Estimating Long-Term Expected Returns

Rui Ma
La Trobe Business School
La Trobe University
m.ma@latrobe.edu.au

Ben R. Marshall*
Massey Business School
Massey University
b.marshall@massey.ac.nz

Nhut H. Nguyen
Department of Finance
Auckland University of Technology
nhut.nguyen@aut.ac.nz

Nuttawat Visaltanachoti
Massey Business School
Massey University
n.visaltanachoti@massey.ac.nz

Abstract

Estimating long-term expected returns as accurately as possible is of critical importance. Researchers typically base their estimates on yield and growth, valuation, or a combined yield, growth, and valuation framework. We run a horse race of the abilities of different frameworks and input proxies within each framework to estimate 10- and 20-year out-of-sample returns over 140-year and more recent time periods. Our results indicate that several approaches strongly outperform estimates based on historical mean benchmark returns, with mean square error improvements exceeding 40%. Using these approaches in asset allocation decisions results in an improvement in Sharpe ratios of more than 50%.

JEL Classification Codes: G11, G12
Keywords: Long-Term Expected Returns, Asset Allocation

First Version: 19 June 2023
This Version: 28 June 2023

Acknowledgments: We thank participants at the Massey University and La Trobe University seminar series. All errors are our own. Corresponding author: Ben Marshall, School of Economics and Finance, Massey Business School, Massey University, Private Bag 11-222, Palmerston North, New Zealand. Ph: 646 951 7033, E-mail: B.Marshall@Massey.ac.nz

Estimating Long-Term Expected Returns

Abstract

Estimating long-term expected returns as accurately as possible is of critical importance. Researchers typically base their estimates on yield and growth, valuation, or a combined yield, growth, and valuation framework. We run a horse race of the abilities of different frameworks and input proxies within each framework to estimate 10- and 20-year out-of-sample returns over 140-year and more recent time periods. Our results indicate that several approaches strongly outperform estimates based on historical mean benchmark returns, with mean square error improvements exceeding 40%. Using these approaches in asset allocation decisions results in an improvement in Sharpe ratios of more than 50%.

JEL Classification Codes: G11, G12
Keywords: Long-Term Expected Returns, Asset Allocation

First Version: 19 June 2023
This Version: 28 June 2023

1. Introduction

The expected return on the equity market—E(R)—over multi-year horizons is one of the most important variables in finance. Small changes in the E(R) can have material impacts on factors ranging from company investment decisions to the prices consumers pay for the services of regulated monopolies, to estimates of the amount that individuals need to save to reach their retirement goals. Fama and French (1988) and Campbell and Shiller (1988) made important early contributions to this literature. More recently, Golez and Koudijs (2018) document long-term predictability in a range of markets and time periods and Atanasov, Møller, and Priestley (2020) introduce consumption variation as a long-term return predictor. However, despite these studies, much less is known about long-term return predictability than the predictability over shorter horizons.[1]

We run a horse race of the various frameworks and proxies used to generate long-term E(R) forecasts and document the performance of the approaches to estimating expected returns that have largely been considered in isolation. We show that 10- to 20-year E(R)s can be estimated ex ante. Out-of-sample (OOS) forecast improvements over historical mean forecasts are as large as 40% even in the most recent period. Importantly, these gains exist within a range of time periods.

The equity valuation model of Gordon (1962) suggests that $P_0 = D_1 / [E(R) - g]$. In other words, today's price (P_0) is related to next year's dividend (D_1), future growth in dividends (g), and the required or expected return on equities in perpetuity (E(R)). Rearranging this formula results in $E(R) = D_1 / P_0 + g$. However, it is important to note that this is the E(R) on the equity market in perpetuity. Over shorter time horizons, there are two reasons why E(R) may be time-varying. The first is rational. E. suggest that investor risk aversion varies over time, which

---

[1] Long-term return forecasts are more relevant to a range of stakeholders, including investors, businesses, and governments. However, data limitations present econometric issues that have impacted this literature (e.g., Boudoukh, Israel, and Richardson, 2022). It is therefore unsurprising that most return predictability literature has focused on monthly return predictability (e.g., Rapach, Ringgenberg, and Zhou, 2016).

implies that different levels of E(R) are required to entice individuals to invest in the equity market. The second is behavioral. Shiller (2016) suggests that there are times of overvaluation and low E(R), as well as times of undervaluation and high E(R), due to investor psychological bias. This suggests that the E(R) for finite horizons is best expressed as $E(R) = D_1 / P_0 + g + \Delta V$, where D_1 / P_0 is the yield, g is the growth, and ΔV is the valuation change.

A range of different proxies has been used for each of the three expected return components (D_1 / P_0, g, and ΔV). Furthermore, while some researchers use proxies for the three components together, others forecast E(R) using yield or valuation change alone. We run a horse race of approaches using "yield alone," "valuation alone," "yield and growth," and a combination of all three inputs, which we refer to as "three components."[2] We also consider different ways of estimating the inputs to these frameworks.

Our evaluation framework addresses various issues that have been raised in the literature. Most long-term return prediction papers focus on in-sample analysis, but as Foster, Smith, and Whaley (1997) point out, these can be susceptible to data mining. Furthermore, overlapping observations are typically used, which can result in bias being introduced into the regression analysis. Statistical techniques, such as those developed by Hansen and Hodrick (1980), Newey and West (1987), and Hjalmarsson (2011), have been employed to mitigate these biases. However, Boudoukh, Israel, and Richardson (2022) show that these widely used measures do not completely remove bias from the analysis. We, therefore, focus on OOS analysis. As Boudoukh, Israel, and Richardson (2022) note, OOS forecasts and statistics, such as the mean square error, are unaffected by overlapping observation bias.[3] It is common to evaluate the accuracy of forecasted returns by examining their correlation with actual returns

[2] Campbell and Shiller (1998) also use the dividend-to-price ratio as a valuation ratio. We adopt the Gordon growth framework and thus classify this as "yield alone" rather than "valuation alone," but this classification has no impact on the reported results.

[3] Boudoukh, Israel, and Richardson (2022) propose an in-sample approach that is free from overlapping sample bias. However, we do not apply this, as a large focus of our work is to compare the performance of various predictive approaches, and this is more readily achieved in an out-of-sample setting.

(e.g., Engle, Focardi, and Fabozzi, 2016; Damodaran, 2022). We focus our analysis on mean absolute errors (MAEs) and mean square errors (MSEs). We suggest that both the average magnitude of the differences between the forecast and actual returns and the extent to which forecasted returns track actual returns are important. Both MAEs and MSEs capture these, whereas correlations do not reflect the average difference between the forecast and actual returns.

Our results indicate that the three-component framework is superior for 10-year forecasts, although not by a large margin. The three-component model that generates ΔV estimates based on the wealth portfolio composition of Rintamaki (2023), denoted as ΔV_{WPC}, outperforms other three-component models in forecasting 10-year returns for the entire sample period from 1891 to 2020. However, in two more-recent sub-sample periods, it is not statistically different from other three-component models, such as the one that assigns an equal weight to four proxies for ΔV. We suggest that this latter model is superior overall, as it performs better in an asset allocation setting. It generates a 34.91% reduction in MAEs and a 57.70% increase in OOS-R^2 compared to the historical mean model for 10-year forecasts over the 1891–2020 sample period. Furthermore, a stock-bond portfolio with weights allocated based on these E(R) forecasts has an approximately 65.56% higher Sharpe ratio and a 50.06% improvement in value at risk (VaR) over the 1891–2020 period. Importantly, this model also leads to improvement gains in more recent periods. Twenty-year returns are typically more difficult to forecast. However, several approaches, such as the three-component model with the total return cyclically adjusted price-to-earnings ratio (TRCAPE) proxy significantly enhance the accuracy of these predictions.

We contribute to several strands of the long-term return predictability literature. Fama and French (1988) use a yield-alone approach and show that dividend yields explain more than 25% of the variance of two- to four-year returns. Campbell and Shiller (1998) contribute to the valuation-alone literature by focusing on predicting 10-year returns using a price-to-earnings ratio that is derived from the average earnings over the last 10 years. They suggest that accounting for earnings fluctuations over the business cycle is important and show that this metric, which is widely referred to as the cyclically adjusted price-to-earnings ratio (CAPE), is effective at predicting stock returns. Bogle (1991a, b) introduces the three-component approach and suggests that the forecasts of the 10-year returns give "a remarkably precise replication of the actual total returns realized."

There have been advances in each of these three approaches. The literature on yield is mixed. Boudoukh, Richardson, and Whitelaw (2008) and Goyal and Welch (2008) suggest that dividend yields are not useful predictors of stock returns for periods of up to five years. However, Cochrane (2008) shows that dividend yields have predictive information for stock returns over the subsequent 1–25 years. More recently, Golez and Koudijs (2018) find that dividend yields predict equity returns over intervals of up to five years in the Netherlands, UK, and US.

The valuation-alone literature has focused on refining measures of CAPE and introducing new proxies. Several studies point out that CAPE has underperformed recently, which provides motivation for considering modifications. Philips and Ural (2016) suggest several modifications, including using cash flows rather than earnings in the valuation ratio calculation. Siegel (2016) points out that changes in the calculation of GAAP earnings may impact CAPE and recommends using alternative earnings data. Arnott, Chaves, and Chow (2017) suggest that adjusting the CAPE based on macroeconomic conditions leads to prediction accuracy improvements for short-term forecasts. More recently, Philips and Kobor (2020) propose that using one year's quarterly earnings results in better predictions than the average of the last 10 year's earnings in CAPE, while Waser (2021) finds that variation in the CAPE can be explained by variation in the economic variables.

Numerous variables have also been considered as valuation proxies. Goyal and Welch (2008) conduct a comprehensive evaluation of the ability of a range of variables to predict one-month to five-year equity returns. These include long-term returns, default return spread, inflation, long-term yield, stock variance, dividend payout ratio, default yield spread, treasury-bill rate, earning price ratio, term spread, equity issuance, book-to-market ratio, net equity expansion, and investment-capital ratio. They conclude that none of these generate consistent in-sample and OOS predictability. We, therefore, do not include these variables as valuation proxies.

More recently, several papers document effective valuation proxies. Atanasov, Møller, and Priestley (2020) document the predictive ability of using cyclical consumption as a proxy. They suggest that in good (bad) times with above (below) trend consumption, investors are willing (unwilling) to forgo consumption and to invest; therefore, current prices rise (decline) and expected returns are lower (higher). They show that cyclical consumption predicts market returns up to five years in advance in in-sample tests. Swinkels and Umlauft (2022) test what they refer to as "the Buffett indicator" following Warren Buffett's observation that the market capitalization of publicly traded stocks to economic output is an extremely effective valuation indicator. Swinkels and Umlauft (2022) show that the Buffett indicator is an effective valuation timing tool over a range of horizons in the US and internationally. Finally, Rintamaki (2023) documents the ability of the ratio of the value of the stock market to bond and residential housing assets to predict equity market returns. He finds strong evidence of predictability in in-sample tests for a range of horizons and OOS tests for one year.

Our contributions are as follows. First, we test the relative performance of the alternative frameworks of "yield alone," "valuation alone," "yield and growth," and all "three components." Second, we run OOS tests that are free from look-ahead bias. Third, we consider all input variables and frameworks in forecasting 10-year and 20-year long-term returns.

Fourth, we consider whether various valuation proxies can be combined to generate superior forecasts.

The rest of this paper is organized as follows. Section 2 contains a description of the data. The method and results are described in Section 3, while Section 4 concludes the paper.

2. Variable Construction, Data, and Methods

We run a horse race of approaches across four frameworks, namely, "yield alone" (*YLD*), "yield and growth" (referred to as "Gordon" or *GOR*), "valuation alone" (ΔV), and "three components" ($GOR + \Delta V$). We start with the "yield alone" framework using a standard predictive regression model:

$$r_{t:t+h} = \alpha + \beta x_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T\text{-}h \tag{1}$$

where $r_{t:t+h} = (1/h)(r_{t+1} + \ldots + r_{t+h})$ with h = 10 or 20 years, r_t is the S&P 500 log return for year t, and x_t is one of our four yield predictors, namely, dividend yield, total yield, net total yield, and cyclically adjusted total yield (*CATY*) as per Straehl and Ibbotson (2017). We focus on OOS analysis and follow Goyal and Welch (2008) to compute our OOS forecasts. The OOS forecasts begin 20 years after the data are available. To generate the h-period ahead OOS forecast, we first estimate α and β in Eq. (1) by regressing on the data up to time t. We then insert regression estimates back to Eq. (1) and use the value of the predictor variable x_t at the end of the in-sample period to compute the forecasting value, denoted as $\hat{r}_{t:t+h}$. We continue our calculation by adding one more observation each time in Eq. (1) and using expanding windows (e.g., Chiang and Hughen, 2017; Gao and Nardari, 2018) to compute a time series of OOS forecasts.

For the "yield and growth" or the "Gordon" approach, we employ the classic Gordon growth model and calculate the expected return as the sum of a current "yield" and a historical averaged "growth" rate over the entire period. We consider the four yields used in Eq. (1). The growth rates we use are earnings growth, dividends growth, total yield growth, and *CATY* growth, respectively.[4]

The "valuation alone" approach is similar to the "yield alone" approach except we replace the four yield predictors in Eq. (1) with four proxies for ΔV, which include: (i) the total return CAPE[5] (*TRCAPE*) as per Jivraj and Shiller (2018); (ii) the total wealth portfolio composition (*WPC*) as per Rintamaki (2023); (iii) the Buffett indicator (*BUF*), calculated as the equity market value scaled by gross domestic product (Swinkels and Umlauft, 2022); (iv) cyclical consumption, or *CON* (Atanasov, Møller, and Priestley, 2020). *TRCAPE* scales the real total return price for the average real earnings over the prior 10 years. It is similar to the cyclically adjusted price-to-earnings ratio (*CAPE*; correlation > 0.99), but it takes dividends into account and assumes dividends to be reinvested into the price index. *WPC* measures the value of stock market wealth relative to the value of other assets including residential housing and government bonds.[6] High *WPC* ratios predict low future stock market returns (Rintamaki, 2023). *BUF* reflects the ratio of the market value of stocks to gross domestic product and it is an indicator for equity market mispricing. Swinkels and Umlauft (2022) show that low *BUF* ratios predict above-average 10-year returns. Atanasov, Møller, and Priestley (2020) find an inverse relation between aggregation consumption and expected stock market returns. We, therefore, use *CON* as our fourth proxy for ΔV.

[4] In unreported results, we also follow Damodaran (2022) and calculate growth as being equal to the risk-free rate. This method does not have a material impact on our key conclusions.
[5] We thank Robert Shiller for making these data available: http://www.econ.yale.edu/~shiller/data.htm
[6] We thank Jordà, Schularick, and Taylor for making their data available: https://www.macrohistory.net/database/

As noted in Rapach, Strauss, and Zhou (2010), forecast combinations incorporate information from individual forecasts and can outperform individual forecasts. We, therefore, apply four forecast combining methods that differ in the weights assigned to each forecast based on the ΔV predictors: (i) the simple average, taking the arithmetic mean of four individual forecasts in each year; (ii) the inverse variance-weighted average (Bates and Granger, 1969); (iii) the Granger and Ramanathan (1984) constrained regression approach; (iv) the Bayesian model averaging method (Min and Zellner, 1993). We generate composite ΔV forecasts using the four combining methods and denote these as ΔV_{EW}, ΔV_{IVW}, ΔV_{GR}, and ΔV_{BIC}.

With the "three components" approach, we expect stock returns for finite horizons to vary with "yield," "growth," and change in "valuation" (i.e., $YLD_{Div} + g_{Div} + \Delta V$ or $GOR_{Div,Div} + \Delta V$). We run the following predictive regression to forecast the OOS h-period-ahead ΔV:

$$\Delta V_{t:t+h} = \gamma + \delta z_t + \varepsilon_{t:t+h} \text{ for } t = 1, \ldots, T\text{-}h \tag{2}$$

where $\Delta V_{t:t+h}$ is the actual h-period change in valuation, calculated as dividend yield at time t and historical dividend growth rate subtracted from actual h-period return $r_{t:t+h}$. The predictor z_t is one of our four proxies for ΔV (i.e., *TRCAPE*, *WPC*, *BUF*, and *CON*). To generate the h-period-ahead OOS ΔV forecast, we first estimate γ and δ in Eq. (2) by regressing on the data up to time t. We then insert regression estimates back to Eq. (2) and use the value of the predictor variable z_t at the end of the in-sample period to compute the forecasting value, denoted as $\widehat{\Delta V}_{t:t+h}$. We then calculate the h-period-ahead OOS return forecast as the sum of dividend yield and historical dividend growth rate at the end of the in-sample period and predicted change in valuation, $\widehat{\Delta V}_{t:t+h}$. We continue our calculation by adding one more observation each time in the regression and using expanding windows (e.g., Chiang and Hughen, 2017; Gao and Nardari, 2018) to compute a time series of OOS forecasts. We also generate composite “three components” forecasts using the four combining methods. We provide a description of the proxies and approaches we use in Appendix 1.

3. Results

We present summary statistics in Table 1. The average annual returns are 10.66%, 11.76%, and 12.27% for the 1872–2020, 1955–2020, and 1988–2020 periods, respectively. Returns have negative skewness across all three sample periods. Kurtosis is negative for the entire period, but positive in the more recent periods. In Panel B, we present mean geometric and log returns for 10-year and 20-year intervals rolling forward one year at a time. It is these annualized log returns that we use in our model forecasts. For the 10-year interval, average annualized log returns are 8.65%, 9.40%, and 8.57% for the three periods, respectively, while for the 20-year interval, these are 8.73%, 9.68%, and 7.55%, respectively. Similarly, Panel C reports the standard deviation of geometric returns and log returns for 10-year and 20-year intervals rolling forward one year at a time.

[Please Insert Table 1 About Here]

In Table 2, we report results for 10-year forecasts. We calculate the MAE as the average absolute difference between the forecast and actual returns. We also calculate the difference in MAEs between each prediction model and the historical mean forecast.[7] We measure the statistical significance of this difference using the moving block bootstrap method, which accounts for autocorrelation in the time series. The optimal block length is determined as per Patton, Politis, and White (2009). For each prediction model, we generate 1000 bootstrap

[7] We believe that MAEs better reflect the performance of a forecast than correlations as they account for the magnitude of errors between forecasted and actual returns. We provide an example in Appendix 2 of a scenario where one forecast can have a larger correlation than another forecast (indicating outperformance) and also have a larger MAE (indicating underperformance).

resamples and report statistical significance based on the one-sided bootstrap *p*-value (i.e., the proportion of the bootstrap sample prediction model MAEs that exceed the historical mean model MAE in the same bootstrap sample).

We are interested in determining whether the lowest model MAE is statistically significantly less than the next lowest model MAE across all four frameworks. The procedure is as follows. First, we sort our 25 prediction models and the historical mean model based on their realized MAEs from smallest to largest. Then, we use the moving block bootstrap method to test the statistical significance of the difference in MAEs of the models with the lowest and second lowest MAEs. If the difference in MAEs is statistically insignificant at the 1% significance level, we continue to test the statistical significance in MAEs of the prediction models with the lowest and third lowest MAEs until the difference in MAEs of the two models is statistically significant. For example, if the difference in MAEs of the models with the lowest and third lowest MAEs is statistically significant at the 1% level, we group the two models with the lowest and second lowest MAEs as Tier 1 models and then continue this procedure to test the difference in MAEs of the models with the third lowest and fourth lowest MAEs, until we group all 26 models into sub-categories.[8]

We also compare models using the OOS-R^2 metric. We calculate OOS-R^2 for each prediction model as per Goyal and Welch (2008) as follows:

$$R^2_{OOS} = 1 - \frac{MSE_A}{MSE_N} \tag{3}$$

[8] In Tables 2 and 3, the values in bold are MAEs of Tier 1 models, which have the lowest MAEs.

where MSE_N is the mean squared forecast error of the historical mean model and MSE_A is the mean squared forecast error of our alternative prediction model over the OOS period. We then use the Clark and West (2007) approach to test H_0: $R^2_{OOS} \leq 0$ versus H_1: $R^2_{OOS} > 0$.

The results in Table 2 indicate that the three-component framework is the best-performing framework. For 10-year forecasts, this framework has the lowest average MAE in the 1981–2020, 1955–2020, and 1988–2020 periods. To ascertain the statistical difference between the "three components" framework and the other three frameworks (i.e., "yield alone," "Gordon," and "valuation alone"), we first compute the yearly average absolute error for each framework over time. Utilizing the resulting four time series, we then apply the Diebold–Mariano test (Diebold and Mariano, 1995). As shown in Appendix 3, the average absolute error of "three components" is statistically significantly lower than "Gordon" in all three time periods and statistically significantly lower than "valuation alone" and "yield alone" in two of the three time periods.

The three-component model with the change in valuation driven by *WPC* and the valuation-alone model based on *WPC* are the best performers over the entire sample period based on both the MAE and OOS-R^2 metrics. However, as highlighted in bold in Table 2, in the two more recent sub-periods, the MAEs of these two measures are not statistically different from the MAEs of the four three-component models: $(GOR_{Div,Div} + \Delta V_k)_{EW}$, $(GOR_{Div,Div} + \Delta V_k)_{IVW}$, $(GOR_{Div,Div} + \Delta V_k)_{GR}$, and $(GOR_{Div,Div} + \Delta V_k)_{BIC}$. They are also not different to several of the valuation-alone approaches based on averaging in the two most recent periods.

The models we highlight generate a meaningful improvement compared to the historical mean forecasts. We will focus on the three-component model with valuation changes measured using an equal-weighted approach, $(GOR_{Div,Div} + \Delta V_k)_{EW}$, to demonstrate this point. We suggest that this model is the best overall performer for 10-year forecasts when asset allocation considerations are also considered.[9] The MAE for the entire period is 0.0273, which is 0.0146 lower than the equivalent period historical average MAE and represents a 34.91% improvement. Furthermore, the R^2_{OOS} indicates a 57.70% improvement, consistently showing that the $(GOR_{Div,Div} + \Delta V_k)_{EW}$ model delivers a significantly lower forecasting error than the historical mean model. Importantly, these performance gains are also evident in more recent sub-periods, with the R^2_{OOS} being 52.30% and 39.71% in the 1955–2020 and 1988–2020 periods, respectively. We depict the improvements in forecasting using this model compared to historical mean model forecasts in Figures 1a, 2a, and 3a.

[Please Insert Figures 1a, 2a, and 3a About Here]

In unreported results, we measure the proportion of times that each model generates a forecasted return that is either greater or smaller than the actual return, along with the average error when the forecast is higher or lower than the actual return. The errors of some models are particularly asymmetric. For instance, the three-component model based on *BUF* underestimates actual returns 96.97% of the time over the entire period, with an average error of 4.86%. When the model overestimates actual returns, the average error is 1.30%. The three-component model with valuation changes measured using an equal-weighted approach generates more symmetrical errors. Throughout the entire period, it overestimates 30.91% of the time. However, the average error from overestimation is 2.79%, compared to an average error of 2.70% when there is an underestimation of returns.

[Please Insert Table 2 About Here]

[9] The asset allocation results are reported later in the section.

In Table 3, we report equivalent results for a 20-year forecast period. The MAEs from the historical mean approach are considerably lower for the 20-year forecast period compared to the 10-year forecast period. For instance, the average 20-year forecast period MAE for the 1988–2020 period is just 0.0129, compared to 0.0403 for the 10-year forecasts. This is consistent with the results in Table 1 Panel C, which show lower standard deviations for 20-year returns.

However, while the levels of MAEs for each framework and model are typically lower than their Table 2 equivalents, the gains over the historical mean approach are also generally lower. As with the 10-year forecast results in Appendix 3, the "three components" framework generates the strongest results in the 1891–2020 and 1955–2020 periods. However, the differences between the frameworks on average are not large.

As highlighted in bold, the MAEs of several models within each framework are not statistically distinguishable from each other. However, we focus on the three-component model with valuation changes determined by *TRCAPE*, as we believe that this model is the best performer across all metrics, including asset allocation. During the periods of 1891–2020, 1955–2020, and 1988–2020, the model generates R^2_{OOS} improvements of 37.23%, 51.45%, and 57.05%, respectively. Figures 1b, 2b, and 3b depict the forecasted returns, forecast errors, and absolute forecast errors of this model and the historical mean model.

[Please Insert Table 3 About Here]

[Please Insert Figures 1b, 2b, and 3b About Here]

Prior studies show return predictability is time-varying (e.g., Devpura, Narayan, and Sharma, 2018; Jurdi, 2022). In Table 4, we report the MAEs in different market states over time. For each prediction model, we run the following time-series regression with Newey–West (1987) standard errors:

$$MAE_{i,t} = \alpha_i + \beta_i\, MKT\_STATE_t + \varepsilon_{i,t} \qquad (4)$$

where MAE_i is the 10- or 20-year MAEs for prediction model i, and MKT_STATE is one of our four market state proxies, calculated over the same 10- or 20-year period as MAE_i. The four market state proxies include market return, market volatility, the Amihud (2002) illiquidity ratio, and a market recession indicator. Market return and the Amihud (2002) ratio are the average annual market return and the average annual value-weighted stock Amihud (2002) ratio, respectively. Market volatility is the standard deviation of annual returns over the same period as MAE_i. The market recession proxy is determined by calculating the proportion of months (within a 10- or 20-year period) that fall within recessionary phases of the National Bureau of Economic Research (NBER) business cycle.

The results in Table 4 indicate that forecasts tend to be more accurate (i.e., MAEs are lower) when returns are lower. Furthermore, forecasts are more accurate when volatility is higher. There is no consistent relation between forecast accuracy and liquidity or the business cycle. The business cycle result differs from shorter horizon predictability, which is stronger in economic contractions (e.g., Henkel, Martin, and Nadari, 2011). In Appendix 4, we present results using mean squared errors as the dependent variable. These are very similar to the results in Table 4. It is worth noting that, in terms of 10-year forecasts, a number of "three components" models perform similarly well across different market states.

[Please Insert Table 4 About Here]

In Tables 5 and 6, we compare the different models from an asset allocation perspective. We allocate the portfolio between stocks and bonds using data on the S&P 500 Index and the US 10-year government bond total return index. We employ the mean–variance approach and consider optimal portfolio weights as the asset weights that maximize the portfolio Sharpe ratio. Following the derivation in Smith (2019), we calculate optimal weights and rebalance the portfolio annually based on the expected Sharpe ratios of the two assets, their historical standard deviations, and the historical correlation between them. To calculate the expected Sharpe ratio, $[E(R) - R_f]/\sigma$, for the S&P 500 (which serves as one input for determining optimal portfolio weights), we use our OOS S&P 500 return forecasts from each of our prediction models (i.e., E(R)), historical risk-free rates sourced from the updated Goyal and Welch (2008) dataset (i.e., R_f), and historical standard deviations of S&P 500 returns (i.e., σ). Accordingly, optimal weights and realized portfolio returns differ across our models in Tables 5 and 6.

We then generate three performance metrics for realized portfolio returns: 5% value at risk (VaR), ex post alpha (alpha), and ex post Sharpe ratio (Sharpe). We employ the aforementioned moving block bootstrap approach to bootstrap realized portfolio returns and determine whether realized VaR of portfolios constructed based on each of our prediction models is significantly improved compared to the historical mean model. Similarly, we also examine whether realized alpha values and Sharpe ratios of portfolios based on our prediction models are significantly higher than those based on the historical mean model.

The results in Table 5 indicate that there are important gains from an asset allocation perspective. For instance, for the entire period, the VaR for the three-component model with equal-weight valuation changes is −8.10%, compared to −16.21% for the historical mean model. Furthermore, the Sharpe ratio of this model is 0.3201, compared to 0.1933 for the historical mean model. Both differences are statistically significant and economically meaningful. These results are not specific to the entire period. For the more recent period of

1988–2020, the VaR for the three-component model with equal-weight valuation changes is −3.19%, compared to −19.12% for the historical mean model. The alpha for this model for the most recent period is 5.00%, compared to 1.81% for the historical mean model.

[Please Insert Table 5 About Here]

Strong gains from an asset allocation perspective are also evident in the 20-year forecast results, as shown in Table 6. The three-component model with valuation changes determined by *TRCAPE* generates a Sharpe ratio of 0.3669 for the entire period, compared to 0.2040 for the historical mean model. The Sharpe ratio generated by this model is also larger than that of the historical mean model in each of the two most recent periods, but the differences are not statistically significant. However, the three-component model with valuation changes determined by *TRCAPE* exhibits significantly superior VaRs in both recent periods when compared to the historical mean model. For instance, the VaR of this model stands at just −3.19% in the 1988–2020 period, compared to −10.53% for the historical mean model.

[Please Insert Table 6 About Here]

4. Conclusions

Accurately estimating long-term expected returns of equity markets is important for both corporate entities and individual investors. We investigate the ability of different frameworks and input proxies to estimate 10- and 20-year OOS returns over long historical time periods and more recent periods. We document that several approaches generate meaningful improvements compared to historical mean model forecasts. OOS-R^2 can be as significant as 40% even in the most recent period, and asset allocation based on our model forecasts can improve a portfolio's Sharpe ratio and VaR by over 50%. We hope that our results are of interest to those who require accurate long-term expected return forecasts.

References

Amihud, Yakov (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.

Arnott, Robert D., Chaves, Denis B., and Chow, Tzee-man (2017). King of the mountain: The Shiller P/E and macroeconomic conditions. *Journal of Portfolio Management*, 44(1), 55-68.

Atanasov, Victoria, Møller, Stig V., and Priestley, Richard (2020). Consumption fluctuations and expected returns. *Journal of Finance*, 75(3), 1677-1713.

Bates, John M., and Granger, Clive W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451-468.

Bogle, John C. (1991a). Investing in the 1990s. *Journal of Portfolio Management*, 17(3), 5-14.

Bogle, John C. (1991b). Investing in the 1990s: Occam's razor revisited. *Journal of Portfolio Management*, 18(1), 88-91.

Boudoukh, Jacob, Israel, Ronen, and Richardson, Matthew (2022). Biases in long-horizon predictive regressions. *Journal of Financial Economics*, 145(3), 937-969.

Boudoukh, Jacob, Richardson, Matthew, and Whitelaw, Robert F. (2008). The myths of long-horizon predictability. *Review of Financial Studies*, 21(4), 1577-1605.

Campbell, John Y., and Cochrane, John H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2), 205-251.

Campbell, John Y., and Shiller, Robert J. (1988). Stock prices, earnings, and expected dividends. *Journal of Finance*, 43(3), 661-676.

Campbell, John Y., and Shiller, Robert J. (1998). Valuation ratios and the long-run stock market outlook. *Journal of Portfolio Management*, 24(2), 11-26.

Chiang, I-Hsuan E., and Hughen, Walker K. (2017). Do oil futures prices predict stock returns? *Journal of Banking and Finance*, 79, 129-141.

Clark, Todd E., and West, Kenneth D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291-311.

Cochrane, John H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4), 1533-1575.

Damodaran, Aswath (2022). Equity risk premiums (ERP): Determinants, estimation, and implications. SSRN Working Paper. Available at: https://ssrn.com/abstract=4066060

Devpura, Neluka, Narayan, Pareshm, and Sharma, Susan (2018). Is stock return predictability time-varying? *Journal of International Financial Markets, Institutions and Money*, 52, 152-172.

Diebold, Francis X., and Mariano, Roberto S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 134-144.

Engle, Robert, Focardi, Sergio, and Fabozzi, Frank (2016). Issues and applying financial econometrics to factor-based modeling in investment management. *Journal of Portfolio Management*, 42(5), 94-106.

Fama, Eugene F., and French, Kenneth R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1), 3-25.

Foster, F. Douglas, Smith, Tom, and Whaley, Robert E. (1997). Assessing goodness-of-fit of asset pricing models: The distribution of the maximal. *Journal of Finance*, 52(2), 591-607.

Gao, Xin, and Nardari, Federico (2018). Do commodities add economic value in asset allocation? New evidence from time-varying moments. *Journal of Financial and Quantitative Analysis*, 53(1), 365-393.

Golez, Benjamin, and Koudijs, Peter (2018). Four centuries of return predictability. *Journal of Financial Economics*, 127(2), 248-263.

Gordon, Myron J. (1962). The savings, investment, and valuation of a corporation. *Review of Economics and Statistics*, 44(1), 37-51.

Goyal, Amit, and Welch, Ivo (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.

Granger, Clive W. J., and Ramanathan, Ramu (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2), 197-204.

Hansen, Lars P., and Hodrick, Robert J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, 88(5), 829-853.

Henkel, Sam J., Martin, J. Spencer, and Nadari, Federico (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99, 560-580.

Hjalmarsson, Erik (2011). New methods for inference in long-horizon regressions. *Journal of Financial and Quantitative Analysis*, 46(3), 815-839.

Jivraj, Farouk, and Shiller, Robert J. (2018). The many colors of CAPE. Yale ICF Working Paper No. 2018-22. Available at: https://ssrn.com/abstract=3258404

Jurdi, Doureige (2022). Predicting the Australian equity risk premium. *Pacific-Basin Finance Journal*, 71, 101683.

Min, Chung-ki, and Zellner, Arnold (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, 56(1-2), 89-118.

Newey, Whitney K., and West, Kenneth D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3), 777-787.

Patton, Andrew, Politis, Dimitris N., and White, Halbert (2009). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 28(4), 372-375.

Philips, Thomas K., and Kobor, Adam (2020). Ultra-simple Shiller's CAPE: How one year's data can predict equity market returns better than ten. *Journal of Portfolio Management*, 46(4), 140-155.

Philips, Thomas, and Ural, Cenk (2016). Uncloaking Campbell and Shiller's CAPE: A comprehensive guide to its contraction and use. *Journal of Portfolio Management*, 43(1), 109-125.

Rapach, David E., Ringgenberg, Matthew C., and Zhou, Guofu (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121(1), 46-65.

Rapach, David E., Strauss, Jack K., and Zhou, Guofu (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2), 821-862.

Rintamaki, Paul (2023). Equity wealth cycles. SSRN Working Paper. Available at: https://ssrn.com/abstract=3924180

Shiller, Robert J. (2016). Irrational exuberance. Available at: https://press.princeton.edu/books/paperback/9780691173122/irrational-exuberance

Siegel, Jeremy J. (2016). The Shiller CAPE ratio: A new look. *Financial Analysts Journal*, 72(3), 41-50.

Smith, Thomas (2019). The Sharpe ratio ratio. Available at: https://riskyfinance.com/wp-content/uploads/2019/07/The_Sharpe_Ratio_Ratio_ThomasSmith-FINAL.pdf

Straehl, Philip U., and Ibbotson, Roger G. (2017). The long-run drivers of stock returns: Total payouts and the real economy. *Financial Analysts Journal*, 73(3), 32-52.

Swinkels, Laurens, and Umlauft, Thomas S. (2022). The Buffett indicator: International evidence. SSRN Working Paper. Available at: https://ssrn.com/abstract=4071039

Waser, Otto (2021). Modelling the Shiller CAPE ratio, mean reversion, and return forecasts. *Journal of Portfolio Management*, 47(3), 155-171.

Table 1
Summary Statistics

Panel A: Summary Statistics

Years	Mean	Std Dev	Skewness	Kurtosis
1872–2020	0.1066	0.1825	−0.2421	−0.0095
1955–2020	0.1176	0.1648	−0.5684	0.1002
1988–2020	0.1227	0.1687	−0.8856	0.9094

Panel B: Average of Geometric and Log Returns

	Geometric Returns		Log Returns	
Years	10-Year	20-Year	10-Year	20-Year
1872–2020	0.0914	0.0917	0.0865	0.0873
1955–2020	0.0997	0.1021	0.0940	0.0968
1988–2020	0.0908	0.0784	0.0857	0.0755

Panel C: Standard Deviation of Geometric and Log Returns

	Geometric Returns		Log Returns	
Years	10-Year	20-Year	10-Year	20-Year
1872–2020	0.0497	0.0323	0.0454	0.0293
1955–2020	0.0506	0.0319	0.0463	0.0287
1988–2020	0.0549	0.0118	0.0505	0.0109

This table presents summary statistics for stock returns over the entire sample period and sub-periods. In Panel A, we present the mean, standard deviation, skewness, and kurtosis of annual stock market returns. Stock market returns are simple returns, including dividends, of the S&P 500. Panel B (Panel C) shows the mean (standard deviation) of geometric and log returns for 10-year and 20-year intervals rolling forward one year at a time.

Table 2
10-Year Forecasts

	1891–2020			1955–2020			1988–2020		
	MAE	MAE Diff	OOS-R^2	MAE	MAE Diff	OOS-R^2	MAE	MAE Diff	OOS-R^2
[1] Historical Mean	0.0416	0.0000	0.00%	0.0406	0.0000	0.00%	0.0403	0.0000	0.00%
Panel A: Yield Alone, YLD_i									
[2] YLD_{Div}	0.0445	0.0026	−11.54%	0.0428	0.0022	−4.83%	0.0428	0.0026	−4.20%
[3] YLD_{Ttl}	0.0411	−0.0008	3.99%*	0.0383	−0.0023	13.65%**	**0.0375**	−0.0028	16.71%*
[4] YLD_{NTtl}	0.0384	−0.0075**	51.45%**	0.0317	−0.0089***	39.31%***	**0.0276**	−0.0127	53.82%**
[5] YLD_{CATY}	0.0413	−0.0006	1.15%	0.0381	−0.0024	9.05%*	**0.0363**	−0.0040	16.31%
Panel B: Gordon, $GOR_{i,j} = YLD_i + g_j$									
[6] $GOR_{Div,E} = YLD_{Div} + g_E$	0.0415	0.0000	4.28%*	0.0413	0.0007	0.43%	0.0429	0.0026	−2.12%
[7] $GOR_{Div,Div} = YLD_{Div} + g_{Div}$	0.0424	0.0008	2.55%*	0.0430	0.0024	−5.84%	0.0444	0.0041	−4.98%
[8] $GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	0.0414	−0.0002	5.83%**	0.0397	−0.0009	6.67%*	0.0383	−0.0020	13.22%
[9] $GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	0.0479	0.0021	12.25%*	0.0453	0.0047	−19.66%	0.0396	−0.0007	8.31%*
[10] $GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	0.0433	0.0017	−3.13%	0.0417	0.0011	−1.17%	0.0390	−0.0013	13.54%
Panel C: Valuation Alone, ΔV_k									
[11] ΔV_{TRCAPE}	0.0375	−0.0045*	20.35%***	0.0347	−0.0059	30.31%***	0.0492	0.0089	−9.73%*
[12] ΔV_{WPC}	**0.0210**	−0.0235***	75.64%***	**0.0220**	−0.0185***	69.95%***	**0.0255**	−0.0148***	63.53%***
[13] ΔV_{BUF}	0.0506	0.0043	0.38%***	0.0504	0.0098	−25.42%**	0.0530	0.0127	−33.61%
[14] ΔV_{CON}	–	–	–	0.0607	0.0166	−92.32%	0.0560	0.0157	−85.74%
[15] ΔV_{EW}	0.0275	−0.0144***	55.38%***	**0.0278**	−0.0127***	46.22%***	**0.0301**	−0.0102*	37.20%*
[16] ΔV_{IVW}	0.0314	−0.0105***	40.94%***	0.0298	−0.0107***	42.64%***	**0.0359**	−0.0044	26.15%*
[17] ΔV_{GR}	0.0248	−0.0172***	62.56%***	0.0297	−0.0109**	42.69%***	0.0391	−0.0012	16.97%*
[18] ΔV_{BIC}	0.0320	−0.0100***	36.96%***	**0.0238**	−0.0167***	66.93%***	**0.0255**	−0.0147***	63.52%***
Panel D: Three Components, $GOR_{Div,Div} + \Delta V_k$									
[19] $GOR_{Div,Div} + \Delta V_{TRCAPE}$	0.0352	−0.0068**	30.51%***	0.0298	−0.0108**	48.06%***	0.0395	−0.0008	24.21%**
[20] $GOR_{Div,Div} + \Delta V_{WPC}$	**0.0246**	−0.0199***	61.71%***	**0.0202**	−0.0204***	72.68%***	**0.0265**	−0.0138**	60.93%***

[21] $GOR_{Div,Div} + \Delta V_{BUF}$	0.0475	0.0011	10.21%***	0.0477	0.0072	−11.55%**	0.0508	0.0105	−20.64%*
[22] $GOR_{Div,Div} + \Delta V_{CON}$	−	−	−	0.0469	0.0028	−37.70%	0.0427	0.0024	−31.53%
[23] $(GOR_{Div,Div} + \Delta V_k)_{EW}$	0.0273	−0.0146***	57.70%***	**0.0268**	−0.0137***	52.30%***	**0.0310**	−0.0093*	39.71%**
[24] $(GOR_{Div,Div} + \Delta V_k)_{IVW}$	0.0285	−0.0134***	52.45%***	**0.0264**	−0.0141***	52.37%***	**0.0307**	−0.0096**	38.94%**
[25] $(GOR_{Div,Div} + \Delta V_k)_{GR}$	0.0255	−0.0165***	55.77%***	**0.0243**	−0.0163***	53.30%***	**0.0312**	−0.0091*	36.06%**
[26] $(GOR_{Div,Div} + \Delta V_k)_{BIC}$	0.0336	−0.0083***	34.14%***	**0.0268**	−0.0138***	56.06%***	**0.0324**	−0.0079	41.33%**

Table 2 reports results for 10-year forecasts over the entire sample period, as well as half and quarter sub-periods. MAE is the mean absolute difference between the forecast and actual returns. MAE Diff is the difference in MAEs between each prediction model and the historical mean model. We measure the statistical significance of MAE Diff using the moving block bootstrap method, which accounts for autocorrelation in the time series. For each prediction model, we generate 1000 bootstrap resamples and report statistical significance based on the one-sided bootstrap p-value. We also use the moving block bootstrap method in determining the statistical significance of the differences in MAEs across our prediction models. MAEs in bold are MAEs of Tier 1 models with the lowest MAEs. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

Table 3
20-Year Forecasts

	1891–2020			1955–2020			1988–2020		
	MAE	MAE Diff	OOS-R^2	MAE	MAE Diff	OOS-R^2	MAE	MAE Diff	OOS-R^2
[1] Historical Mean	**0.0282**	0.0000	0.00%	0.0256	0.0000	0.00%	0.0129	0.0000	0.00%
Panel A: Yield Alone, YLD_i									
[2] YLD_{Div}	**0.0291**	−0.0014	12.00%***	0.0251	−0.0005	7.01%*	**0.0084**	−0.0045**	56.96%**
[3] YLD_{Ttl}	**0.0275**	−0.0029*	17.63%***	0.0246	−0.0010	9.84%**	**0.0071**	−0.0058**	66.45%**
[4] YLD_{NTtl}	–	–	–	**0.0182**	−0.0073**	49.69%***	0.0126	−0.0003	−1.49%**
[5] YLD_{CATY}	**0.0285**	−0.0019*	12.06%***	0.0252	−0.0004	5.87%*	**0.0102**	−0.0027***	38.26%**
Panel B: Gordon, $GOR_{i,j} = YLD_i + g_j$									
[6] $GOR_{Div,E} = YLD_{Div} + g_E$	0.0320	0.0038	−13.79%	0.0311	0.0055	−19.79%	0.0138	0.0009	5.40%**
[7] $GOR_{Div,Div} = YLD_{Div} + g_{Div}$	0.0326	0.0044	−14.89%	0.0335	0.0080	−35.78%	0.0166	0.0037	−26.05%**
[8] $GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	0.0317	0.0035	−13.48%	0.0314	0.0059	−26.57%	0.0108	−0.0021	36.82%**
[9] $GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	0.0468	0.0122	−38.06%	0.0446	0.0191	−111.64%	0.0293	0.0164	−292.82%**
[10] $GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	0.0338	0.0056	−27.68%	0.0335	0.0080	−45.88%	0.0118	−0.0011	25.30%**
Panel C: Valuation Alone, ΔV_k									
[11] ΔV_{TRCAPE}	**0.0257**	−0.0048***	31.64%***	0.0213	−0.0042*	40.58%***	0.0167	0.0037	−76.68%**
[12] ΔV_{WPC}	**0.0262**	−0.0062***	38.25%***	**0.0178**	−0.0077**	57.51%***	**0.0087**	−0.0043	54.68%**
[13] ΔV_{BUF}	**0.0227**	−0.0026	27.00%***	0.0228	−0.0028	27.39%***	**0.0066**	−0.0064**	73.57%**
[14] ΔV_{CON}	–	–	–	–	–	–	0.0316	0.0154	−202.83%
[15] ΔV_{EW}	**0.0249**	−0.0055***	34.95%***	0.0193	−0.0063**	47.21%***	**0.0075**	−0.0055**	69.61%**
[16] ΔV_{IVW}	**0.0249**	−0.0055***	35.39%***	0.0192	−0.0063**	47.97%***	**0.0077**	−0.0052**	61.41%**
[17] ΔV_{GR}	**0.0271**	−0.0033*	27.52%***	0.0236	−0.0019	32.53%***	0.0207	0.0077	−107.91%
[18] ΔV_{BIC}	**0.0257**	−0.0048***	31.64%***	0.0213	−0.0042*	40.58%***	0.0167	0.0037	−76.68%**
Panel D: Three Components, $GOR_{Div,Div} + \Delta V_k$									
[19] $GOR_{Div,Div} + \Delta V_{TRCAPE}$	**0.0248**	−0.0056*	37.23%***	**0.0183**	−0.0073***	51.45%***	**0.0086**	−0.0043*	57.05%*
[20] $GOR_{Div,Div} + \Delta V_{WPC}$	**0.0271**	−0.0053**	25.69%***	**0.0164**	−0.0092**	63.81%***	**0.0103**	−0.0026	38.19%*

[21] $GOR_{Div,Div} + \Delta V_{BUF}$	**0.0205**	−0.0048	41.91%**	0.0203	−0.0052	43.33%**	**0.0076**	−0.0053*	66.95%**
[22] $GOR_{Div,Div} + \Delta V_{CON}$	–	–	–	–	–	–	0.0302	0.0140	−179.67%
[23] $(GOR_{Div,Div} + \Delta V_k)_{EW}$	**0.0257**	−0.0047	34.22%***	0.0187	−0.0069*	53.47%***	0.0111	−0.0018	43.29%*
[24] $(GOR_{Div,Div} + \Delta V_k)_{IVW}$	**0.0266**	−0.0038	28.48%***	0.0193	−0.0063*	51.33%***	0.0130	0.0001	22.41%*
[25] $(GOR_{Div,Div} + \Delta V_k)_{GR}$	**0.0287**	−0.0017	16.84%***	0.0216	−0.0040	45.53%***	0.0260	0.0131	−193.01%
[26] $(GOR_{Div,Div} + \Delta V_k)_{BIC}$	**0.0248**	−0.0056*	37.23%***	**0.0183**	−0.0073***	51.45%***	**0.0086**	−0.0043	57.05%*

Table 3 reports results for 20-year forecasts over the entire sample period, as well as half and quarter sub-periods. MAE is the mean absolute difference between the forecast and actual returns. MAE Diff is the difference in MAEs between each prediction model and the historical mean model. We measure the statistical significance of MAE Diff using the moving block bootstrap method, which accounts for autocorrelation in the time series. For each prediction model, we generate 1000 bootstrap resamples and report statistical significance based on the one-sided bootstrap p-value. We also use the moving block bootstrap method in determining the statistical significance of the differences in MAEs across our prediction models. MAEs in bold are MAEs of Tier 1 models with the lowest MAEs. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

Table 4
MAEs in Different Market States

	10-Year Forecasts				20-Year Forecasts			
	Return	Volatility	Amihud	Recession	Return	Volatility	Amihud	Recession
Panel A: Yield Alone, YLD_i								
YLD_{Div}	0.3597***	−0.1742**	−0.128	−0.0264	0.6678***	−0.3508***	0.1513	−0.0328
YLD_{Ttl}	0.3453***	−0.1515**	−0.0523	−0.0090	0.6543***	−0.3529***	0.1362	−0.0122
YLD_{NTtl}	0.3681***	−0.2419***	1.9298***	0.0352	0.4211***	−0.4575***	−0.3971	−0.0571
YLD_{CATY}	0.3455***	−0.1523*	−0.0512	0.0150	0.6620***	−0.3511***	0.1448	−0.0387
Panel B: Gordon, $GOR_{i,j} = YLD_i + g_j$								
$GOR_{Div,E} = YLD_{Div} + g_E$	0.3418***	−0.1640*	−0.0834	−0.0218	0.6437***	−0.3049***	0.1918	−0.0155
$GOR_{Div,Div} = YLD_{Div} + g_{Div}$	0.3632***	−0.1739**	−0.1129	−0.0234	0.6604***	−0.3530***	0.0797	−0.0397
$GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	0.3656***	−0.1695**	−0.0764	−0.0070	0.6964***	−0.4114***	0.0489	−0.0186
$GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	0.5573***	−0.3121***	−0.2384	−0.0657*	0.6105***	−0.4293***	−0.0477	−0.0434
$GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	0.3701***	−0.1761**	−0.0608	−0.0041	0.7309***	−0.4088***	0.0912	−0.0242
Panel C: Valuation Alone, ΔV_k								
ΔV_{TRCAPE}	0.2608***	−0.0962*	−0.1335	−0.0469	0.5284***	−0.2817***	0.0916	−0.0303
ΔV_{WPC}	−0.0320	−0.0158	−0.0978	−0.0156	0.4760***	−0.2117***	0.2646*	0.0305
ΔV_{BUF}	0.3612***	−0.3817***	−0.0165	−0.1594***	0.5898***	−0.6623***	0.0737	−0.0629
ΔV_{CON}	0.2255	−0.192	9.2521*	0.0144	0.3495	0.7750	11.9036	−1.8281***
ΔV_{EW}	0.2579***	−0.1236*	−0.1134	−0.0443	0.5595***	−0.2779***	0.1907	0.0004
ΔV_{IVW}	0.2120**	−0.0862	−0.1177	−0.0276	0.5264***	−0.2447***	0.2376*	0.0164
ΔV_{GR}	0.0818	−0.0618	−0.2222**	−0.0676*	0.5385***	−0.2826***	0.0425	−0.0619
ΔV_{BIC}	0.2016**	−0.0721	0.0797	0.0179	0.5284***	−0.2817***	0.0916	−0.0303
Panel D: Three Components, $GOR_{Div,Div} + \Delta V_k$								
$GOR_{Div,Div} + \Delta V_{TRCAPE}$	0.1708	−0.0071	0.0942	0.0099	0.3774***	−0.1098	0.3130***	0.0801*
$GOR_{Div,Div} + \Delta V_{WPC}$	−0.0480	0.0628	0.0246	0.0190	0.2944**	−0.0699	0.2759**	0.1105***
$GOR_{Div,Div} + \Delta V_{BUF}$	0.2814***	−0.3210***	0.2089	−0.1498***	0.4500***	−0.5137***	0.3743	0.0286
$GOR_{Div,Div} + \Delta V_{CON}$	0.2190	−0.1949	8.7832*	0.0154	0.6304**	1.2579**	19.0762**	−1.2821**

$(GOR_{Div,Div} + \Delta V_k)_{EW}$	0.1321	−0.0210	0.0317	−0.0058	0.3410***	−0.0991	0.2567**	0.0864**
$(GOR_{Div,Div} + \Delta V_k)_{IVW}$	0.1110	−0.0005	0.0824	0.0140	0.3450***	−0.1019	0.2333**	0.0765**
$(GOR_{Div,Div} + \Delta V_k)_{GR}$	0.0475	0.0087	−0.0335	−0.0157	0.2496**	−0.0397	0.1681	0.0543
$(GOR_{Div,Div} + \Delta V_k)_{BIC}$	0.1714	−0.0135	0.1521	0.0244	0.3774***	−0.1098	0.3130***	0.0801*

In Table 4, we report the MAEs in different market states. For each prediction model, we run the following time-series regression with Newey–West (1987) standard errors: $MAE_{i,t} = \alpha_i + \beta_i\, MKT\_STATE + \varepsilon_{i,t}$, where MAE_i is the 10- or 20-year mean absolute errors for prediction model i, and MKT_STATE is one of our four market state proxies, calculated over the same 10- or 20-year period as MAE_i. The four market state proxies include market return, market volatility, market Amihud (2002) ratio, and market recession. Market return and the Amihud (2002) ratio are the average annual market return and the average annual value-weighted stock Amihud (2002) ratio, respectively. Market volatility is the standard deviation of annual returns over the same period as MAE_i. The market recession proxy is determined by calculating the proportion of months (within a 10- or 20-year period) that fall within recessionary phases of the NBER business cycle. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

Table 5
Asset Allocation 10-Year Forecasts

	1891–2020			1955–2020			1988–2020		
	VaR	Alpha	Sharpe	VaR	Alpha	Sharpe	VaR	Alpha	Sharpe
[1] Historical Mean	−0.1621	0.0027	0.1933	−0.1912	0.0060	0.2618	−0.1912	0.0181	0.3181
Panel A: Yield Alone, YLD_i									
[2] YLD_{Div}	−0.0695***	0.0194***	0.3543**	−0.0695***	0.0173*	0.3116	−0.0552*	0.0332*	0.4747
[3] YLD_{Ttl}	−0.1183**	0.0073	0.2457	−0.0838***	0.0143	0.3043	−0.0397*	0.0288	0.4494
[4] YLD_{NTtl}	−0.0869	0.0096	0.3521	−0.1449	0.0128	0.2829	−0.0781	0.0273	0.3726
[5] YLD_{CATY}	−0.1128***	0.0050*	0.2357	−0.1128***	0.0125*	0.3031	−0.0368**	0.0252*	0.4439*
Panel B: Gordon, $GOR_{i,j} = YLD_i + g_j$									
[6] $GOR_{Div,E} = YLD_{Div} + g_E$	−0.0828***	0.0102**	0.2618*	−0.0882***	0.0195*	0.3228	−0.0373*	0.0423**	0.5045
[7] $GOR_{Div,Div} = YLD_{Div} + g_{Div}$	−0.0711***	0.0114*	0.2691*	−0.0752***	0.0227**	0.3489	−0.0477*	0.0469**	0.5500
[8] $GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	−0.0862***	0.0104***	0.2862**	−0.0748***	0.0152*	0.3192	−0.0319**	0.0286	0.4617
[9] $GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	−0.0826	0.0093	0.2541	−0.0781***	0.0144	0.2723	−0.0781*	0.0288	0.4005
[10] $GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	−0.0660***	0.0104**	0.2721*	−0.0698***	0.0210**	0.3630	−0.0319**	0.0411**	0.5675
Panel C: Valuation Alone, ΔV_k									
[11] ΔV_{TRCAPE}	−0.0789***	0.0126*	0.2751	−0.0947***	0.0150	0.2759	−0.0781*	0.0358	0.4278
[12] ΔV_{WPC}	−0.1314	0.0167	0.3198	−0.1107**	0.0269	0.3862	−0.0319**	0.0613**	0.7585**
[13] ΔV_{BUF}	−0.0688***	0.0332***	0.4393	−0.0688***	0.0284**	0.3368	−0.0781*	0.0481**	0.4366
[14] ΔV_{CON}	–	–	–	−0.2045	−0.0014	0.2810	−0.2045	0.0037	0.2756
[15] ΔV_{EW}	−0.1046*	0.0124*	0.2738	−0.0819***	0.0239**	0.3643	−0.0120**	0.0519**	0.6627*
[16] ΔV_{IVW}	−0.1183	0.0054	0.2359	−0.1015	0.0058	0.2441	−0.0781	0.0117	0.3000
[17] ΔV_{GR}	−0.1183	0.0108	0.2620	−0.1107**	0.0229	0.3494	−0.0426*	0.0527**	0.6396
[18] ΔV_{BIC}	−0.0789***	0.0183**	0.3306**	−0.0797***	0.0260	0.3833	−0.0319**	0.0613***	0.7585*
Panel D: Three Components, $GOR_{Div,Div} + \Delta V_k$									
[19] $GOR_{Div,Div} + \Delta V_{TRCAPE}$	−0.0781**	0.0171**	0.3108*	−0.0781***	0.0182	0.3057	−0.0414*	0.0449**	0.5311
[20] $GOR_{Div,Div} + \Delta V_{WPC}$	−0.1076	0.0195*	0.3285	−0.1107**	0.0264	0.3755	−0.0319**	0.0609**	0.7211*

[21] $GOR_{Div,Div} + \Delta V_{BUF}$	−0.0624***	0.0357***	0.4549	−0.0743***	0.0300**	0.3477	−0.0781*	0.0474**	0.4290
[22] $GOR_{Div,Div} + \Delta V_{CON}$	−	−	−	−0.1100	0.0119	0.3701	−0.1100	0.0165	0.3639
[23] $(GOR_{Div,Div} + \Delta V_k)_{EW}$	−0.0810**	0.0183**	0.3201**	−0.0781***	0.0207	0.3326	−0.0319**	0.0500**	0.6147*
[24] $(GOR_{Div,Div} + \Delta V_k)_{IVW}$	−0.1046*	0.0145**	0.2998*	−0.0892***	0.0133	0.2933	−0.0781**	0.0311	0.4529
[25] $(GOR_{Div,Div} + \Delta V_k)_{GR}$	−0.1107	0.0137*	0.2731	−0.1107**	0.0257	0.3699	−0.0319**	0.0611**	0.7231*
[26] $(GOR_{Div,Div} + \Delta V_k)_{BIC}$	−0.0781**	0.0199**	0.3344**	−0.0781***	0.0236	0.3534	−0.0319**	0.0583**	0.6885*

In Table 5, we compare different models from an asset allocation perspective for 10-year forecasts. We allocate the portfolio between stocks and bonds using data on the S&P 500 Index and the US 10-year government bond total return index. We employ the mean–variance approach and consider optimal portfolio weights as the asset weights that maximize the portfolio Sharpe ratio. Following the derivation in Smith (2019), we calculate optimal weights and rebalance the portfolio annually, based on the expected Sharpe ratios of the two assets, their historical standard deviations, and the historical correlation between them. To calculate the expected Sharpe ratio, $[E(R) – R_f]/\sigma$, for the S&P 500 (which serves as one input for determining optimal portfolio weights), we use our OOS S&P 500 return forecasts from each of our prediction models, historical risk-free rates sourced from the updated Goyal and Welch (2008) dataset, and historical standard deviations of S&P 500 returns. Accordingly, optimal weights and realized portfolio returns differ across our models. We then generate three performance metrics for realized portfolio returns: 5% value at risk (VaR), ex post alpha (Alpha), and ex post Sharpe ratio (Sharpe). We employ the aforementioned moving block bootstrap approach to bootstrap realized portfolio returns and determine whether realized VaR of portfolios constructed based on each of our prediction models is significantly lower than that based on the historical mean model. Similarly, we also examine whether realized alpha and Sharpe ratio of portfolios based on our prediction models are significantly higher than those based on the historical mean model. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

Table 6
Asset Allocation 20-Year Forecasts

	1891–2020			1955–2020			1988–2020		
	VaR	Alpha	Sharpe	VaR	Alpha	Sharpe	VaR	Alpha	Sharpe
[1] Historical Mean	−0.1344	−0.0021	0.2040	−0.1053	0.0025	0.3025	−0.1053	−0.0009	0.5063
Panel A: Yield Alone, YLD_i									
[2] YLD_{Div}	−0.1046	0.0039	0.2581	−0.0867***	0.0032	0.3018	−0.0722**	0.0047	0.5353
[3] YLD_{Ttl}	−0.1087	0.0043	0.2608	−0.0846***	0.0031	0.2956	−0.0674**	0.0051	0.5337
[4] YLD_{NTtl}	–	–	–	−0.0999	0.0101	0.3348	−0.0683	0.0372	0.6400
[5] YLD_{CATY}	−0.1046	0.0032	0.2517	−0.0912***	0.0025	0.2980	−0.0912*	0.0006	0.5121
Panel B: Gordon, $GOR_{i,j} = YLD_i + g_j$									
[6] $GOR_{Div,E} = YLD_{Div} + g_E$	−0.0774***	0.0052	0.2495	−0.0774**	0.0076	0.3046	−0.0373	0.0202	0.5365
[7] $GOR_{Div,Div} = YLD_{Div} + g_{Div}$	−0.0752***	0.0058	0.2504	−0.0752**	0.0102	0.3161	−0.0493	0.0252	0.5309
[8] $GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	−0.0748**	0.0082**	0.2794*	−0.0690***	0.0085	0.3137	−0.0319*	0.0160	0.5323
[9] $GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	−0.0826*	0.0052	0.2471	−0.0656**	0.0088	0.2625	−0.0781	0.0242	0.4561
[10] $GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	−0.0698***	0.0061*	0.2529	−0.0698***	0.0115	0.3311	−0.0319*	0.0234	0.5558
Panel C: Valuation Alone, ΔV_k									
[11] ΔV_{TRCAPE}	−0.1046	0.0058	0.2623	−0.0846**	0.0060	0.2957	−0.0781	0.0284	0.5233
[12] ΔV_{WPC}	−0.1107	0.0203***	0.4150***	−0.1107	0.0078	0.3354	−0.0359***	0.0188	0.6061
[13] ΔV_{BUF}	−0.1083	0.0079	0.3544	−0.1083	0.0082	0.3301	−0.0319**	0.0264	0.6211
[14] ΔV_{CON}	–	–	–	–	–	–	−0.0257**	0.0341**	0.7264**
[15] ΔV_{EW}	−0.1112	0.0095	0.2980	−0.0923***	0.0108	0.3490	−0.0319***	0.0372	0.7081
[16] ΔV_{IVW}	−0.1183	0.0096	0.2963	−0.0977***	0.0106	0.3414	−0.0319**	0.0396	0.6799
[17] ΔV_{GR}	−0.1046	0.0093	0.2986	−0.0869***	0.0119**	0.3665**	−0.0319***	0.0333*	0.7427
[18] ΔV_{BIC}	−0.1046	0.0058	0.2623	−0.0846***	0.0060	0.2957	−0.0781	0.0284	0.5233
Panel D: Three Components, $GOR_{i,j} + \Delta V_k$									
[19] $GOR_{Div,Div} + \Delta V_{TRCAPE}$	−0.1046	0.0184*	0.3669*	−0.0920**	0.0059	0.3123	−0.0319*	0.0380	0.6897
[20] $GOR_{Div,Div} + \Delta V_{WPC}$	−0.1046	0.0221*	0.4272**	−0.0981**	0.0057	0.3194	−0.0319***	0.0286	0.6549

[21] $GOR_{Div,Div} + \Delta V_{BUF}$	−0.0806**	0.0056	0.3460	−0.0806**	0.0052	0.3145	−0.0319*	0.0259	0.6515
[22] $GOR_{Div,Div} + \Delta V_{CON}$	−	−	−	−	−	−	−0.0093**	0.0405*	0.7753*
[23] $(GOR_{Div,Div} + \Delta V_k)_{EW}$	−0.1046	0.0193*	0.3761**	−0.0893**	0.0066	0.3271	−0.0319***	0.0334	0.7145
[24] $(GOR_{Div,Div} + \Delta V_k)_{IVW}$	−0.1046	0.0193*	0.3760**	−0.0869**	0.0067	0.3275	−0.0319***	0.0339	0.7158
[25] $(GOR_{Div,Div} + \Delta V_k)_{GR}$	−0.1046	0.0197*	0.3798**	−0.0981**	0.0074	0.3390	−0.0319***	0.0300*	0.7226
[26] $(GOR_{Div,Div} + \Delta V_k)_{BIC}$	−0.1046	0.0184*	0.3669*	−0.0920**	0.0059	0.3123	−0.0319*	0.0380	0.6897

In Table 6, we compare different models from an asset allocation perspective for 20-year forecasts. We allocate the portfolio between stocks and bonds using data on the S&P 500 Index and the US 10-year government bond total return index. We employ the mean–variance approach and consider optimal portfolio weights as the asset weights that maximize the portfolio Sharpe ratio. Following the derivation in Smith (2019), we calculate optimal weights and rebalance the portfolio annually, based on the expected Sharpe ratios of the two assets, their historical standard deviations, and the historical correlation between them. To calculate the expected Sharpe ratio, $[E(R) – R_f]/\sigma$, for the S&P 500 (which serves as one input for determining optimal portfolio weights), we use our OOS S&P 500 return forecasts from each of our prediction models, historical risk-free rates sourced from the updated Goyal and Welch (2008) dataset, and historical standard deviations of S&P 500 returns. Accordingly, optimal weights and realized portfolio returns differ across our models. We then generate three performance metrics for realized portfolio returns: 5% value at risk (VaR), ex post alpha (Alpha), and ex post Sharpe ratio (Sharpe). We employ the aforementioned moving block bootstrap approach to bootstrap realized portfolio returns and determine whether realized VaR of portfolios constructed based on each of our prediction models is significantly lower than that based on the historical mean model. Similarly, we also examine whether realized alpha and Sharpe ratio of portfolios based on our prediction models are significantly higher than those based on the historical mean model. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

Appendix 1
Predictor Description

Model	Description
Historical Mean	Historical average return on the S&P 500 Index is used as the forecast.
YLD_{Div}	We use the log dividend yield (e.g., Ferreira and Santa-Clara, 2011; Rapach, Ringgenberg, and Zhou, 2016) as the predictor to compute a standard predictive regression forecast. Dividend yield in year t is calculated as dividend in year t divided by the S&P 500 Index value at the end of year $t-1$.
YLD_{Ttl}	We use the log total yield as the predictor to compute a standard predictive regression forecast. Total yield is the dividend yield plus the buyback yield as per Straehl and Ibbotson (2017).
YLD_{NTtl}	We use the log net total yield as the predictor to compute a standard predictive regression forecast. Net total yield is the dividend yield less net issuance as per Straehl and Ibbotson (2017).
YLD_{CATY}	We use the cyclically adjusted total yield (CATY) as the predictor to compute a standard predictive regression forecast. CATY is computed as per Straehl and Ibbotson (2017).
$GOR_{Div,E} = YLD_{Div} + g_E$	The Gordon growth model with log dividend yield and average historical growth in earnings is used to compute a return forecast.
$GOR_{Div,Div} = YLD_{Div} + g_{Div}$	The Gordon growth model with log dividend yield and average historical growth in dividends is used to compute a return forecast.
$GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	The Gordon growth model with log total yield and average historical total yield growth is used to compute a return forecast. To calculate average historical total yield growth, we first multiply yearly total yield in year t by the S&P 500 Index value at the end of year $t-1$, and then we calculate the average historical growth rate in the resulting time series. .
$GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	The Gordon growth model with log net total yield and average historical total yield growth is used to compute a return forecast.
$GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	The Gordon growth model with CATY and average historical CATY growth is used to compute a return forecast. To calculate average historical CATY growth, we first multiply yearly CATY in year t by the S&P 500 Index value at the end of year $t-1$, and then we calculate the average historical growth rate in the resulting time series.
ΔV_{TRCAPE}	We use the total return cyclically adjusted price-to-earnings ratio (TRCAPE) as the predictor to compute a standard predictive regression forecast. The TRCAPE data are sourced from the Shiller website: http://www.econ.yale.edu/~shiller/data.htm
ΔV_{WPC}	We use the wealth portfolio composition (WPC) indicator as the predictor to compute a standard predictive regression forecast. The WPC indicator is calculated as per Appendix A of Rintamaki (2023) with data collected from the Jordà-Schularick-Taylor Macrohistory Database.
ΔV_{BUF}	We use the Buffett indicator, the equity market capitalization scaled by gross domestic product, as the predictor to compute a standard predictive regression forecast.
ΔV_{CON}	We use the detrended cyclical consumption as the predictor to compute a standard predictive regression forecast (e.g., Atanasov, Møller, and Priestley, 2020). We detrend consumption data in quarter q as per Eq. (2) of Atanasov, Møller, and Priestley (2020), and we use only the data available up to quarter q to avoid look-forward bias from the detrending process.
ΔV_{EW}	We calculate the simple average of the forecasts of ΔV_{TRCAPE}, ΔV_{WPC}, ΔV_{BUF}, and ΔV_{CON} in each year and form a composite forecast.

ΔV_{IVW}	We calculate the inverse variance-weighted average of the forecasts of ΔV_{TRCAPE}, ΔV_{WPC}, ΔV_{BUF}, and ΔV_{CON} in each year (e.g., Bates and Granger, 1969) and form a composite forecast.
ΔV_{GR}	We combine the forecasts of ΔV_{TRCAPE}, ΔV_{WPC}, ΔV_{BUF}, and ΔV_{CON} in each year using the Granger and Ramanathan (1984) constrained regression approach. To form the composite forecast, we regress actual returns on the forecasts of ΔV_{TRCAPE}, ΔV_{WPC}, ΔV_{BUF}, and ΔV_{CON} with no constant, subject to the constraints that all regression beta coefficients are non-negative and the sum of beta coefficients is equal to 1. Next, the beta coefficients are used as the weights assigned to the forecasts of ΔV_{TRCAPE}, ΔV_{WPC}, ΔV_{BUF}, and ΔV_{CON}.
ΔV_{BIC}	We use the Bayesian model averaging method (e.g., Min and Zellner, 1993) to combine the forecasts of ΔV_{TRCAPE}, ΔV_{WPC}, ΔV_{BUF}, and ΔV_{CON}, and form a composite forecast.
$GOR_{Div,Div} + \Delta V_{TRCAPE}$	The return forecast is the sum of the Gordon growth model (with dividend yield and dividend growth) forecast and expected change in valuations. We use TRCAPE as the predictor to forecast the OOS change in valuations. The TRCAPE data are sourced from the Shiller website: http://www.econ.yale.edu/~shiller/data.htm
$GOR_{Div,Div} + \Delta V_{WPC}$	The return forecast is the sum of the Gordon growth model (with dividend yield and dividend growth) forecast and expected change in valuations. We use WPC as the predictor to forecast the OOS change in valuations. The WPC indicator is calculated as per Appendix A of Rintamaki (2023).
$GOR_{Div,Div} + \Delta V_{BUF}$	The return forecast is the sum of the Gordon growth model (with dividend yield and dividend growth) and expected change in valuations. We use the Buffett indicator, the equity market capitalization scaled by gross domestic product, as the predictor to forecast the OOS change in valuations.
$GOR_{Div,Div} + \Delta V_{CON}$	The return forecast is the sum of the Gordon growth model (with dividend yield and dividend growth) and expected change in valuations. We use the detrended cyclical consumption (e.g., Atanasov, Møller, and Priestley, 2020) as the predictor to forecast the OOS change in valuations. We detrend consumption data in quarter *q* as per Eq. (2) of Atanasov, Møller, and Priestley (2020), and we use only the data available up to quarter *q* to avoid look-forward bias from the detrending process.
$(GOR_{Div,Div} + \Delta V_k)_{EW}$	We calculate the simple average of the forecasts of $(GOR_{Div,Div} + \Delta V_{TRCAPE})$, $(GOR_{Div,Div} + \Delta V_{WPC})$, $(GOR_{Div,Div} + \Delta V_{BUF})$ and $(GOR_{Div,Div} + \Delta V_{CON})$ in each year and form a composite forecast.
$(GOR_{Div,Div} + \Delta V_k)_{IVW}$	We calculate the inverse variance-weighted average of the forecasts of $(GOR_{Div,Div} + \Delta V_{TRCAPE})$, $(GOR_{Div,Div} + \Delta V_{WPC})$, $(GOR_{Div,Div} + \Delta V_{BUF})$ and $(GOR_{Div,Div} + \Delta V_{CON})$ in each year and form a composite forecast.
$(GOR_{Div,Div} + \Delta V_k)_{GR}$	We combine the forecasts of $(GOR_{Div,Div} + \Delta V_{TRCAPE})$, $(GOR_{Div,Div} + \Delta V_{WPC})$, $(GOR_{Div,Div} + \Delta V_{BUF})$ and $(GOR_{Div,Div} + \Delta V_{CON})$ in each year using the Granger and Ramanathan (1984) constrained regression approach. To calculate the composite forecast, we regress actual returns on the forecasts of $(GOR_{Div,Div} + \Delta V_{TRCAPE})$, $(GOR_{Div,Div} + \Delta V_{WPC})$, $(GOR_{Div,Div} + \Delta V_{BUF})$ and $(GOR_{Div,Div} + \Delta V_{CON})$ with no constant, subject to the constraints that all regression beta coefficients are non-negative, and the sum of beta coefficients is equal to 1. Then the beta coefficients are used as the weights assigned to the forecasts of $(GOR_{Div,Div} + \Delta V_{TRCAPE})$, $(GOR_{Div,Div} + \Delta V_{WPC})$, $(GOR_{Div,Div} + \Delta V_{BUF})$ and $(GOR_{Div,Div} + \Delta V_{CON})$.
$(GOR_{Div,Div} + \Delta V_k)_{BIC}$	We use the Bayesian model averaging method (Min and Zellner, 1993) to combine the forecasts of $(GOR_{Div,Div} + \Delta V_{TRCAPE})$, $(GOR_{Div,Div} + \Delta V_{WPC})$, $(GOR_{Div,Div} + \Delta V_{BUF})$ and $(GOR_{Div,Div} + \Delta V_{CON})$ and form a composite forecast.

We conduct an ordinary least squares (OLS) out-of-sample (OOS) forecasting approach (e.g., Goyal and Welch, 2008). All OOS forecasts use only the data available up to the year in which the forecast is calculated.

Appendix 2
Correlations and MAEs

Year	Actual Return	Rule 1 Predicted Return	Rule 2 Predicted Return	Rule 1 Absolute Error	Rule 2 Absolute Error
1	6.00%	8.00%	6.00%	2.00%	0.00%
2	−8.00%	−6.00%	−5.00%	2.00%	3.00%
3	1.00%	3.00%	0.00%	2.00%	1.00%
4	12.00%	14.00%	12.00%	2.00%	0.00%
5	8.00%	10.00%	8.00%	2.00%	0.00%
6	7.00%	9.00%	8.00%	2.00%	1.00%
Pearson Correlation		1.0000	0.9854		
Spearman Correlation		1.0000	0.9786		
Mean Absolute Error				2.00%	0.83%

This table presents the correlations and MAEs of two hypothetical return prediction rules. These rule returns demonstrate differences between correlations and MAEs.

Appendix 3
Average MAEs

	1891–2020	1955–2020	1988–2020
Panel A: 10-Year Forecasts			
Yield Alone	0.0413	0.0377	0.0361
Gordon	0.0433	0.0422	0.0408
Valuation Alone	0.0321	0.0349	0.0393
Three Components	0.0317	0.0311	0.0356
Three Components versus:			
	Diebold–Mariano Statistics		
Yield Alone	−4.82***	−3.99***	−0.19
Gordon	−4.90***	−5.34***	−2.11**
Valuation Alone	−0.26	−3.96***	−3.52***
Panel B: 20-Year Forecasts			
Yield Alone	0.0284	0.0233	0.0096
Gordon	0.0354	0.0348	0.0165
Valuation Alone	0.0253	0.0208	0.0145
Three Components	0.0255	0.0190	0.0144
Three Components versus:			
	Diebold–Mariano Statistics		
Yield Alone	−0.58	−2.53**	1.75
Gordon	−3.84***	−7.81***	−1.03
Valuation Alone	0.40	−1.86*	0.15

This table presents the average MAEs for each of the four forecasting frameworks discussed in Section 2. We also employ the Diebold–Mariano (DM) test to evaluate the statistical differences in MAEs between the three-component framework and the other three frameworks. Panel A shows the results for 10-year forecasts; Panel B shows the results for 20-year forecasts. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

Appendix 4
MSEs in Different Market States

	10-Year Forecasts				20-Year Forecasts			
	Return	Volatility	Amihud	Recession	Return	Volatility	Amihud	Recession
Panel A: Yield Alone, YLD_i								
YLD_{Div}	0.0375***	−0.0168**	−0.0115	−0.0031	0.0482***	−0.0259***	0.0079	−0.0018
YLD_{Ttl}	0.0345***	−0.0136**	−0.0053	−0.0003	0.0464***	−0.0258***	0.0061	−0.0008
YLD_{NTtl}	0.0355***	−0.0195***	0.2341***	0.0045	0.0246***	−0.0270***	−0.0231	−0.0035
YLD_{CATY}	0.0350***	−0.0140*	−0.0036	0.0014	0.0492***	−0.0266***	0.0074	−0.0020
Panel B: Gordon, $GOR_{i,j} = YLD_i + g_j$								
$GOR_{Div,E} = YLD_{Div} + g_E$	0.0345***	−0.0153*	−0.0065	−0.0020	0.0508***	−0.0250***	0.0143	−0.0010
$GOR_{Div,Div} = YLD_{Div} + g_{Div}$	0.0380***	−0.0169**	−0.0089	−0.0023	0.0537***	−0.0294***	0.0062	−0.0023
$GOR_{Ttl,Ttl} = YLD_{Ttl} + g_{Ttl}$	0.0372***	−0.0158**	−0.0063	−0.0003	0.0552***	−0.0321***	0.0041	−0.0011
$GOR_{NTtl,Ttl} = YLD_{NTtl} + g_{Ttl}$	0.0576***	−0.0304***	−0.0209	−0.0053	0.0622***	−0.0423***	−0.0037	−0.0037
$GOR_{CATY,CATY} = YLD_{CATY} + g_{CATY}$	0.0390***	−0.0170**	−0.0035	0.0000	0.0613***	−0.0348***	0.0059	−0.0017
Panel C: Valuation Alone, ΔV_k								
ΔV_{TRCAPE}	0.0254***	−0.0083*	−0.0081	−0.0035	0.0344***	−0.0180***	0.0075	−0.0003
ΔV_{WPC}	−0.0011	−0.0010	−0.0048	−0.0010	0.0308***	−0.0132***	0.0162*	0.0026
ΔV_{BUF}	0.0371***	−0.0393***	−0.0124	−0.0175***	0.0364***	−0.0401***	−0.0131	−0.0043
ΔV_{CON}	0.0573	−0.0494	1.5557*	−0.0079	0.0173	0.0374	0.5203	−0.0920***
ΔV_{EW}	0.0205**	−0.0094*	−0.0102	−0.0046	0.0362***	−0.0179***	0.0112	0.0009
ΔV_{IVW}	0.0186**	−0.0072	−0.0090	−0.0031	0.0342***	−0.0158***	0.0138	0.0017
ΔV_{GR}	0.0089	−0.0054	−0.0161**	−0.0058*	0.0362***	−0.0187***	0.0055	−0.0016
ΔV_{BIC}	0.0190**	−0.0054	0.0087	0.0021	0.0344***	−0.0180***	0.0075	−0.0003
Panel D: Three Components, $GOR_{Div,Div} + \Delta V_k$								
$GOR_{Div,Div} + \Delta V_{TRCAPE}$	0.0134	0.0020	0.0142	0.0028	0.0237***	−0.0080	0.0175**	0.0050*
$GOR_{Div,Div} + \Delta V_{WPC}$	−0.0027	0.0037	0.0020	0.0013	0.0180**	−0.0054	0.0143**	0.0060***
$GOR_{Div,Div} + \Delta V_{BUF}$	0.0297***	−0.0327***	0.0318	−0.0150***	0.0240***	−0.0272***	0.0134	0.0012
$GOR_{Div,Div} + \Delta V_{CON}$	0.0487	−0.0419	1.2584	−0.0080	0.0321**	0.0681**	0.9994**	−0.0726***

$(GOR_{Div,Div} + \Delta V_k)_{EW}$	0.0102	−0.0009	0.0020	−0.0006	0.0214***	−0.0074	0.0145**	0.0053**
$(GOR_{Div,Div} + \Delta V_k)_{IVW}$	0.0085	0.0016	0.0071	0.0015	0.0220***	−0.0079	0.0131*	0.0049**
$(GOR_{Div,Div} + \Delta V_k)_{GR}$	0.0074	−0.0016	−0.0053	−0.0025	0.0171**	−0.0044	0.0104	0.0039*
$(GOR_{Div,Div} + \Delta V_k)_{BIC}$	0.0129	0.0019	0.0178	0.0039	0.0237***	−0.0080	0.0175**	0.0050*

In Appendix 4, we report the MSEs in different markets states. For each prediction model, we run the following time-series regression with Newey–West (1987) standard errors: $MSE_{i,t} = \alpha_i + \beta_i\ MKT\_STATE + \varepsilon_{i,t}$, where MSE_i is the 10- or 20-year mean squared errors for prediction model i, and MKT_STATE is one of our four market state proxies, calculated over the same 10- or 20-year period as MSE_i. The four market state proxies include market return, market volatility, market Amihud (2002) ratio, and market recession. Market return and the Amihud (2002) ratio are the average annual market return and the average annual value-weighted stock Amihud (2002) ratio, respectively. Market volatility is the standard deviation of annual returns over the same period as MSE_i. The market recession proxy is determined by calculating the proportion of months (within a 10- or 20-year period) that fall within recessionary phases of the NBER business cycle. ***, **, and * indicate the significance levels of 1%, 5%, and 10%, respectively.

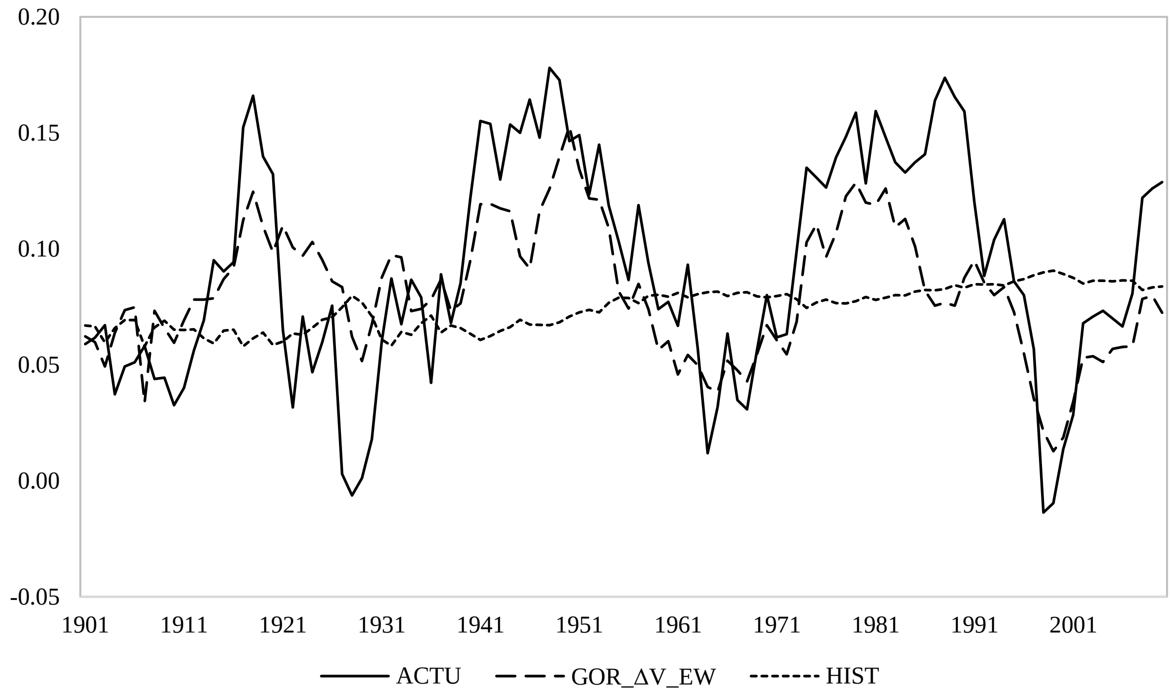

Figure 1a: 10-Year Forecasts. This figure plots the annualized 10-year forecasts based on the three-component approach with equal weights assigned to proxies for ΔV, along with historical mean returns and actual annualized 10-year returns.

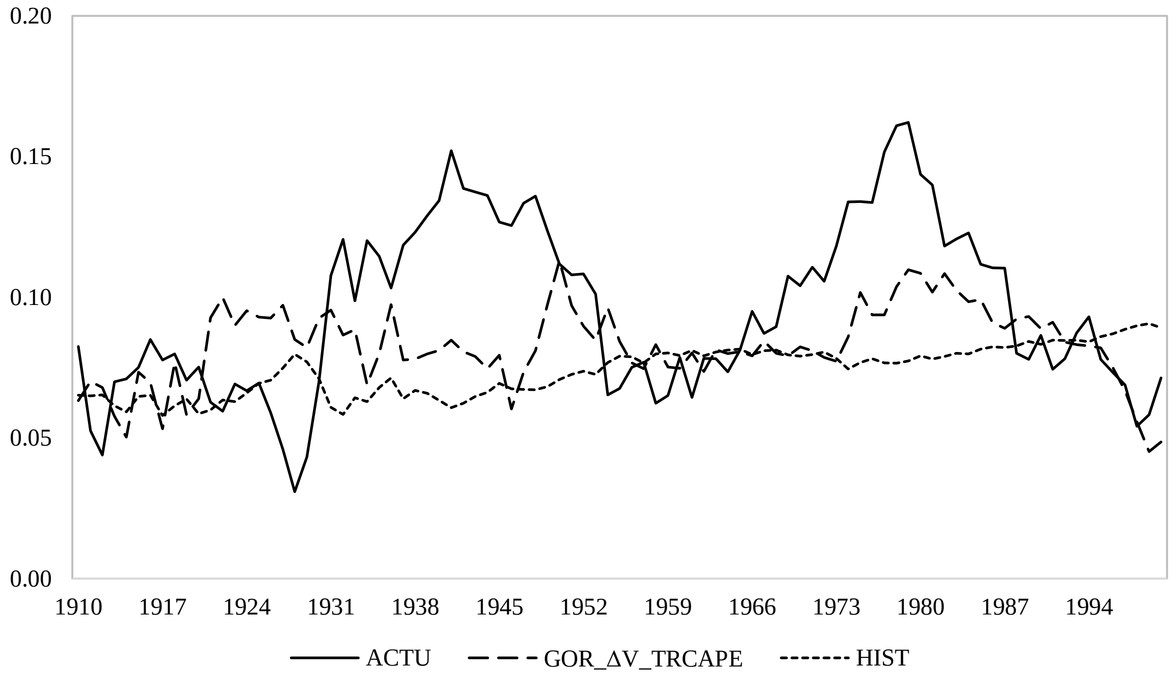

Figure 1b: 20-Year Forecasts. This figure plots the annualized 20-year forecasts based on the three-component approach with the TRCAPE proxy for ΔV, along with historical mean returns and actual annualized 20-year returns.

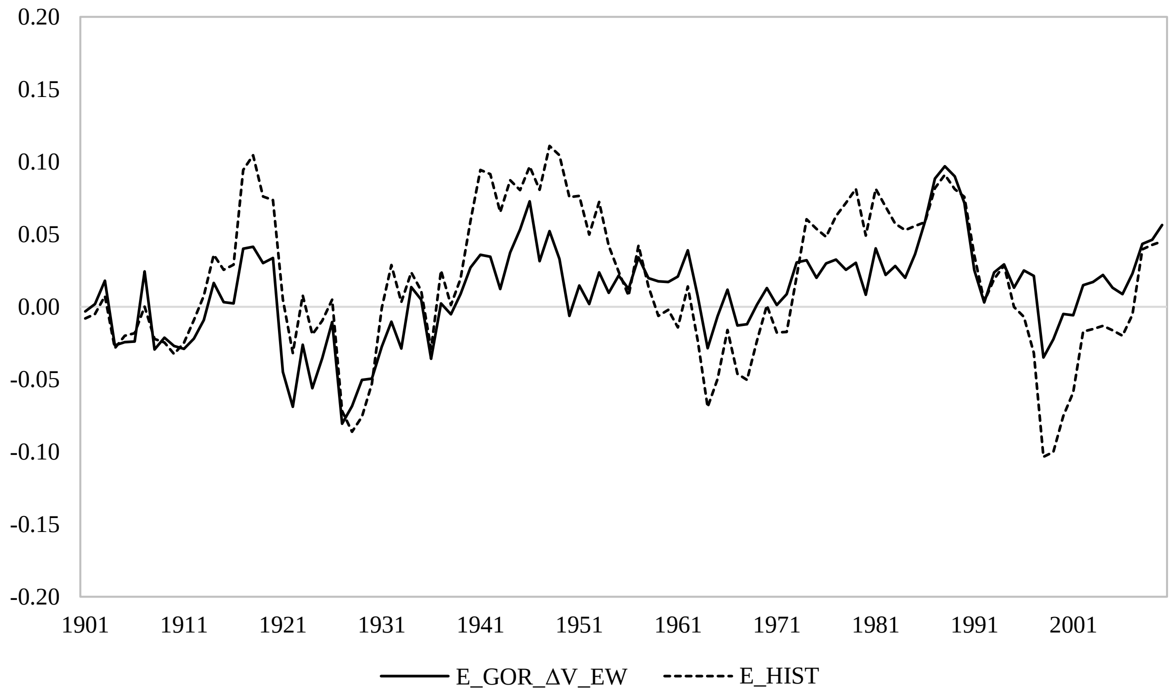

Figure 2a: 10-Year Forecast Errors. This figure plots the forecast errors for 10-year forecasts based on the three-component approach with equal weights assigned to proxies for ΔV, along with forecast errors for the historical mean model.

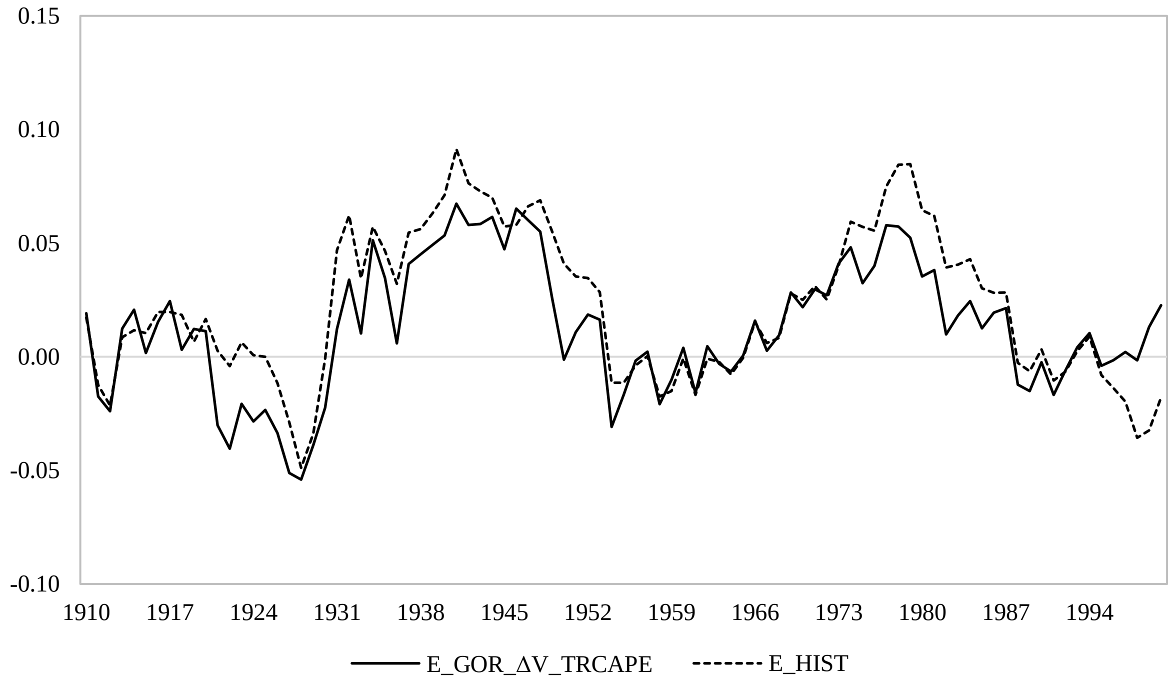

Figure 2b: 20-Year Forecast Errors. This figure plots the forecast errors for 20-year forecasts based on the three-component approach with the TRCAPE proxy for ΔV, along with forecast errors for the historical mean model.

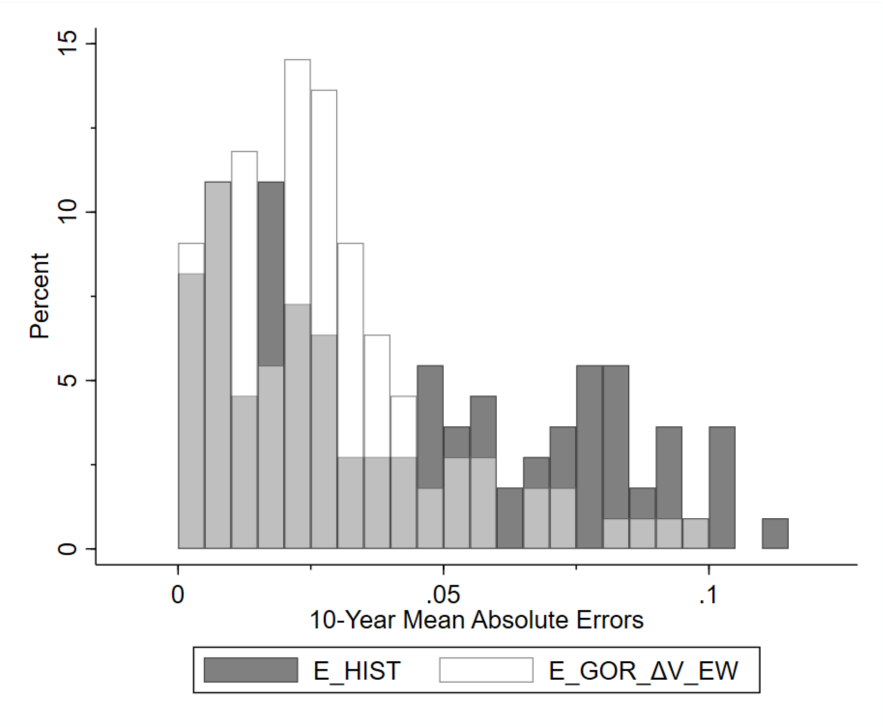

Figure 3a: 10-Year Absolute Forecast Errors. This figure displays the distributions of absolute forecast errors for two models: the three-component model with equal weights assigned to proxies for ΔV (depicted in white), and the historical mean model (depicted in dark gray). Areas of overlap between the two distributions are shown in light gray.

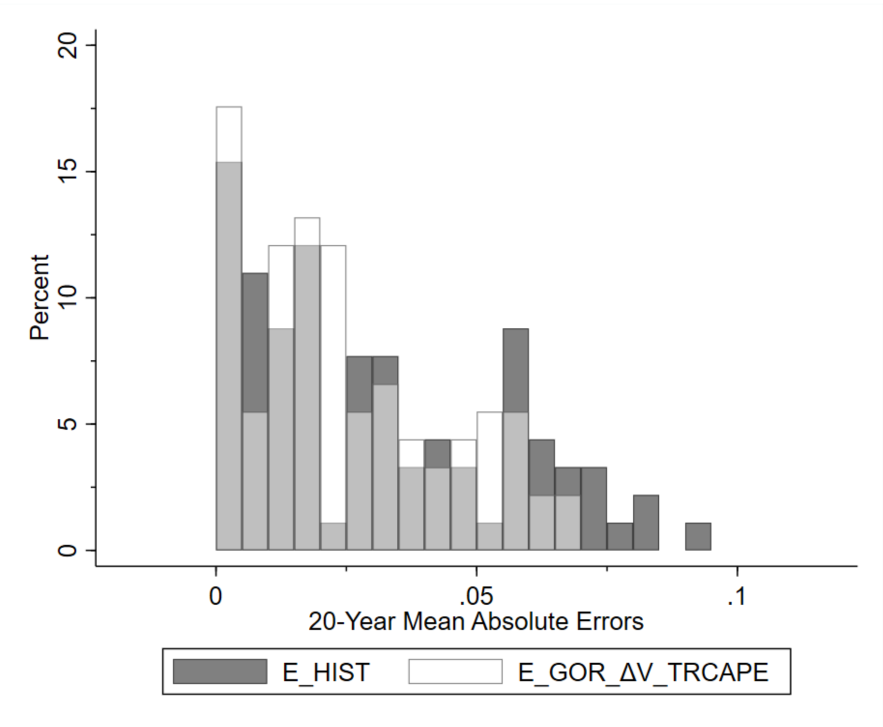

Figure 3b: 20-Year Absolute Forecast Errors. This figure displays the distributions of absolute forecast errors for two models: the three-component model with the TRCAPE proxy for ΔV (depicted in white), and the historical mean model (depicted in dark gray). Areas of overlap between the two distributions are shown in light gray.