Reconciling the Return Predictability Evidence under Structural Breaks

Cheolbeom Park*

Abstract: This study shows that the poor out-of-sample performance of the real-time adjusted dividend-price ratio reported in Lettau and Nieuwerburgh (2008) is mainly a result of the gap period between the occurrence of a break and its detection, which implies that the poor out-of-sample performance of the adjusted dividend-price ratio is due to the requirement in Bai and Perron's (1998) procedure that breaks must be away from the boundaries of the sample. A substantial improvement in the out-of-sample performance of the adjusted dividend-price ratio during the gap period is shown with the use of Andrews's (2003) procedure in the real-time adjustment of the dividend-price ratio. The newly suggested procedure for the adjusted dividend-price ratio in this study has better out-of-sample performance than the simple sample mean, although it is not significant.

Keywords: stock-return predictability, structural break, out-of-sample forecast

INTRODUCTION

In a 2008 article published by the *Review of Financial Studies*, Martin Lettau and Stijn Van Nieuwerburgh provide an explanation intended to reconcile the ongoing inconsistent results in statistical tests designed to show stock-return predictability. Lettau and Van Nieuwerburgh claim that the structural change in the U.S. economy has caused the shift in the long-run mean of the dividend-price ratio and that this shift in the mean of the dividend-price ratio has made the ratio look nonstationary and obscured the predictive relation between the dividend-price ratio and future stock returns after the break. Having adjusted the dividend-price ratio with the use of Bai and Perron's structural break test procedure, Lettau and Van Nieuwerburgh show a significant and stable in-sample predictive relation between the dividend-price ratio

---

* Cheolbeom Park is a professor in the Department of Economics at Korea University, Seoul. E-mail: cbpark_kjs@kiorea.ac.kr. I am grateful to anonymous referees for helpful comments. The usual disclaimers apply.

Manuscript received June 24, 2016; out for review July 8, 2016; review completed August 12, 2016; accepted August 13, 2016.

The Korean Journal of Policy Studies, Vol. 31, No. 2 (2016), pp. 71-81.

and stock returns.

However, Lettau and Van Nieuwerburgh report that the forecast based on the real-time adjusted dividend-price ratio with the use of Bai and Perron's (1998) procedure has a relatively poor out-of-sample forecast accuracy when compared with the forecast based on the simple sample mean.[1] Lettau and Van Nieuwerburgh argue that the difficulty in estimating the new regime mean is the reason for the poor out-of-sample performance of the real-time adjusted dividend-price ratio. They state that "in real time, however, the changes in the steady-state are not only difficult to detect but also estimated with significant uncertainty, making the return forecastability hard to exploit. Out-of-sample tests performed in real time reflect this difficulty. While adjusted price ratios have superior out-of-sample forecasting power relative to their unadjusted counterparts, they do not outperform the simple random walk model" (2008, p. 3).

However, Lettau and Van Nieuwerburgh's interpretation of the results of the comparison of out-of-sample forecasting power is vague and misleading. I show here that the inherent requirement of Bai and Perron's procedure that a break be bounded from boundaries of the sample is a big disadvantage in real-time adjustment. This requirement of Bai and Perron's procedure makes it impossible to detect a break immediately after its occurrence in real-time adjustment. In other words, there is always a time lag between the occurrence of a break and the detection of it due to this requirement, and the accuracy of out-of-sample forecastability is particularly low during this gap period, which is the main source of the poor out-of-sample performance reported in Lettau and Van Nieuwerburgh. Here I provide empirical evidence for this explanation through a subsample analysis and a comparison of forecasts based on Bai and Perron's procedure and Andrews's (2003) end-of-sample instability test procedure.

In section 2, a detailed interpretation of the out-of-sample forecast results is provided. Empirical results for this interpretation are presented in section 3 and concluding remarks are in section 4.

INTERPRETATION OF OUT-OF-SAMPLE FORECAST PERFORMANCE

The algorithm for examining the out-of-sample predictability of the adjusted dividend-price ratio proposed by Lettau and Van Nieuwerburgh may be written as follows:

1. Perron 1989 is erroneously cited in Lettau and Van Nieuwerburgh's reference. They agreed in a personal communication that Perron 1989 in their reference should be replaced with Bai and Perron 1998.

1. Conduct Bai and Perron's sequential sup $F(l + 1|l)$ test to estimate break points in the mean of df_t for $t = 1,\dots,t_0$.[2]
2. Adjust df_t according to the results obtained in step 1 and equation 8 in Lettau and Van Nieuwerburgh 2008.
3. Run the predictive regression 1 in Lettau and Van Nieuwerburgh 2008 with $\tilde{r}_t$ and $\tilde{z}_t$ for $t = 1,\dots,t_0 - 1$ where $\tilde{r}_t$ is the demeaned log stock return and $\tilde{z}_t$ is the adjusted df_t in step 2.
4. Form a forecast for r_{t_0+1} based on $r_{t_0+1} = \tilde{r}_{t_0} + \hat{k}\tilde{z}_{t_0}$ where $\overline{r}_{t_0}$ is the mean of log stock return at t_0 and $\hat{k}$ is the estimate of the slope coefficient of the predictive regression in step 3.
5. Repeat the above steps after updating the data to $t_0 + 1$.

Bai and Perron's structural-break test is popular in the literature, since their method allows multiple structural breaks to occur at unknown points. While Lettau and Van Nieuwerburgh find a significant in-sample predictive relation between the adjusted dividend-price ratio and one-period-ahead stock returns, they report relatively poor out-of-sample forecastability of the adjusted dividend-price ratio. Having compared the out-of-sample forecastability based on the above algorithm with the pseudo-out-of-sample forecastability based on a regime-switching model (equipped with regime means estimated from the full sample), Lettau and Van Nieuwerburgh attribute the relatively poor out-of-sample performance to the difficulty in estimating the new mean of the dividend-price ratio immediately after the break. "Clearly," they observe, "estimating that new long-run mean based on a few data points incurs a lot of measurement error. . . . The difficulty in estimating this mean is what accounts for the increase in prediction error between the Hamilton and the Perron procedure" (2008, p. 20).

However, Lettau and Van Nieuwerburgh's interpretation could be misleading because Bai and Perron's method requires that break points be bounded from boundaries of the sample by at least εT(T is the sample size) to make each break distinct. In other words, when a break point is i_0, the break point can be estimated only after at least $\varepsilon \cdot i_0$ periods have passed since i_0 in the real-time adjustment of the dividend-price ratio. At least approximately $(1+\varepsilon)\cdot i_0$ observations are needed for the break at i_0 to be detected. As a result, once a break is detected at $(1+\varepsilon)\cdot i_0$, whether there can be enough observations

2. Although Lettau and Van Nieuwerburgh use the Bayesian information criterion (BIC) proposed by Yao (1988) and the Schwarz information criterion proposed by Liu, Wu, and Zidek (1997) to estimate break points in addition to the sequential sup $F(l + 1|l)$ test, I focus on just the sequential sup $F(l + 1|l)$ test because Bai and Perron (2003) report that the sequential sup $F(l + 1|l)$ test performs better than the other two methods, especially in the presence of serial correlation in the errors.

to estimate the new mean under the new regime depends on the magnitudes of i_0 and ε in the real-time adjustment. Since Lettau and Van Nieuwerburgh provide statistical evidence that the dividend-price ratio experiences a shift in the mean during the 1990s, which is close to the end of the full sample (which implies a large value of i_0), the new regime mean of the dividend-price ratio can be estimated reasonably well even when the break during the 1990s is detected for the first time.

For these reasons, this study considers another possible explanation of the poor-out-of-sample forecastability of the adjusted dividend-price ratio reported by Lettau and Van Nieuwerburgh. I focus on the requirement in Bai and Perron's method that ε be positive even it is small. This requirement can be a disadvantage in the real-time adjustment of the dividend-price ratio because the dividend-price ratio will be adjusted with the mixture of the old and new regime means between i_0 and $(1+\varepsilon)\cdot i_0$, due to the inability of Bai and Perron's method to detect the break occurring at i_0 before $(1+\varepsilon)\cdot i_0$. Since the wrongly adjusted dividend-price ratio affects the outcome of the predictive regression and the forecast of the stock return eventually, the relatively poor out-of-sample performance reported by Lettau and Van Nieuwerburgh may be a result of this problem, especially when the break occurs near the end of the sample, rather than owe to the difficulty in estimating the new mean of the dividend-price ratio immediately after the break, as Lettau and Van Nieuwerburgh suggest.

This study employs two strategies to address the question of which explanation might be correct. The first strategy is to divide the full sample period into subsample periods and then examine during which of the subsample periods the root mean squared error (RMSE) is particularly high. For example, if only one break point (i_0) is estimated from the full sample, and this break is detected at $i_1(>i_0)$ for the first time in the real-time adjustment, then the sample period is divided into three subsample periods, the first encompassing observations from 1 through i_0-1, the second observations from i_0 through i_1-1, and the third observations from i_1 through T. Then, the out-of-sample forecast errors from the real-time adjustment of the dividend-price ratio are computed over these subsample periods. If Lettau and Van Nieuwerburgh's interpretation is correct (i.e., if the uncertainty in estimating the new regime mean is what matters most), then the RMSE will be higher during the third subsample period (after the break is detected, as stated in Lettau and Van Nieuwerburgh 2008) in the example. However, if the poor out-of-sample forecastability results from the fact that the break point occurs at a different time from when it is detected for the first time in the real-time adjustment, then the RMSE will be higher during the second subsample period. This is because the dividend-price ratio is adjusted with the wrong regime mean contaminated by the first regime observations during the second sub-sample period. The number of subsample periods can be larger if multiple breaks are detected from the full sample.

The second strategy used in this study is to employ a structural-break test procedure that can reduce or remove the gap between i_0 and i_1 in the above example. If the new procedure provides a lower RMSE during the subsample period between i_0 and i_1, then that could be another piece of evidence that warrants reconsideration of the out-of-sample results Lettau and Van Nieuwerburgh obtained. For this purpose, I adapt Andrews's (2003) end-of-sample instability test for real-time adjustment. Unlike most existing structural-break test methods, including Bai and Perron's, for which asymptotics requires that both the numbers of observations before a potential break point and in the potential breakdown period go to infinity, the Andrews test method is designed to detect a structural break when the number of observations during the period of a potential structural breakdown is relatively small—possibly as small as one. In other words, Andrews's method is asymptotically valid so long as the number of observations that do not fall in a potential breakdown period goes to infinity while the number of observations in the period of the potential breakdown remains fixed.

This property of Andrews's method can be particularly useful in real-time adjustment, because one can apply the method to the last observation (with one-period duration) in a real-time adjustment if no break is detected with previous observations. Since the method is applied to the last observations of the sample as data are updated, there is no difference between a given break point and the time when the break is first detected. Theoretically, the break is immediately detected by an investor or an econometrician with the use of Andrews's method. Once a break is detected with the last observation at t_0, the Andrews test with a two-period duration and t_0 break point can be applied to updated observations 1 through t_0+1 to see whether the break continues as observations are updated. That is, the length of the duration of the new regime should increase accordingly as observations are updated in real-time adjustment.

Although Andrews's method is asymptotically valid when errors in the regression model are nonnormal, (conditionally) heteroskedastic, and/or autocorrelated and when the regressor is not strictly exogenous, a problem can arise from the fact that it is designed for a one-time temporary break. As a result, if multiple breaks exist in the data or the duration of the second regime is large relative to that of the first regime, then the use of Andrews's method in a real-time adjustment may not offer a reliable result. However, one possible way to sidestep these problems is that if there is another break at the end of the second regime after excluding the observations during the first regime, so long as the second regime is long enough. In other words, when the duration of the second regime is large relative to that of the first regime, one can carry out the procedure as if the data start from the first observations of the second regime.[3]

3. Only one break, which lasts from October 1992 to the end of the sample, is detected from

EMPIRICAL RESULTS

While Lettau and Van Nieuwerburgh use the annual data from the Center for Research in Security Prices (CRSP), this study uses the monthly data from the CRSP. There are several reasons why I chose to use monthly data in this study. First, it is often hard to obtain a reasonable number of observations for a subsample between the break point and the time when the break point is first detected or for a subsample after the break. For example, 1995 is estimated as a break point with the use of the annual data, but the break is detected for the first time in 2004 in real-time adjustment. As a result, only nine observations are available during a subsample between the break point and the time immediately before the break point is first detected (i.e., between 1995 and 2003), and three observations are available during a subsample after the detection of the break (i.e., between 2004 and 2006).

Second, Bai and Perron (2003) report that a small value for could cause substantial size distortions unless there are many observations. Hence, I have raised the number of observations so as to maintain a low value for $\varepsilon(0.05)$. This is favorable to the forecast formed by the use of Bai and Perron's procedure, since the gap between the break occurrence point and the detection point is expected to be short when ε is low.[4]

Despite the difference in the frequency of the data, the variables have been constructed following Lettau and Van Nieuwerburgh. Subtracting the Consumer Price Index inflation rate from the log returns of the CRSP value-weighted market portfolio yields the aggregate real stock return, while log dividend-price ratio has been generated using monthly returns both with and without dividends of the CRSP value-weighted market portfolio. The full sample period is from January 1926 to December 2006, and the forecast horizon considered in this study is one month.

Following the approach taken by Lettau and Van Nieuwerburgh, this study uses the first 20 years of data for the first forecasting regression. Four forecast methods are examined. The first is the current sample mean that can be inferred from the random walk model for the stock price. The second is the unadjusted dividend-price ratio, which is widely examined in the stock- return predictability literature. The third is the adjusted dividend-price ratio based on Lettau and Van Nieuwerburgh's method, which

---

the full sample of the monthly dividend-price ratio using Bai and Perron's method, although this information is not used in a real-time adjustment. Hence, multiple breaks and/or the long duration of the second regime are not expected to cause a problem with the use of the Andrews method in this article.

4. With the monthly frequency data, the empirical results reported in this study are not sensitive when the value of ε is 0.15.

uses the Bai and Perron sequential sup$F(l + 1|l)$ test in a real-time adjustment. The fourth is the adjusted dividend-price ratio based on Andrews's end-of-sample structural-break test in a real-time adjustment, which is designed to eliminate the gap between the break point and the time when the break point is first detected. The first three methods are examined by Lettau and Van Nieuwerburgh using annual data.[5]

Table 1 reports the percentage RMSE by the four forecasting methods. The RMSE from the full sample is shown in the first row of table 1. Consistent with Goyal and Welch (2008) and Lettau and Van Nieuwerburgh, the random walk model has a lower RMSE than forecasts based on the unadjusted dividend-price ratio or the adjusted dividend-price ratio that results from using the Bai and Perron method in real-time adjustment. However, the RMSE based on Bai and Perron's method is slightly higher than that based on the unadjusted dividend-price ratio computed using the monthly data. Overall, the forecast based on the Andrews method has the lowest RMSE, which implies that the superior out-of-sample performance of the forecast based on the Andrews method might be related to the gap between the break point and the first detection time point.

To check this possibility further, this study conducted a subsample analysis. With the full-sample observations, only one break, occurring in October 1992, is estimated

Table 1. Out-of-Sample Predictability: RMSE

	Random Walk	Unadjusted dp_t	Bai and Perron's sequential supF	Andrews end-of-sample
Full sample	4.2387	4.2407	4.2432	4.2312
December 1945-September 1992	3.5566	3.5453	3.5540	3.5504
October 1992-August 2001	4.2896	4.3694	4.3687	4.2923
September 2001-December 2006	3.9364	3.9407	3.8976	3.9097
September 2001-August 2002	5.9861	5.7589	5.7912	5.9333
September 2001-August 2003	5.6898	5.5962	5.5723	5.6461
September 2001-August 2004	4.9089	4.8583	4.8225	4.8734

5. The sample period is the same as that used by Lettau and Van Nieuwerburgh. However, the frequency of data is monthly in this article instead of annual.

at the 5% significance level using Bai and Perron's sequential sup$F(l + 1|l)$ test method. However, this break point is detected for the first time in August 2001 (almost 10 years after the occurrence of the break) in the real-time adjustment using the Bai and Perron procedure. As a result, the out-of-sample forecast errors are divided into three subsample periods: observations from December 1945 through September 1992, observations from October 1992 through August 2001, and observations from September 2001 through December 2006 (we assume that the true break point is October 1992 as estimated).[6]

Although the RMSEs from the four forecasting methods are very close during the first subsample period (i.e.. before the break occurs), the RMSE based on Bai and Perron's method deteriorates more than the RMSE from the random walk model or that based on Andrews's method during the second subsample period, which is between October 1992 and August 2001. However, in contrast to what Lettau and Van Nieuwerburgh's interpretation predicts, the forecast based on Bai and Perron's procedure is most accurate during the third subsample period, which implies that the uncertainty in estimating the new regime mean is not the reason for the poor out-of-sample performance based on Bai and Perron's method. At the time when the break is detected for the first time in a real-time adjustment (i.e., August 2001), there appears to be a sufficient number of observations falling into the new regime, which results in the lowest RMSE for the forecast based on Bai and Perron's method during the third subsample period. The last three rows of table 1 present out-of-sample performance based on the RMSE immediately after August 2001 (when the uncertainty in estimating the new regime mean is presumably highest due to the smaller number of observations falling into the new regime). The RMSE based on Bai and Perron's method is not much different from that derived from the other methods during the first (two or three) year(s) immediately after the break is detected, which is not consistent with Lettau and Van Nieuwerburgh's interpretation.

In order to examine whether the differences in the RMSEs are significant, this study conducted the Diebold and Mariano (1995) test for the sequences of three forecasts, forecasts based on the simple sample mean, the adjusted dividend-price ratio that uses Bai and Perron's method, and the adjusted dividend-price ratio that uses Andrews's method. The results are provided in table 2. Although the forecast using Andrews's

6. The real-time adjustment with Andrews's method reports that the test statistic is significant at the 5% level from February 1995 and on. Hence, if the true break occurred in October 1992, then Andrews's method detects it more quickly than Bai and Perron's in a real-time adjustment, and the benefit of this quick detection seems to overcome the cost of the imprecise estimate of the break point.

Table 2. Out-of-Sample Predictability: Diebold and Mariano Test

	Random walk vs. Bai and Perron	Random walk vs. Andrews	Bai and Perron vs. Andrews
Full sample	-0.3843	0.8997	1.3884
December 1945-September 1992	0.5742	0.9602	0.2484
October 1992-August 2001	**-1.7334**	-0.1062	**2.7786**
September 2001-December 2006	**1.6722**	1.3884	-0.5661
September 2001-August 2002	1.4637	1.3143	-1.2896
September 2001-August 2003	1.7725	1.1614	-0.9144
September 2001-August 2004	1.5990	1.1752	-0.7697

Note: This table reports the results of the Diebold and Mariano test that this study used to compare the out-of-sample forecastability of the four forecasting methods. The Diebold and Mariano test compares the null of equal forecast accuracy. Positive signs in the Diebold and Mariano test statistics indicate that out-of-sample forecast errors from the first forecasting method are larger than those from the second forecasting method in the comparison. Results that are significant at the 10% level or lower are shown in boldface.

method in a real-time adjustment shows the best out-of-sample performance in the full sample analysis, the differences are not statistically significant. A similar tendency can be found for the first subsample period. However, the forecast using Bai and Perron's method in a real-time adjustment shows significantly poor out-of-sample forecastability compared with the forecast using Andrews's method or the simple sample mean during the second subsample period, which is the gap period between the break point and the time when the break point is first detected. The null hypothesis of equal forecast accuracy is rejected at the 1% level for forecasts based on Bai and Perron's procedure and Andrews's procedure, while the null hypothesis can be rejected at the 10% level for forecasts based on Bai and Perron's procedure and the sample mean.

During the third subsample period, the forecast based on Bai and Perron's method shows the best out-of-sample performance. Its out-of-sample performance is significantly better than the forecast based on the simple random walk model at the 10% level. Since the forecast based on Bai and Perron's method is better than other forecasts even during the first (two or three) year(s) immediately after the break is first detected, the imprecise estimate of the new regime mean is not the reason for the overall poor out-of-sample performance of the forecast based on Bai and Perron's method.

CONCLUSION

Although the mean-shift adjusted dividend-price ratio is an attractive option for reconciling the conflicting pieces of evidence on stock-return predictability, its usefulness in real-time stock return forecasting is reported to be limited by Lettau and Van Nieuwerburgh. However, Lettau and Van Nieuwerburgh's interpretation that the poor out-of-sample performance of the adjusted dividend-price ratio results from the difficulty in estimating a new regime mean with few observations is flawed. This study shows that the poor out-of-sample performance of the adjusted dividend-price ratio arises mainly from the gap period between the occurrence of a break and its detection, which implies that the poor out-of-sample performance of the adjusted dividend-price ratio is due to the requirement in Bai and Perron's procedure that breaks must be away from the boundaries of the sample. Significant improvement in the out-of-sample performance of the adjusted dividend-price ratio during the gap period is shown with the use of Andrews's procedure in the real-time adjustment of the dividend-price ratio. The newly suggested procedure for the adjusted dividend-price ratio in this article has a better out-of-sample performance than the simple sample mean, although it is not significant.

A large number of previous studies show that financial markets are interrelated with political and economic systems.[7] Hence, the results of this study, although it examines stock-return predictability under structural breaks, have implications for policy makers. The 2008 global financial crisis made it clear that turbulences in financial markets have serious impacts on the real economies in the world. The ability to see when such crises in financial markets are occurring and to make precise forecasts during the crisis is crucial, and the algorithm with Andrews's test used in this study offers a small step in that direction.

REFERENCES

Andrews, D. 2003. End-of-sample instability tests. *Econometrica*, 71(6): 1661-1694.

Bai, J., & Perron, P. 1998. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1): 47-78.

Bai, J., & Perron, P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1): 1-22.

Diebold, F. X., & Mariano, R. S. 1995. Comparing predictive accuracy. *Journal of*

7. For example, Park and An (2015) show that election cycles also have an impact on the volatilities of stock markets in several advanced countries.

Business and Economics Statistics, 13(3): 253-262.
Goyal, A., & Welch, I. 2008. A comprehensive look at the empirical performance of the equity premium prediction. *Review of Financial Studies*, 21(4): 1455-1508.
Lettau, M., & Nieuwerburgh, S. V. 2008, Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4): 1607-1652.
Liu, J., Wu, S., & Zidek, J.V. 1997. On segmented multivariate regressions. *Statistica Sinica*, 7(2): 497-525.
Park, C., & An, J. 2015. Election cycles and stock market reaction: International evidence. *Korea Journal of Policy Studies*, 30(3): 23-40.
Perron, P. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6): 1361-1401.
Yao, Y.-C. 1988. Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters*, 6(3): 181-189.