Journal of Econometrics 186 (2015) 160–177

Contents lists available at ScienceDirect

Journal of Econometrics

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

Nested forecast model comparisons: A new approach to testing equal accuracy

Todd E. Clark [a,*], Michael W. McCracken [b]

[a] Economic Research Department, Federal Reserve Bank of Cleveland, P.O. Box 6387, Cleveland, OH 44101, United States

[b] Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166, United States

ARTICLE INFO

Article history:
Received 4 January 2012
Received in revised form
19 December 2013
Accepted 27 June 2014
Available online 22 July 2014

JEL classification:
C53
C12
C52

Keywords:
Mean square error
Prediction

ABSTRACT

We develop methods for testing whether, in a finite sample, forecasts from nested models are equally accurate. Most prior work has focused on a null of equal accuracy in population — basically, whether the additional coefficients of the larger model are zero. Our asymptotic approximation instead treats the coefficients as non-zero but small, such that, in a finite sample, forecasts from the small and large models are expected to be equally accurate. We derive the limiting distributions of tests of equal mean square error, and develop a bootstrap for inference. Simulations show that our procedures have good size and power properties.

1. Introduction

In this paper we examine the asymptotic and finite-sample properties of bootstrap-based tests of equal accuracy of out-of-sample forecasts from a baseline nested model and an alternative nesting model. In our analysis, we address two forms of the null hypothesis of equal predictive ability. One hypothesis, considered in such studies as Clark and McCracken (2001, 2005a), Corradi and Swanson (2002), Inoue and Kilian (2004), and McCracken (2007), is that the models have equal population-level predictive ability. This situation arises when the coefficients associated with the additional predictors in the nesting model are zero and hence at the population level, the forecast errors are identical and thus the models have equal predictive ability.

However, this paper focuses on a different null hypothesis, one that arises when some of the additional predictors have non-zero coefficients associated with them, but the marginal predictive content is small. In this case, addressed in Trenkler and Toutenburg (1992), Giacomini and White (2006), Hjalmarsson (2009) and Clark and McCracken (2009), the two models can have equal predictive ability at a fixed forecast origin (say time R) due to a bias–variance trade-off between a more accurately estimated, but misspecified, nested model and a correctly specified, but imprecisely estimated, nesting model. Building upon this insight, we derive the asymptotic distributions associated with standard out-of-sample tests of equal predictive ability between estimated models with weak predictors. We then develop a bootstrap-based method for imposing the null of equal predictive ability upon these distributions and conducting asymptotically valid inference. In our results, the forecast models may be estimated either recursively or with a rolling sample. Giacomini and White (2006) use a different asymptotic approximation to testing equal forecast accuracy in a given sample, but their asymptotics apply only to models estimated with a rolling window of fixed and finite width.

Our approach to modeling weak predictors is identical to the standard Pitman drift used to analyze the power of in-sample tests against small deviations from the null of equal population-level predictive ability. It has also been used by Inoue and Kilian (2004) in the context of analyzing the power of out-of-sample tests. In that sense, some (though not all) of our analytical results are quite similar to those in Inoue and Kilian (2004).

We differ, though, in our focus. While Inoue and Kilian (2004) are interested in examining the power of out-of-sample tests against the null of equal population-level predictive ability, we are interested in using out-of-sample tests to test the null hypothesis of equal finite sample predictive ability. This distinction arises because the estimation error associated with estimating unknown

* Corresponding author.
E-mail addresses: todd.clark@clev.frb.org (T.E. Clark), michael.w.mccracken@stls.frb.org (M.W. McCracken).

http://dx.doi.org/10.1016/j.jeconom.2014.06.016
0304-4076/

regression parameters can cause a misspecified, restricted model to be as accurate or more accurate than a correctly specified unrestricted model when the additional predictors are imprecisely estimated (or, in our terminology, are "weak"). We use Pitman drift simply as a tool for constructing an asymptotic approximation to the finite sample problem associated with estimating a regression coefficient when the marginal signal associated with it is small.

The lengthy literature evaluating direct, multi-step (DMS) forecasts from nested models indicates our results for these forecasts should be useful to many researchers. Applications considering DMS forecasts from nested linear models include, among others: many of the studies cited above; Diebold and Rudebusch (1991); Mark (1995); Kilian (1999); Lettau and Ludvigson (2001); Stock and Watson (2003); Bachmeier and Swanson (2005); Butler et al. (2005); Cooper and Gulen (2006); Giacomini and Rossi (2006); Guo (2006); Rapach and Wohar (2006); Bruneau et al. (2007); Bordo and Haubrich (2008); Inoue and Rossi (2008); Molodtsova and Papell (2009); Chen et al. (2010); and Ferreira and Santa-Clara (2011).

The remainder proceeds as follows. Section 2 uses a very simple illustrative data-generating process to flesh out the intuition behind our analysis and results — including the precise nature of the null hypothesis, the bootstrap algorithm, and the validity of the bootstrap. With that foundation, the paper then turns to the more general case. Section 3 introduces the notation, assumptions, and presents our theoretical results and bootstrap for testing the null of equal forecast accuracy in the finite sample. Proofs are provided in a supplementary online appendix (see Appendix A). Section 4 presents Monte Carlo results on the finite-sample performance of the asymptotics and the bootstrap. Section 5 applies our tests to evaluate the predictability of US stock returns and core PCE inflation. Section 6 concludes.

2. An illustrative example

We begin by using a simple example to first clarify how our results differ from those obtained in Giacomini and White (2006) and to then illustrate our essential ideas. This example uses a simple DGP: $y_{t+1} = \mu + u_{t+1}$, where μ is non-stochastic and u_{t+1} forms a homoskedastic martingale difference sequence with variance σ^2.

2.1. Simple version of our test of equal forecast accuracy in the finite sample

Consider comparing the finite sample forecast accuracy of two nested models, with accuracy measured under quadratic loss. In this simple example, Model 0 is a no-change model, such that $\hat{y}_{0,t+1} = 0$. Model 1 is an OLS-estimated location model, corresponding to the form of the DGP: $\hat{y}_{1,t+1} = \bar{y}_t$, where $\bar{y}_t$ equals $t^{-1}\sum_{s=1}^{t} y_s$ and $R^{-1}\sum_{s=t-R+1}^{t} y_s$ under the recursive (expanding window) or rolling window estimation schemes, respectively. From these models, we produce a total of P forecasts, take the difference in the squared forecast errors, and average across the forecast origins $t = R, \ldots, R+P-1$. The expectation of this difference in average squared errors, $P^{-1}\sum_{t=R}^{R+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2)$, equals $\mu^2 - P^{-1}\sum_{t=R}^{R+P-1}\frac{\sigma^2}{t}$ and $\mu^2 - \frac{\sigma^2}{R}$ for the recursive and rolling schemes, respectively. The μ^2 (bias) term of the difference in mean square errors (MSEs) arises due to misspecification in model 0, while the second term arises due to marginally greater estimation risk (variance) in model 1. We say the two models are expected to exhibit equal finite sample accuracy when the tradeoff between the bias and variance terms implies $P^{-1}\sum_{t=R}^{R+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2) = 0$.

The goal then becomes to develop the distribution of a statistic when this moment condition forms the null hypothesis. To avoid strong assumptions about the predictors and the model errors, we focus on asymptotic distributions. Unfortunately, as the number of forecasts P and initial sample size R diverge to infinity, $P^{-1}\sum_{t=R}^{R+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2)$ converges to μ^2 and hence in large samples the two models will be equally accurate only in the trivial case in which $\mu = 0$. Since the null of interest is one of equal finite sample accuracy, we cannot simply proceed with this formulation of the DGP and the null implication that $\mu = 0$. The reason is that with $\mu = 0$, the models will not be equally accurate in the finite sample; if $\mu = 0$, model 1 has to be less accurate than model 0 in the finite sample, because model 1 introduces estimation risk of a parameter that is 0 in population. Put another way, when $\mu = 0$ there can be no bias–variance tradeoff that makes the models equally accurate in the finite sample.

To develop a test of equal accuracy in the finite sample, Giacomini and White (2006) worked with a different null hypothesis. They departed from a null hypothesis formulated using $\lim_{R,P\to\infty} P^{-1}\sum_{t=R}^{R+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2) = 0$ because, if the asymptotics allow R to increase, the estimation risk component σ^2/R converges to zero, precluding the bias-tradeoff needed for the models to be equally accurate in forecasting in the finite sample. They instead assumed the estimation window size R to be fixed and finite, with model parameters and forecasts produced using a rolling window scheme. They then formulated the null hypothesis as $\lim_{P\to\infty} P^{-1}\sum_{t=R}^{R+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2) = 0$. With the model estimation sample size held fixed and this version of the null hypothesis of equal accuracy in the finite sample, the null-implied hypothesis $\mu^2 - \frac{\sigma^2}{R} = 0$ is viable even when P diverges to infinity. Note that in this asymptotic framework, parameter estimation error remains "large" because the parameter estimates do not converge in probability.

In this paper, to permit environments in which the parameter estimates are estimated using an expanding, or recursive, window as we proceed across forecast origins, we must take a different approach to the asymptotics and null hypothesis. The main problem is that, with recursive estimation and standard large R, P asymptotics, estimation error eventually becomes "small" in the sense that the parameter estimates converge in probability. To avoid this problem, and yet still allow a bias–variance tradeoff to exist, we model the bias as being equally small. In the context of our current example, consider modeling the unconditional mean μ as being local-to-zero such that $\mu = \mu_w/R^{1/2}$. Let $\lim_{P,R\to\infty} P/R = \lambda_P \in (0,\infty)$. If we then restate the null hypothesis as $\lim_{R,P\to\infty}\sum_{t=R}^{R+P-1} E(\hat{u}_{0,t+1}^2 - \hat{u}_{1,t+1}^2) = 0$, we find that this is equivalent to $\lambda_P\mu_w^2 - \ln(1+\lambda_P)\sigma^2 = 0$ and $\mu_w^2 - \sigma^2 = 0$ under the recursive and rolling estimation schemes, respectively.

Under this null hypothesis we derive the asymptotic distribution of two tests of equal mean square error. The simpler one we will focus on in this example is an F-type test of equal MSE, given by

$$\text{MSE-F} = P \times \frac{\text{MSE}_0 - \text{MSE}_1}{\text{MSE}_1}.$$

Because the asymptotic distribution of the statistic is non-standard, we use a bootstrap to obtain asymptotic critical values. In this simple example, the bootstrap proceeds as follows.

1. (a) Estimate the model $y_s = m + u_{1,s}$, $s = 1, \ldots, T$, by OLS. Save the residuals $\hat{u}_{1,s}$ and residual variance $\hat{\sigma}_1^2$. (b) Estimate the ridge regression

$$\tilde{\mu}_{w,T} = \arg\min_m \sum_{s=1}^{T}(y_s - m)^2 \quad \text{s.t. } m^2 = \hat{d}/R,$$

where $\hat{\lambda}_P = P/R$ and $\hat{d}$ equals $\frac{\ln(1+\hat{\lambda}_P)}{\hat{\lambda}_P}\hat{\sigma}_1^2$ and $\hat{\sigma}_1^2$ for the recursive and rolling schemes, respectively. Save $\tilde{\mu}_{w,T}$. This