Treasury Rates No Longer Predict Returns: A Reappraisal of Breen, Glosten and Jagannathan (1989)

Philip Gray*
Department of Banking and Finance
Monash Business School
Monash University

Thanh Huynh
Department of Banking and Finance
Monash Business School
Monash University

v4: April 2019

Abstract

Breen, Glosten, and Jagannathan (1989) show that the negative relation between excess stock returns and Treasury bill rates is economically important. From 1954 to 1986, the predictive ability of interest rates facilitated a trading strategy that generated average returns at least on par with a buy-and-hold market investment but with significantly lower risk. The services of a portfolio manager using this predictive model to invest justified a management fee of nearly 2% per annum. Using currently-available data, we can nearly perfectly replicate Breen et al.'s (1989) key findings in sample. However, the success of Treasury bill rates as a predictor of equity returns appears to be specific to the time period studied. When the same methodology is applied out of sample from 1987 to 2018, there is little statistical or economic evidence of predictability. Additional out-of-sample analysis of G20 countries shows only sporadic support for the notion that interest rates predict equity returns.

JEL Classification: G11; G12; G14

Keywords: Return predictability; Treasury bill rates; trading strategy; out-of-sample forecasts

*Correspondence: Department of Banking and Finance, Monash Business School, Monash University, Caulfield East VIC 3145, Australia, philip.gray@monash.edu. The authors gratefully acknowledge the guidance of Ivo Welch (the editor) throughout this project.

A negative relation between excess stock returns and short-term interest rates is well documented, particularly in data through to the mid 1980s (Fama and Schwert, 1977; Campbell, 1987; Breen et al., 1989). The predictability literature predominantly assesses the importance of assorted predictors with reference to the statistical significance of coefficients in predictive regressions and/or R^2. Breen, Glosten, and Jagannathan (1989, hereafter BGJ) take a different tack by studying the performance of a trading strategy based on signals from a predictive regression. Using Treasury bill returns to forecast one-month-ahead excess market returns, funds are switched between equities and fixed income according to the predicted sign of the market risk premium. This strategy generates average returns at least on par with a buy-and-hold market investment, but with substantially lower variance. Further, a portfolio manager who invests according to the signals from the predictive model can justify a management fee of nearly 2% per annum. As such, BGJ conclude that the short rate's ability to predict market returns is economically significant.

This paper re-visits the economic significance of Treasury bill rates as a predictor of market returns. As a necessary starting point, we endeavor to replicate BGJ's key findings using the same data and empirical methodology over the same time period. Summary statistics for our data are nearly identical to those reported by BGJ. Safe in the knowledge that we are working with identical data, we can then ensure that our implementation of the empirical methodology is without flaw. We proceed to show that the main findings of BGJ can be nearly perfectly replicated. This is true for the predictive regressions, the performance of the portfolio switching strategy and the market timing tests. As such, the integrity of BGJ's original findings is beyond reproach.

The fact that we can so closely replicate BGJ's data and in-sample findings allows an extremely objective assessment of whether the key findings also manifest out of sample. Over 1987 to 2018, there is little statistical or economic evidence that Treasury bill rates predict excess market returns. Statistically, the negative slope in the predictive regression has a T-statistic of -0.26 (compared to T=-2.91 for the original 1954 to 1986 period). Formal market timing tests show no support for the notion that the short rate predicts either the magnitude or the sign of the market risk premium. In terms of economic significance, the return on the managed portfolio that switches between equities and fixed income based on signals from the predictive model significantly underperforms a buy-and-hold market investment. This is the case across the full out-of-sample period (1987 to 2018), within two subperiods and across the combined in- and out-of-sample horizon (1954 to 2018). Over the most recent 32 years, the management fee justified for the services of a portfolio manager who invests according to model predictions is now a small fraction of that during the in-sample period. In summary, although there is no reason to question the integrity of BGJ's results, our out-of-sample analysis strongly suggests that the original finding that the short rate is an economically significant predictor of equity returns simply does not hold for the last 32 years.

The current paper is the latest in a modest but emerging literature that examines whether evidence that stock market returns are predictable continues to apply stably in post-publication data. Goyal and Welch (2008) and Campbell and Thompson (2008) study the predictive ability of assorted variables, including interest rates. Given the large number of predictors studied, Goyal and Welch (2008) adopt a common metric to assess performance; namely, the out-of-sample R^2. The Treasury bill rate is one of the few predictors that demonstrates significant performance both in and out of sample through to 2005. However, Goyal and Welch (2008, p.1492) caution that these findings are heavily driven by data up to and including the oil shock years of 1973 to 1975. In contrast, Campbell and Thompson (2008, p.1510) are less disposed to judging predictability via out-of-sample statistics. While their focus is primarily on valuation ratios as predictors, they report a highly statistically significant in-sample association between Treasury bill returns and equity returns.

Our paper complements the work of Goyal and Welch (2008) and Campbell and Thompson (2008), but differs in a number of respects. By studying a single predictor (Treasury bill returns), we can conduct an in-sample replication and out-of-sample assessment of robustness using exactly the same performance metrics and market timing tests as BGJ. Further, like BGJ, our emphasis is on the economic significance of predictability. The passage of time also allows us to study an out-of-sample period through to December 2018, a non-trivial 32-year extension of BGJ's sample and an additional 13 years beyond the sample period of Goyal and Welch (2008) and Campbell and Thompson (2008). Importantly, the out-of-sample period from 1987 to 2018 can be partitioned into subperiods that capture distinct interest-rate regimes. BGJ find stronger predictability in their second subperiod, which they attribute to greater variability in Treasury bill returns. The fact that our out-of-sample period comprises regimes of low and high interest rate variability allows us to scrutinize this conjecture. Finally, we apply BGJ's methodology to assess evidence of predictability in major global markets included in the G7 and G20.

The paper proceeds as follows. Section 1 briefly reviews BGJ's empirical methodology and describes their data. We then describe our efforts to source the same data and replicate their key findings. Section 2 presents our study of whether BGJ's finding that Treasury bill returns are an economically important predictor of equity returns holds over an extended out-of-sample period, across disparate interest-rate regimes and within other major global markets. Section 3 draws our conclusions.

1 In-sample replication

BGJ adopt a very simple predictive model that utilizes information available at time t to forecast the excess market return over period $t+1$. Let r_{mt} and r_{ft} denote the month t return on the market index and riskfree asset, respectively. The predictive model is

$$x_{t+1} = \beta_0 + \beta_1 r_{ft} + \varepsilon_{t+1}, \tag{1}$$

where $x_t = r_{mt} - r_{ft}$ is the return on the market index in excess of the riskfree rate.[1] At the end of each month t, model (1) is estimated using data from the last 36 months (up to and including month t). Given the fitted parameters $(\hat{\beta}_0, \hat{\beta}_1)$ and the time-t riskfree rate, the one-month-ahead excess market return $(\hat{x}_{t+1})$ is forecast. BGJ's 'managed portfolio' is essentially a trading strategy whereby funds are fully invested in either the market portfolio if $\hat{x}_{t+1} > 0$ or Treasury bills if $\hat{x}_{t+1} \leq 0$. This process is repeated at the end

[1] Our predictive model (1) is presented precisely as shown in BGJ eqn (1). However, successful replication of their empirical results requires careful implementation of this regression. We re-visit this important issue shortly.

of each month, using a rolling 36-month window to fit predictive model (1) and generate a series of trading signals based on the forecasted excess market return.

The data requirements for this study are minimal. The riskfree asset is proxied by the monthly return on Treasury bills, sourced from Ken French's data library. Regarding the risky asset, BGJ study both value- and equal-weighted market indexes. To replicate their results, it is imperative to note that BGJ utilize indexes of NYSE-only stocks, as opposed to more common indexes based on the NYSE/NASDAQ/AMEX population. We source NYSE-only market indexes from the Center for Research in Security Prices.[2]

BGJ study the 393-month period spanning April 1954 through December 1986. They further partition this sample into two subperiods: (i) from April 1954 through July 1970 (196 months) and (ii) from August 1970 through December 1986 (197 months). We adopt the same partitions in presenting our in-sample replication. Our out-of-sample analysis spans January 1987 through December 2018 (384 months). As such, our tables contain three distinct panels. The first displays results copied verbatim from BGJ; the second panel represents our attempt to replicate BGJ's findings over the identical in-sample period; and the third panel, which we discuss in Section 2, contains new results for our out-of-sample period.

Table 1 reports summary statistics for monthly returns on the value-weighted NYSE market index (r_{mt}), the one-month Treasury bill (r_{ft}), and excess market returns (x_t).[3] Comparing the BGJ and Replication panels, it is readily apparent that we have nearly perfectly replicated the original summary statistics. The summary statistics for monthly returns on the riskfree asset also match precisely. The only observable differences in Table 1 are negligible discrepancies in estimates of first-order autocorrelation.[4]

BGJ's study is premised on the empirical observation that future excess stock returns are negatively correlated with the current short-term interest rate. Although their trading strategy is based on forecasts from a series of rolling regressions using the past 36 months, BGJ verify the existence of the negative relationship in their sample. Model (1) is estimated for the entire in-sample period and two subperiods, with standard errors estimated after correcting for heteroscedasticity and autocorrelation (20 lags in the full sample and 14 lags in each subperiod). Table 2 summarizes BGJ's findings and presents our attempt to replicate them.

Comparing the BGJ and Replication panels, our estimates of the model parameters (β_0, β_1), T-statistics, and R^2 are extremely close but not identical. Accordingly, the strong statistically significant negative relation between stock returns and the short-term interest rate reported by BGJ is unambiguously corroborated in our replication. BGJ also document conditional heteroscedasticity by regressing the squared residuals from model (1) on the short rate. Although this is not of first-order importance in our assessment of the economic significance of the trading strategy, Table 2 Panel B demonstrates that we can very closely replicate the existence of conditional heteroscedasticity over BGJ's sample period.

[2] In WRDS: library `crspa`, file `msi`, market `nyse`, variables `vwretd` and `ewretd`.

[3] To keep the exposition manageable, we only tabulate findings in relation to the value-weighted market index. BGJ's key takeaway of economically significant predictability was confined to the value-weighted market. Nevertheless, for completeness, we also tabulate the results of all analysis for the equal-weighted index in an accompanying appendix.

[4] With respect to $\rho(1)$ for stock returns in the second subperiod, BGJ report autocorrelation of 0.07 whereas our estimate is -0.007. Given the accuracy with which we can replicate all other summary statistics, BGJ's 0.07 may be a typo.

Before proceeding to study the economic significance of trading using forecasts from rolling predictive regressions, it is imperative to clearly articulate how BGJ implement their model. The time subscripts in BGJ's predictive model (copied verbatim in our model (1)) give the impression that the month-t riskfree return is used to forecast the excess market return in month $t + 1$. However, proceeding with this interpretation, we could not satisfactorily replicate the in-sample regression results of Table 2 or the plethora of market timing tests that follow shortly. Our preliminary results were qualitatively similar to BGJ and their key takeaways remained intact; however, the discrepancies were disconcerting given that Table 1 shows we are using identical data.

The resolution lies in paying more attention to BGJ's words and less attention to their notation. BGJ (1989, p.1179) define r_{ft} as the riskfree return in month t and their predictive model depicts r_{ft} as forecasting x_{t+1}. However, the wording surrounding their model (1) clearly (and correctly) states that x_{t+1} is forecast using the month $t + 1$ Treasury bill return which is known at the end of month t. When we re-estimate model (1) by regressing x_{t+1} on r_{ft+1}, our results in Table 2 nearly perfectly match BGJ. We stress that regressing x_{t+1} on r_{ft+1} does *not* invalidate the predictive aspect of BGJ's model or the trading-strategy feature of their paper, because r_{ft+1} is known at the end of month t. However, because this issue proved to be the major stumbling block in our preliminary efforts to replicate BGJ, we believe it is worth clarifying.

We now turn to the main focus of BGJ; namely, whether the negative relation between the short-term interest rate and future market returns is economically significant. The first prediction is for April 1954. Model (1) is fit using 36 months of data from April 1951 through March 1954. The parameter estimates and the riskfree return for April 1954 (known at the end of March 1954) are used to forecast the excess market return in April 1954. The managed portfolio is assumed to be fully invested in either equities if the one-month-ahead forecast is positive, or Treasury bills if the forecast is negative. This procedure is repeated each month using rolling 36-month estimation windows. Table 3 reports summary statistics for the returns on the managed portfolio and success of the trading strategy.

Over the full in-sample period (393 months), BGJ's model forecasts 122 down markets (271 up markets), 62 (165) of which prove to be correct predictions. Our replication forecasts 123 down markets (270 up markets), 62 (164) of which are correct. As such, the difference amounts to a single month in which we forecast a down market when BGJ forecast an up market. Given that the BGJ and Replication panels in Table 3 both show 62 correctly forecasted down markets, the extra down prediction from our model proved incorrect. This single difference flows through to a marginally lower average return for our managed portfolio (0.52) compared to BGJ (0.55). Closer examination of the subperiod results reveals that the one extra down prediction of our model occurs in the second subperiod. During the first subperiod, we have precisely replicated the findings of BGJ, except for a difference in the mean return to the managed portfolio of 1 basis point per month.

The findings in Table 3 are the headline result of BGJ. The signals generated by the predictive model facilitate the switching of funds between equities and fixed income such that the managed portfolio (0.55% per month) marginally outperforms a buy-and-hold market investment (0.53% per month from Table 1), but with only about 60% of the variance of the latter $((3.24/4.16)^2)$. It is worth noting that the lower variance of returns on the managed portfolio is unsurprising. Compared to a buy-and-hold investment that is always fully invested in the market, any switching strategy that spends any time in fixed income must have lower variance.[5] The value of the predictive model, therefore, lies in generating enough correct predictions to allow the managed portfolio to at least match the average return on the buy-and-hold investment.

BGJ augment their main finding with a series of formal market timing tests. Following Cumby and Modest (1987), they test whether the excess market return is higher when the predictive model forecasts an up market

$$x_{t+1} = a_0 + a_1 I_t + \nu_{t+1}, \tag{2}$$

where x_{t+1} is the excess market return in month $t+1$ and I_t is unity (zero) if the predictive model forecasts an up (down) market in month $t+1$. Table 4 Panel A shows that our replication corroborates the main takeaway of BGJ, especially in relation to the key parameter a_1.[6]

Extending the Cumby and Modest (1987) approach to the second moment of the return distribution, the slope (b_1) in the following regression tests whether the predictive model is useful in forecasting the variance of excess returns

$$\nu_{t+1}^2 = b_0 + b_1 I_t + \eta_{t+1}. \tag{3}$$

Table 4 Panel B again documents close replication of the original in-sample findings. As a variation of the first market timing test, BGJ consider whether signals from the predictive model are related to the probability of an up market

$$y_{t+1} = c_0 + c_1 I_t + \omega_{t+1}, \tag{4}$$

where $y_{t+1} = 1$ when $x_{t+1} > 0$ and zero otherwise. A comparison of the BGJ and Replication estimates in Table 4 Panel C shows near-perfect replication.

Finally, as an alternate metric of the economic significance of their predictive model, BGJ estimate the "value at margin" following Merton (1981) and Henriksson and Merton (1981). In brief, value at margin is the management fee that a naïve investor is prepared to pay to have their funds managed according to the switching strategy that follows the signals from a predictive model. The key parameter (α_1) is estimated by

$$I_t = \alpha_0 + \alpha_1 y_{t+1} + u_{t+1}. \tag{5}$$

The value of the predictive model is equal to α_1 one-period call options on the market index with a current value of one dollar and a strike price equal to one plus the riskfree rate. Table 4 Panel D shows that we are able to nearly perfectly replicate BGJ's estimates. Over the full in-sample period, $\alpha_1 = 0.10$, which equates

[5] This follows from the fact that the variation in riskfree returns is an order of magnitude lower than the variation in equities.
[6] In assessing the market timing tests, one must bear in mind that our predictive model generates one different signal to BGJ over the 393-month in-sample period. Accordingly, we will not be able to perfectly replicate their findings.

to value at margin of 1.97% per annum.[7]

Taken together, the empirical analysis in this section demonstrates a very high level of integrity in BGJ's original findings. Using their methodology and currently available data sources, we are able to very closely replicate all the key findings. In summary, when estimated across the entire in-sample period, there is a statistically significant negative relation between the short-term interest rate and one-month-ahead excess market returns. Compared to a buy-and-hold market investment, an investment strategy that switches funds between equities and fixed income according to signals from a predictive regression using rolling 36-month estimation windows yields similar mean returns but at substantially lower risk. In reference to Henriksson and Merton's (1981) value at margin, a naïve investor would be prepared to pay fees in the vicinity of 2% per annum to have her funds managed according to this switching strategy.

The formal tests of market timing ability are less convincing. The null that excess market returns are higher following a model prediction of an up market is rejected only at the 10% significance level and, even then, only across the entire in-sample period (i.e., the result does not manifest in either subperiod). Although the model's signal has significant predictive ability for the sign of future market returns, this finding appears to be driven by BGJ's second subperiod.

2 Out-of-sample extension and global analysis

Having nearly perfectly replicated the original BGJ findings, we can now make an objective assessment of whether the key findings hold out of sample. For this purpose, we have an additional 32 years over which to study a number of issues raised by BGJ regarding both disparate interest-rate regimes and the necessary time horizons for precise parameter estimation.

For perspective, Figure 1 plots the annualized Treasury bill rates from 1954 to 2018. The first two panels depict BGJ's subperiods, the first spanning April 1954 to July 1970 (196 months) and the second spanning August 1970 to December 1986 (197 months). We partition the 1987 to 2018 out-of-sample period into two unequal subperiods divided by the onset of the global financial crisis (GFC), the first spanning January 1987 through July 2007 (247 months) and the second spanning August 2007 through December 2018 (137 months).[8] The 'Update' panel within each table reports findings using BGJ's empirical methodology for: (i) the full out-of-sample period from 1987 to 2018, (ii) the two out-of-sample subperiods shown in Figure 1, and (iii) the combined in- and out-of-sample period spanning 1954 to 2018.

Table 1 shows that the first out-of-sample subperiod is qualitatively similar to BGJ's sample, with monthly

[7]The Black and Scholes (1973) model prices a call option on the NYSE-only index at \$0.0164, with $d_1 = 0.0205$ and $d_2 = -0.0205$ (a spreadsheet demonstration of this calculation is included in the accompanying ZIP file). On an annual basis, therefore, the value at margin is $0.10 \times 1.64 \times 12 = 1.97\%$. This calculation of value at margin is also detailed by BGJ (1989, p.1186).

[8]With the benefit of hindsight, events that are now regarded as contributing to the GFC occurred progressively throughout 2007 and 2008. Noting that growing concerns over liquidity and confidence prompted the Federal Open Market Committe to begin reducing the Federal funds rate in August 2007, we partition the out-of-sample period in July 2007.

value-weighted market returns having a mean and standard deviation of 1.02% and 3.96%, respectively. The latter subperiod, however, is characterized by relatively low mean returns of 0.55% and a marginally higher standard deviation (4.43%).[9] More importantly, the subperiods capture distinct interest-rate regimes. In particular, Treasury bill returns in the latter subperiod have an extremely low mean (0.05%) and standard deviation (0.08%).

The Update panel in Table 2 allows an assessment of whether the negative relation between nominal interest rates and excess market returns is robust over an extended period. BGJ document a highly significant negative relation from 1954 to 1986. However, this finding is heavily driven by their earlier subperiod. The magnitude of their estimated regression slope (β_1) declines from -8.05 to -2.65 across their subperiods. Further, over our updated sample, there is no evidence of any relation from 1987 to 2018 ($\beta_1 = -0.29, \mathrm{T} = -0.26$). Excess returns are unrelated to the short rate over the first 247-month subperiod. The subperiod from 2007 to 2018 does exhibit a non-trivial negative slope ($\beta_1 = -9.43, \mathrm{T} = -2.12$). Arguably, this is unsurprising for a period during which interest rates were at unprecedented lows while the market enjoyed a nearly unbroken bull run. However, despite the apparent strength of the negative relation during this period, we will shortly see that it does not facilitate profitable market timing. Combining the in- and out-of-sample periods, the overall relation across 1954 to 2018 is modest but significant ($\beta_1 = -1.39, \mathrm{T} = 2.20$).

BGJ's headline result is that their predictive model guides a switching strategy that generates average returns (at least) on par with a buy-and-hold market investment, but with substantially lower risk. The Update Panel of Table 3 shows that the success of the switching strategy is confined to BGJ's sample period. From 1987 to 2018, the average excess return on the managed portfolio (0.46%) significantly underperforms a buy-and-hold market investment (0.60% from Table 1). This is also the case in each out-of-sample subperiod. In fact, combining in- and out-of-sample periods, the switching strategy (0.49%) underperforms the market investment (0.56%). Therefore, despite sporadic episodes of apparently strong relations between stock-market returns and nominal interest rates over the past 64 years, Table 3 suggests that the degree of predictability does not consistently enable successful market timing.

The inferences from formal market timing tests were underwhelming in-sample and do not improve out of sample (see Table 4). With respect to Cumby and Modest's (1987) test of whether excess market returns are higher when the predictive model forecasts an up market, BGJ (1989, p.1184) conjecture that their 'inability to find the α_1's in model (2) to be statistically significantly different from zero may be due in part to the fact that stock returns are extremely volatile and rather long time periods of observations are needed for precise estimates.' Table 4 Panel A (last column) demonstrates that, even with a 64-year horizon, the T-statistic on α_1 barely reaches unity. Similarly, Panel C suggests that the signals from the predictive model have no out-of-sample ability to forecast the sign of the one-month-ahead market premium.

The fact that the out-of-sample period comprises two distinct interest-rate regimes also allows us to re-visit the notion that equity returns are more predictable when interest rates are volatile. BGJ (1989, p.1186)

[9] Despite an extended bull market over much of the post-GFC period, the mean return is influenced downward by severe market corrections at both ends of the latter subperiod.

motivate the partition of their sample into subperiods with reference to the observation that interest rates were more volatile in their second subperiod: 'since there are theoretical reasons for there to be more variation in stock index expected returns during periods of greater variation in expected inflation, we may expect the forecasting model to perform better in the second half of the sample period.' This prior is supported by the fact that all (in sample) regression slopes reported in Table 4 have greater magnitude in BGJ's second subperiod.

In our out-of-sample period, while the first subperiod exhibits far greater variation in Treasury bill returns than the second subperiod (see Table 1), we do not observe any reliable evidence of stronger predictability in the first subperiod (or *any* predictability for that matter). Further, note that the standard deviation of Treasury bill returns during BGJ's second subperiod (0.24) is not dissimilar from the volatility of riskfree returns over our full out-of-sample period (0.21). Again, this high volatility in the short rate does not translate into meaningful predictability from 1987 to 2018. Accordingly, we conjecture that the heightened predictability during BGJ's second subperiod could have been time specific.

With respect to Henriksson and Merton's (1981) value at margin, Table 4 Panel D shows that the key parameter (α_1) is statistically indistinguishable from zero. Over the full out-of-sample period, the services of a manager who switches between equities and fixed income according to the predictive model are worth 40 basis points per annum, a small fraction of that reported by BGJ during the in-sample period. In the entire analysis, the only encouraging predictive aspect is the relation between the short rate and the future variance of excess stock returns (Table 4 Panel B). Across long time horizons, the predictive model appears to have reasonable ability to forecast stock market volatility.

As noted, BGJ's conclusion that Treasury bill rates predict market returns is confined to the value-weighted NYSE index. In contrast, their model's ability to forecast the equal-weighted NYSE index is "miserable" (BGJ, 1989, p.1187). In our accompanying appendix, we again very closely replicate BGJ's in-sample findings for the equal-weighted NYSE market index. However, the misery associated with any attempt to forecast equally weighted market returns continues over our extended 32-year out-of-sample period. Despite some statistical support for a negative relation between the short rate and the equal-weighted market index (Table A2), this finding does not translate into market timing ability or superior trading returns.

To complete our robustness analysis, we examine the predictable variation in equity returns for a number of non-US markets; namely, the countries included in the G7 and G20. The empirical methodology of BGJ is applied to the longest sample available for each country (full details are provided in the accompanying appendix). Inevitably, some metrics in some countries are consistent with the existence of predictability. In South Africa, for example, the managed portfolio significantly outperforms a buy-and-hold investment from 1986 to 2019. There is a modest negative relation between the short rate and market returns ($\beta_1 = -1.83, \mathrm{T} = -1.67$), and the predictive model has significant ability to forecast the magnitude of future excess returns ($a_1 = 1.61, \mathrm{T} = 2.41$), leading to value at margin of 2.16% per annum. However, even a quick glance at Table A6 does not provide confidence of consistent support for predictability across different metrics and countries. Using a 5% level of significance, β_1 from model (1) is significantly negative in only two of 18 countries examined (Argentina and India). Similarly, significant market timing ability (a_1 from model (2)) is present for a single country (South Africa). At the very least, our analysis of major global markets suggests a healthy dose of skepticism is warranted before inferring predictable variations in equity returns.

3 Conclusion

Whether stock returns are predictable is an issue that continues to attract debate. Presently, there is no consensus on which metric is best suited to judge the evidence of predictability and the conclusions often differ according to the chosen model diagnostic. One way this literature can advance is through careful replication of important prior findings, to both verify the integrity of the original findings and to explore whether similar findings manifest in alternate data sets and over different time horizons. This paper re-examines the finding of BGJ that Treasury bill returns are an economically significant predictor of the one-month-ahead excess stock market return. Using currently available data sources, we are able to nearly perfectly replicate all of BGJ's key findings over their horizon from 1954 to 1986. The integrity of their findings is beyond reproach.

To assess whether similar evidence of predictability persists beyond BGJ's original horizon, we draw on a 32-year period from 1987 to 2018. Given BGJ's conjecture that predictability is likely to be stronger when interest rates are more volatile, we partition the out-of-sample period into distinct regimes that exhibit low and high variation in Treasury bill returns. Using precisely the same empirical methodology as BGJ, we find little evidence that Treasury bill rates predict equity returns over the out-of-sample period. Although numerous studies have documented a negative association between interest rates and equity returns, our regression slope is statistically indistinguishable from zero between 1987 to 2018. Combining the in- and out-of-sample periods, we find a modest negative relation that is significant at the 5% level. However, this does not translate into economically meaningful returns to a trading strategy based on the predictive regression. Regarding economic significance, the headline results of BGJ do not persist in more recent data. The returns to a managed portfolio that trades on signals from the predictive model significantly underperform a buy-and-hold investment in the market index. This is the case across the full out-of-sample period (1987 to 2018), within two recent subperiods that exhibit disparate interest-rate regimes and even across the combined in- and out-of-sample periods (1954 to 2018). Over the last 32 years, value at margin is a small fraction of that reported during the in-sample period. Taken together, our study strongly suggests that the apparent economic significance of the predictability documented by BGJ is specific to the particular period studied.

References

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–654.

Breen, W., L. R. Glosten, and R. Jagannathan. 1989. Economic significance of predictable variations in stock index returns. *Journal of Financial Economics* 44:205–258.

Campbell, J. Y. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18:373–399.

Campbell, J. Y., and S. B. Thompson. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:1509–1531.

Cumby, E., and D. Modest. 1987. Testing for market timing ability: A unified framework for evaluation. *Journal of Financial Economics* 19:169–189.

Fama, E. F., and G. Schwert. 1977. Asset returns and inflation. *Journal of Financial Economics* 5:115–146.

Goyal, A., and I. Welch. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21:1455–1508.

Henriksson, R. D., and R. C. Merton. 1981. On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. *Journal of Business* 54:513–533.

Merton, R. C. 1981. On market timing and investment performance I: An equilibrium theory of value for market forecasts. *Journal of Business* 54:363–403.

Table 1: Summary Statistics

Description: This table reports summary statistics for monthly returns to the CRSP value-weighted NYSE-only market index, the one-month Treasury bill and the market return in excess of the riskfree rate. The full period spans April 1954 through December 2018, with various subperiods as indicated. The time-series mean and standard deviation of monthly returns is reported, along with the first-order autocorrelation in returns. The BGJ panel displays results copied verbatim from Breen et al. (1989). The Replication panel reflects our attempt to replicate the BGJ findings. The Update panel reports summary statistics for the out-of-sample period.

Interpretation: Comparing the BGJ and Replication panels demonstrates that we have sourced nearly identical data to the original study over the in-sample period. The Update panel shows that the out-of-sample period exhibits qualitatively similar characteristics to the in-sample period, whilst the subperiod breakdown suggests the existence of two distinct interest-rate regimes over which to study predictability.

	BGJ			Replication			Update			
	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	87:1 to 18:12	87:1 to 07:7	07:8 to 18:12	54:4 to 18:12
	$N = 393$	$N = 196$	$N = 197$	$N = 393$	$N = 196$	$N = 197$	$N = 384$	$N = 247$	$N = 137$	$N = 777$
					Monthly Stock Return (%)					
Mean	0.98	0.89	1.07	0.98	0.89	1.07	0.85	1.02	0.55	0.92
Std dev	4.13	3.70	4.53	4.13	3.70	4.53	4.14	3.96	4.43	4.13
$\rho(1)$	0.03	0.08	0.07	0.03	0.08	-0.007	0.10	0.07	0.15	0.06
					Monthly Nominal Interest Rate (%)					
Mean	0.45	0.28	0.63	0.45	0.28	0.63	0.25	0.37	0.05	0.35
Std dev	0.26	0.13	0.24	0.26	0.13	0.24	0.21	0.16	0.08	0.26
$\rho(1)$	0.95	0.93	0.92	0.96	0.94	0.92	0.98	0.95	0.90	0.97
					Monthly Excess Stock Return (%)					
Mean	0.53	0.61	0.44	0.53	0.61	0.44	0.60	0.65	0.50	0.56
Std dev	4.16	3.74	4.55	4.16	3.74	4.56	4.13	3.96	4.45	4.15
Kurtosis	0.77	0.17	0.90	0.76	0.15	0.89	3.70	4.89	2.30	2.16
$\rho(1)$	0.05	0.10	0.03	0.04	0.10	0.00	0.10	0.07	0.15	0.07

Table 2: Regression Estimates

Description: Panel A reports estimates of model (1) which regresses the excess return on the CRSP value-weighted index of NYSE-only stocks on the Treasury bill rate. Panel B reports estimates where the squared error term from model (1) are regressed on the Treasury bill return. T-statistics that correct for heteroscedasticity and autocorrelation are reported in parentheses. Heteroscedasticity-corrected T-statistics are reported in square brackets. The BGJ panel displays results copied verbatim from Breen et al. (1989). The Replication panel reflects our attempt to replicate the BGJ estimates. The Update panel reports new regression estimates for the out-of-sample period.

Interpretation: A comparison of the BGJ and Replication panels demonstrates that the original regression estimates can be nearly perfectly replicated. The Update panel shows that the negative relation between Treasury bill returns and future excess market returns has not continued stably over the 32-year period following the original findings.

	BGJ			Replication			Update			
	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	87:1 to 18:12	87:1 to 07:7	07:8 to 18:12	54:4 to 18:12
	Panel A: $x_t = \beta_0 + \beta_1 r_{ft} + \varepsilon_t$									
β_0	1.60	2.85	2.10	1.59	2.75	2.16	0.67	0.71	0.93	1.05
	(4.28)	(4.60)	(2.60)	(4.22)	(4.42)	(2.64)	(1.95)	(1.18)	(2.23)	(4.20)
β_1	−2.37	−8.05	−2.65	−2.35	−7.73	−2.74	−0.29	−0.16	−9.43	−1.39
	(−2.91)	(−3.56)	(−2.11)	(−2.87)	(−3.41)	(−2.16)	(−0.26)	(−0.10)	(−2.12)	(−2.20)
R^2	0.02	0.08	0.02	0.02	0.07	0.02	0.00	0.00	0.02	0.01
	Panel B: $\varepsilon_t^2 = \gamma_0 + \gamma_1 r_{ft} + \xi_t$									
γ_0	9.15	5.39	13.77	9.32	5.79	13.77	16.16	6.18	18.58	13.52
	(3.36)	(1.51)	(1.93)	(3.30)	(1.53)	(2.12)	(2.98)	(0.87)	(3.20)	(3.91)
	[4.20]	[1.74]	[2.56]	[4.24]	[1.86]	[2.53]	[5.19]	[1.07]	[5.83]	[6.88]
γ_1	17.11	26.58	10.28	16.81	25.65	10.31	3.46	25.38	10.51	9.93
	(2.86)	(2.41)	(1.18)	(2.58)	(2.14)	(1.17)	(0.22)	(1.17)	(0.43)	(1.38)
	[3.26]	[2.37]	[1.28]	[3.17]	[2.26]	[1.27]	[0.27]	[1.26]	[0.42]	[1.93]
R^2	0.02	0.03	0.00	0.02	0.03	0.00	0.00	0.01	0.00	0.00

Table 3: Summary Statistics for Managed Portfolio

Description: This table reports summary statistics for monthly returns to the managed portfolio. At the end of each month, model (1) is fit using the a rolling window of 36 months prior data. The parameter estimates and the current riskfree rate generate a forecast of the one-month-ahead excess market return. The managed fund is a trading strategy that fully invests funds in either the market portfolio (if the forecast is positive) or Treasury bills (if the forecast is negative). The time-series mean and standard deviation of monthly returns is reported, along with the first-order autocorrelation in returns. The BGJ panel displays results copied verbatim from Breen et al. (1989). The Replication panel reflects our attempt to replicate the BGJ findings. The Update panel reports summary statistics for the out-of-sample period.

Interpretation: A comparison of the BGJ and Replication panels demonstrates that the original in-sample performance of the managed portfolio can be nearly perfectly replicated. Across 393 month, our replication differs for a single month. Comparing the summary statistics in the Update panel to those in Table 1 for a buy-and-hold market investment suggests that the predictability is not economically significant out of sample.

	BGJ			Replication			Update			
	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	87:1 to 18:12	87:1 to 07:7	07:8 to 18:12	54:1 to 18:12
	$N = 393$	$N = 196$	$N = 197$	$N = 393$	$N = 196$	$N = 197$	$N = 384$	$N = 247$	$N = 137$	$N = 777$
Monthly Excess Return on the Managed Portfolio (%)										
Mean	0.55	0.59	0.51	0.52	0.58	0.46	0.46	0.47	0.43	0.49
Std dev	3.24	2.70	3.70	3.20	2.71	3.63	3.35	3.49	3.09	3.27
$\rho(1)$	0.07	0.19	0.01	0.08	0.19	0.02	0.05	0.09	-0.03	0.06
Forecasted Down Markets										
N	122	75	47	123	75	48	84	52	32	207
Correct forecasts	62	33	29	62	33	29	33	20	13	95
Forecasted Up Markets										
N	271	121	150	270	121	149	300	195	105	570
Correct forecasts	165	81	84	164	81	83	191	123	68	355

Table 4: Market Timing Tests

Description: This table reports assorted tests of the market timing ability of the predictive model (1). Panel A and Panel B are Cumby and Modest (1987)-style tests of whether the model is useful in predicting the first and second moments of the return distribution respectively. Panel C assesses whether the model predicts the one-month ahead sign of the market return. Panel D relates to estimating Henriksson and Merton's (1981) value at margin. The BGJ panel displays results copied verbatim from Breen et al. (1989). The Replication panel reflects our attempt to replicate the BGJ findings. The Update panel reports our estimates for the out-of-sample period.

Interpretation: A comparison of the BGJ and Replication panels demonstrates that our in-sample replication of market timing tests are extremely close but not identical to the original estimates. The Update panel suggests that the predictive model has no market timing ability over the out-of sample period, other than to forecast the variance of excess market returns.

	BGJ			Replication			Update			
	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	54:4 to 86:12	54:4 to 70:7	70:8 to 86:12	87:1 to 18:12	87:1 to 07:7	07:8 to 18:12	54:4 to 18:12
					Panel A: $x_{t+1} = a_0 + a_1 I_t + \nu_{t+1}$					
a_0	−0.08	0.06	−0.30	0.02	0.08	−0.08	0.65	0.86	0.30	0.27
	(−0.21)	(0.13)	(−0.39)	(0.05)	(0.18)	(−0.11)	(0.99)	(1.50)	(0.19)	(0.79)
a_1	0.87	0.88	0.98	0.74	0.87	0.69	−0.07	−0.27	0.25	0.39
	(1.80)	(1.55)	(1.14)	(1.76)	(1.37)	(0.97)	(−0.10)	(−0.42)	(0.16)	(1.04)
					Panel B: $\nu_{t+1}^2 = b_0 + b_1 I_t + \eta_{t+1}$					
b_0	21.91	17.36	29.11	22.61	17.34	30.84	26.95	16.62	43.55	24.47
	(4.98)	(5.81)	(3.11)	(4.88)	(5.95)	(3.07)	(3.40)	(4.40)	(3.37)	(5.77)
	[6.91]	[7.15]	[3.99]	[7.07]	[7.19]	[4.29]	[5.14]	[5.28]	[3.69]	[8.78]
b_1	−6.96	−5.92	−11.36	−7.92	−5.86	−13.58	−12.69	−1.32	−31.21	−9.99
	(−2.01)	(−1.68)	(−1.31)	(−1.72)	(−1.55)	(−1.29)	(−1.52)	(−0.26)	(−2.28)	(−2.25)
	[−1.98]	[−1.99]	[−1.48]	[−2.23]	[−1.98]	[−1.80]	[−2.24]	[−0.30]	[−2.60]	[−3.22]
					Panel C: $y_{t+1} = c_0 + c_1 I_t + \omega_{t+1}$					
c_0	0.49	0.56	0.38	0.50	0.56	0.40	0.61	0.61	0.59	0.54
	(10.87)	(9.77)	(5.40)	(9.91)	(10.66)	(5.54)	(11.53)	(9.70)	(5.66)	(13.99)
c_1	0.12	0.11	0.18	0.11	0.11	0.16	0.03	0.02	0.05	0.08
	(2.16)	(1.53)	(2.13)	(1.91)	(1.37)	(2.15)	(0.52)	(0.22)	(0.47)	(1.86)
					Panel D: $I_{t-1} = \alpha_0 + \alpha_1 y_t + \eta_t$					
α_0	0.63	0.55	0.69	0.63	0.55	0.69	0.77	0.78	0.74	0.69
	(16.95)	(9.40)	(14.70)	(8.17)	(5.92)	(6.57)	(10.44)	(9.89)	(5.51)	(12.62)
				[16.95]	[9.41]	[14.70]	[21.66	[18.20]	[11.93]	[26.49]
α_1	0.10	0.11	0.13	0.10	0.11	0.12	0.02	0.01	0.04	0.07
	(2.16)	(1.53)	(2.13)	(1.78)	(1.34)	(1.67)	(0.51)	(0.22)	(0.46)	(1.78)
				[2.06]	[1.53]	[1.95]	[0.49]	[0.20]	[0.55]	[2.03]
VaM	1.97%	1.95%	2.79%	1.97%	1.95%	2.79%	0.40%	0.19%	0.86%	1.39%

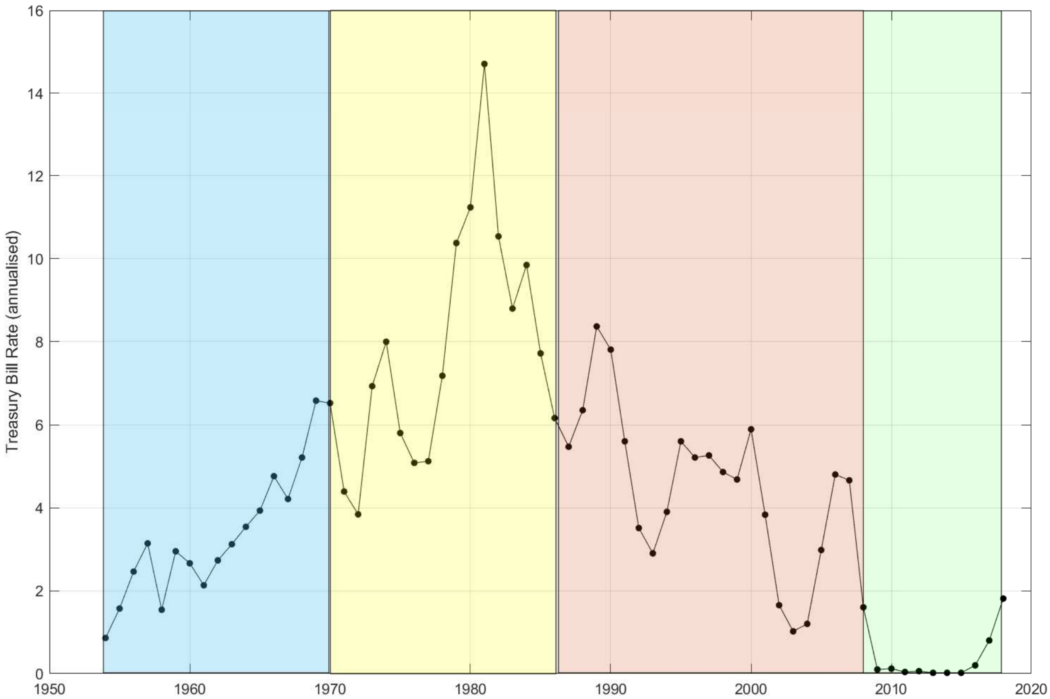

Figure 1: Treasury Bill Rates

Decsription: This figure plots annualized Treasury Bill rates from 1954 to 2018 (source: Ken French data library). The first two panels depict BGJ's two subperiods (April 1954 through July 1970; August 1970 through December 1986). The final two panels divide the the 32-year out-of-sample period into two subperiods loosely divided by the onset of the GFC (January 1987 through July 2007; August 2007 through December 2018).

Interpretation: The in- and out-of-sample periods can each be partitioned into distinct regimes of low and high volatility in Treasury bill returns.