Long-Horizon Stock Valuation and Return Forecasts Based on Demographic Projections *

Chaoyi Chen[‡]
Magyar Nemzeti Bank & John von Neumann University

Nikolay Gospodinov[§]
Federal Reserve Bank of Atlanta

Alex Maynard[¶]
University of Guelph

Elena Pesavento[‖]
Emory University

April 27, 2022

Abstract

We incorporate low-frequency information from demographic variables into a simple predictive model to forecast stock valuations and returns using demographic projections. The demographics appear to be an important determinant of stock valuations, such as the dividend-price ratio. The availability of long-term demographic projections allows us to provide (very) long-horizon forecasts of stock market valuations and returns. We also exploit the strong contemporaneous correlation between returns and valuations to improve return forecasts – something which is not possible in a predictive regression with only lagged predictors. Extensive pseudo out-of-sample forecast comparisons and tests demonstrate the predictive value that an accurate demographic projection can deliver. Although the availability of historical Census Bureau projections is limited, we demonstrate that they could have been employed in real time to improve true long-horizon stock return prediction. We show how the model can be used to adjust predictions under alternative demographic assumptions, incorporating, for example, the demographic impact of COVID-19 or recent changes to immigration policy.

*The views expressed here are the authors'and not necessarily those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the Magyar Nemzeti Bank (Central Bank of Hungary). This paper has benefited greatly from extensive comments by an anonymous referee. We also thank Stephen Bonnar, Ibrahim Jamali, Jaideep Oberoi, Adriana Robertson, participants at the 2016 Waterloo International Workshop on the Implications of Aging on Asset Values and the participants at the 2017 Canadian Economic Association Meeting for useful comments and suggestions. We thank the SSHRC for research funding under grant number 430417.

[†]Magyar Nemzeti Bank (Central Bank of Hungary), Budapest, 1054, Hungary; Email: chenc@mnb.hu.

[‡]MNB Institute, John von Neumann University, Budapest, 1117, Hungary.

[§]Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St. N.E., Atlanta, Georgia, 30309–4470, USA; Email: nikolay.gospodinov@atl.frb.org.

[¶]Department of Economics and Finance, Gordon S. Lang School of Business and Economics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: maynarda@uoguelph.ca.

[‖]Department of Economics, Emory University, Atlanta, Georgia, 30322–2240, USA; Email: epesave@emory.edu.

1 Introduction

Dating back to seminal work on tests of stock market predictability (Fama and French (1988); Campbell and Shiller (1988a,b); Hodrick (1992)), there is a long empirical tradition of using valuation ratios to predict stock returns at both short and long horizons. Nonetheless, the apparently strong predictive results from the early literature have since been questioned on several grounds and subjected to further scrutiny. First, it has been widely documented that the Stambaugh (1999) bias, and its resulting size distortion, may overstate the statistical significance at short horizons (Mankiw and Shapiro (1986); Stambaugh (1986); Cavanagh et al. (1995); Stambaugh (1999)). The inference distortion becomes even more pronounced at longer horizons (Valkanov (2003); Boudoukh et al. (2008); Hjalmarsson (2011)). Meanwhile, Goyal and Welch (2003, 2008) demonstrate that the evidence on out-of-sample predictive power of valuation ratios is tenuous and unstable.

Another strand of literature argues that demographic dependency-type ratios may explain some of the long-term trends in asset market valuations. This is supported by both theory (Yoo (1994); Brooks (2002); Poterba (2004); Geanakoplos et al. (2004)) and empirical work (Favero et al. (2011); Liu and Spiegel (2011)). Savings rates and possibly risk preferences vary substantially over the life-cycle, with savings rates peaking in middle age and then being drawn down in old age. These savings directly impact the pool of funds available for investment in the stock market.

Recent work by Favero et al. (2011) connects these two strands of literature by employing lagged demographic ratios in a predictive regression context. They find that the persistence in valuation ratios stems from slowly evolving demographic trends. By removing this demographic component, they obtain an adjusted dividend-price ratio which exhibits less persistence – thereby addressing the Stambaugh (1999) bias – and better out-of-sample forecast ability. In essence, Favero et al. (2011) improve the unconditional stock return forecasts by including both the lagged dividend-price ratio and lagged demographic information as predictors.

Liu and Spiegel (2011) also employ demographic projections to produce very long-horizon forecasts of stock price valuations. Unlike Favero et al. (2011), they do not employ a predictive regression-type framework to forecast returns. Instead, they use the demographic variables to predict the earnings-price ratio and to forecast separately earnings, and then back out future price changes. This approach has the advantage of imposing more structure on the forecasts, but also relies on the accuracy of very long-horizon earnings forecasts, which may be inherently harder to predict than demographics.

The point of departure that underlies these arguments is the potential low-frequency co-movement in stock returns, valuation ratios and demographic variables. To evaluate informally the empirical support of such a co-movement, we use annual data for the period 1946-2016 for the S&P500 returns, changes in S&P500 price-dividend ratio, changes in S&P500 dividend yield, and changes in S&P500 earnings-price ratio. As in Bai and Ng (2004), we extract the first principal component of these series and then integrate (and linearly detrend) the resulting process. We interpret this as the common factor in stock returns and valuation ratios. We also use annual demographic data from the Census Bureau and construct the middle-young (MY) ratio as the ratio of middle-aged (40-49) and young (20-29) cohorts (see Section 3 for more details). The common factor in stock prices and the middle-young ratio, including its projections by the Census Bureau until 2040, are plotted in Figure 1. It is striking how closely this demographic variable matches the common dynamics in the stock returns and the stock valuation ratios.

In light of this evidence, we propose a simple forecast model and demonstrate its usefulness for prediction of stock return valuations and returns based on demographic projections. Our forecast model capitalizes on two insights from the previous literature. First, the innovations to valuations and returns are strongly contemporaneously correlated. For example, an increase in the dividend-price ratio due a fall in stock prices automatically implies a negative return. Thus, any variable that is predictive for future dividend-price innovation is likely to have predictive power for returns.

Secondly, the aforementioned literature on the potential impact of demographics on financial markets suggests that future stock valuations depend, in part, on future trends in demographic ratios. A special property of demographic ratios is that they can be predicted with reasonable accuracy at very long horizons. While birth rates, death rates, and immigration rates can all change in unexpected ways – e.g., due to a health shock (COVID-19) or immigration policy shift – the dominant changes to demographic ratios are often simply due to the entirely predictable aging of the existing population. We therefore use the demographic projections of the Census Bureau to project future stock valuations, which are in turn used to predict returns. Because demographic projections are available far into the future, this can be done at very long horizons. By incorporating future projected demographics into the model, we contribute to a growing literature on conditional forecasting (see Waggoner and Zha (1999), Faust and Wright (2008), among others).

While we build in several ways on Favero et al. (2011), one key difference is that we exploit projections of <u>future</u> demographic changes to construct and include forecasts of <u>future</u> stock valuations in our predictive model. This exploits information available in the official forecasts, especially when forecasting at long, multi-year horizons, which begin to match the frequency at which demographic changes may occur.

Secondly, we show, by incorporating future demographic projections, that our model is well-suited to provide forecasts which are contingent on either the official demographic forecasts or alternative scenario forecasts. This type of scenario analysis can be useful to researchers, policy makers and pension fund managers who are interested in the implications of demographic projections for market returns and valuations. For example, we use the model to explore the return forecast implications of alternative demographic projections regarding the impact of COVID-19 and changing immigration policies on future demographics.

Thirdly, we compile and employ the limited number of actual historical demographic forecasts recorded by the Census Bureau in order to explore whether the projections can help improve true predictions of returns in real time. The historical forecasts available for this exercise are somewhat restricted and entail some gaps. Nonetheless, we find that the historical middle-to-young age ratio forecasts were predictive for returns at the five-year horizon out-of-sample, especially when using rolling sample forecast methods. This demonstrates that the predictive value of (projected) future demographic information is not simply due to look-ahead bias and thus supports the use of demographic projections in real-time, long-horizon return forecasting. Indeed, we find that a model that includes Census Bureau projections of future demographics provides more accurate forecasts than a model with only current demographic information.

Finally, we employ recent econometric methods that account for the persistence of the low-frequency predictors. While these methods have been applied before to valuation ratios, we are not aware of their previous application to demographic predictors.

The rest of the paper is organized as follows. Section 2 presents and discusses the empirical forecasts models. Section 3 describes the data and provides in-sample results. Section 4 provides both pseudo and true out-of-sample (OOS) analysis on historical data for forecasts made using either the ex-post demographic measurements (pseudo OOS) or the latest ex-ante historical Census forecast available at the time of the forecast (true OOS). Section 5 presents valuation ratio and return forecasts at very long horizons and studies their sensitivity to alternative assumptions on immigration and the demographic impact of COVID-19. Section 6 concludes. Some technical details on the justification and derivation of our proposed demographic-augmented model are collected in the Appendix. The main tables and figures are included at the end of the paper, with additional results relegated to a separate not-for-publication appendix.

2 Empirical Forecast Models

The empirical forecast models that we consider in this paper are the classical predictive regression of future returns using the *current* valuation ratio as a predictor as well as two predictive models that are augmented with the current or future projected values of demographic ratios. To introduce the main notation and intuition behind our augmented predictive regression approach, it is instructive to start with the standard k-period (long-horizon) predictive regression model given by

$$r_{t+k}(k) = \beta_0^{pr}(k) + \beta_1^{pr}(k)x_t + \varepsilon_{1,t+k}^{pr}(k), \tag{1}$$

where $r_{t+k}(k) = \sum_{h=1}^{k} r_{t+h}$ is a long-horizon (k-period) return, r_{t+h} is the log one-period stock return (between time $t+h-1$ and $t+h$) and x_t denotes a generic lagged predictor. The superscript pr denotes predictive regression and is used to distinguish these coefficients from those of alternative specifications below. Valuation ratios are a common choice for x_t and we focus on the dividend-price ratio dp_t, constructed from the natural logs of the stock price p_t and its corresponding dividend d_t.[1]

Two well known characteristics of valuation ratios is that they are both persistent and endogenous. For example, it is common in the literature to model the valuation ratios individually as first-order (or higher-order) autoregressive models:

$$x_t = \rho_0^{ar} + \rho_1^{ar} x_{t-1} + \varepsilon_{2,t}^{ar}. \tag{2}$$

Estimates of ρ_1 in (2) are generally close to one (high persistence), while the estimated correlation between the contemporaneous innovations $\varepsilon_{1,t+k}^{pr}(k)$ and $\varepsilon_{2,t}^{ar}$ are often close to negative one (strong endogeneity). These are salient features of the data that give rise to the well known Stambaugh bias, which has led the literature to question the significance of the positive estimates of $\beta_1^{pr}(k)$ reported in the previous literature.

This same negative correlation between $\varepsilon_{1,t+k}^{pr}(k)$ and $\varepsilon_{2,t}^{ar}$ also suggests a much stronger, more robust contemporaneous relationship between x_{t+k} and $r_{t+k}(k)$. At an intuitive level, holding dividends constant, a smaller value of Δx_{t+1} corresponds to a larger price increase and thus a larger return. In other words, if the goal was to explain rather than predict returns, then augmenting (1) with x_{t+k} as in

$$\begin{aligned} r_{t+k}(k) &= \beta_0^{aug}(k) + \beta_1^{aug}(k)x_t + \beta_2^{aug}(k)x_{t+k} + \varepsilon_{1,t+k}^{aug}(k), \\ &= \beta_0^{aug}(k) + (\beta_1^{aug}(k) + \beta_2^{aug}(k))\, x_t + \beta_2^{aug}(k)(x_{t+k} - x_t) + \varepsilon_{1,t+k}^{aug}(k), \end{aligned} \tag{3}$$

[1] Other popular valuation ratios include the dividend yield, earnings-price ratio and book-to-market ratio.

would yield a much tighter fit. Of course, (3) cannot be used for prediction or forecasting since x_{t+1} is unknown at time t. From a forecasting perspective, this is an infeasible specification since it uses future information on x_t. On the other hand, any information that can be used to (partially) predict x_{t+k} may also be useful for predicting returns. For example, if $\hat{x}_{t+k|t}$ is a time t forecast of x_{t+k}, then this forecasted predictor could be used to augment the predictive regression using:[2]

$$\begin{aligned} r_{t+k}(k) &= \beta_0^{aug}(k) + \beta_1^{aug}(k)x_t + \beta_2^{aug}(k)\hat{x}_{t+k|t} + \varepsilon_{1,t+k}^{aug}(k), \\ &= \beta_0^{aug}(k) + \left(\beta_1^{aug}(k) + \beta_2^{aug}(k)\right)x_t + \beta_2^{aug}(k)\left(\hat{x}_{t+k|t} - x_t\right) + \varepsilon_{1,t+k}^{aug}(k). \end{aligned} \tag{4}$$

The usefulness of (4) depends on the quality of the additional information provided by the forecast of $x_{t+k} - x_t$ above and beyond that already contained in x_t itself.

The next two predictive models build (partially or fully) on this augmented regression approach with an appropriate choice of an auxiliary predictor for $x_{t+k} - x_t$, such as slowly varying demographic trends. For example, Favero et al. (2011) (denoted by FGT) employ the middle-young (MY) demographic ratio to improve return forecasts by augmenting the predictive regression (1) as

$$r_{t+k}(k) = \beta_0^{FGT}(k) + \beta_1^{FGT}(k)x_t + \beta_2^{FGT}(k)MY_t + \varepsilon_{1,t+k}^{FGT}(k). \tag{5}$$

The approach by Favero et al. (2011) is motivated by their important finding that dividend-price ratio and MY ratio share a common persistent or slowly varying component, such that the residual persistence of the dividend-price ratio is reduced after the demographic trend in the MY ratio is removed.[3] The intuition for this argument can be understood from the bottom line of (3). The ability of a lagged valuation predictor x_t to predict the future return $r_{t+k}(k)$ depends mainly on ability to predict the future change in valuation $x_{t+k} - x_t$, or equivalently, to add predictive content to x_{t+k} above and beyond its own past. However, if the valuation predictor is close to being a random walk ($\rho_1^{ar} = 1$ in (2)), then $x_{t+k} - x_t$ is close to unpredictable. By reducing the persistence in the (demographically adjusted) dividend-price, Favero et al. (2011) improve its ability to forecast its own future value and thus the future return. Put another way, the slow moving, persistent MY ratio captures and refines the forecast of the time-varying mean of the dividend-price ratio (see again Favero et al. (2011)).

[2] Note that the definition of the population regression coefficients differ across the three specifications: (1), (3), and (4).

[3] In the more sophisticated version of their model, Favero et al. (2011) replace the dividend-price ratio by the error-correction term, say $d_t - c_3 p_t - c_4 MY_t$ in the cointegrated VAR for (d_t, p_t, MY_t). When $c_3 = 1$, this collapses to $dp_t - c_4 MY_t$ which is the residual after removing the demographic trend in MY_t from the dividend-price ratio. They refer to this residual dividend-price ratio as the MY-adjusted dividend-price ratio and show that it has more predictive power than the dividend-price ratio itself.

Favero et al. (2011) focus on five-year ahead forecasts ($k = 5$), which employ the current x_t and MY_t as predictors in (5). However, for their real out-of-sample forecasts they consider much longer horizons. In this case, they utilize Census Bureau projections of MY_{t+k-5} in order to forecast $r_{t+k}(5)$. For example, to forecast ten years ahead, they use a five-year ahead Census projection of MY_{t+5} as a predictor for $r_{t+10}(5)$ in (5).[4] We take this to the next logical step by fully incorporating Census projections of MY at all forecast horizons.

The approach that we advance in this paper can also be understood as using the demographic ratios to improve forecasts of future valuation changes. However, we do this by directly employing forecasts of future demographics to predict future valuation change, rather than by using lagged demographic predictors. Our prediction model for x_t is based on three insights from the literature:

1. Future demographic trends, as captured by the middle-young ratio MY_{t+j} (or any other demographic ratio), are far more predictable than most economic and financial variables. While shocks to birth, death, and immigration rates produce prediction errors, the aging of the remaining population is entirely predictable. In fact, the U.S. government produces official demographic forecasts to the year 2060.

2. Current (and possibly future) demographic variables influence current stock market valuations due to their impact on aggregate savings.

3. The persistence in demographics explains part of the persistence in stock valuations. After controlling for demographics, the serial correlation in the valuation ratios is reduced but not completely removed (Favero et al. (2011)).

Capitalizing on these observations, we incorporate the Census (time-t) projections of the future demographic ratio MY_{t+1} in the predictive model for returns. Let these projections be denoted by $\widehat{MY}_{t+k-h|t}$ (for $h = 0, ..., k$) with $\widehat{MY}_{t|t} = MY_t$. Then, our proposed model has the form:

$$r_{t+k}(k) = \alpha_0(k) + \alpha_1(k)x_t + \sum_{h=0}^{k} \alpha_{2,h}(k)\widehat{MY}_{t+k-h|t} + \epsilon_{1,t+k}(k), \tag{6}$$

[4] Favero et al. (2011) define the adjusted dividend-price ratio as the error-correction term in a cointegrated VAR on (d_t, p_t, MY_t). They then employ it as the lagged predictor (x_t) in the direct long-horizon regression in (1). They use this to produce five-year ahead return forecasts. When conducting forecasts at horizons beyond five years, they use forecasts of MY_t as an exogenous forcing variable in their cointegrated VAR to update their lagged predictor in the five-year predictive regression. For example, to forecast ten years ahead, they would first forecast the adjusted dividend-price ratio five years ahead and then use it in a five-year long-horizon regression to forecast the five-year return for the period starting five years into the future and ending ten years into the future. This makes partial use of demographic forecasts, employing the demographic forecast for the first, but not the second, five-year period.

where the coefficients $\alpha_0(k)$, $\alpha_1(k)$, and $\alpha_{2,h}(k)$ (for $h = 0, ..., k$) are obtained either in iterative (equation (A1.7) in Appendix A1.1) or direct (equation (A1.8) in Appendix A1.1) manner. As equation (6) illustrates, the key difference between our approach and that of Favero et al. (2011) is that we employ projected future demographics rather than realized lagged demographics. Since demographics are slowly evolving, this distinction may not be too important at short horizons but its importance can be expected to increase with the forecast horizon. Typical long-horizon regressions can be at five- or even ten-year horizons and a unique feature of demographic data is the availability of projections ten to twenty years into the future. To foreshadow our later results, despite similar performance at shorter horizons, at the five-year horizon, we find that (6), which employs the projected future path of MY_t outperforms (5), which uses only current MY_t, for the purpose of both pseudo and true-of-sample forecasting. In Appendix A1, we provide further details on how we arrive at the predictive model in (6).

3 Data and In-Sample Results

3.1 Data

For both our in-sample and pseudo out-of-sample results, we employ annual stock return data from 1901 to 2015 on stock returns, valuations and demographics, and also consider the post World War II (WWII) sub-period from 1947-2015 (post World War II, or post-WWII, period). Our data comes from two sources. Our stock return data is from Amit Goyal's webpage, which updates the Goyal and Welch (2008) data set until 2015. We employ annual, continuously compounded stock returns including dividends on the S&P 500. Dividends are defined as the twelve month moving sums of the dividends paid on the S&P 500 index. The dividend-price ratio $dp_t = d_t - p_t$ is constructed as the difference between the log dividend (d_t) and the log stock price (p_t).

Our demographic data is from the U.S. Census Bureau. We define: Y_t (age 20-29), M_t (age 40-49), and O_t (age 60-69) as the young, middle, and old U.S. population sizes. We use the middle-to-young ratio ($MY_t = M_t/Y_t$) as a predictor for persistent stock market valuations, as measured by the dividend-price ratio.[5] We employ historical census data until the end of 2015 and 2015 Census Bureau projections extending from 2015 until 2060. The historical census data is employed for estimation, whereas the projections are used in the forecast exercise. Additional historical Census Bureau forecasts used for our true out-of-sample results are described in detail later in Section 4.3

[5] Our separate (non-for-publication) Appendix also reports results for the middle-to-old ratio ($MO_t = M_t/O_t$) on its own and in conjunction with the MY ratio. Overall, these results reveal that the MY ratio is a more informative predictor than the MO ratio.

and Appendix A2.

3.2 Unit root and Cointegration Test Results

The data is first tested for the presence of a unit root with the Augmented Dickey-Fuller Generalized Least Squares (ADF-GLS) test of Elliott et al. (1996). Detailed results are available in an additional (not-for-publication) Appendix. The results are somewhat dependent on the lag length and sample period. Overall, we find strong evidence that we cannot reject the null hypothesis of a unit root in dp in both samples considered. For the MY ratios, we fail to reject the null (at all but the BIC-selected lag length) in the full sample but we reject the presence of a unit root in the post-WWII subsample. As expected, we strongly reject a unit root for all cases in the return series. The failure to reject a unit root in the valuation and, in some cases, the demographic ratios does not necessarily imply a true or exact unit root in these series. The power of unit root tests is well known to be low when the roots are close, but not equal to, unity. Also, there are good a priori reasons to rule out an exact unit root in a ratio variable, which must be bounded between zero and one. Nonetheless, at the very least, the test results confirm that all these variables (valuation and demographic ratios) are highly persistent.

Favero et al. (2011) report evidence of cointegration between the dp and MY ratios. We next test for cointegration among dp and MY using the two-step Engle-Granger and Johansen cointegration tests.[6] The results are again somewhat sensitive to the lag length and sample choice. Overall, the Engle-Granger test does not provide strong evidence of cointegration between the dp ratio and the demographic variable. The Johansen test results are somewhat more supportive of a possible cointegrating relationship in the post-WWII sample. This is to be expected since the Engle-Granger test is known to have lower power in the presence of strong endogeneity (Pesavento (2004)). However, none of the cointegration tests can be reliable given the possibility of very persistent but not exactly unit root variables.

We address the persistence of the regressors and the uncertainties regarding both the orders of integration and presence of cointegration in several ways. First, we note that our specification in linking the dividend-price ratio and demographic ratios in (A1.1) includes a lag of the dividend-price ratio, guaranteeing the stationarity of the residual and thus ruling out the possibility of a spurious regression. Secondly, as argued by Favero et al. (2011), controlling for the demographics may reduce the high persistence in the dividend-price ratio, mitigating the Stambaugh bias in predictive regressions. However, since this argument is only partially supported by our cointegration tests, we

[6] Detailed results can be found in the separate (not-for-publication) Appendix.

also confirm the robustness of our results by employing the IVX method of Kostakis et al. (2015).

3.3 In-Sample Estimation

In this section, we provide in-sample estimation and testing of our models for the dividend-price ratio and stock market returns when using the actual ex-post demographic trajectory. This offers a preliminary assessment of the usefulness of demographic projections in both pseudo and true out-of-sample forecasts, since a model which is not supported in in-sample evaluation is unlikely to be successful for out-of-sample prediction. Likewise, even the best demographic forecast would have little use in return predictions if future stock price movements are independent of future demographic trends. Below, we establish that this is not the case, thereby motivating our out-of-sample forecast analysis in the sections that follow.

3.3.1 In-Sample Estimation of Models for the Dividend-Price Ratio

Table 1 reports the estimates of our predictive models for the dividend-price ratio. Panels A and B correspond to the sample periods starting in 1901 and 1947, respectively. Column 2-3 provide the estimates from a simple AR(1) model for the dividend-price ratio as in (2). Columns 4-5 show the results from an AR(1) model augmented with projected demographic MY ratios as in (A1.1), but with the actual future values of MY_{t+1} replacing its projections as discussed above. For both regression specifications, we include standard OLS results alongside IVX estimation. The IVX was performed following Kostakis et al. (2015) using $c = -1$ and $\alpha = 0.95$.[7].

The coefficient on the lag dividend-price ratio is both large and highly significant. The p-value for the overall test of significance when using just one lagged predictor is already quite low. This is not surprising given the well known persistence of the dividend-price ratio. As anticipated by Favero et al. (2011), the value of the lagged dividend-price ratio coefficient drops slightly after inclusion of the demographic projections. However, it remains large and significant. This confirms that even after controlling for demographics the lagged dividend-price ratio is a highly important predictor for the future dividend-price ratio.

Comparing Columns 2-3 to Columns 4-5 helps us to assess the improvements that come from adding the demographic variable to the prediction equation for the dividend-price ratio. The coefficient on MY is significant in both sample periods. Although, the evidence is weaker using IVX than OLS, the IVX coefficient for MY remains significant at the 5% level in the post-WWII

[7]Robustness checks with different values for c and α can be found in the not-for-publication Appendix

sample and at the 10% level in the full sample, confirming the robustness of the finding. Overall, the results indicate that the MY ratio provides a modest improvement for in-sample prediction of the dp ratio, even after controlling for the lagged dp ratio.

3.4 In-Sample Estimation of Models for the Return Regression

Table 2 reports in-sample estimation of the return regression models in equations (1) for $k = 1$, (3), and (4) for both the full (Panel A) and post-WWII (Panel B) samples. We again present results for both OLS and IVX. Column 3-4 provide the results for the predictive regression in (1) for $k = 1$, using the lagged dividend-price ratio as a predictor. Using conventional significance levels (column 2), the lagged divided price-ratio is highly significant at the 1% level in the post-WWII sample, but not significant in the full sample. However, this evidence is likely to be overstated due to the Stambaugh bias in the predictive regression. Using IVX (column 3), the dp ratio is again insignificant in the full sample and less highly significant in the post-WWII sample. Overall, the results indicate that most of the future movement in stock prices is unpredictable.

The strong negative contemporaneous relationship between dp_{t+1} and r_{t+1} suggest that any additional information that helps to predict dp_{t+1}, even marginally, may help us to predict returns. Table 1 indicated that projected demographic ratios, particularly MY_{t+1}, were modestly helpful in predicting dp_{t+1}. In Columns 4-5, we next ask whether a correct demographic projection improves predictions of dp_{t+1} enough to lead to in-sample prediction improvements for r_{t+1}. In other words, we estimate (4), using dp_t and $\hat{d}p_{t+1|t}$ as regressors, where $\hat{d}p_{t+1|t}$ is itself predicted using (A1.1), with the results reported in Table 2.

In comparison with the traditional predictive regression in Column 2, the p-value of the F test for both the full and post-WWII samples improves notably with the use of MY-based predictor. When using MY to predict dp_{t+1} the coefficient on $\widehat{dp}_{t+1|t}$ is highly significant in both samples. Moreover, the dp ratio becomes much more highly significant after the inclusion of the MY-based forecast of $\widehat{dp}_{t+1|t}$, supporting the earlier results of Favero et al. (2011). Finally, the significance of the in-sample results employing the MY demographics (columns 4-5) appears quite robust to the Stambaugh bias, with similarly strong results across both OLS and IVX estimation.

Overall, there seems to be a clear in-sample return prediction improvement when including the predicted dividend-price ratio based on the actual future MY demographic trends. The results of Table 2 illustrate that even the modest improvements to the prediction of the future dividend-price ratio seen in Table 2 can lead to non-trivial improvements in in-sample return predictions. In the next section, we ask whether these improvements hold up out-sample using either true future demographic projections (pseudo OOS) or real-time Census Bureau projections (true OOS).

4 Out-of-Sample Forecasts

In this section, we investigate the usefulness of the model to provide both pseudo out-of-sample predictions using the ex-post future demographic trends and true out-of-sample predictions using ex-ante historical Census Bureau projections.

Since our ability to examine the true out-of-sample predictions is limited by the availability of historical Census Bureau projections, we start in Sections 4.1-4.2 by evaluating pseudo out-of-sample forecasts that employ the actual future demographic trends for MY. This serves several purposes. First, it allows us the use of a larger, more regular data set. Secondly, it is a natural precursor to the true out-of-sample analysis. If even the actual future demographic trends are not predictive for stock returns, then it is unlikely that any forecast of these trends will prove useful. Thirdly, this may be of independent interest in evaluating the usefulness of the model for scenario analysis involving alternative demographic projections. Suppose that a policy maker or pension administrator wants to analyze the stock market implications of several alternative future demographic scenarios. Her goal then is to predict the counterfactual returns that would occur under a particular demographic scenario; i.e., in the state that the given demographic scenario represents the true future demographic trend. For this purpose, she would need to relate future demographic ratios to future returns. The pseudo out-of-sample exercise evaluates the ability of our model to perform this task.

In Section 4.3, we then employ additional data to examine the usefulness of true historical Census Bureau projections in providing true out-of-sample predictions for stock returns. This addresses the usefulness of the model to a policy maker or pension fund manager interested in providing real baseline stock return forecasts and complements the scenario analysis discussed above.

4.1 Pseudo Out-of-Sample Recursive Forecast Results

In Table 3, we provide long-run pseudo out-of-sample forecasting results. We run recursive forecasting regressions for the five-year ahead horizon ($k = 5$).[8] Panel A and Panel B correspond to the

[8]We have obtained the pseudo out-of-sample results using both OLS and IVX estimation. It is not obvious a priori which should perform better in term of out-of-sample mean squared error, as IVX reduces the bias in OLS at the expense of greater variance. In practice, we found that the OLS based forecasts perform as well or better than the IVX based forecasts in most cases, although the results were often similar. In order to conserve space, we present

full and post-WWII samples. To see the sensitivity of the out-of-sample performance to the time windows, results with three different initial training periods are provided in each panel. Specifically, we use 30, 40, and 60 years for the full sample and 20, 25, and 30 years for the post-WWII samples as the training window. For the purpose of illustrating our findings, we divide each panel into two parts. The first part measures the forecast improvements provided by the demographic projections by reporting the out-of-sample MSEs and R^2s. Then, the following part reports the out-performance testing results.

We run out-of-sample forecasting recursively with a five-year ahead horizon. Therefore, the long-horizon forecast is the cumulative sum of 5 one-year forecasts. Due to the cumulative nature of the long-horizon forecast, we may predict the long-horizon return either in a recursive way as defined in (A1.4) and (A1.5) or in a direct way as defined in (A1.8). The two methods are asymptotically equivalent under correct specification. However, they may have different finite-sample performance. Similar to the presentation of the previous results, we separate each panel into two parts. The first part shows the out-of-sample mean squared error and the R^2 for each model. The second part reports the out-of-sample performance test results.

Column 1 lists the candidate models. HM denotes the historical mean. PR is the long-horizon version of the predictive regression model in (1) since $k > 1$. FGT provides the estimates of (5) with $k = 5$. Model MY_R provides estimates of (A1.4) and (A1.5) by replacing the projected demographic $\widehat{MY}_{t+h|t}$ in (A1.3) by the actual realized values of MY_{t+h}.[9] This extends the one-period ahead MY based forecast to the long-horizon by recursive forward substitution of the one-period ahead forecasts. Alternatively, MY_D instead provides a direct long-horizon forecast based on (A1.8), where we again replace $\widehat{MY}_{t+h|t}$ by the actual future demographic MY_{t+h}.[10]

Within the first part of each panel, columns 2, 4, and 6 of Table 3 show the out-of-sample mean square errors (OOS MSE) of each model under various training periods. Columns 3, 5, and 7 present the out-of-sample R^2 (OOS R^2), introduced by Campbell and Thompson (2007).[11] For the OOS MSE of both panels, either MY_D or MY_R reports the lowest value in all six cases and both

the OLS OOS results here and include the IVX counterpart in the additional appendix.

[9]For the iterative out-of-sample forecasts, IVX estimation can be performed in three ways. 1. We can use IVX approach to estimate (A1.3) and OLS method to estimate (A1.4). 2. We can estimate (A1.3) by OLS method and (A1.4) by IVX method. 3. We can use IVX approach to estimate both (A1.3) and (A1.4). In the separate appendix, we report the IVX results using method 1 (IVX-OLS) due to the fact we found large biases when using the other two approaches.

[10]In Section 4.3 below, we assess the accuracy of the unconditional forecast, using the real-time Census Bureau historical forecasts available for MY.

[11]The OOS R^2 is defined as $R^2 = 1 - \frac{\text{OOS MSE}}{\text{HM MSE}}$, where $\frac{\text{OOS MSE}}{\text{HM MSE}}$ is the ratio of the out-of-sample mean squared error (OOS MSE) of the model to that of a baseline forecast projecting the historical mean (HM MSE).

models beat the historical mean in five of six cases. This is consistent with the findings from the OOS R^2s, which are positive with only one exception for MY_R and MY_D, but often negative for the other forecast models. Our results strongly indicate that the MY demographic ratio improves the long-run forecast.

The second part of each panel of Table 3 reports tests of the demographic models against two benchmark forecasts: the historical mean (HM) and the predictive regression (PR). The demographic models nest both benchmarks. The Diebold-Mariano (DM) test is known to follow a non-standard distribution when the forecast models are nested (Clark and McCracken (2001), and McCracken (2007)). This results in a bias that inflates the OOS MSE of the larger model under the null hypothesis that the two models provide equal forecasts in large sample. Clark and West (2007) provide a bias correction to the DM test resulting in a t test with a standard normal distribution. It is this Clark and West (CW) correction to the DM test that we employ to examine the significance of the out-performance.[12]

As shown in the second part of each panel, with both the historical mean and the predictive regression model as benchmarks, MY_D and MY_R are significant at, at least, the 10% level in all but one case. For the full sample in Panel A, many are also significant at the 5% and 1% levels. This confirms the significance of the MY-based forecasts in out-performing OOS MSE and OOS R^2 from the first part of the panel. We also observe a strong distinction of p values between MY and FGT in almost every evaluation, which confirms the usefulness of using future middle-young demographic ratios to improve the additional long-run forecasting. Intuitively, MY_t captures well the slow changes in the dividend-price ratio. However, it is reasonable to expect that its ability to predict the future of this slowly evolving component in the dp ratio gradually diminishes over time. Hence, with longer horizon, MY, which employs the future path of MY_t performs better than FGT, which uses only current MY_t.

4.2 Pseudo Out-of-Sample Rolling Forecast Results

To complement our recursive out-of-sample forecasting, we run rolling conditional forecasting regressions at the five-year ahead horizons in Tables 4. As seen in the first part of each panel, either MY_R or MY_D always shows the lowest MSE regardless of the size of the sample and training window, whereas the FGT model shows higher MSE and is quite sensitive to the training window.

The CW test used earlier is designed for recursive forecasts. To examine the significance of

[12]For some caveats on applying tests of equal forecast accuracy to conditional forecast models see Clark and McCracken (2017) and Faust and Wright (2008).

the out-performance of the rolling window, we use the Giacomini and White (2006) test, which is designed for rolling windows and applicable to nested models. It is worth noting that the object of evaluation of the Giacomini and White (2006) test is not simply the forecasting model as in the DMW approach, but the forecasting method. Thus, the null hypothesis of Giacomini and White (2006) test depends on the parameter estimates, that differs from the null in DMW which is defined in terms of the population parameters. The null hypothesis of the GW test is therefore less favorable to larger nesting models, which involve extra parameters to be estimated with error in small rolling samples. Therefore it provides a more stringent test than the CW in which it is considerably harder to out-perform the smaller benchmark models.

The second parts of the panels in Table 4, show the GW test for out-performance in rolling samples relative to the historical mean predictive regression. We observe some evidence on the ability of MY to predict long-run stock returns even using this more stringent GW test. In particular, we find that MY_R significantly outperforms all of the historical mean and predictive regression models at the 10% level or better in the post-WWII sample of Table 1 Panel B. In some cases it also outperforms at the 5% or 1% levels.

To further explore the sensitivity of our findings to the training periods and window size, we show in Figures 2 and 3 the OOS R^2 and p-values for out-performance against the historical mean (vertical axis) for a range of training periods (horizontal axis). Figure 2 shows five-year ahead recursive forecasts, expanding on Table 3 while Figure 3 shows the corresponding rolling forecasts and expands on Table 4. The graphical evidence demonstrates the robustness of the results reported earlier in the tables to the choice of training period (recursive) and window size (rolling). The good performance of both the MY_D and MY_R appears quite robust, with both showing an OOS R^2 that remains stable across training periods and window lengths. Moreover, the forecasting results provide evidence that an accurate demographic ratio projection can improve return prediction, particularly in the long run. Consistent with Favero et al. (2011), the results support the use of the middle-to-young ratio as a strong predictor. Moreover, in this pseudo OOS exercise, we find that the demographic ratio projection performs better than the FGT model due to the extra information contained in the future path of the MY ratio.

4.3 True Out-of-Sample Forecasts using Historical Census Bureau Predictions

The preceding pseudo out-of-sample analysis employed the true future projected demographic ratio. As noted above, this establishes the usefulness of the model for scenario analysis. We now ask whether the same approach can be useful for true, real-time forecasts. This requires us to replace the actual future demographic ratios by the corresponding Census Bureau projections that would be available to the forecaster at the time of the prediction. To this end, we collect, compile, and electronically record, historical forecasts from the U.S. Census Bureau. This is complicated by the many years in which the Census Bureau did not update their projections, thereby creating gaps in their historical forecast record. We address these gaps by building our forecasts with the latest available demographic projection that would have been available to the forecaster at the time that the forecast was produced, as described in detail in Appendix A2.

Table 5 presents our five-year ahead real-time forecast using both the recursive (Panel A) and rolling (Panel B) methods. As in our earlier tables, these show the results for three different choices of training periods (recursive) or window sizes (rolling). Figure 4 shows both the OOS R^2s (left panels) and OOS p-values (right panels) for a range of training periods or window sizes. Panels a-b show the results for the five-year ahead recursive forecasts, while Panels c-d show the corresponding results for the rolling forecasts.

Turning first to the recursive results in Panel A of Table 5, MY_D and MY_R continue to out-perform all other forecast methods for each of the three training periods considered. Indeed, they are the only two methods to out-perform the historical mean. The out-performance relative to both the historical mean and predictive regression are weakly significant for both MY_D and MY_R using a training period of 20 and for at least one of the two forecasts using a training period of 25. For the longest training period, none of the forecasts significantly outperform the others, yet MY_D and MR_Y still dominate the point-wise comparison. Similarly, Panel (a) of Figure 4 displays a robust out-performance of MY_D and MY_R. While Panel (b) indicates that the significance of the out-performance is sensitive to the training period, the OOS p-values associated with both MY_D and MY_R remain substantially lower than the other p-values across the full range of training periods.

The out-of-sample rolling forecast results are even stronger. Panel (B) of Table 5, shows that in terms of MSE, the two MY-based forecasts again out-perform the other forecast methods. None of the other rolling forecasts out-perform the historical mean for any of the three window sizes, whereas the MY-based forecasts show a positive OOS R^2 in five of six cases.[13] Moreover, the OOS MSEs of the competing methods are often close to twice as large as those for either MY_D and/or MY_R. For example, in Column 4 of Panel (B), the MSE for MY_D and MY_R are 0.1000 and 0.1075 respectively, whereas the predictive regression (PR) and FGT forecasts have MSEs of 0.2697 and 0.3228. Panel (c) of Figure 4 shows that the improvement in OOS R^2s is robust across window

[13]The only exception is MY_D when using a training sample period of 20 observations.

sizes.

The out-performances of the MY based forecasts are also robustly significant. MY_D significantly improves on the historical mean for a window size of 25 periods and weakly out-performs for a 30 period window. The improvements using MY_R are even more significant, with strong rejections at the 1% level for windows of size 20 and 25 and a weak rejection for size 30. Panel (d) of Figure 4, further confirms the robust significance of MY_R across window widths and of MY_D for all but the smaller window sizes.

Overall, these results confirm the predictive content of the Census Bureau MY forecast for long-horizon stock return forecasts. For the recursive forecasts, the true OOS forecast results using real Census forecasts from 1950-2015 are only modestly weaker than the pseudo OOS recursive forecast results for the period 1947-2015. Due to data limitations, there is no true forecast equivalent to the much stronger pseudo OOS recursive results obtained over the period 1901-2015. The rolling sample results hold up even better when we move from the pseudo to true OOS forecasting. For example, comparing panels (c) and (d) of Figures 3 and 4, we observe a similar pattern for both the OOS R^2 and p-values, except for MY_D at low window widths. It is not at all surprising that pseudo OOS forecasts using the actual trajectory of MY produce stronger results than those using historical Census forecasts, some of which were already dated (stale) by the time of the forecast. To the contrary, it is quite encouraging to observe how much of the improvements from the pseudo OOS forecast exercise carry over to the true OOS forecast. This shows that the strong pseudo OOS forecast out-performance was not simply a result of a look-ahead bias and supports the utility of our approach, whether employed for either scenario analysis or real prediction.

5 Very Long-Horizon Forecasts

In this Section, we employ our framework to provide actual forecasts until 2060. Our baseline forecasts using the Census Bureau projections are discussed in Section 5.1. We then explore the sensitivity of our forecasts to alternative assumptions with respect to immigration in Section 5.2 and COVID-19 deaths in Section 5.3.

5.1 Baseline Forecasts

The official projections for the MY_t demographic ratio until 2060, made by the Census Bureau in 2015, are plotted together with its historical values in the top panel of Figure 5. The middle-age population peaked a little after 2000 relative to the young cohort. It is projected to keep declining until the early 2020s, at which point it is projected to start growing slowly, reaching a second smaller "echo" peak at around 2040. Currently, the MY ratio is slightly above its long-run mean.

The resulting forecasts for the dividend-price ratio, together with their historical values, are shown in the middle plot of Figure 5. With a brief exception around the 2008 crisis, the dividend-price ratio remains below its long-term average since the early 1990s. Using (2) with an estimated value of $\hat{\rho}_1^{ar} < 1$, would lead us to forecast a mean-reverting increase in the dividend-price ratio on this basis alone. However, the empirical estimate of $\hat{\rho}_1^{ar}$ is quite close to one (see Table 1) and unit root tests do not reject the hypothesis ($\rho_1^{ar} = 1$) that is non-mean reverting. In addition, the dividend-price ratio may be subject to structural breaks (Lettau and Nieuwerburgh (2008)).

Favero et al. (2011) show that much of the low-frequency trend in the dividend-price ratio is explained by the demographics and we obtain lower estimates of the coefficient on the lagged dividend-price ratio after controlling for MY_t as in (A1.1) (see Table 1). Therefore, a better question to ask may be whether, according to the dividend-price ratio, stocks appear over-valued relative to their mean, conditional on demographic factors.

The MY ratio ends the sample slightly above its mean. However, this discrepancy is small and does not appear large enough to justify the low dividend-price ratio. Thus, part of the projected near-term rise in the dividend-price ratio is simply a predicted correction to its low current value relative to the long-run value implied by the demographics.

Starting in the early 2020s, the relative size of the middle-age population is projected to increase again causing the MY ratio to rise. This leads to a projected decline in the dividend-price ratio. The decline in the dividend-price ratio is projected to drop below its current value by 2040, when MY reaches a new peak.

The bottom graph in Figure 5 shows the net returns with dividends averaged over a five-year rolling window along with the demographic-based projections for this same five-year average returns out to 2060. The near-term forecast calls for still positive, but below average returns. This is due to both a projected correction to the currently low dividend-price ratio and the continued fall of the predicted MY ratio. The latter effect is picked up primarily by the current predictor $\widehat{dp}_{t+1}$, which predicts a below average return as an almost mechanical consequence of a rising dividend-price ratio. Both are caused by a (relative) fall in price.

Returns are subsequently projected to slowly recover over the next 15 years during much of the 2020s and 2030s once the relative size of the high-saving middle-age population starts to increase again and after the dividend-price ratio has finished its earlier projected correction and has become more accommodative of higher returns. This last effect is captured by the lagged predictor $\widehat{dp}_t$ which reflects the prediction that returns are higher when markets are less highly valued (higher dividend-price ratio).

We should acknowledge the large degree of model and data uncertainty when accessing the likely impact of either valuation corrections or demographic trends on future stock valuations and returns. First, quantifying the model uncertainty is not straight-forward. Even characterizing the forecast uncertainty for a particular model is complicated since the forecasts have three steps, each with its own forecast error. In the first step, the official population projections are themselves subject to forecast error. This creates a generated regressor problem in the second stage when these projections are used to forecast the dividend-price ratio. This is also subject to error and creates a second generated regressor problem when using the dividend-price ratio predictions to forecast returns – a third prediction which is subject to forecast error. We conjecture that the cumulative effect of these forecast errors, combined with possible model misspecification, would result in a substantial forecast uncertainty of these predictions.

Second, the Census projections that we use in our analysis embed particular assumptions about population growth and immigration flows that are affected by demographic/health shocks as well as policy. Regarding immigration, some estimates suggest that after the end of our sample period in 2015, the cumulative shortfall in immigration is in excess of 2 million. An even larger potential problem to the baseline projections and the uncertainty around these projections is posed by the COVID-19 pandemic and its future dynamics.

We explore the robustness of our results to these two recent developments in the next subsections. More specifically, we show that our main findings are largely unaffected by these material shifts in the demographic dynamics. How is it possible that these large – in magnitude and scope – shocks had only a limited impact on our analysis? We believe that the main reason for this is our choice of demographic predictors. First, while the COVID-19 increased substantially (directly and indirectly) the mortality rate, the worst affected part of the population (aged 50 and above) is not part of the MY ratio. Second, the fact that the demographic variable is a ratio helps to alleviate some of the large absolute impact of the shock and focuses on its relative (distributional) aspect. We note that mutations of the virus and vaccination rates will continue to affect differentially the various age cohorts but the relative and smooth nature of the MY ratio makes this impact more muted. Similar arguments apply to the effect of immigration flows on the MY ratio. Finally, the extraordinary fiscal and monetary support during the initial stages of the COVID-19 pandemic has led to elevated stock market valuations whose source is not fully controlled for in our predictive framework. However, the long-horizon focus of our analysis tends to be somewhat immune to such short- to medium-term variations and suggests that the low-frequency co-movements between demographic and financial variables may remain robust even when subjected to large, but potentially transitory, health or policy shocks.

5.2 Immigration

We first consider the sensitivity of our baseline forecast above to two alternative immigration scenarios provided by the Census Bureau. The first is their high immigration projection, in which the baseline Census projection is adjusted by increasing the foreign-born immigration projections by 50 percent. The second is their low immigration projection, in which the Census Bureau instead adjusts foreign born immigration down in a manner which is log-symmetrical to the upward projection in the first series. These alternative scenarios cover the 2016-2060 projections and were published by the Census Bureau in 2017 in a series entitled "Projected Population by Single Year of Age, Sex, Race, and Hispanic Origin for the United States."[14]

The top plot of Figure 6 shows the MY ratio projection by Census Bureau for both the baseline and the high and low immigration projections. Since immigration levels are higher for adults in their twenties than for adults in their forties, the high immigration scenario results in the lowest trajectory for MY, whilst the low immigration scenario corresponds to the highest projections for MY. The differences are nonetheless fairly modest throughout most of the series, with greater divergence towards the end of the forecast horizon. Even this large change in immigration rate assumptions has only a moderate impact on the MY ratio and does not much change the overall pattern of the trajectory.

The bottom-left plot in Figure 6 provides forecasts for the dividend-price ratio based on the Census Bureau alternative immigration projections. The higher immigration scenario results in a higher dividend-price ratio projection (lower valuation), due to the lower ratio of the high-saving middle age population resulting from the younger immigrant population inflow. Likewise, the low immigration scenario results in a higher dividend-price ratio. In both cases, the resulting forecast changes are modest, widening slightly at the end of the forecast horizons.

The bottom-right plot in Figure 6 presents forecasts for the five-year rolling average return (including dividends) on the S&P 500 index using the Census Bureau alternative immigration

[14]Evidence from the Census since the end of our sample period lends support to the low immigration scenario with a cumulative shortfall (between 2016 and 2021) in excess of 2.3 million in net international migration, with notable declines during the COVID-19 pandemic; see `https://www.census.gov/library/stories/2021/12/net-international-migration-at-lowest-levels-in-decades.html`.

projections. The increased dividend-price ratio trajectory of the high immigration scenario has two contradictory effects on stock returns: a negative contemporaneous effect as an a drop in valuations (increase in the dividend price ratio) is generally accompanied by a negative return and a positive lagged effect in which lower valuations tend to be followed by higher future returns. The higher immigration projections lower the projected returns in the early forecast period, raise them after the mid-2030s and lower them again towards the end of the sample. Overall, the change to the forecast remain small, demonstrating their relative insensitivity to immigration rates.

5.3 COVID-19

We next assess the possible impacts of COVID-19 deaths on our demographic projections. We obtain COVID-19 deaths by age categories for 2020 and 2021 from the CDC (Centers for Disease Control and Prevention) COVID-19 Death Data and Resources.[15] We model the impact of the 2020-2021 deaths along with four alternative scenarios for the future course of the pandemic. All four scenarios envision a linear transition from the current pandemic phase to a milder, but permanent endemic phase. However, the scenarios differ in both the speed of the transition and the endemic death rate post-transition. For transition speed, we consider both a quick transition that is completed by the start of 2023, and a long transition that is not completed until 2030. Likewise, we consider a low-mortality endemic death rate of just ten percent of 2021 COVID-19 deaths and a high-mortality endemic mortality rate equal to thirty percent of 2021 deaths. Thus, in our most optimistic scenario, COVID-19 death rates would fall to just ten percent of 2021 levels by 2023, whereas in our most pessimistic scenario, they would fall to 30 percent of 2021 mortalities by 2030.

Since COVID-19 has a higher mortality rate for the middle-age population than for the young, it can be expected to lower the MY ratio. We calculate separately the reduction in M and Y for each year under each scenario. It is important to note that each death can reduce M or Y for several years, depending on the age at death. For example, a 41 year old who passed away in 2021, could otherwise have been counted as part of M up though 2029. Similarly, although the 2021 death of a 39 year old would have no impact on M in 2021, it could reduce M by one for a full decade starting in 2022. To track these cumulative impacts, we require an accounting of deaths by both age and year. This in turn requires that we make two simplifying assumptions. First, we assume that in the absence of COVID-19, the individuals that died of COVID-19 would

[15] We acknowledge that "excess deaths," that also include deaths indirectly attributed to COVID-19 (e.g., arising from a lack of timely diagnosis or treatment due to the COVID-overwhelmed health system), would provide a more accurate mortality measure but we were unable to obtain age-specific data of excess deaths which is required for the demographic projection adjustment.

otherwise have stayed in the Census count until turning 50. Secondly, since the CDC reports death by multi-year cohorts, whereas we need to track deaths by single-year cohorts, we assume an even distribution of COVID-19 fatalities within each CDC age category range. We acknowledge that both assumptions could be made more realistic at the cost of additional model complexity, but view them as reasonable simplifications given the under-50 age population with which we work. The detailed scenario models, parameter values and dynamic equations to account for the yearly reductions in M and Y under each scenario are provided in the Appendix A3.

For the sake of brevity, we focus our discussion of the results on our worst case scenario described above. The remaining three COVID-19 scenarios are included in Section B7 of our not-for-publication appendix. The top plot in Figure 7 shows the original baseline Census Bureau projection for MY alongside the same projection after being adjusted for our worst case COVID-19 scenario. We have three observations. First, as anticipated, the projected path for MY falls in response to COVID-19. Secondly, the drop is long-lived, and the size of the effect is increasing over time. This results from the late transition to the endemic state, the relatively high death rate in the endemic state of this worst case scenario, and most importantly, the cumulative impact of COVID-19 deaths on population cohorts. Thirdly, the magnitude of the change in MY is quite small. This reflects that even for the 40-49 age group, the number of COVID-19 deaths, while tragically large, are nonetheless small relative to the overall size of the middle age population.

Next, the bottom-left plot in Figure 7 illustrates our baseline forecast for the dividend-price ratio alongside the forecast adjusted for our worst case COVID-19 scenario. The downward adjustment to the MY projection results in an upward adjustment in the forecasted dividend-price ratio. This results from the negative sign on MY in the dividend price ratio regression. It is also consistent with economic intuition, in which a smaller middle age population implies lover savings and thus lower valuations. The overall impact of this demographic adjustmentment on the valuation forecast is again long-lasting and increasing with time, but overall quite small in magnitude. Finally, the bottom-right plot in Figure 7 compares our five-year rolling return forecast using the baseline Census projection, to the equivalent forecast after adjusting the Census projection for our worst case COVID-19 scenario. The impact on the projected returns is again persistent, yet quite small.

6 Conclusion

We confirm previous evidence that demographic ratios can help to predict valuation ratios, such as the dividend-price ratio. Rather than using current demographic ratios, we employ Census

Bureau projections, since they are arguably more informative about future demographic and, hence, valuation trends. It is well known that the contemporaneous valuation ratios are much more strongly correlated with returns than their lagged counterparts employed in predictive regressions. Since they contain future information, they cannot be employed for prediction purposes. However, using demographic projections, the traditional predictive regression can be augmented to include predicted values of the contemporaneous valuation predictor. We show that this can improve both pseudo and true out-of-sample stock return prediction at the five-year horizon. This demonstrates that the value of projected future demographic information is not simply due to look-ahead bias.

Furthermore, we find that a model including future demographic projections can outperform a model containing only current demographic information. Since the Census produces projections until 2060, this allows us to provide very long-horizon predictions for both valuation ratios and returns. We assess the sensitivity of these long-horizon predictions under alternative scenarios that include the demographic impact of the COVID-19 pandemic or recent changes in net international migration. It should be recognized, of course, that future stock returns naturally depend on many other factors and unforeseeable events, adding a great deal of uncertainty to these forecasts. Nonetheless, all else equal, our proposed predictive approach can prove useful by providing an indication of the likely effect of both current and projected demographic trends on future stock market valuations.

There is a wide range of potential applications where future demographic projections can elicit important information about the underlying dynamics of key macroeconomic variables. Recent work by Favero et al. (2021), Lunsford and West (2019), and Del Negro et al. (2019), among others, emphasize the secular role of demographics in determining the equilibrium level of safe real interest rates and the common trends across the term structure and across countries. Furthermore, Goodhart and Pradhan (2020) discuss the effects of the expected demographic shifts on underlying inflation, with possible extensions to the low-frequency movements in inflation risk and stock-bond correlation. Adapting our proposed approach, by leveraging the information content of long-horizon demographic projections, to these setups appears to be a promising direction for future research.

References

Bai, J. and Ng, S. (2004), 'A panic attack on unit roots and cointegration', Econometrica **72**, 1127–
–1177.

Boudoukh, J., Richardson, M. and Whitelaw, R. (2008), 'The myth of long-horizon predictability', The Review of Financial Studies **21**(4), 1577–1605.

Brooks, R. (2002), 'Asset-Market Effects of the Baby Boom and Social-Security Reform', American Economic Review **92**(2), 402–406.

Campbell, J. Y. and Shiller, R. (1988a), 'The dividend-price ratio and expectations of future dividends and discount factors', Review of Financial Studies **1**(3), 195–228.

Campbell, J. Y. and Shiller, R. J. (1988b), 'Stock prices, earnings, and expected dividends', Journal of Finance **43**, 661–76.

Campbell, J. Y. and Thompson, S. B. (2007), 'Predicting excess stock returns out of sample: Can anything beat the historical average?', Review of Financial Studies **21**(4), 1509–1531.

Cavanagh, C. L., Elliott, G. and Stock, J. H. (1995), 'Inference in models with nearly integrated regressors', Econometric Theory **11**, 1131–1147.

Clark, T. E. and McCracken, M. W. (2001), 'Tests of equal forecast accuracy and encompassing for nested models', Journal of Econometrics **105**, 85–110.

Clark, T. E. and McCracken, M. W. (2017), 'Tests of predictive ability for vector autoregressions used for conditional forecasting', Journal of Applied Econometrics **32**(3), 533–553.

Clark, T. E. and West, K. D. (2007), 'Approximately normal tests for equal predictive accuracy in nested models', Journal of Econometrics **138**, 219–311.

Del Negro, M., Giannone, D., Giannoni, M. P. and Tambalotti, A. (2019), 'Global trends in interest rates', Journal of International Economics **118**, 248–262.

Elliott, G., Rothenberg, T. J. and Stock, J. (1996), 'Efficient forecast tests for conditional policy forecasts', Econometrica **64**, 813–836.

Fama, E. and French, K. (1988), 'Permanent and temporary components of stock prices', Journal of Political Economy **96**, 246–273.

Faust, J. and Wright, J. H. (2008), 'Efficient forecast tests for conditional policy forecasts', Journal of Econometrics **146**, 293–303.

Favero, C., Gozluklu, A. E. and Tamoni, A. (2011), 'Demographic trends, the dividend-price ratio, and the predictability of long-run stock market returns', Journal of Financial and Quantitative Analysis **46**(05), 1493–1520.

Favero, C., Melone, A. and Tamoni, A. (2021), 'Monetary policy and bond prices with drifting equilibrium prices', C.E.P.R. Discussion Papers October 2021. . CEPR Discussion Papers 16629.

Geanakoplos, A.-N. J., Magill, M. and Quinzii, M. (2004), 'Demography and the Long-Run Predictability of the Stock Market', Brookings Papers on Economic Activity **35**(1), 241–326.

Giacomini, R. and White, H. (2006), 'Tests of conditional predictive ability', Econometrica **74**(6), 1545–1578.

Goodhart, C. A. and Pradhan, M. (2020), The Great Demographic Reversal: Ageing societies, waning inequality and an inflation revival, Palgrave MacMillan.

Goyal, A. and Welch, I. (2003), 'Predicting the equity premium with dividend ratios', Management Science **49**(5), 639–654.

Goyal, A. and Welch, I. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', Review of Financial Studies **48**(2), 663–670.

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

Hodrick, R. (1992), 'Dividend yields and expected stock returns: Alternative procedures for inference and measurement', Review of Financial Studies **5**, 357–386.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M. P. (2015), 'Robust econometric inference for stock return predictability', The Review of Financial Studies **28**(5), 1506–1553.

Lettau, M. and Nieuwerburgh, S. V. (2008), 'Reconciling the Return Predictability Evidence', Review of Financial Studies **21**(4), 1607–1652.

Liu, Z. and Spiegel, M. M. (2011), 'Boomer retirement: Headwinds for U.S. equity markets?', FRBSF Economic Letter **2011-26**.

Lunsford, K. G. and West, K. D. (2019), 'Some evidence on secular drivers of us safe real rates', American Economic Journal: Macroeconomics **11**(4), 113–39.

Mankiw, N. G. and Shapiro, M. (1986), 'Do we reject too often? small sample properties of tests of rational expectations models', Economics Letters **20**, 139–145.

McCracken, M. W. (2007), 'Asymptotics for out of sample tests of granger causality', Journal of Econometrics **140**, 719–752.

Pesavento, E. (2004), 'An analytical evaluation of the power of tests for the absence of cointegration', Journal of Econometrics **122**(2), 375–421.

Poterba, J. M. (2004), 'Impact of population aging on financial markets in developed countries', Economic Review (Q IV), 43–53.

Stambaugh, R. F. (1986), Bias in regressions with lagged stochastic regressors. Center for Research in Security Prices Working Paper 156, University of Chicago.

Stambaugh, R. F. (1999), 'Predictive regressions', Journal of Financial Economics **54**, 375–421.

Valkanov, R. (2003), 'Long-horizon regressions: Theoretical results and applications', Journal of Financial Economics **68**, 201–232.

Waggoner, D. F. and Zha, T. (1999), 'Conditional forecasts in dynamic multivariate models', Review of Economics and Statistics **81**(4), 639–651.

Yoo, P. S. (1994), Age distributions and returns of financial assets, Working Papers 1994-002, Federal Reserve Bank of St. Louis.

A1 Appendix: Details and Derivation of Forecast Models

In this section of the appendix, we provide further details on how we arrive at the predictive model in (6). Section A1.1 outlines our main approach and Section A1.2 provides a derivation of a technical result used in Section A1.1.

A1.1 Iterative and Direct Forecast Approach

Incorporating the official time-t demographic projections in the usual autoregressive specification (2) generates a one-step forecast for the valuation predictor given by

$$\hat{x}_{t+1|t} = \hat{\rho}_0 + \hat{\rho}_1 x_t + \hat{\rho}_2 \widehat{MY}_{t+1|t}, \tag{A1.1}$$

or, by shifting (A1.1) h periods, an $(h+1)$-step forecast

$$\hat{x}_{t+h+1|t+h} = \hat{\rho}_0 + \hat{\rho}_1 x_{t+h} + \hat{\rho}_2 \widehat{MY}_{t+h+1|t+h}, \tag{A1.2}$$

where $\hat{\rho} = (\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2)'$ denote the standard least squares estimates and the notation $t+h+1|t+h$ in $\widehat{MY}_{t+h+1|t+h}$ indicates a projection for a $t+h+1$ variable produced or made available at time $t+h$. Since the forecast $\widehat{MY}_{t+h+1|t+h}$ are not available until $t+h$, when working with historical and current Census forecasts as in Sections 4.3 and 5, we replace $\widehat{MY}_{t+h+1|t+h}$ by the most recent Census forecasts available at the time that the prediction is made, namely $\widehat{MY}_{t+h+1|t}$.[16]

We employ an iterative forecast strategy to replace x_{t+h} by its forecast value. The resulting (partially recursive) valuation forecast is obtained as

$$\hat{x}_{t+h+1|t} = \begin{cases} \hat{\rho}_0 + \hat{\rho}_1 \hat{x}_{t+h|t} + \hat{\rho}_2 \widehat{MY}_{t+h+1|t} & \text{for } h > 0, \\ \hat{\rho}_0 + \hat{\rho}_1 x_t + \hat{\rho}_2 \widehat{MY}_{t+1|t} & \text{for } h = 0. \end{cases} \tag{A1.3}$$

Similarly, a feasible $(h+1)$-period return forecast is given by:

$$\hat{r}_{t+h+1|t} = \begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{t+h|t} + \hat{\beta}_2 \hat{x}_{t+h+1|t} & \text{for } h > 0, \\ \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\beta}_2 \hat{x}_{t+1|t} & \text{for } h = 0. \end{cases} \tag{A1.4}$$

Forecasts of the multi-period return $r_{t+k}(k)$ can then be formed by a cumulative sum of one period forecasts:

$$\hat{r}_{t+k|t}(k) = \sum_{h=0}^{k-1} \hat{r}_{t+h+1|t}. \tag{A1.5}$$

[16] $\widehat{MY}_{t+h+1|t+h}$ can also be replaced by an alternative demographic projection, which could be useful in conducting sensitivity analysis or assessing forecast risks. When evaluating stock valuation/return forecasts that are meant to be conditional on a given future demographic trajectory, we should replace $\widehat{MY}_{t+h+1|t+h}$ by MY_{t+h+1}, the actual future value. This way, we separate the errors in the forecast conditional on the demographic trajectory from errors in the trajectory itself.

The forecast formulas provided above are partially iterative, due to the residual persistence in x_t even after controlling for demographics. It is also informative to solve for the explicit h-period ahead forecasts. Iterative forward substitution of (A1.3) yields the following iterated forecast of the valuation ratio:

$$\hat{x}_{t+h+1|t} = \hat{\rho}_0 \sum_{j=0}^{h} \hat{\rho}_1^j + \hat{\rho}_2 \sum_{j=0}^{h} \hat{\rho}_1^j \widehat{MY}_{t+h+1-j|t} + \hat{\rho}_1^{h+1} x_t. \tag{A1.6}$$

As the formula illustrates, the valuation forecast depends both on the current valuation x_t and the entire trajectory of the projected demographic trends over the forecast horizon.

Finally, by substituting (A1.6) into (A1.4) and (A1.5) and after simplifying, we can express the iterated long-horizon forecast $r_{t+k}(k)$ as [17]

$$\hat{r}_{t+k}(k) = \hat{\alpha}_0(k) + \hat{\alpha}_1(k) x_t + \sum_{h=0}^{k} \hat{\alpha}_{2,h}(k) \widehat{MY}_{t+k-h|t}, \tag{A1.7}$$

where, defining $\mathbf{1}_{h<k}$ as an indicator taking the value of one when $h < k$, the coefficients in (A1.7) are given by

$$\begin{aligned}
\hat{\alpha}_0(k) &= k\left(\widehat{\beta}_0 + \widehat{\beta}_2 \widehat{\rho}_0\right) + \frac{\left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_0 \left(k(1-\widehat{\rho}_1) - 1 + \widehat{\rho}_1^k\right)}{(1-\widehat{\rho}_1)^2} \\
\widehat{\alpha}_1(k) &= \frac{\left(\hat{\beta}_1 + \hat{\beta}_2 \hat{\rho}_1\right)\left(1 - \hat{\rho}_1^k\right)}{1-\hat{\rho}_1}, \\
\hat{\alpha}_{2,h}(k) &= \left(\widehat{\beta}_2 \widehat{\rho}_2\right) + \frac{\left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_2}{1-\widehat{\rho}_1} \left(1 - \widehat{\rho}_1^h\right) \mathbf{1}_{h>0}.
\end{aligned}$$

There are several implications of (A1.7) that are worth mentioning. The increase in the intercept $\hat{\alpha}_1(k)$ (if non-zero) as a function of horizon k is a standard result, since a small expected return will accumulate more over long horizons. In line with the long-horizon regression literature, the coefficient $\hat{\alpha}_1(k)$ on the current valuation x_t (if non-zero) increases in k. Its effect on future one-period returns diminishes as the horizon increases, but this effect is more than offset by the cumulative effect over longer periods. Perhaps most interestingly, the long-horizon return depends in a complicated way on the entire trajectory of the demographic variable throughout the full horizon of the return. As an alternative to the iterated forecast, this suggests a direct long-horizon projection on all of the predicted demographic ratios throughout the return horizon

$$\tilde{r}_{t+k}(k) = \tilde{\alpha}_0(k) + \tilde{\alpha}_1(k) x_t + \sum_{h=0}^{k} \tilde{\alpha}_{2,h}(k) \widehat{MY}_{t+k-h|t}, \tag{A1.8}$$

[17] The derivation is given in the Appendix A1.2 and includes (A1.9) as an intermediate step.

where the coefficients $\tilde{\alpha}_0(k), \tilde{\alpha}_1(k), \tilde{\alpha}_{2,0}(k), ..., \tilde{\alpha}_{2,k}(k)$ are directly estimated from the regression in (A1.8). In the empirical section, we refer to model (A1.8) as MY_D and to model (A1.7) as MY_R.

A1.2 Derivation of Equation A1.7

Plugging (A1.6) into (A1.4) and simplifying gives us the iterated forecast formula for the one period ahead return $h+1$ periods ahead as

$$\hat{r}_{t+h+1|t} = \begin{cases} \hat{\beta}_{0,h} + \hat{\beta}_2 \hat{\rho}_2 \widehat{MY}_{t+h+1|t} + \left(\hat{\beta}_1 + \hat{\beta}_2 \hat{\rho}_1\right) \hat{\rho}_2 \sum_{j=0}^{h-1} \hat{\rho}_1^j \widehat{MY}_{t+h-j|t} \\ \qquad + \left(\hat{\beta}_1 + \hat{\beta}_2 \hat{\rho}_1\right) \hat{\rho}_1^h x_t & \text{for } h > 0 \\ \widehat{\beta}_0 + \hat{\beta}_2 \hat{\rho}_0 + \widehat{\beta}_2 \widehat{\rho}_2 \widehat{MY}_{t+1|t} + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) x_t & \text{for } h = 0, \end{cases} \tag{A1.9}$$

Next, note that the $h+1$-period ahead return forecast can be expressed as

$$\begin{aligned}
\widehat{r}_{t+h+1|t} &= \widehat{\beta}_0 + \widehat{\beta}_1 \widehat{dp}_{t+h|t} + \widehat{\beta}_2 \widehat{dp}_{t+h+1|t} \\
&= \widehat{\beta}_0 + \widehat{\beta}_2 \widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{dp}_{t+h|t} + \left(\widehat{\beta}_2 \widehat{\rho}_2\right) \widehat{MY}_{t+h+1|t} \\
&= \widehat{\beta}_0 + \widehat{\beta}_2 \widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \left[\widehat{\rho}_0 \sum_{j=0}^{h-1} \widehat{\rho}_1^j + \widehat{\rho}_2 \sum_{j=0}^{h-1} \widehat{\rho}_1^j \widehat{MY}_{t+h-j|t} + \widehat{\rho}_1^h dp_t\right] + \left(\widehat{\beta}_2 \widehat{\rho}_2\right) \widehat{MY}_{t+h+1|t} \\
&= \widehat{\beta}_0 + \widehat{\beta}_2 \widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_0 \sum_{j=0}^{h-1} \widehat{\rho}_1^j + \left(\widehat{\beta}_2 \widehat{\rho}_2\right) \widehat{MY}_{t+h+1|t} + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_2 \sum_{j=0}^{h-1} \widehat{\rho}_1^j \widehat{MY}_{t+h-j|t} \\
&+ \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_1^h dp_t.
\end{aligned}$$

Hence, the sum of future returns is given as

$$\begin{aligned}
\sum_{h=1}^{k} r_{t+h} &= \sum_{h=0}^{k-1} r_{t+h+1} \\
&= \sum_{h=0}^{k-1} \Big\{ \widehat{\beta}_0 + \widehat{\beta}_2 \widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_0 \sum_{j=0}^{h-1} \widehat{\rho}_1^j + (\widehat{\beta}_2 \widehat{\rho}_2) \widehat{MY}_{t+h+1|t} + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_2 \sum_{j=0}^{h-1} \widehat{\rho}_1^j \widehat{MY}_{t+h-j|t} \\
&\quad + (\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1) \widehat{\rho}_1^h dp_t \Big\} \\
&= k\left(\widehat{\beta}_0 + \widehat{\beta}_2 \widehat{\rho}_0\right) + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_0 \sum_{h=0}^{k-1} \sum_{j=0}^{h-1} \widehat{\rho}_1^j + \left(\widehat{\beta}_2 \widehat{\rho}_2\right) \sum_{h=0}^{k-1} \widehat{MY}_{t+h+1|t} \\
&\quad + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_2 \sum_{h=0}^{k-1} \sum_{j=0}^{h-1} \widehat{\rho}_1^j \widehat{MY}_{t+h-j|t} + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \sum_{h=0}^{k-1} \widehat{\rho}_1^h dp_t \\
&= \widehat{\alpha}_0 + \widehat{\alpha}_1 dp_t + \left(\widehat{\beta}_2 \widehat{\rho}_2\right) \sum_{h=1}^{k} \widehat{MY}_{t+h|t} + \left(\widehat{\beta}_1 + \widehat{\beta}_2 \widehat{\rho}_1\right) \widehat{\rho}_2 \sum_{h=1}^{k-1} \sum_{j=1}^{h} \widehat{\rho}_1^{j-1} \widehat{MY}_{t+h-j+1|t},
\end{aligned}$$

where

$$
\begin{aligned}
\widehat{\alpha}_0 &= k\left(\widehat{\beta}_0+\widehat{\beta}_2\widehat{\rho}_0\right)+\left(\widehat{\beta}_1+\widehat{\beta}_2\widehat{\rho}_1\right)\widehat{\rho}_0\sum_{h=0}^{k-1}\sum_{j=0}^{h-1}\widehat{\rho}_1^{j} \\
&= k\left(\widehat{\beta}_0+\widehat{\beta}_2\widehat{\rho}_0\right)+\frac{\left(\widehat{\beta}_1+\widehat{\beta}_2\widehat{\rho}_1\right)\widehat{\rho}_0\left(k(1-\widehat{\rho}_1)-1+\widehat{\rho}_1^{k}\right)}{(1-\widehat{\rho}_1)^2} \\
\widehat{\alpha}_1 &= \frac{\left(\widehat{\beta}_1+\widehat{\beta}_2\widehat{\rho}_1\right)\left(1-\widehat{\rho}_1^{k}\right)}{1-\widehat{\rho}_1}.
\end{aligned}
$$

For the double sum above, we can use the change of variable $q = h - j$ to substitute in for j to obtain (note that if $j = 1$, then $q = h - j = h - 1$, if $j = h$, then $q = h - h = 0$, and if $j = h - q$ then $j - 1 = h - q - 1$)

$$
\sum_{h=1}^{k-1}\sum_{j=1}^{h}\widehat{\rho}_1^{j-1}\widehat{MY}_{t+h-j+1|t} = \sum_{h=1}^{k-1}\sum_{q=0}^{h-1}\widehat{\rho}_1^{h-q-1}\widehat{MY}_{t+q+1|t}
$$

Notice that $h = 1, 2, \ldots k-1$ means $1 \leq h < k-1$ and $q = 0, 1, \ldots h-1$ implies $q \leq h-1$ or $h \geq q+1$. Also $q \geq 0$ so $q+1 \geq 1$. Therefore $q+1 \leq h \leq k-1$. Also $0 \leq q \leq h-1 \leq k-2$. This allows us to rewrite the double sum as

$$
\begin{aligned}
\sum_{h=1}^{k-1}\sum_{q=0}^{h-1}\widehat{\rho}_1^{h-q-1}\widehat{MY}_{t+q+1|t} &= \sum_{q=0}^{k-2}\sum_{h=q+1}^{k-1}\widehat{\rho}_1^{h-q-1}\widehat{MY}_{t+q+1|t} \\
&= \sum_{q=0}^{k-2}\widehat{MY}_{t+q+1|t}\sum_{h=q+1}^{k-1}\widehat{\rho}_1^{h-q-1}
\end{aligned}
$$

For the inner sum, we can now use the change of variables $v = h - q - 1$ to write (note if $h = q+1$ then $v = 0$, and if $h = k-1$ then $v = k - q - 2$)

$$
\begin{aligned}
\sum_{q=0}^{k-2}\widehat{MY}_{t+q+1|t}\sum_{h=q+1}^{k-1}\widehat{\rho}_1^{h-q-1} &= \sum_{q=0}^{k-2}\widehat{MY}_{t+q+1|t}\sum_{v=0}^{k-q-2}\widehat{\rho}_1^{v} \\
&= \sum_{q=0}^{k-2}\widehat{MY}_{t+q+1|t}\frac{1-\widehat{\rho}_1^{k-q-1}}{1-\widehat{\rho}_1} \\
&= \frac{1}{1-\widehat{\rho}_1}\sum_{q=0}^{k-2}\left(1-\widehat{\rho}_1^{k-q-1}\right)\widehat{MY}_{t+q+1|t},
\end{aligned}
$$

which completes our derivation.

A2 Appendix: Details of Historical Census Forecast Collection and Utilization

We employ data from the Census Bureau on the following seven historical projections: (1) 1950 forecasts, annual projection available from 1951 to 1960; (2) 1963 forecasts, annual projection available from 1963 to 1985; (3) 1970 forecasts, annual projection available from 1970 to 1985; (4) 1983 forecasts, annual projection available from 1983 to 2000; (5) 1988 forecasts, annual projection available from 1988 to 2005; (6) 2001 forecasts, annual projection available from 2001 to 2020; (7) 2012 forecasts, annual projection available from 2012 to 2020. The historical projection includes three series: high, middle, and low, based on different fertility assumptions. In this study, we use the high series since it is the only series that is complete for all of historical projections.[18]

Three practical issues arise when using the real time historical Census forecasts: stale forecasts, multiple (stale) forecasts, and missing forecasts. These issues arise because the Census Bureau updates its forecast infrequently. Let t denote the time of the forecast. Let h denote the forecast horizon. In other words, we make a forecast for time $t+h$ sitting at time t. We have a historical forecast available to us at time $t-J$ where $J \geq 0$. Let's call this $\widetilde{MY}_{t+h|t-J}$. This is the forecast by the Census Bureau taken at time $t-J$ but forecasting date $t+h$.

The first issue that arises is the problem of multiple forecasts. For example, the forecast for the year 1975 is available in both the 1963 forecasts and 1970 forecasts. To formalize this, suppose that we have two forecasts for the same year, $\widetilde{MY}_{t+h|t-J}$ and $\widetilde{MY}_{t+h|t-J'}$ where $J < J'$. In this case, we simply chose $\widetilde{MY}_{t+h|t-J}$ as the more recent of the two forecasts. In the example above, we therefore use the prediction from the more recent 1970 Census forecast series to forecast the year 1975.

We next confront the problem of missing forecasts. We have two types of missing forecast problems. First, there are two years for which we have no Census forecast: 1961 and 1962. This impacts both our one- and five-year forecasts. The second type of missing forecasts matters only for the five-year forecasts. In some cases, we have Census forecasts extending one, but not five years ahead. For instance, sitting in the year 1982, we need unconditional predictions from 1983 to 1987 for our five-year forecast. However, in 1982, our latest forecasts were made in 1970 and available only until 1985. In total, twenty-three of our five-year forecasts miss one or more of the single-year projections from which they are constructed.[19] Table 6 shows the projection availability

[18] Middle and low series are not available in 1950 forecasts.

[19] Since the last observation of the 1950 forecasts is 1960 and the first observation of the 1960 forecasts is 1963, we lack one year for the 1956 five-year forecast (1961), two years for the 1957 five-year forecast (1961-1962), three years

yearly break down.

To rationally handle the missing data using only information known to the forecaster at the time of the forecast, we propose the following approach. Suppose that we wish to forecast MY_{t+h} at time t, but the latest date for which we have a forecast is $t+h'$ where $-J < h' < h$, which we denote by $\widetilde{MY}_{t+h'|t-J}$. If $h' > 0$ then we simply use $\widetilde{MY}_{t+h'|t-J}$ in place of $\widetilde{MY}_{t+h|t-J}$. If $h' \leq 0$, we instead use the realized value MY_t in place of $\widetilde{MY}_{t+h'|t-J}$. As an example, recall that in 1960 there was no Census projection available for 1961. Therefore, we use the actual realized demographic ratio for the year 1960 as the forecast for 1961: $\widetilde{MY}_{1961|1960} = MY_{1960}$. As a contrasting example when $h' > 0$, to perform a five-year prediction in 1982, we require yearly forecasts for both 1986 and 1987. However, in 1982 the latest available Census forecast is the forecast for 1985 that was produced in 1970. Since 1985 is more recent than 1982, we use the 1985 forecast to replace the missing 1986 and 1987 projections: $\widetilde{MY}_{1986|1982} = \widetilde{MY}_{1987|1982} = MY_{1985|1970}$.

The final issue to be addressed are the stale forecasts. When $J \geq 1$, $\widetilde{MY}_{t+h,t-J}$ is a stale forecast. Although we do not have $\widetilde{MY}_{t+h|t-0}$ available, we do have MY_t available – this is the actual value at time t. We have explored two possibilities for addressing this. The simplest solution is to employ the stale forecast. In this case, we forecast, MY_{t+h} by $\widetilde{MY}_{t+h|t-J}$. In our tables and figures we refer to this as a "stale" forecast.

Our second approach is motivated by breaking MY_{t+h} into a (time-t) level and a change (between t and $t+h$) as follows:

$$MY_{t+h} = MY_t + [MY_{t+h} - MY_t] \tag{A2.1}$$

We can do the same thing with the Census Bureau forecast

$$\widetilde{MY}_{t+h|t-J} = \widetilde{MY}_{t|t-J} + [\widetilde{MY}_{t+h|t-J} - \widetilde{MY}_{t|t-J}] \tag{A2.2}$$

We recognize that at time t, the forecast $\widetilde{MY}_{t|t-J}$ is no longer required since MY_t is already known. Therefore we replace $\widetilde{MY}_{t|t-J}$ by MY_t in (A2.2) to obtain a new forecast defined as

$$\widehat{MY}_{t+h|t-J} = MY_t + [\widetilde{MY}_{t+h|t-J} - \widetilde{MY}_{t|t-J}] \tag{A2.3}$$

The first term on the RHS is the current value of MY_t which we know and therefore do not need to forecast. The second term is the most recent forecast of the change between now (t) and the

for the 1958 five-year forecast (1961-1963), four years for the 1959 five-year forecast (1961-1964), and five years for the 1960-1962 five-year forecast (1961-1965, 1962-1966, 1963-1967). Similarly, since the last observation of the 1970 forecasts is 1985 and the first observation of the 1980 forecast is 1983, we lack one year for the 1981 five-year forecast (1986) and two years for the 1982 five-year forecast (1986-1987).

date we want to forecast $(t + h)$. In our tables and figures, we refer to this second approach as the "real time" forecast. Since both methods yield similar results in practice, we include only the simpler stale forecast results in the paper, relegating the real time forecast results to the additional appendix.

A3 Appendix: Details of COVID-19 Scenarios

A3.1 COVID-19 Fatalities by Age Category

We define $D_{A,t}$ to be the number of U.S. COVID-19 deaths in age category A during year t. Our data from the CDC on COVID-19 deaths for individuals under 50 are grouped into the following age categories, where we add the category name and symbol:

The following 4 age relevant categories are included in the COVID-19 data:

Category	Variable	Age Range
Children 0-17 (CD17)	$D_{C17,t}$	0-17
Young 18-29 (Y18)	$D_{Y18,t}$	18-29
Thirty-Something (TS)	$D_{TS,t}$	30-39
Middle (M)	$D_{M,t}$	40-49

Since the CDC definition of young does not match the Census Bureau definition, we proportionally rescale deaths from both the Children and Young categories to work with the following four age categories.

Category	Abbreviation	Age Range
Children 0-19 (C)	$D_{C,t} = 19/17 D_{C17,t}$	0-19
Young 20-29 (Y)	$D_{Y,t} = 10/12 D_{Y18,t}$	20-29
Thirty-Something (TS)	$D_{TS,t}$	30-39
Middle (M)	$D_{M,t}$	40-49

A3.2 COVID-19 Scenarios Model by Age Category

We have data on COVID-19 deaths only for 2020 and 2021. For future years, we must consider different possible scenarios. We introduce a scenario model which is simple, flexible, and transparent. It entails a transition from a pandemic to an endemic. The transition involves a gradual linear decline in COVID-19 deaths from the 2021 death level to a lesser but non-zero endemic death level. This requires the specification of only two parameters: an end-date to the transition level that determines its speed and the level of deaths in the endemic state relative to the 2021 pandemic state. This quantifies the severity of the endemic state to which we transition. Since we are working at a yearly level, we don't model seasonality or waves.

To formalize this, define t^* as the year in which COVID-19 becomes fully endemic instead of pandemic. Define $\psi_{A,0} = D_{A,t^*}$ as average death level in age category A once COVID-19 has become endemic. Define

$$\psi_{A,1} = \frac{D_{A,t^*} - D_{A,2021}}{t^* - 2021} = \frac{\psi_{A,0} - D_{A,2021}}{t^* - 2021}$$

as the slope parameter that determines the yearly decline in deaths for age category A. Note that $\psi_{A,1}$ is fully determined by $(\psi_{A,0}, t^*)$ so that only two parameters determine the linear death trajectory. The implied numbers of deaths for age category A is thus given by:

$$D_{A,t} = \left\{ \begin{array}{ll} D_{A,2021} + \psi_{A,1}(t - 2021) & 2022 \leq t < t^* \\ \psi_{M,0} & t \geq t^* \end{array} \right\}. \tag{A3.1}$$

Different future scenarios can be considered via alternative choices for t^* and $\psi_{A,0}$. In our main analysis, we simplify our scenario model further by setting the yearly endemic death levels for all four age categories as the same fraction (θ) of 2021 pandemic death levels, such that[20]

$$\psi_{A,0} = \theta D_{A,2021} \tag{A3.2}$$

Now we can define scenarios in terms of two easily interpretable parameters: t^* controls the speed of the transition from pandemic to endemic, θ controls the severity of the eventual endemic. We consider four scenarios that combine slow and fast transitions with mild and severe endemics as follows

Scenario 1 (Good Case)	Quick End, Mild Endemic	$t^* = 2023$	$\theta = 0.10$
Scenario 2 (Mixed)	Quick End, Severe Endemic	$t^* = 2023$	$\theta = 0.30$
Scenario 3 (Mixed)	Slow End, Mild Endemic	$t^* = 2030$	$\theta = 0.10$
Scenario 4 (Bad Case)	Slow End, Severe Endemic	$t^* = 2030$	$\theta = 0.30$

To save space, we present the results from Scenario 4 in the paper, and the results from Scenarios 1-3 in Section B7 of our not-for-publication appendix.

A3.3 Deaths by age and year

Due to population ageing, a death of a 39 year-old in 2020 will result in one fewer 40 year old in 2021. To handle this progression in a general, flexible way, we need to track deaths by age and year. Since we do not have data on deaths by yearly age, this requires us to make an assumption on the distribution of deaths within each age category. Due to its simplicity and transparency, we assume

[20]In additional results available upon request, we have also explored one case in which θ depends on A.

an equal distribution of deaths within age categories. Denoting $d_{a,t}$ as the number of COVID-19 deaths for individuals of age a in year t, we thus define:

$$d_{t,a} = \begin{cases} \frac{1}{19}D_{C,t} & \text{for } t \geq 2020 \text{ and } 1 \leq a \leq 19 \\ \frac{1}{10}D_{Y,t} & \text{for } t \geq 2020 \text{ and } 20 \leq a \leq 29 \\ \frac{1}{10}D_{TS,t} & \text{for } t \geq 2020 \text{ and } 30 \leq a \leq 39 \\ \frac{1}{10}D_{M,t} & \text{for } t \geq 2020 \text{ and } 40 \leq a \leq 49 \\ 0 & \text{otherwise} \end{cases}$$

Similar analysis could be conducted replacing this definition by a more complicated and, perhaps more realistic, functional form.

A3.4 Cumulative Removals from Middle and Young Age Ranges

Next, we denote $R_{M,t}$ and $R_{Y,t}$ as the cumulative size of the population that we remove from the original M_t and Y_t projections, respectively, in order to adjust for both current and prior year COVID-19 deaths. To adjust M_t, we remove all COVID-19 deaths for the age range 40-49 in year t, for the age range 39-48 in year $t-1$, for the age range 38-47 in year $t-2$, etc. Defining $J_t = t-2020$ as the number of years since the onset of the pandemic this implies:

$$R_{M,t} = \sum_{a=40}^{49} d_{t,a} + \sum_{a=40-1}^{49-1} d_{t-1,a} + \ldots + \sum_{a=40-J_t}^{49-J_t} d_{t-J_t,a} = \sum_{j=1}^{J_t} \sum_{a=40-j}^{49-j} d_{t-j,a}$$

Similarly, we can define:

$$R_{Y,t} = \sum_{a=20}^{29} d_{t,a} + \sum_{a=20-1}^{29-1} d_{t-1,a} + \ldots + \sum_{a=20-J_t}^{29-J_t} d_{t-J_t,a} = \sum_{j=1}^{J_t} \sum_{a=20-j}^{29-j} d_{t-j,a}$$

Finally, we define our COVID-19 adjusted projections for M_t and Y_t as the original Census projections minus the removals: as $M_{t,\text{Adusted}} = M_{t,\text{Census}} - R_{M,t}$ and $Y_{t,\text{Adjusted}} = Y_{t,\text{Census}} - R_{Y,t}$ where $M_{t,\text{Census}}$ and $Y_{t,\text{Census}}$ are the original census projections.

Table 1: **Dividend-Price Ratio Model: OLS and IVX Estimation Results**

Model:	AR(1)		Augmented AR(1)	
			MY and	
	lag dp		lag dp	
	OLS	IVX	OLS	IVX
		Panel A: 1901-2015		
Const	-0.368^{**}	-0.368^{**}	-0.395^{***}	-0.395^{***}
lag dp	0.889^{***}	0.886^{***}	0.757^{***}	0.776^{***}
MY			-0.504^{***}	-0.400^{*}
Model Test	395.415^{***}	374.89^{***}	222.308^{***}	409.87^{***}
p-value	0	0	0	0
		Panel B: 1947-2015		
Const.	-0.309^{**}	-0.309^{**}	-0.335^{***}	-0.335^{***}
lag dp	0.915^{***}	0.904^{***}	0.829^{***}	0.777^{***}
MY			-0.312^{**}	-0.408^{**}
Model Test	377.506^{***}	170.61^{***}	200.331^{***}	228.78^{***}
p-value	0	0	0	0

$***$ significantly different from zero at the 1% level, $**$, significantly different from zero at the 5% level, $*$ significantly different from zero at the 10% level. This table reports OLS and IVX estimates of (2) & (A1.1). Columns 2-3 provide estimates for the pure AR(1) process in (2). Columns 4-5 provide estimates for the augmented AR(1) process including the MY ratio in (A1.1) as demographic controls. Results for the time period 1901-2015 are shown in Panel A and results for 1947-2015 period are shown in Panel B. For the IVX estimation, we set $c = -1$, $\alpha = 0.95$. Sensitivity results with respect to these IVX tuning parameters are included in the additional (not-for-publication) Appendix.

Table 2: **Return Regression Models: OLS and IVX Estimation Results**

Model:	lag dp		MY and lag dp	
	OLS	IVX	OLS	IVX
Panel A: 1901-2015				
Const	0.257***	0.257**	-0.022	-0.022
$\widehat{dp}_{t+1\|t}$			-0.755^{***}	-0.649^{**}
dp_t	0.062	0.062	0.733**	0.642**
Model Test	2.615	2.537	5.789***	7.817**
p-value	0.109	0.111	0.004	0.020
Panel B: 1947-2015				
Const	0.402**	0.402**	0.032	0.032
$\widehat{dp}_{t+1\|t}$			-1.198^{***}	-1.549^{***}
dp_t	0.098**	0.114*	1.194***	1.550***
Model test	5.008**	3.561*	6.509***	12.250***
p-value	0.029	0.0589	0.003	0.0022

$***$ significantly different from zero at the 1% level, $**$, significantly different from zero at the 5% level, $*$ significantly different from zero at the 10% level. This table reports OLS and IVX estimation results for equations (1) for $k = 1$, (3), and (4). The dependent variable in all cases are annual log returns including dividends. Columns 2-3 (lag dp), provide the estimates of (1) for $k = 1$ in which only on the past dp_t is employed as a predictor. Columns 4-5 provide estimates of (4), using MY as the demographic ratio projection $dp_{t+1|t}$ in (A1.1). Results for the time period 1901-2015 are shown in Panel A and results for 1947-2015 period are shown in Panel B. Sensitivity results with respect to these IVX tuning parameters are included in the additional (not-for-publication) Appendix.

Table 3: **Results of Return Regression Models (Pseudo Out-of-Sample, Recursive, Five-Year Ahead Forecast)**

Panel A: 1901-2015: using an initial training period of tp years.

training period (tp)	$tp = 30$		$tp = 40$		$tp = 60$	
Model	OOS MSE	OOS R^2	OOS MSE	OOS R^2	OOS MSE	OOS R^2
HM	0.1438	0	0.1384	0	0.1171	0
PR	0.1336	0.0708	0.1333	0.0370	0.1175	-0.0037
FGT	0.1278	0.1110	0.1239	0.1047	0.1366	-0.1668
MY_D	0.1429	0.0058	0.1279	0.0757	**0.0786**	**0.3287**
MY_R	**0.0997**	**0.3065**	**0.0876**	**0.3672**	0.0881	0.2471
Model	p-value (HM, CW)	p-value (PR, CW)	p-value (HM, CW)	p-value (PR, CW)	p-value (HM, CW)	p-value (PR, CW)
PR	0.077*		0.1186		0.1744	
FGT	0.0253**	0.0495**	0.0518*	0.0847*	0.1480	0.3101
MY_D	0.0063***	0.0075***	0.0142**	0.0147**	0.0560*	0.0658*
MY_R	0.0291**	0.00258**	0.0320**	0.0210**	0.0856*	0.0588*

Panel B: 1947-2015: using an initial training period of tp years.

training period (tp)	$tp = 20$		$tp = 25$		$tp = 30$	
Model	OOS MSE	OOS R^2	OOS MSE	OOS R^2	OOS MSE	OOS R^2
HM	0.1569	0	0.1286	0	0.1168	0
PR	0.1611	-0.0269	0.1634	-0.2704	0.1769	-0.5148
FGT	0.1649	-0.0511	0.1709	-0.3293	0.1875	-0.6058
MY_D	**0.0942**	**0.3993**	**0.0788**	**0.3870**	**0.0856**	**0.2670**
MY_R	0.0990	0.3689	0.1079	0.1613	0.1200	-0.0275
Model	p-value (HM, CW)	p-value (PR, CW)	p-value (HM, CW)	p-value (PR, CW)	p-value (HM, CW)	p-value (PR, CW)
PR	0.1178		0.1697		0.2675	
FGT	0.0937*	0.1418	0.1156	0.1514	0.2004	0.2003
MY_D	0.0532*	0.0785*	0.0521*	0.0729*	0.0810*	0.0919*
MY_R	0.0826*	0.0415**	0.0895*	0.0570*	0.1632	0.0758*

$***$, $**$, $*$ significantly out-performs the benchmark forecasts at the 1% level, 5% level, and 10% level, respectively. Reported p-values are one-sided. This table provides five-year out-of-sample forecasting results. The dependent variable in all cases are annual log returns including dividends. Results for the time period 1901-2015 are shown in Panel A and results for 1947-2015 period are shown in Panel B. HM stands for historical mean. PR denotes the five-year ahead forecast obtained by forward recursion from the one-year ahead predictive regression forecast. FGT provides the estimates of (A1.4) in which the $\hat{x}_{t+1}$ is estimated by MY_t. MY_D shows the results from (A1.8) while MY_R refers to the iterative model (A1.7). The OOS R^2 uses the HM as its benchmark and is defined in Footnote 11. CW is the Clark and West (2007) test. The columns marked (HM, CW) and (PR, CW) provide p-values for the CW test using HM and PR, respectively, as benchmarks.

Table 4: **Results of Return Regression Models (Pseudo Out-of-Sample, Rolling, Five-Year Ahead Forecast)**

Panel A: 1901-2015: using a rolling window of length w

window (w)	$w = 30$		$w = 40$		$w = 60$	
Model	OOS MSE	OOS R^2	OOS MSE	OOS R^2	OOS MSE	OOS R^2
HM	0.1698	0	0.1492	0	0.1266	0
PR	0.1564	0.0787	0.1509	-0.0112	0.1254	0.0099
FGT	0.1572	0.0744	0.1113	0.2540	0.1566	-0.2367
MY_D	0.2374	-0.3983	0.1292	0.1338	0.0979	0.2270
MY_R	**0.1286**	**0.2428**	**0.1080**	**0.2757**	**0.0940**	**0.2578**
Model for $\widehat{dp}_{t+1\|t}$	p-value (HM, GW)	p-value (PR, GW)	p-value (HM, GW)	p-value (PR, GW)	p-value (HM, GW)	p-value (PR, GW)
PR	0.3555		0.5228		0.4866	
FGT	0.3896	0.5129	0.2151	0.0882*	0.6722	0.7733
MY_D	0.6919	0.7232	0.3732	0.3708	0.2369	0.2613
MY_R	0.2571	0.3258	0.1964	0.1582	0.1842	0.1384

Panel B: 1947-2015: using a rolling window of length w

window (w)	$w = 20$		$w = 25$		$w = 30$	
Model for $\widehat{dp}_{t+1\|t}$	OOS MSE	OOS R^2	OOS MSE	OOS R^2	OOS MSE	OOS R^2
HM	0.1941	0	0.1694	0	0.1465	0
PR	0.2331	-0.2008	0.2492	-0.4712	0.2082	-0.4218
FGT	0.2181	-0.1236	0.2962	-0.7487	0.2354	-0.6073
MY_D	0.1506	0.2242	**0.0778**	**0.5409**	**0.0637**	**0.5652**
MY_R	**0.0914**	**0.5291**	0.0848	0.4991	0.0943	0.3561
Model for $\widehat{dp}_{t+1\|t}$	p-value (HM, GW)	p-value (PR, GW)	p-value (HM, GW)	p-value (PR, GW)	p-value (HM, GW)	p-value (PR, GW)
PR	0.6840		0.8382		0.9082	
FGT	6741	0.4196	0.9458	0.8523	0.9531	0.7360
MY_D	0.0786*	0.1570	0.0026***	0.0454**	0.0033***	0.0215**
MY_R	0.0075***	0.0393**	0.0004***	0.0344**	0.0025***	0.0229**

$***$, $**$, $*$ significantly out-performs the benchmark forecasts at the 1% level, 5% level, and 10% level, respectively. Reported p-values are one-sided. This table provides five-year out-of-sample forecasting results based on the rolling forecast method. It also provides the out-performance test results. The dependent variable in all cases are annual log returns including dividends. Results for the time period 1901-2015 are shown in Panel A and results for 1947-2015 period are shown in Panel B. HM stands for historical mean. PR denotes the five-year ahead forecast obtained by forward recursion from the one-year ahead predictive regression forecast. FGT provides the estimates of (A1.4) in which the $\hat{x}_{t+1}$ is estimated by MY_t. MY_D shows the results from (A1.8) while MY_R refers to the iterative model (A1.7). The OOS R^2 uses the HM as its benchmark and is defined in Footnote 11. GW is the adjusted one-sided Giacomini and White (2006) test. The columns marked (HM, GW) and (PR, GW) provide p-values for the GW test using HM and PR, respectively, as benchmarks.

Table 5: **Results of Return Regression Models (True Out-of-Sample, Five-Year Ahead Forecast, Stale)**

Panel A: Recursive, 1951-2015: using an initial training period of tp years.

training period (tp)	tp=20		tp=25		tp=30	
Model	OOS MSE	OOS R^2	OOS MSE	OOS R^2	OOS MSE	OOS R^2
HM	0.1270	0	0.1157	0	0.1273	0
PR	0.1531	-0.2048	0.1629	-0.4074	0.1684	-0.3233
FGT	0.1784	-0.4043	0.1871	-0.6169	0.2036	-0.6000
MY_D	**0.0945**	**0.2563**	**0.1034**	**0.1063**	**0.1163**	**0.0862**
MY_R	0.1020	0.1974	0.1087	0.0611	0.1216	0.0443
Model	p-value (HM, CW)	p-value (PR, CW)	p-value (HM, CW)	p-value (PR, CW)	p-value (HM, CW)	p-value (PR, CW)
PR	0.1996		0.3341		0.3110	
FGT	0.1395	0.1341	0.2681	0.1931	0.3625	0.4877
MY_D	0.0525*	0.0833*	0.0844*	0.1151	0.1083	0.1665
MY_R	0.0811*	0.0563*	0.1371	0.0854*	0.1730	0.1475

Panel B: Rolling, 1951-2015: using an initial training period of tp years.

training period (tp)	tp=20		tp=25		tp=30	
Model	OOS MSE	R^2	OOS MSE	R^2	OOS MSE	R^2
HM	0.1754	0	0.1593	0	0.1550	0
PR	0.2414	-0.3761	0.2697	-0.6934	0.2137	-0.3786
FGT	0.2309	-0.3164	0.3228	-1.0265	0.2653	-0.7116
MY_D	0.1834	-0.0453	**0.1000**	**0.3723**	**0.0915**	**0.4097**
MY_R	**0.1009**	**0.4251**	0.1075	0.3251	0.1219	0.2140
Model for $\widehat{dp}_{t+1\|t}$	p-value (HM, GW)	p-value (PR, GW)	p-value (HM, GW)	p-value (PR, GW)	p-value (HM, GW)	p-value (PR, GW)
PR	0.7981		0.9099		0.8639	
FGT	0.9163	0.4489	0.9822	0.8576	0.9807	0.8920
MY_D	0.5803	0.2738	0.0155**	0.0481**	0.0657*	0.0841*
MY_R	0.0018***	0.0546*	0.0100***	0.0455**	0.0775*	0.0769*

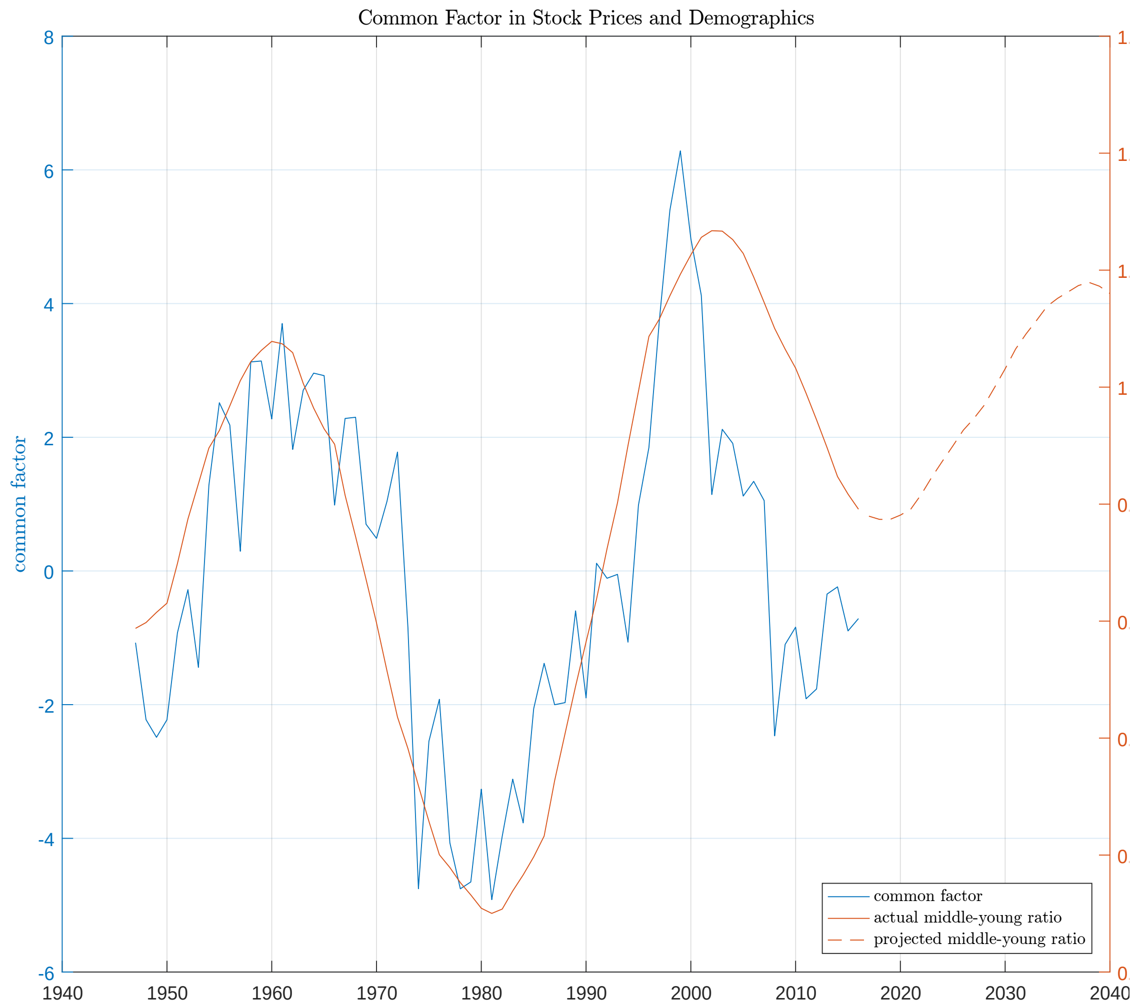
$***$, $**$, $*$ significantly out-performs the benchmark forecasts at the 1% level, 5% level, and 10% level, respectively. Reported p-values are one-sided. This table provides five-year out-of-sample unconditional forecasting results using the stale forecasts. The dependent variable in all cases are annual log returns including dividends. Results for the recursive method are shown in Panel A and results for the rolling method are shown in Panel B. Columns 2-3 report the out-of-sample mean square error (OOS MSE) and out-of-sample R^2 (OOS R^2) with a 20-year training period for panel A/panel B. Columns 4-5 show OOS MSE and OOS R^2 with a 25-year training period for panel A/panel B. Columns 6-7 give OOS MSE and OOS R^2 with a 30-year training period for panel A/panel B. This table also provides the out-performance test results. HM is the out-of-sample historical mean. PR is the predictive regression model. FGT provides the estimates of (A1.4) in which the $\hat{x}_{t+1}$ is estimated by MY_t. MY_D shows the results from (A1.8) while MY_R refers to the iterative model (A1.7). The OOS R^2 uses the HM as its benchmark and is defined in Footnote 11. CW is the Clark and West (2007) test. GW is the adjusted one-sided Giacomini and White (2006) test. The columns marked (HM, CW) and (PR, CW) provide p-values for the CW test using HM and PR, respectively, as benchmarks. Similarly, (HM,GW) and (PR,GW) denote p-values with HM or PR as benchmark using the GW test.

Table 6: Projection Availability Coverage Broken Down by Year

Year	1950 forecasts	1963 forecasts	1970 forecasts	1983 forecasts	1988 forecasts	2000 forecasts	2012 forecasts
1950	X						
1951	X						
1952	X						
1953	X						
1954	X						
1955	X						
1956	X						
1957	X						
1958	X						
1959	X						
1960	X						
1961							
1962							
1963		X					
1964		X					
1965		X					
1966		X					
1967		X					
1968		X					
1969		X					
1970		X	X				
1971		X	X				
1972		X	X				
1973		X	X				
1974		X	X				
1975		X	X				
1976		X	X				
1977		X	X				
1978		X	X				
1979		X	X				
1980		X	X				
1981		X	X				
1982		X	X				
1983		X	X	X			
1984		X	X	X			
1985		X	X	X			
1986				X			
1987				X			
1988				X	X		
1989				X	X		
1990				X	X		
1991				X	X		
1992				X	X		
1993				X	X		
1994				X	X		
1995				X	X		
1996				X	X		
1997				X	X		
1998				X	X		
1999				X	X		
2000				X	X		
2001					X	X	
2002					X	X	
2003					X	X	
2004					X	X	
2005					X	X	
2006						X	
2007						X	
2008						X	
2009						X	
2010						X	
2011						X	
2012						X	X
2013						X	X
2014						X	X
2015						X	X

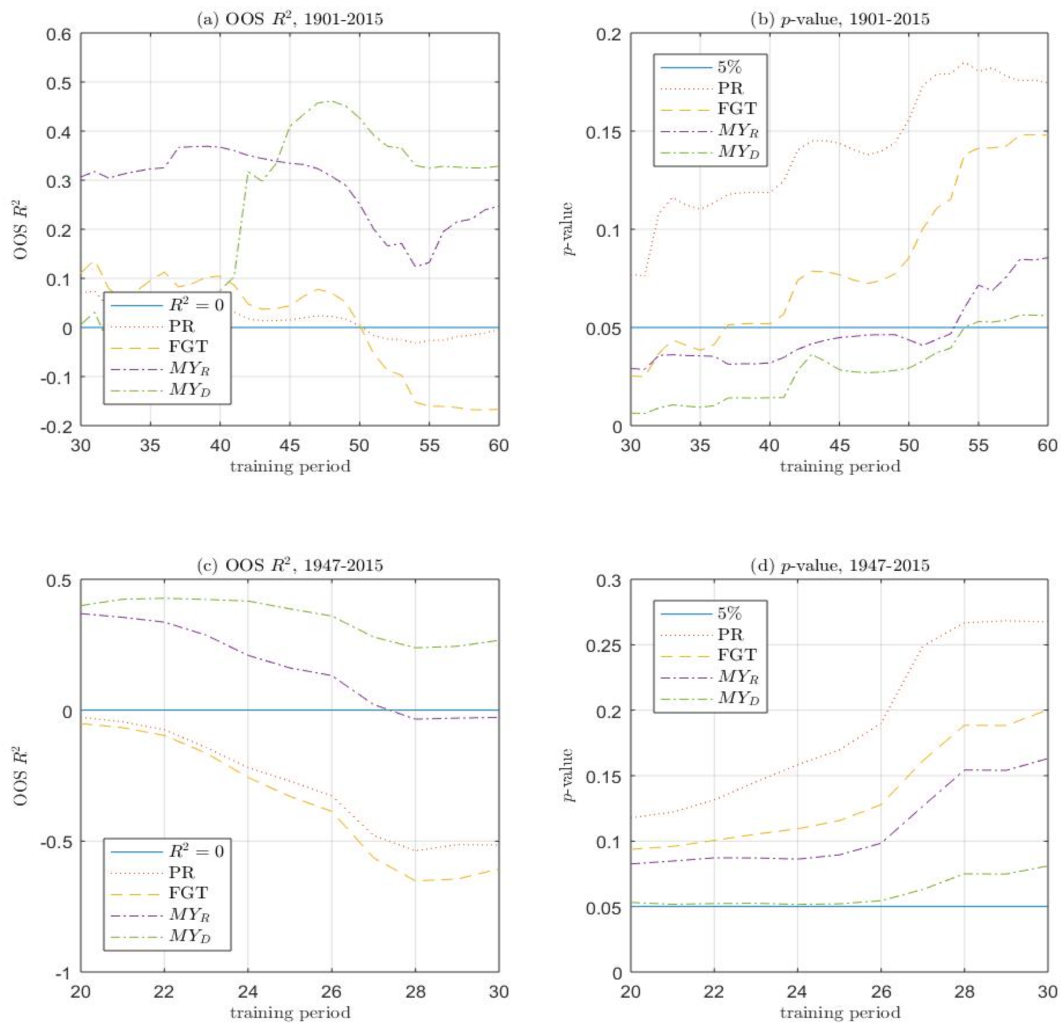
Note: An ‘x’ denotes coverage of a given year (row) by a given Census projection series (column)

Figure 1: Common Factor in Stock Returns and Valuation Ratios, and Demographics



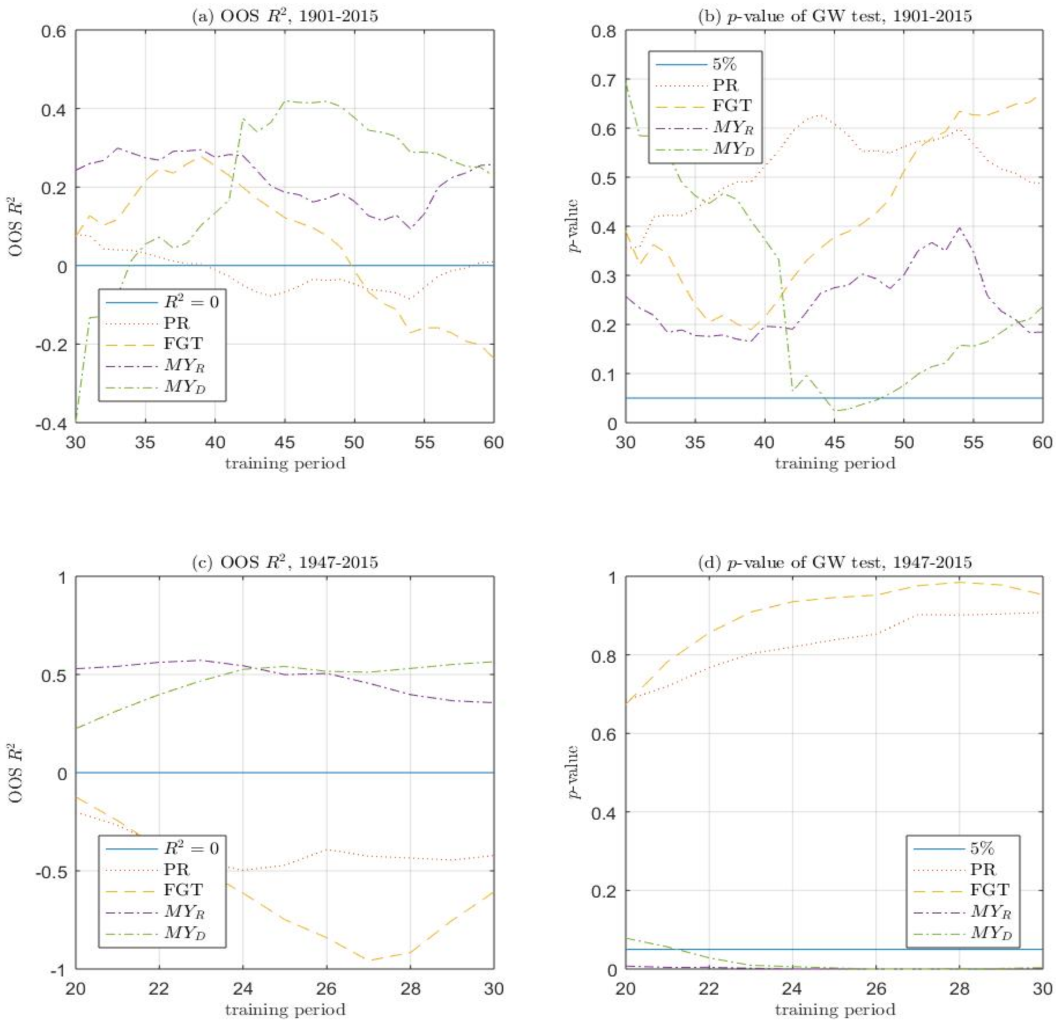
This figure plots the (actual and projected) middle-young ratio and the common factor in stock returns and valuation ratios, constructed as the detrended cumulative sum of the first principal component of S&P500 returns, changes in S&P500 price-dividend ratio, changes in S&P500 dividend yield, and changes in S&P500 earnings-price ratio.

Figure 2: Pseudo Out-of-Sample Recursive Forecasts, Five-Year Ahead



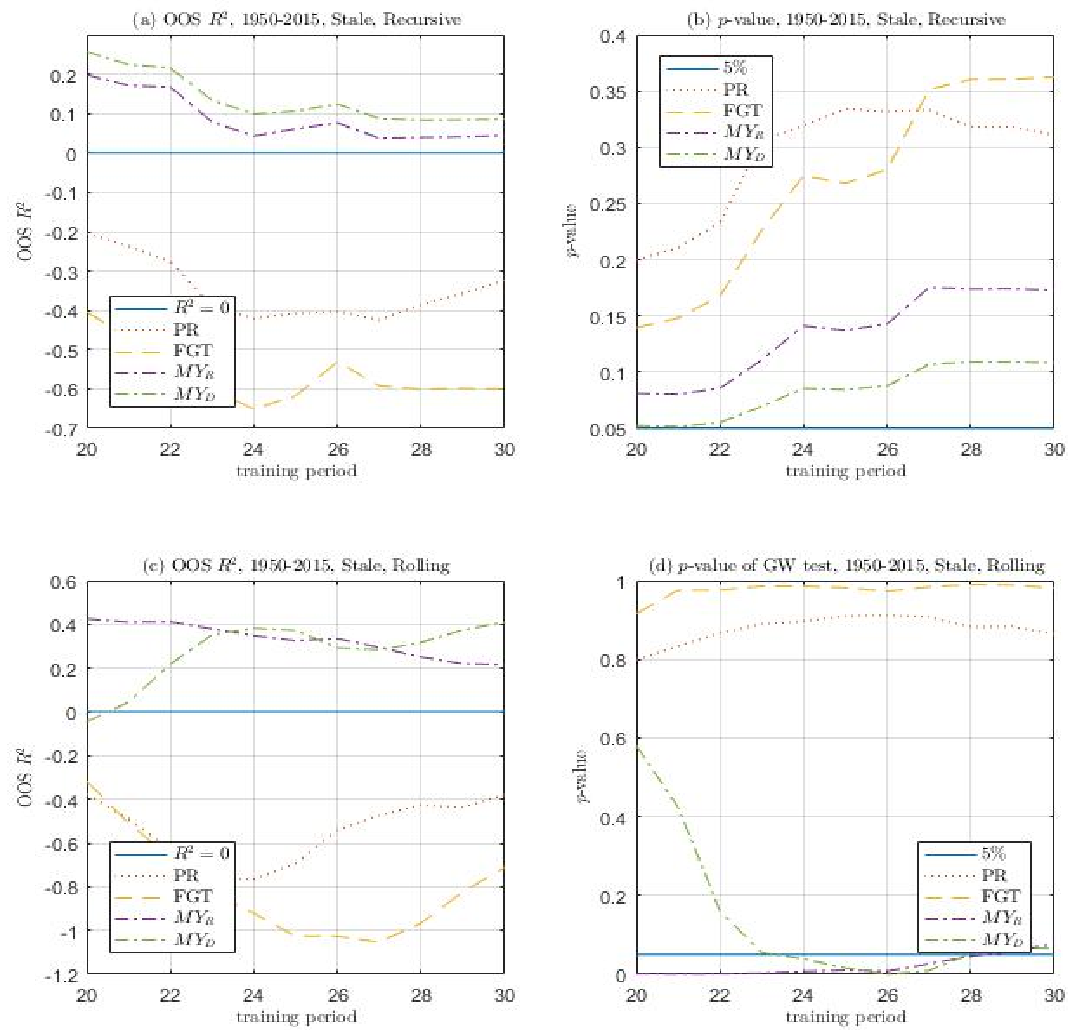
This figure provides five-year ahead out-of-sample forecasting results with respect to different window size based on the recursive method. (a)-(b) show the OOS R^2 and one-sided p-values for the CW test in the full sample period 1901-2015 with the window size varying from 30-60 years. (c)-(d) show the analogous result for the post WWII period 1947-2015 with the window size varying from 20-30 years. The green, purple, red, and yellow dashed lines stand for the MY_D, MY_R, predictive regression model, and FGT model respectively. The blue solid line stands for the historical mean model in (a) and (c) and the 5% significant level in (b) and (d). The OOS R^2 uses the HM as its benchmark and is defined in Footnote 11.

Figure 3: Pseudo Out-of-Sample Rolling Forecasts, Five-Year Ahead



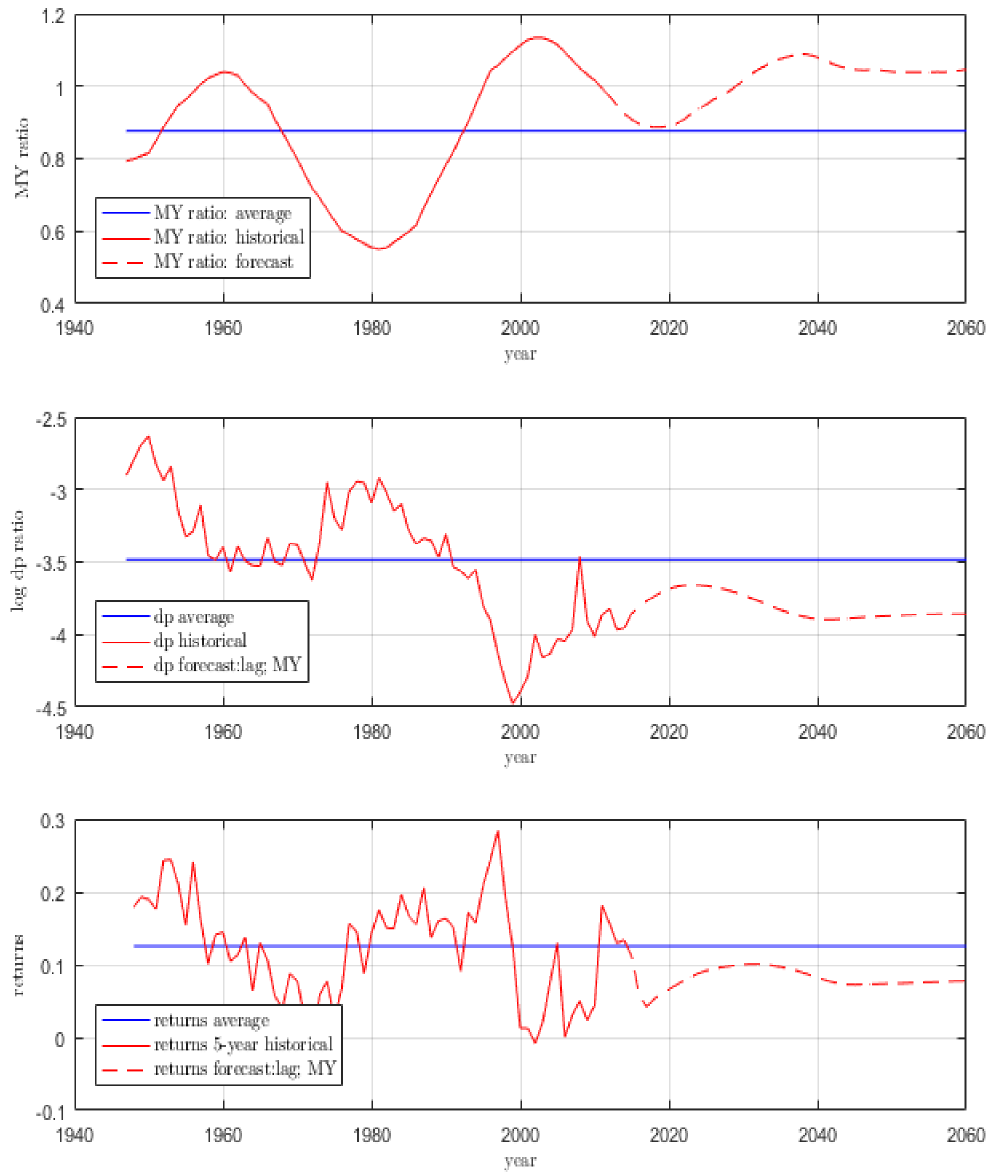
This figure provides five-year ahead out-of-sample forecasting results with respect to different window size based on the rolling method. (a)-(b) show the OOS R^2 and p-values for GW test for the full sample period 1901-2015 with the window size varying from 30-60 years. (c)-(d) show the analogous result for the post WWII period 1947-2015 with the window size varying from 20-30 years. The green, purple, red, and yellow dashed lines stand for the MY_D, MY_R, predictive regression model, and FGT model respectively. The blue solid line stands for the historical mean model in (a) and (c) and the 5% significant level in (b) and (d). The OOS R^2 uses the HM as its benchmark and is defined in Footnote 11.

Figure 4: True Out-of-Sample Forecasts, Five-Year Ahead, Stale



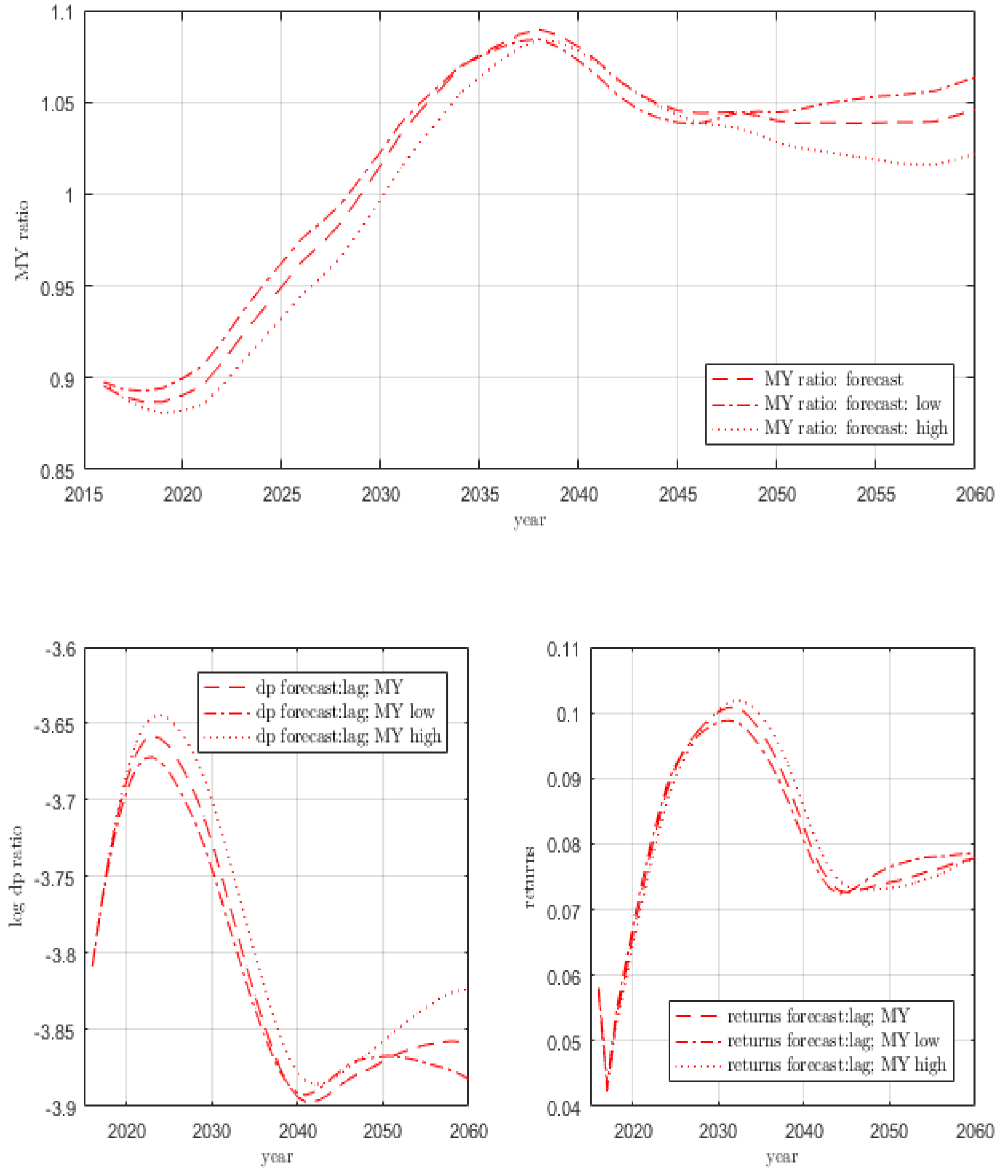
This figure provides five-year ahead out-of-sample forecasting results with respect to different window sizes using stale forecasts. (a)-(b) show the OOS R^2s and one-sided p-values for the CW test using the recursive method. (c)-(d) show the analogous results for the OOS R^2 GW test p-value using the rolling method. The green, purple, red, and yellow dashed lines stand for the MY_D, MY_R, predictive regression model, and FGT model respectively. The blue solid line stands for the historical mean model in (a) and (c) and the 5% significant level in (b) and (d). The OOS R^2 uses the HM as its benchmark and is defined in Footnote 11.

Figure 5: Historical Series and Projections for MY Ratio, dp Ratio and Five-Year Returns



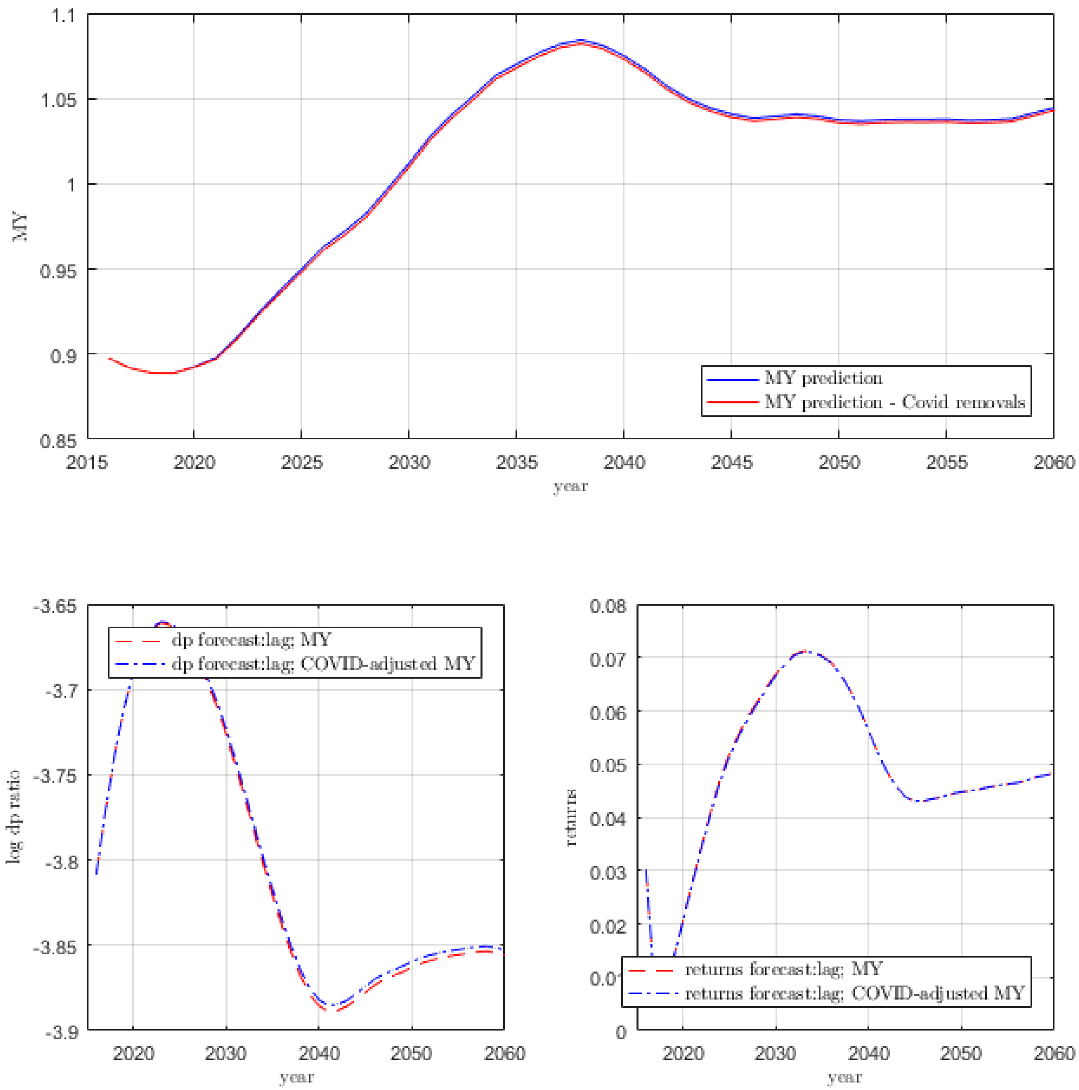
The top plot shows historical values followed by Census Bureau projections for the MY ratio. The middle plot provides historical values followed by model-implied forecasts for the dividend-price ratio, based on the Census Bureau MY projections and model (A1.1). The bottom plot presents historical values for the five-year average excess returns (including dividends) on the S&P 500 index followed by the five year rolling average of the model-implied forecasts based on the Census Bureau projections using (A1.4) and (A1.5). In all graphs, the solid red lines show the historical values, the dashed-dotted red lines signify the Census projections, and the horizontal blue lines indicate the historical means.

Figure 6: Historical Series and Projections for MY Ratio, dp Ratio and Five-Year Rolling Returns, Based on Different Immigration Scenarios



This figure shows only the forecast period. The top plot displays three different Census Bureau projections for the MY demographic ratio. The dashed line shows the same baseline Census Bureau projection as in Figure 5. The dashed-dotted and dotted lines show the alternative MY projection based on the Census Bureau's low and high immigration scenario projections, respectively. The bottom left and right plots show the three resulting forecasts for the dividend price ratio and five year rolling average excess returns (including dividends) on the S&P 500 index, respectively. The dashed line forecasts use the baseline Census Bureau projection and are the same forecasts in Figure 5. The dashed-dotted (dotted) line forecasts are constructed identically to those in Figure 5, except that they employ the Census Bureau's low (high) immigration scenario projection in place of the baseline projection.

Figure 7: Original and COVID-19 Adjusted Projections of MY Ratio, dp Ratio and Five-Year Returns



The figures show only the forecast period. The top plot shows the same baseline Census Bureau MY ratio projection as in Figure 5 (red dashed line), alongside an alternative MY projection adjusted for our bad-case COVID-19 scenario (blue dashed-dotted line). The bottom left and right plots show the resulting dividend-price ratio and and five year rolling average excess return (including dividends) forecast, respectively, for the S&P 500 index. The red dashed lines again show the same forecasts as in Figure 5. The blue dashed-dotted line forecasts are constructed identically to those in Figure 5, except that they employ our alternative MY projection adjusted for our bad-case COVID-19 scenario in place of the baseline Census Bureau projection. The COVID-19 adjustment is described in Appendix A3, with additional details provided Section B7 of our not-for-publication appendix. The bad case COVID-19 scenario, is based on values $\theta = 0.30$ and $t^* = 2030$ in (A3.1) and (A3.2) (see Scenario 4 in Appendix A3.2).