International Journal of Forecasting 33 (2017) 1–10

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Monte Carlo forecast evaluation with persistent data

Lynda Khalaf[a,b,c,*], Charles J. Saunders[d]

[a] Department of Economics and Centre for Monetary and Financial Economics (CMFE), Carleton University, Canada

[b] Centre interuniversitaire de recherche en économie et analyse quantitative (CIREQ), Canada

[c] Centre de Recherche en économie de l'Environnement, de l'Agroalimentaire, des Transports et de l'Énergie (CREATE), Université Laval, Canada

[d] Department of Economics, University of Western Ontario, Canada

ARTICLE INFO

Keywords:
Evaluating forecasts
Model selection
Stationarity
Statistical tests
Monte Carlo
Finite sample tests

ABSTRACT

Persistent processes, including local-to-unity and random walks, are commonly considered as forecasting models of interest. However, the associated forecast errors follow non-standard distributions that complicate forecast evaluation tests. We propose a finite sample simulation-based solution to this problem. The method requires a flexible parametric null model that can be simulated as long as a finite dimension nuisance parameter can be specified. The size control of our method is robust to non-standard limiting distributions, such as degenerate asymptotic distribution problems that arise from nested and unit root models. Our simulation studies demonstrate that many of the existing forecast evaluation methods, including various bootstraps, over-reject for highly persistent data. In contrast, our method is level correct and has good power. We extend our approach to the inversion of forecast evaluation statistics in order to construct exact confidence sets for the benchmark model. Confidence sets provide much more information than tests, particularly in the case of the persistence-adjusted relevance of predictive regressors (Rossi, 2005).

1. Introduction

Forecast evaluation methods and statistics allow for the ranking and comparison of models, but inference in time series contexts is related strongly to the degree of persistence. Local-to-unity and unit roots models are persistent processes that are considered commonly as models of interest; see Alquist and Kilian (2010), Baumeister et al. (Forthcoming), and Bernard et al. (2012) for some applications to commodity prices and macroeconomic data.

The forecast errors from persistent processes are known to follow non-standard distributions, see Kemp (1999) and Phillips (1998). Diebold and Kilian (2000) suggest using a unit root pre-test to choose a linear or first-difference forecasting model design. Their method provides some improvement over an arbitrary selection of the model structure, but the improvements rely on low-power tests for unit roots. The forecast evaluation tests for cointegrated and unit root models that were examined by Berkowitz and Giorgianni (2001) and Corradi et al. (2001) rely on non-standard critical values for inference.

Rossi (2005) employs a Bonferroni method, based on the work of Cavanagh et al. (1995) and Stock and Watson (1988), to account for the non-standard distribution of forecast evaluation statistics. Rossi's method focuses on a local-to-unity definition of the autoregressive parameters underlying the predictive model, which approaches the random walk forecast near the boundary even when the predictive covariates are not irrelevant. More broadly,

* Corresponding author at: Department of Economics and Centre for Monetary and Financial Economics (CMFE), Carleton University, Canada.
E-mail addresses: Lynda.Khalaf@carleton.ca (L. Khalaf), csaund9@uwo.ca (C.J. Saunders).

http://dx.doi.org/10.1016/j.ijforecast.2016.06.004
0169-2070/

Bonferroni bounds are known to suffer from low power, due to their conservative nature.

Outside the context of forecast evaluation, the methods for constructing confidence intervals for autoregressive parameters at or near unity are challenged by Phillips (2014). Specifically, the confidence intervals based on local-to-unity approaches surveyed by Phillips are shown to be invalid in the stationary case, with zero asymptotic coverage probability. This includes methods of the form considered by Rossi (2005). Such discontinuities provide the motivation for this paper.

Forecasting evaluations under the assumption of stationarity have resulted in several successful bootstrap approaches; see for example Giacomini and White (2006), Hansen (2005), Harvey and Newbold (2000), Hubrich and West (2010) and White (2000). However, stationary and strongly persistent series cause a deterioration of the properties of these bootstrap methods, leading to forecast evaluation tests that are severely oversized. The poor performances of bootstrap methods for local-to-unity and unit root processes are not unexpected; Andrews (2000) and Mikusheva (2007) demonstrate that bootstrap and subsampling methods can be inconsistent.

This paper proposes a finite sample motivated approach for addressing the above problems, building on the Monte Carlo (MC) test methods of Dufour (2006). This simulation-based procedure is exact when the null distribution of the statistic considered can be simulated under the null hypothesis. Complicated finite or limiting distributions, which covers various asymptotic discontinuities, cause no concern.

Thus, our approach leads to exact *p*-values for forecast evaluation statistics, independent of the degree of persistence of the data, whether stationary, local-to-unity, or a unit root process, in spite of possible underlying non-standard, asymmetric or degenerate distributions, and regardless of whether the alternatives consist of a single or multiple models, in which case sup-type statistics are simulated. The limiting distribution of a forecast evaluation statistic does not even need to be known *a priori*.

In addition to testing, we also use the MC method to produce confidence intervals for intervening parameters. For example, for the problem analyzed by Rossi (2005), we provide simultaneous confidence sets for the persistence parameter and for the coefficient of the predictors under test. As was argued by Rossi (2005), near-unit roots confound the contributions of predictors, even when the latter are relevant. The joint MC confidence intervals that we propose provide much more information than tests, an advantage that we illustrate empirically via the well-known Meese-Rogoff puzzle.

Outside the forecasting context, the MC and MMC test methods have also been shown to solve complications that arise from unidentified nuisance parameters, see Dufour et al. (2004). The MMC method has been applied successfully in a range of areas of econometrics, with a focus on level correction or the computation of confidence sets; refer to Beaulieu et al. (2007), Beaulieu et al. (2010b), Beaulieu et al. (2013), Bernard et al. (2012), and Bernard et al. (2007). To the best of our knowledge, this is the first extension of the MMC method to the forecasting of evaluation statistics.

Our second contribution is two simulation studies that demonstrate the rejection frequency properties of our proposed approach. The first study examines the predictive ability of a random walk null model, where the forecast evaluation statistic is based on a pair of models: the random walk benchmark model and a single alternative model. The simulation design is inspired by the work of Rossi (2005). In this case, our MC test method is applied to a scaled Diebold and Mariano (1995) type statistic (denoted MSE*t*) and the encompassing statistic outlined by Clark and West (2007) (denoted ENC*t*). To maintain a focus on the forecast evaluation statistics, we implement the benchmark method defined by Rossi (2005, p. 83) as the "infeasible test", where the confidence interval for the local-to-unity parameter is assumed to include only the true value. We find that all methods provide level control in terms of size. In terms of power, the ENC*t* MC test has a higher power than either the MSE*t* MC test or the infeasible Rossi approach for all of our simulation settings. The MSE*t* MC test method dominates the infeasible Rossi method for larger sample sizes and for lower persistent processes.

The second simulation study considers a highly persistent and parsimonious benchmark model and compare it against multiple alternative models. The design is based on the work of Hubrich and West (2010). Under the null of our proposed MMC method, the rejection frequency demonstrates level control. In contrast, alternative methods, including those of Giacomini and White (2006), Hansen (2005), Harvey and Newbold (2000), Hubrich and West (2010), and White (2000), and two encompassing reality checks of Clark and McCracken (2012), are over-sized. The MMC methods demonstrate good power that improve as the sample size increases.

Lastly, we propose to invert the forecast evaluation statistic, based on the MC test method, in order to obtain exact confidence intervals on the parameters in the forecast period. Traditional approaches to obtaining confidence intervals for the benchmark model parameters assume that the in-sample estimation properties are suitable for the construction of out-of-sample bands, which may be problematic as the persistence approaches unity. In contrast, our confidence set is constructed by collecting the set of 'benchmark' models that satisfy the data. Test inversion theory has been applied to time series and for forecasting, as well as when at or approaching unity, see Cavanagh et al. (1995), Stock (1991), and Stock and Watson (1988), with the in-sample properties being exploited in each case. The idea of constructing a model confidence set was put forward by Hansen et al. (2011) in the context of a finite and discrete set of models. Our inversion produces confidence sets for the parameters of one class of alternative models, namely a class that can be defined broadly as parameterizing the "alternative" to the null hypothesis under test. To the best of our knowledge, with the notable exception of the study by Hansen et al. (2011), this is the first study to construct out-of-sample confidence sets using MC and inversion theories applied to forecast evaluation statistics.

We apply our MC test inversion to the well-known Meese-Rogoff puzzle, and confirm that the Deutsche Mark to US exchange rate fails to reject the null of a random walk. The confidence intervals based on MSE*t*, ENC*t* and our