Tests of Conditional Predictive Ability: A Comment

Michael W. McCracken

Working Paper 2019-018A
https://doi.org/10.20955/wp.2019.018

July 2019

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166
Tests of Conditional Predictive Ability: A Comment *

Michael W. McCracken
Federal Reserve Bank of St. Louis
July 2019

Abstract

We investigate a test of equal predictive ability delineated in Giacomini and White (2006; Econometrica). In contrast to a claim made in the paper, we show that their test statistic need not be asymptotically Normal when a fixed window of observations is used to estimate model parameters. An example is provided in which, instead, the test statistic diverges with probability one under the null. Simulations reinforce our analytical results.

JEL Nos.: C53, C12, C52
Keywords: prediction, out-of-sample, inference

---

*McCracken: Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; michael.w.mccracken@stls.frb.org. We are grateful to Ken West, Tatevik Sekhposyan, and Rafaella Giacomini for helpful comments as well as participants at the 2019 International Symposium on Forecasting and the 2019 International Association for Applied Econometrics. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System.
1 Introduction

In Giacomini and White (2006), “Tests of Conditional Predictive Ability,” the authors develop a new approach to tests of out-of-sample predictive ability that preserves finite-sample estimation error under the null. As an example, suppose we want to compare the accuracy of $h$-step ahead point forecasts from two parametric models ($i = 1, 2$) under quadratic loss. These forecasts are functions of estimated parameters $\hat{\beta}_{i,t}$ and imply forecast errors $\hat{u}_{i,t+h} = u_{i,t+h}(\hat{\beta}_{i,t})$. Much of the literature, exemplified by the work in West (1996), uses sample averages of the loss differential $\hat{\omega}_t^2 = \hat{u}_{1,t+h}^2 - \hat{u}_{2,t+h}^2$, $t = R, ..., T - h = R + P - 1$ to test a null hypothesis of the form $E(u_{1,t+h}^2 - u_{2,t+h}^2) = 0$ where $u_{i,t+h} = u_{i,t+h}(\beta^*_i)$ and $\hat{\beta}_{i,t} \rightarrow_{a.s.} \beta^*_i$. Instead, Giacomini and White consider a hypothesis of the form $E(\hat{\omega}_t^2) = 0$. The first hypothesis is a statement about the population loss differential while the second is a statement about the finite-sample loss differential.\(^1\)

This new hypothesis required a new approach to asymptotically valid inference that prevents the parameter estimates ($\hat{\beta}_t$) from converging to their population counterparts ($\beta^*$). Giacomini and White achieve this by requiring that a finite number of observations ($R$) are always used to estimate the parameters. One such approach is to estimate the parameters using a rolling window of observations $Z$ such that $\hat{\beta}_t = \beta(Z_t, ..., Z_{t-R+1})$. Along with other technical assumptions, they prove that when $P \rightarrow \infty$, a statistic of the form $P^{-1/2} \sum_{t=R}^{T-h} \hat{d}_{t+h}/\hat{\omega}$ is asymptotically standard Normal, under this new hypothesis, so long as a consistent estimate of the long-run variance of $\hat{d}_{t+h}$, $\hat{\omega}^2$, is used to form $\hat{\omega}$.\(^2\)

Giacomini and White also claim that their results hold if a fixed window of observations is used to estimate model parameters and hence, $\hat{\beta}_t = \beta(Z_{R}, ..., Z_{1})$ for all $t$. This claim is incorrect. In the proof of their Theorem 4, the authors use Lemma 2.1 of White and Domowitz (1984) to establish that the loss differentials are mixing when the underlying data is mixing. Recall that White and Domowitz prove that for a mixing sequence $Z_t$, and measurable function $g_t(.)$, $X_t = g_t(Z_t, ..., Z_{t-\tau})$ is mixing of the same order as $Z_t$ so long as the window size $\tau$, is finite. While it is true that, under the fixed scheme, the loss differential $\hat{d}_{t+h}$ is a function with a finite number of arguments, the window size $\tau$ is not finite as $P$ increases, and hence Lemma 2.1 of White and Domowitz doesn’t apply. As such, Theorem

---

\(^1\)Giacomini and White also consider the null hypothesis $E(\hat{d}_{t+h}|Z_t) = 0$ for a non-trivial time $t$ information set $Z_t$. We do not address this hypothesis in this comment.

\(^2\)In Giacomini and White, the hypotheses apply to any well-behaved loss function and are not exclusive to point forecasts. Even so, for exposition purposes we will focus on the most common applications in which quadratic loss is used to evaluate point forecasts.
4 in Giacomini and White is not valid, as currently proven, when using the fixed scheme. It remains valid under the rolling scheme.

2 An Example

Consider an application in which, under quadratic loss, the accuracy of a no-change forecast of \( yt+h \) is compared to the accuracy of a forecast based on an estimated location model. That is, \( \hat{y}_{t+h} = 0 \) while \( \hat{y}_{2,t+h} = \bar{y}_t \) where \( \bar{y}_t = \bar{y}_R \) for all \( t \) under the fixed scheme and \( \bar{y}_t = R^{-1} \sum_{s=t-R+1}^{t} y_s \) under the rolling scheme. Straightforward algebra reveals that if \( y_t = \mu + \eta_t, \eta_t = \varepsilon_t + \sum_{j=1}^{h-1} \theta_j \varepsilon_{t-j} \) with \( \varepsilon_t \sim i.i.d. N(0, \sigma^2) \), then \( E\hat{d}_{t+h} = 0 \) for all \( t = R, ..., T - h \) when \( \mu = R^{-1/2}[\gamma_0 + 2 \sum_{j=1}^{R-1} \left( \frac{R-j}{R} \right) \gamma_j]^{1/2} \) and \( \gamma_j = E\eta_t \eta_{t-j} \).

Since \( \hat{d}_{t+h} = (y_{t+h} - \bar{y}_t)^2 - (y_{t+h} - \bar{y}_t)^2 = 2y_{t+h}y_t - \bar{y}_t^2 \) we find that, in the notation of White and Domowitz, \( \tau \) equals \( t+h-1 \) for the fixed scheme and \( (t+h)-(t-R+1) = R+h-1 \) for the rolling scheme. As \( P \) diverges, \( \tau \) remains finite for the rolling scheme but does not for the fixed; hence Lemma 2.1 implies \( \hat{d}_{t+h} \) is mixing for the rolling scheme but is silent for the fixed.

For the fixed scheme we can show the loss differentials are not mixing without relying on White and Domowitz. To do so we calculate the first and second moments of the loss differential but, for brevity, do so only for \( h = 1 \). Straightforward algebra reveals that \( E\hat{d}_{t+1} = 0, E\hat{d}_{t+1}^2 = 4\sigma^2 E\bar{y}_R^2 + E(\mu^2 - (R^{-1} \sum_{s=1}^{R} \varepsilon_s)^2)^2, \) and \( E\hat{d}_{t+1} \hat{d}_{t+1-j} = E(\mu^2 - (R^{-1} \sum_{s=1}^{R} \varepsilon_s)^2)^2 \) for all \( j \geq 1 \). Evidently, while the loss differentials are covariance stationary, the autocovariances do not vanish as the lag increases. As such, we know they are not mixing.

Unsurprisingly, this implies that asymptotic Normality fails under the fixed scheme. Let \( h = 1 \) and \( \hat{\omega}^2 = P^{-1} \sum_{t=R}^{T-1} \hat{d}_{t+1}^2 \). Rearranging terms we find that under the fixed scheme, \( P^{-1/2} \sum_{t=R}^{T-1} \hat{d}_{t+1}/\hat{\omega} \) equals

\[
\frac{2(P^{-1} \sum_{t=R}^{T-1} \varepsilon_{t+1}) \bar{y}_R + P^{1/2}(\mu^2 - (R^{-1} \sum_{s=1}^{R} \varepsilon_s)^2)}{\sqrt{4\bar{y}_R^2(P^{-1} \sum_{t=R}^{T-1} \varepsilon_{t+1})^2 + 4\bar{y}_R^2(2\mu - \bar{y}_R)(P^{-1} \sum_{t=R}^{T-1} \varepsilon_{t+1}^2) + (\mu^2 - (R^{-1} \sum_{s=1}^{R} \varepsilon_s)^2)^2}}
\]

As \( P \) diverges, holding \( R \) constant, the denominator is \( O_p(1) \) with probability limit \( \sqrt{4\bar{y}_R^2 \sigma^2 + (\mu^2 - (R^{-1} \sum_{s=1}^{R} \varepsilon_s)^2)^2} \). In contrast, while the first part of the numerator is \( O_p(1) \), the second part diverges to \( \pm \infty \) with probability one.

To reinforce our point, in Table 1, and based on the data-generating processes delineated above, we provide finite-sample simulation evidence on the actual size of a two-sided, nominally 5% test. Further details on the simulation design are provided in the Table note.
In the top panel, that associated with the fixed scheme, we see that in all cases the actual size of the test is larger than 5%. More importantly, the size distortions generally increase with the sample size \( P \) with rejection frequencies as high as 70%. In the lower panel, that associated with the rolling scheme, we find that the actual size of the test is closer to what we would expect. At the shortest forecast horizon, \( h = 1 \), rejection frequencies are as high as 9% when \( P = 25 \) but otherwise range between 3% and 7%. At the longer horizons, the actual size of the test is as high as 20% but trend towards 5% as \( P \) increases.

3 Conclusion

In this comment we show that Theorem 4 of Giacomini and White is not valid when the fixed scheme is used and hence their test statistic need not be asymptotically Normal under the null. The reason it is not valid lies in their application of Lemma 2.1 of White and Domowitz which they use to verify mixing properties of the loss differentials. A simple counterexample and supporting simulation evidence are also provided.

It’s worth noting that in this simple example, the test statistic fails to be asymptotically Normal despite the fact that the loss differentials are covariance stationary. As such we also rebut a claim made in Diebold (2015) that so long as loss differentials are covariance stationary, the t-statistic delineated in Diebold and Mariano (1995) must be asymptotically Normal. The issue is not whether or not the loss differentials are covariance stationary but whether or not they are mixing, and this seems unlikely to hold for the fixed scheme.

We would be remiss if we did not emphasize that, to our knowledge, this comment does not overturn any existing empirical applications of the theory developed in Giacomini and White. Seemingly all empirical applications of Theorem 4 use the rolling scheme under which their test statistic remains asymptotically Normal. Nevertheless, the prospect for future applications of the fixed scheme remains. Unfortunately, it is easy to find extensions and applications of Giacomini and White that claim the fixed scheme is a viable option. An example from my own work can be found on p. 1135 of Clark and McCracken (2013) but other examples include Giacomini and Komunjer (2005), Amisano and Giacomini (2007), Giacomini and Rossi (2010), Borup and Thyrsaard (2017), Rossi and Sekhposyan (2019), and Yen and Yen (2019). The list is not long, but persists.
References


Table 1: Tests of Unconditional, Finite-Sample Predictive Ability

<table>
<thead>
<tr>
<th>Scheme</th>
<th>R</th>
<th>P</th>
<th>R</th>
<th>P</th>
<th>R</th>
<th>P</th>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>25</td>
<td>0.15</td>
<td>0.24</td>
<td>0.33</td>
<td>0.39</td>
<td>0.70</td>
<td>0.23</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.11</td>
<td>0.13</td>
<td>0.17</td>
<td>0.19</td>
<td>0.51</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>0.10</td>
<td>0.10</td>
<td>0.13</td>
<td>0.14</td>
<td>0.40</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>175</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.34</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>Rolling</td>
<td>25</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.12</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.03</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>175</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.03</td>
<td>0.13</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Notes: $R$ is the window size used to generate the time series and estimate the model parameters before generating a forecast, $h$ is the forecast horizon, and $P$ is the number of forecast periods. Each entry in the table represents the fraction of 5,000 replications where the null hypothesis was rejected at the 5% level using the standard Normal critical values of a two-sided test for a given $R$, $h$, and $P$. The test statistic takes the form $P^{1/2} \sum_{t=R}^{P-h} \hat{d}_{t+h}/\hat{\omega}$. The long-run variance $\omega^2$ was estimated using the Bartlett kernel with the bandwidth set to $\lfloor 4(P/100)^2/9 \rfloor + 1$. Throughout we set the error variance $\sigma^2$ equal to 1 and the MA coefficients to $\theta_j = (0.5)^j$. 