International Journal of Forecasting 30 (2014) 217–234

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Empirical prediction intervals revisited

Yun Shin Lee [a,*], Stefan Scholtes [b]

[a] KAIST College of Business, Korea Advanced Institute of Science and Technology, 85 Hoegi-Ro, Dongdaemoon-Gu, Seoul 130-722, Republic of Korea

[b] Judge Business School, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, United Kingdom

ARTICLE INFO

Keywords:
Interval forecasting
Probabilistic forecasting
Out-of-sample forecast error
Model uncertainty
Non-Gaussian distribution

ABSTRACT

Empirical prediction intervals are constructed based on the distribution of previous out-of-sample forecast errors. Given historical data, a sample of such forecast errors is generated by successively applying a chosen point forecasting model to a sequence of fixed windows of past observations and recording the associated deviations of the model predictions from the actual observations out-of-sample. The suitable quantiles of the distribution of these forecast errors are then used along with the point forecast made by the selected model to construct an empirical prediction interval. This paper re-examines the properties of the empirical prediction interval. Specifically, we provide conditions for its asymptotic validity, evaluate its small sample performance and discuss its limitations.

1. Introduction

Prediction intervals are valuable complements to point forecasts, as they indicate the precision of the forecasts: future realizations will fall within a prediction interval with a prescribed probability. The problem of constructing prediction intervals has traditionally been studied using a theoretical (model-based) approach, which assumes that the applied forecasting model specifies the underlying stochastic process correctly and that the forecast errors follow a specific distribution (Chatfield, 1993). It is assumed that the chosen forecasting model makes unbiased point forecasts, i.e., the mean of the forecast error is zero. The variance of the forecast error is found using theoretical formulae derived from the chosen forecasting model (see for example Box, Jenkins, & Reinsel, 1994, for ARMA models). Although in principle other error distributions are also possible, it is often assumed that the error distribution is Gaussian, as this facilitates the derivation of theoretical formulae. It has long been known, however, that such theoretical prediction intervals tend to be too narrow if the forecasting model is misspecified, i.e., if the forecast errors have a non-zero mean or if the error distribution is non-normal, see e.g. Chatfield (1993, 1995). If there are doubts about model assumptions, empirically based approaches offer a useful alternative.

The literature on empirical approaches to estimating prediction intervals can be divided into two strands. The first strand has explored the use of *empirical residual errors*, in order to avoid assumptions regarding the spread and shape of the error distribution. They compute the residual errors of a fitted forecasting model at different forecast lead times and apply non-parametric methods, such as Chebyshev's inequality (Gardner, 1988) and kernel density estimators (Wu, 2010), and semi-parametric methods, such as quantile regression (Taylor & Bunn, 1999), to construct prediction intervals. Whilst these approaches relax assumptions on the spread and shape of the error distribution, they remain based on residual errors rather than out-of-sample forecast errors. It is well known, however, that true post-sample forecast errors tend to be larger than the fitted residuals (Makridakis & Winkler, 1989). The fitted residuals – the differences between the observed and fitted values in-sample – measure how well

* Corresponding author. Tel.: +82 2 958 3339; fax: +82 2 958 3359.
E-mail addresses: yunshin@business.kaist.ac.kr (Y.S. Lee), s.scholtes@jbs.cam.ac.uk (S. Scholtes).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.07.018

the chosen model fits the data. Out-of-sample forecast errors – the differences between the realizations (which are not included in the fitting process) and the predictions of the model – indicate the chosen model's true predictive performance. They incorporate all causes of errors in the model predictions simultaneously, including random variations in the data-generating process, parameter estimation errors, and errors due to incorrect model specifications.

The second strand of the literature therefore employs *empirical out-of-sample forecast errors* to construct prediction intervals. This approach is based on the generation of a sample of out-of-sample forecast errors by fitting a chosen point forecasting model successively to a sequence of windows of past observations, and recording the associated deviations of the model predictions from the actual observations out-of-sample. Given a desired nominal coverage rate – the prespecified probability that the interval should contain future observations – the relevant quantiles of the distribution of these empirical forecast errors are used with the point forecasts made by the selected model to calculate an empirical prediction interval. This concept was introduced by Williams and Goodman (1971), and is increasingly applied as an alternative to traditional approaches (see e.g. Cohen, 1986; Jogensen & Sjoerg, 2003; Rayer, Smith, & Tayman, 2009; Isengildina-Massa, Irwin, Good, & Massa, 2011). However, little is known about the theoretical underpinnings of the approach, and some important questions remain unanswered: under what conditions is this empirical approach robust under model uncertainty? What is the finite sample performance of the approach? When is the approach preferable to the alternatives? The purpose of this paper is to focus on the empirical approach that uses out-of-sample forecast errors, and give this approach a full re-examination. Specifically, we consider two sources of model misspecification:

1. incorrect assumptions on the forecast error distribution;
2. incorrect assumptions on the functional form of the point forecasting model, leading to a biased point forecast;

and examine the robustness of the empirical approach against these two types of model uncertainty using asymptotic results, and simulation and empirical studies. We also discuss its limitations.

To illustrate the benefits of using out-of-sample forecast errors to construct prediction intervals, consider the process $Y_t = \mu + u_t$, where $u_t \sim N(0, \sigma_u^2)$. Suppose that the chosen point forecasting model is biased and produces one-step-ahead point forecasts at time t by $\hat{Y}_{t,1} = \hat{\mu}_t = \mu + b_t$, where $b_t \sim N(b, \sigma_b^2)$. This leads to out-of-sample forecast errors $E_{t,1} = Y_t - \hat{Y}_{t,1} = \mu - \hat{\mu}_t + u_t = -b_t + u_t$, and implies that $\mathrm{E}(E_{t,1}) = -b$ and $\mathrm{Var}(E_{t,1}) > \sigma_u^2$. Therefore, we can use the mean of the forecast error to re-center the prediction interval in order to correct for the forecast bias, and also use the larger variance of the forecast error to widen the interval so as to incorporate model uncertainty in addition to the true random variation u_t of the process.

Our asymptotic results show that when the data-generating process is *stationary ergodic*, the mean and variance of the out-of-sample forecast errors can be estimated consistently, and therefore the empirical prediction intervals have asymptotically correct coverages, regardless of the point forecasting model selected. Furthermore, the assumption of Gaussian errors can be avoided by applying the empirical quantiles of the forecast errors when calculating the interval endpoints. Therefore, empirical prediction intervals avoid the assumptions of a correctly specified forecasting model and Gaussian forecast errors. Since empirical prediction intervals are valid for arbitrary point forecasting models, their use also extends to forecasting models that include judgemental aspects that cannot be subsumed in the theoretical approach to estimating prediction intervals.

We evaluate the finite sample performance of the empirical prediction intervals using Monte Carlo experiments, and provide an empirical study of real exchange rate forecasts. The focus of the simulation and empirical studies is on an examination of the robustness of the approach in the face of model misspecification, in comparison with an alternative theoretical (model-based) approach and a purely non-parametric approach. Both simulation and empirical studies indicate that empirical prediction intervals are particularly robust for time series that are nearly non-stationary. In addition, given that the empirical approach relies on the generation of empirical forecast errors, it necessitates the availability of sufficient data. We find that the empirical prediction intervals for up to 10-step-ahead forecasts are fairly robust for sample sizes above 120.

The major limitation of the empirical approach is that the estimated intervals are not conditional on past observations or other predictors. If the point forecasting model contains predictors and produces biased conditional point forecasts, then the empirical approach will not produce asymptotically correct conditional intervals, as the approach widens the intervals by incorporating unconditional model uncertainty. This unconditional aspect of the approach does not cause its performance to deteriorate on average (Chatfield, 1993), but may lead to larger standard deviations of the interval estimates in practical situations, compared to alternative approaches that are conditional on previous observations. This points to a crucial trade-off in applications: the benefit of robustness against the unbiasedness of the point forecasting model must be traded off against the loss in efficiency resulting from the unconditional nature of the approach. However, if the point forecasting model employed is known to produce unbiased point forecasts conditional on predictors, the empirical approach will construct consistent conditional intervals as well.

This paper is organized as follows. In Section 2, we describe the main approaches for obtaining theoretical and empirical prediction intervals. Section 3 specifies assumptions for the asymptotic validity of the empirical approach. Section 4 contains a small-sample Monte Carlo study that compares the relative performances of the theoretical and empirical prediction intervals. An application to real data is presented in Section 5, and Section 6 provides a conclusion. The Appendix contains the main proof of the asymptotic analysis in Section 3.