecasts is small relative to the number of predictor variables that a researcher uses to proxy a forecaster's information set. Random forests, therefore, are a powerful modeling device that is of interest for every researcher who studies the properties of macroeconomic forecasts. Upon estimating random forests on forecasts of four German research institutes, we document that optimality of longer-term inflation forecasts cannot be rejected and that inflation forecasts are weakly efficient. For shorter-term inflation forecasts, our results are heterogeneous across research institutes. When we pool the data across the research institutes, we reject optimality of both shorter-term and longer-term forecasts.

1. Introduction

Building on important earlier work (Granger, 1969; Varian, 1974; Zellner, 1986, for recent contributions, see also Christoffersen and Diebold, 1996, 1997), research on the properties of macroeconomic forecasts under flexible loss has mushroomed in recent years. In this area of research, researchers relax the standard assumption of symmetric (quadratic) loss and instead assume that forecasters form their forecasts under a flexible and potentially asymmetric loss function. Researchers have explored the implications of asymmetric loss in fields such as central banking (Capistrán, 2008; Pierdzioch et al., 2016b, among others), fiscal forecasting (Artis and Marcellino, 2001; Elliott et al., 2005), growth and inflation forecasting (Christodoulakis and Mamatzakis, 2008; Pierdzioch et al., 2016a; Sun et al., 2018), and financial forecasting (Aretz et al., 2011; Fritsche et al., 2015). Much of this research builds on results on optimal forecasts under asymmetric loss derived by Elliott et al. (2005, 2008), who develop a GMM-framework for estimating and testing forecast optimality when the loss function is of the lin-lin or quad-quad form.

Patton and Timmermann (2007) further relax the assumption regarding the form of the loss function and propose a test of forecast optimality under unknown loss. They show (see their Proposition 3) that, for a data-generating process that (i) has dynamics only in the conditional mean and a loss function that depends only on the forecast error, or, (ii) has dynamics in the conditional mean and variance and the loss function is homogeneous in the forecast error, forecast optimality implies that the sequence, $\mathbf{1}_{e_{t+1}\leq 0}$, is independent of the predictor variables known to a forecaster at the time when a forecast is formed, where $\mathbf{1}$ denotes the indicator function and e_{t+1} denotes the forecast error. One way to test forecast optimality, thus, is to estimate a linear-probability or a logit/probit model of the format $\mathbf{1}_{e_{t+1}\leq 0} = \mathbf{X}_t\beta + u_{t+1}$.[1] In this model, β denotes a vector of coefficients, the matrix $\mathbf{X}_t$ contains the predictor variables in a forecaster's period-t information set, and u_{t+1} denotes the error term. Forecast optimality requires that the null

* Corresponding author.
E-mail address: c.pierdzioch@hsu-hh.de (C. Pierdzioch).

[1] For recent research on the directional accuracy of survey forecasts, see also Hutson et al. (2014) and Pierdzioch et al. (2017), among others.

https://doi.org/10.1016/j.econmod.2018.02.004
Received 16 August 2017; Received in revised form 7 February 2018; Accepted 9 February 2018
Available online XXX
0264-9993/

hypothesis that the elements of the vector of coefficients, β, are all zero cannot be rejected.

We use the Patton-Timmermann test to study the optimality of inflation forecasts of four leading German economic research institutes. To this end, we use random forests (on random forests, see Breiman, 2001). Random forests combine a large number of classification trees to model the links between the forecast error and the predictor variables in a forecaster's information set (on classification trees, see Breiman et al., 1984). Forecast optimality can then be interpreted as a classification problem and studied by estimating a model of the format $\mathbf{1}_{e_{t+1}\leq 0} = \sum_T T(\mathbf{X}_t) + u_{t+1}$, where T denotes an individual classification tree.

Classification trees use recursive hierarchical binary splits to subdivide the predictors into non-overlapping hypercubes. For every subspace, the prediction of $\mathbf{1}_{e_{t+1}\leq 0}$ is then computed by means of a majority vote. In this way, classification trees capture nonlinear linkages of unknown form between the forecast error and the predictors (for the pros and cons of trees, see Hastie et al., 2009, p. 351). A nonlinearity arises if the linkage between the forecast error and, for example, the short-term interest rate differs in times of a severe financial crisis such as the one that hit the world economy in 2007/2008 from the corresponding linkage that is characteristic of tranquil economic times. Furthermore, the hierarchical structure of classification trees implies that they capture interaction effects of unknown form between the predictors of the forecast error. Such interaction effects arise if, for example, a forecaster relies on quantity-theory considerations to form an inflation forecasts based on the growth rate of money supply in normal times but uses another model to predict inflation in times when the short-term interest rate reaches the zero-lower bound.

Research on the properties of macroeconomic forecasts for Germany has a long tradition (see, e.g., Neumann and Buscher, 1980; Kirchgässner, 1984; Döpke, 2000). Heilemann and Stekler (2013) analyze the time-varying accuracy of growth and inflation forecasts, Döpke and Fritsche (2006b) study the properties of the dispersion of growth and inflation forecasts, and Kirchgässner and Müller (2006) study the implications of costly forecast revisions. Döpke and Fritsche (2006a) study a panel of growth and inflation forecasts and cannot reject unbiasedness and weak efficiency, where weak efficiency requires that the lagged forecast error has no predictive value for the subsequent forecast error (Kirchgässner, 1993; Timmermann, 2007, among others). Döpke and Fritsche (2006a) reject, however, strong efficiency of forecasts as forecast errors are not orthogonal to predictors in a forecaster's information set (see also Kirchgässner and Savioz, 2001). Using a Bayesian approach, Behrens et al. (2017) also cannot reject weak efficiency of growth and inflation forecasts of German economic research institutes. Their results for the strong efficiency of forecasts are mixed. Döpke et al. (2010) test for forecast optimality under both a symmetric and asymmetric loss function. They cannot reject optimality of growth forecasts of individual research institutes irrespective of the symmetry/asymmetry of the loss function. The evidence against the optimality of inflation forecasts is stronger (for forecasts of the German Council of Economic Experts, see also Krüger and Hoss, 2012), and so we focus in this research on inflation forecasts. They also document the sensitivity of their results to the assumed form of the loss function (lin-lin vs. quad-quad), and the choice of the instruments used to set up the GMM approach of Elliott et al. (2005, 2008). Hence, it is worthwhile to reexamine the optimality of inflation forecasts under a flexible (and unknown) loss function.

We briefly describe random forests in Section 2. Because random forests are a well-known technique in the machine-learning literature (see Hastie et al., 2009, chapter 9), our description is relatively compact. We describe our data and summarize our empirical results in Section 3. We conclude in Section 4.

2. Random forests

A particularly simple approach to implement the Patton and Timmermann (2007) test of forecast optimality is to consider a linear-probability model: $\mathbf{1}_{e_{t+1}\leq 0} = \mathbf{X}_t\beta + u_{t+1}$. Optimality of forecasts cannot be rejected if the hypothesis that the components (including the intercept) of the vector, β, are zero cannot be rejected. A problem with this approach (and its logit/probit counterpart) is that degrees of freedom become a limiting factor when a researcher only has available a relatively small number of forecasts and, at the same time, the matrix $\mathbf{X}_t$ includes several predictors that may enter into a forecaster's information set. If a researcher encounters such a problem, a natural strategy to reduce the complexity of the model is to consider only a subset of $\mathbf{X}_t$, or to estimate several models that only contain a single predictor as explanatory variable. The choice of a subset of $\mathbf{X}_t$, however, is rather arbitrary. Moreover, the single-predictor strategy yields a potentially large number of estimates, which may be difficult to interpret and reconcile. The number of estimates (as well as the complexity of the model) even further increases when the sign of the forecast error is linked in a nonlinear way to the predictors and/or it is important to account for the interplay of predictors. A researcher may also find it difficult to pin down a priori the functional form of a nonlinearity (e.g., quadratic, cubic, or some other form). Similarly, the interplay of predictors may take on various forms, and it would be better to let the data decide on the appropriate form of potential interaction effects rather than to fix a priori any specific form, which then may result in a misspecified model if the data do not support the specific form of the assumed interaction effect.

A tree-based approach to testing forecast optimality avoids the problems of the linear-probability and logit/probit models. Trees have three building blocks: a root, interior nodes, and terminal nodes (the leaves). The nodes partition the space of predictors, $X_t = (x_{1t}, x_{2t}, \ldots)$, into non-overlapping hypercubes in a recursive top-down and binary way. In analogy to a standard least-squares regression, for a *regression* tree the partitioning predictor, s, and the partitioning point, z, that form a node are determined by minimizing the residual sum of squares. Using a notation similar to the one also used by Hastie et al. (2009, chapter 9), we choose s and z at the top level of a regression tree such that the half-planes, $R_1(s,z) = \{x_s | x_s \leq z\}$ and $R_2(s,z) = \{x_s | x_s > z\}$ solve, $\min_{s,z}\{RSS_1 + RSS_2\}$, where $RSS_k = \sum_{x_s \in R_k(s,z)}(e_i - \overline{e}_k)^2$, with $\overline{e}_k = \text{mean}\{e_i | x_s \in R_k(s,z)\}$, $k = 1, 2$, and e_i are the observations of the forecast error sent to region k (where we have dropped the time index). This search-and-partition process continues until a preselected maximal tree size is reached, or every terminal node has a minimum number of observations. For a *classification* tree, the search for optimal nodes is set up in a similar way but the partitioning predictor and the partitioning point are chosen to minimize the Gini index, G. Upon letting p_{km} denote the proportion of the binary-coded forecast error in region k that are from class $m = 0, 1$, we can compute the Gini index as $G = p_{km}(1 - p_{km})$. Once the optimal nodes have been identified, the region-specific prediction of $\mathbf{1}_{e_{t+1}\leq 0}$ is then determined by a simple majority vote: $m(k) = \arg\max_m p_{km}$.

This search-and-partition process gives rise to a hierarchical tree structure that captures both nonlinearities in the links between the forecast errors and the predictors and any interaction effects between the predictors. Fig. 1 plots a simple example that illustrates how the search-and-partition process is used to build a classification tree. At the root of the tree, the predictor x_1 is chosen as the partitioning predictor. The resulting binary split sends data either to the left (if $x_1 < c_1$) or to the right (if $x_1 \geq c_1$) of the node and gives rise to a simple "step-function"-like nonlinear link between the sign of the forecast error and the value of predictor, x_1. To the right-hand side of the top-level node, there are no further splits, and the classification tree predicts that forecast errors are from class $m = 1$. To the left-hand side of the top-level node, the tree has a second node. This second node adds a further layer of complexity to the classification tree, and illustrates the hierarchical structure