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

Chaoyi Chen† University of Guelph

Nikolay Gospodinov‡ Federal Reserve Bank of Atlanta

Alex Maynard§ University of Guelph

Elena Pesavento¶ Emory University

June 22, 2020

Abstract

We incorporate low frequency information from demographic variables into a simple predictive model to forecast stock valuations and returns conditional on 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 conditional forecasts of stock market valuations and returns. We also exploit the strong contemporaneous correlation between returns and valuations to improve conditional return forecasts – something which is not possible in a predictive regression with only lagged predictors. Extensive conditional out-of-sample forecast comparisons and tests demonstrate the predictive value that an accurate demographic projection can deliver. The model also provides a simple way to adjust predictions under alternative demographic assumptions, incorporating, for example, the demographic impact of covid19 or recent changes to immigration policy.

JEL Classification: G14, G17, C22.

Keywords: demographics, stock market valuation, stock return prediction, conditional forecasts, long-horizon forecasts

*The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. We thank Stephen Bonnar, Ibrahim Jamali, Jaideep Oberoi, Adriana Robertson (discussant), 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.

†Department of Economics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: chaoyi@uoguelph.ca.

‡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, 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 4 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 conditional on demographic projections. Our conditional 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 price 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 demographics and 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, 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.

A second key difference is that while Favero et al. (2011) provide real-time unconditional stock return forecasts, we demonstrate that our model is well-suited to provide forecasts that are conditional on either the official demographic forecasts or alternative scenario forecasts. This type of conditional forecast 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, the model could be used to condition the return forecast on alternative demographic projections regarding the impact of covid19 or changing immigration policies on future demographics. Indeed, even if the market is unpredictable in an unconditional sense, it may be possible to provide useful market forecasts conditional on future demographic scenarios.

In future research, it would be interesting to asses whether our model using the Census Bureau projections also improves unconditional forecasts. In other words, rather than drawing out the implications of the Census projections, to ask whether the projections can help in unconditional prediction of returns in real time. Unfortunately, the availability of historical real-time Census Bureau projections are too sparse for us to answer this distinctly different question. This remains an interesting venue for future research.

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 pseudo out-of-sample analysis for on historical data for forecasts made conditional on a correct demographic projection. Section 5 presents and compares valuation ratio and return forecasts at very long horizons. Section 6 concludes. Tables and figures are included at the end of the paper.

2 Empirical Forecast Models

The classic short-horizon predictive regression model is of the form

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

in which r_{t+1} is a log stock return 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. Its long-horizon counterpart is given by

$$r_{t+k}(k) = \beta_0^{pr}(k) + \beta_1^{pr}(k) x_t + \varepsilon_{1,t+k}^{pr}(k), \tag{2}$$

where $r_{t+k}(k) = \sum_{j=1}^{k} r_{t+j}$ is a long-horizon (k-period) return. Of course, when $k = 1$ eq.(2) collapses to eq.(1) and $\beta_1^{pr}(1) = \beta_1^{pr}$. Valuation ratios are a common choice for x_t and we focus on the dividend price ratio includes valuation, where p_t and d_t refer to the natural logs of the stock price and dividend.[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 them individually as first-order (or higher order) autoregressive models as

$$x_t = \rho_0^{ar} + \rho_1^{ar} x_{t-1} + \varepsilon_{2,t}^{ar}. \tag{3}$$

Estimates of ρ_1 in (3) are generally close to one (high persistence), while the estimated correlation between the contemporaneous innovations $\varepsilon_{1,t}^{pr}$ and $\varepsilon_{2,t}^{ar}$ are often close to negative one (strong endogeneity). These are the characteristics which give rise to the well known Stambaugh bias, which has led the literature to question the significance of the positive estimates of $\beta_1(k)$ reported in the previous literature.

This same negative correlation between $\varepsilon_{1,t}^{pr}$ and $\varepsilon_{2,t}^{ar}$ also suggests a much stronger, more robust contemporaneous relationship between x_{t+1} and r_{t+1}. At an intuitive level, holding dividends (or earnings or book value) constant, a larger 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+1} as in

$$\begin{aligned} r_{t+1} &= \beta_0^{\mathbf{aug}} + \beta_1^{\mathbf{aug}} x_t + \beta_2^{\mathbf{aug}} x_{t+1} + \varepsilon_{1,t+1}^{\mathbf{aug}}, \\ &= \beta_0^{\mathbf{aug}} + \left(\beta_1^{\mathbf{aug}} + \beta_2^{\mathbf{aug}}\right) x_t + \beta_2^{\mathbf{aug}} \Delta x_{t+1} + \varepsilon_{1,t+1}^{\mathbf{aug}}, \end{aligned} \tag{4}$$

would yield a much tighter fit. Of course, (4) 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+1} may also be useful for predicting returns. For example, if $\hat{x}_{t+1|t}$ is a time t forecast of x_{t+1}, then this forecasted predictor could be used to augment the predictive regression using:[2]

$$\begin{aligned} r_{t+1} &= \beta_0 + \beta_1 x_t + \beta_2 \hat{x}_{t+1|t} + \varepsilon_{t+1}, \\ &= \beta_0 + (\beta_1 + \beta_2) x_t + \beta_2 \left(\hat{x}_{t+1|t} - x_t\right) + \varepsilon_{t+1}. \end{aligned} \tag{5}$$

The usefulness of (5) depends on the quality of the the additional information by the forecast of x_{t+1} (or Δx_{t+1}) above and beyond that already contained in x_t itself.

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

[2] Note that the definition of the population regression coefficients differ across the three specifications: (2), (4), and (5).

A large and growing literature suggests that slowly varying demographic trends are both (partially) predictable and predictive for stock return valuations.

Our prediction models for x_t are based on three insights from this literature:

1. Future demographic trends, as captured by middle-young and middle-old ratios MY_{t+j} and MO_{t+j}, are far more predictable than most economic and financial variables. While surprises 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 influence 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)).

As our baseline specification for x_t, we therefore modify the usual autoregressive specification as in (3) to incorporate the official time t demographic projections of future demographic ratios, such as MY_{t+1} and MO_{t+1}, which we denote generically by dr_{t+1}, where dr stands for demographic ratio. We refer to these projections as $\widehat{dr}_{t+1|t}$, where the notation $t+1|t$ indicates a projection for a $t+1$ variable produced or made available at time t. For example, in our baseline specification $\widehat{dr}_{t+1|t} = \left(\widehat{MY}_{t+1|t}, \widehat{MO}_{t+1|t}\right)'$. This results in the model:

$$x_{t+1} = \rho_0 + \rho_1 x_t + \rho_2 \widehat{dr}_{t+1|t} + \varepsilon_{2,t+1} \tag{6}$$

generating a forecast for the valuation predictor of the form:

$$\hat{x}_{t+1|t} = \hat{\rho}_0 + \hat{\rho}_1 x_t + \hat{\rho}_2 \widehat{dr}_{t+1|t}, \tag{7}$$

where $\hat{\rho} = (\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2)'$ denote the standard least squares estimates.

The next step in the forecast process is to substitute the valuation ratio prediction $\widehat{x}_{t+1|t}$ from (7) into the augmented predictive regression in (5) for the return forecast r_{t+1}. This produces a one-period ahead forecast. Predictive improvements are based on the inclusion of the demographic predictors. However, since the demographic predictors are slowly varying, as Favero et al. (2011) point out, the distinction between using current and future demographic information may not be large at such short horizons.

Perhaps a more interesting application of the model is to use the long-run predictability to forecast both valuation ratios and stock returns at much longer horizons, including at horizons far exceeding the five- or even ten-year horizons typically employed in the long-horizon predictive regressions in (2).

We consider the return forecast r_{t+h}, which involves forecasting one period returns many periods into the future and forecasts of cumulative long horizon returns $r_{t+k}(k) = \sum_{h=1}^{k} r_{t+h}$. We begin by discussing the prediction for r_{t+h+1}. The first step is to forecast the future valuation predictor x_{t+h}. Shifting (7) h periods ahead gives

$$\hat{x}_{t+h+1|t+h} = \hat{\rho}_0 + \hat{\rho}_1 x_{t+h} + \hat{\rho}_2 \widehat{dr}_{t+h+1|t+h}. \tag{8}$$

However, the official forecasts of $\widehat{dr}_{t+h+1|t+h}$ are not available until $t+h$, making it infeasible for out-of-sample prediction for $h > 0$. Instead, we replace $\widehat{dr}_{t+h+1|t+h}$ by the most recent official forecasts available at the time that the prediction is made, namely $\widehat{dr}_{t+h+1|t}$. Similarly we employ an iterative forecast strategy to replace x_{t+h} by its forecast value. The resulting (partially recursive) valuation forecast is given by

$$\hat{x}_{t+h+1|t} = \begin{cases} \hat{\rho}_0 + \hat{\rho}_1 \hat{x}_{t+h|t} + \hat{\rho}_2 \widehat{dr}_{t+h+1|t} & \text{for } h > 0, \\ \hat{\rho}_0 + \hat{\rho}_1 x_t + \hat{\rho}_2 \widehat{dr}_{t+1|t} & \text{for } h = 0. \end{cases} \tag{9}$$

Similarly, shifting (5) h step ahead gives,

$$r_{t+h+1} = \beta_0 + \beta_1 x_{t+h} + \beta_2 \hat{x}_{t+h+1|t+h} + \varepsilon_{t+h+1},$$

where x_{t+h} and $\hat{x}_{t+h+1|t+h}$ are both infeasible for $h > 0$. We again replace them by the best feasible forecasts, $\hat{x}_{t+h|t}$ and $\hat{x}_{t+h+1|t}$ respectively, which we obtained in the previous step from (9). This yields a feasible h-period ahead return forecast given by:

$$\hat{r}_{t+h+1|t} = \begin{cases} \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{t+h|t} + \hat{\beta}_2 \hat{x}_{t+h+1|t} & \text{for } h > 0, \\ \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\beta}_2 \hat{x}_{t+1|t} & \text{for } h = 0. \end{cases} \tag{10}$$

Forecasts of the multiperiod return $r_{t+k}(k) = \sum_{h=1}^{k} r_{t+h}$ can then be formed by a cumulative sum of one period forecasts to obtain:

$$\hat{r}_{t+k|t}(k) = \sum_{h=0}^{k-1} \hat{r}_{t+h+1|t}. \tag{11}$$

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 (9) yields the following iterated forecast of the valuation ratio:

$$\hat{x}_{t+h+1|t} = \hat{\rho}_0 \sum_{j=0}^{h} \hat{\rho}_1^j + \hat{\rho}_2 \sum_{j=0}^{h} \hat{\rho}_1^j \widehat{dr}_{t+h+1-j|t} + \hat{\rho}_1^{h+1} x_t. \tag{12}$$

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, substitute (12) into (10) and (11) and express, after simplifying, the iterated long-horizon forecast $r_{t+k}(k)$ as [3]

$$\hat{r}_{t+k}(k) = \hat{\alpha}_0(k) + \hat{\alpha}_1(k) x_t + \sum_{h=0}^{k} \hat{\alpha}_{2,h}(k) \widehat{dr}_{t+k-h|t}, \tag{13}$$

where, defining $\mathbf{1}_{h<k}$ as an indicator taking the value of one when $h < k$, the coefficients in (13) are given by

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

There are several implications of (13) 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 cumulate 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 more 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

$$\tilde{r}_{t+k}(k) = \tilde{\alpha}_0(k) + \tilde{\alpha}_1(k) x_t + \sum_{h=0}^{k} \tilde{\alpha}_{2,h}(k) \widehat{dr}_{t+k-h|t}, \tag{14}$$

where the coefficients $\tilde{\alpha}_0(k), \tilde{\alpha}_1(k), \tilde{\alpha}_{2,h}(k)$ are directly estimated from the regression in (14).

[3] The derivation is given in the appendix and includes (A.1) as an intermediate step.

Favero et al. (2011) also employ demographic ratios to improve return forecasts. However, their approach and motivation differs somewhat from ours. They advocate the addition of a lagged demographic variable (the MY ratio) in a direct long-horizon regression, such as (2). In its simplest form, this suggests a predictive regression augmented by current demographic information as in

$$r_{t+k}(k) = \beta_0^{FGT}(k) + \beta_1^{FGT}(k)x_t + \beta_2^{FGT}(k)MY_t + \varepsilon_{1,t+k}^{FGT}(k). \tag{15}$$

Their approach 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. In the more sophisticated version of their model, they 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.

The intuition for their result can be understood from the bottom line of (4). The ability of a lagged valuation predictor x_t to predict the future return r_{t+1} depends mainly on ability to predict the future change in valuation Δx_{t+1}. However, if the valuation predictor is close to being a random walk ($\rho_1^{ar} = 1$ in (3)), then Δx_{t+1} is close to unpredictable. By reducing the persistence in the (demographically adjusted) dividend-price, they improve its ability to forecast its own future change and thus the future return. The same logic applies at longer horizons. Our approach 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 forecast future valuation change,rather than by using lagged demographic predictors.

We can re-express our forecast models in terms of a traditional predictive regression with a generated predictor by replacing the lagged predictor in (1) by $(x_t, \hat{x}_{t+1|t})$ or equivalently by $(x_t, \hat{x}_{t+1|t} - x_t)$ and noting that the prediction $\hat{x}_{t+1|t}$ is itself a time t variable. Alternatively we can substitute (7) into (5) to obtain

$$r_{t+1} = \beta_0 + \beta_2\hat{\rho}_0 + (\beta_1 + \beta_2\hat{\rho}_2)\, x_t + \beta_2\hat{\rho}_2\widehat{dr}_{t+1|t} + \varepsilon_{t+1}. \tag{16}$$

From this perspective, another 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.[4]

3 Data and In-Sample Results

3.1 Data

We employ annual stock return data from 1901 to 2015 on both stock returns and valuations and demographics, and also consider the post World War II (WWII) sub-period from 1947-2015. 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 of dividend (d_t) and the log stock price (p_t).

Our demographic data is from the U.S. Census Bureau. We define: Y_t (20-29), M_t (40-49), and O_t (60-69) as the young, middle, and old U.S. population sizes. And we use the middle-to-young ($MY_t = M_t/Y_t$) and middle-to-old ratios ($MO_t = M_t/O_t$) as predictors for persistent stock market valuation variables, such as the dividend-price ratio. 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.

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 Appendix not for publication. The results are somewhat dependent on the lag length and sample period. Overall we find strong evidence that we cannot reject a unit root in dp in both samples considered. For the MY ratios we fail to reject at all but the BIC selected lag-length in the full sample but we reject in the later part of the samples. The tests for MO reject using both BIC

[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 (2). 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 10 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 return for the period starting 5 years into the future and ending in 10 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.

and MAIC in both samples. 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 a literal 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 three variables are highly persistent.

Favero et al. (2011) report evidence of cointegration between the dp and MY ratios. We next test for cointegration among dp, MO, and MY, both in pairs and all together using Engle-Granger two-step and Johansen cointegration tests.[5] The results are again somewhat sensitive to the lag length and sample. Overall, the Engle-Granger test result does not provide strong evidence of cointegration between the dp ratio and the the two demographic variables. The Joahansen test provides somewhat stronger evidence of cointegration than the Engle-Granger test, with the exception of the (dp,MY) pair over the full 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 possiblity of very persistent but not exactly unit roots 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 (6) 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), after controlling for the demographics, the dividend price ratio may be a less persistent predictor, mitigating the Stambaugh bias in predictive regressions. However, since this argument is only partially supported by our cointegration test results, we also confirm the robustness of our results by employing the IVX method of Kostakis et al. (2015).

3.3 Conditional In-Sample Predictions

In this section, we examine the ability of the model to provide in-sample predictions of both stock market valuations and stock market returns conditional on a given demographic projection. We take the perspective of a policy maker or pension administrator who wants to know the implication of a given demographic projection on future valuations and returns. The policy maker may simply take the Census projection as given or she may examine the implication of several alternative

[5] Detailed results can be found in the separate Appendix.

projections.For example, she may be interested in the financial implications of alternative adjustments to baseline demographic projections due to Covid19. In each case, she conditions her stock valuation/return prediction on the accuracy of the demographic forecast. Consequently, we employ the actual future demographic ratios in place of a specific demographic projection. The residuals from this conditional prediction capture the in-sample error in predicting future valuations given a correct demographic projection, but not the errors in the demographic forecast itself. This is distinct from an unconditional prediction employing a historical real-time demographic forecast.An unconditional prediction exercise that accounts for the error in the Census predictions itself would also be of interest. However, although the Census Bureau data archives include their historical demographic forecasts in a few particular past years, they do not, to our knowledge, provide consistent enough real-time forecast data to undertake such a study.

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

Table 1 reports the estimates of our predictive models for the dividend-price ratio. Panels A and B again 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 (3). Columns 4-9 show the results from an AR(1) model augmented with projected demographic ratios as in (6), but with the actual future values of MY_{t+1} and MO_{t+1} replacing their projections as discussed above. Column 4-5, 6-7, and 8-9 respectively present the estimates using the middle-to-old ratio (MO), the middle-to-young ratio (MY), and both MO and MY together as the demographic ratio(s). For each regression specification, we include standard OLS results alongside IVX estimation. The IVX was performed following Kostakis et al. (2015) using c=-1 and α=0.95.[6].

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 4-9 helps us to assess the improvements that come from adding the demographic variables to the prediction equation for the dividend-price ratio. The

[6]Robustness checks with different values for c and α can be found in the separate appendix

coefficient on MY is significant in both sample periods. Even after including MO, it is still significant in the full-sample period.Although, the evidence is less strong using IVX than OLS, the IVX coefficients for MY remain significant at the ten percent level, confirming the robustness of the finding. On the other hand, the coefficients on MO are generally insignificant. 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. Moreover, it suggests that we should prefer the MY ratio or the combination of the MY and MO ratios over the MO ratio alone.

3.4 In-Sample Estimation of Conditional Predictive Model for Return Regression

Table 2 reports in-sample estimation of the return regression models in equations (1), (4), and (5) 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), 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 percent level in the post-WWII sample, but not significant in the full sample. However, this evidence is likely to be overstated ude to the Stambaugh bias in the predictive regression. Using IVX (column 3), the dp ratio is again insigificant 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 4-9, 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 (5), using dp_t and $\hat{d}p_{t+1|t}$ as regressors, where $\hat{d}p_{t+1|t}$ is itself predicted using (6), with the results reported in Table 1 . Columns 4-5, 6-7, and 8-9 respectively employ $MO_{t+1|t}$, $MY_{t+1|t}$ and $(MO_{t+1|t}, MY_{t+1|t})$ in place of the demographic projections in (6).

In comparison with the traditional predictive regression in Column 2, the p-value of the F-test the for the full and post-WWII samples improves depending on the choices of demographic variables, most notably for the MY and (MO, MY) specifications. The post-WWII sample also witnesses an improvement. When using MY and (MO, MY) to predict dp_{t+1} the coefficient on $\widehat{dp}_{t+1|t}$ is highly significant in both samples while its coefficient is insignificant when including just MO. 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 6-9) 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 (conditional) prediction improvement when including the predicted dividend-price ratio based on demographic projections, especially using the middle-young (MY) ratio. As explained earlier, we do not have the data available to assess the impact of Census Bureau projection errors on the the return forecast. Thus the results confirm the usefulness of the model in producing predictors conditional on a given demographic projection, but are silent the quality of the projection itself.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.

4 Out of Sample Forecasts

In this section, we investigate the usefulness of the model to provide out-of-sample forecasts conditional on a given demographic projection. Specifically, we address the following question. Suppose that the forecaster's demographic projections were correct. Then, how accurate would the model be in predicting asset returns? This allows us to evaluate the usefulness of the model in predicting implications of a given demographic forecast on future stock returns and valuations. Such a conditional forecasting exercise can be used both to draw out the implications of the official demographic projections, as well as the implications of alternative demographic assumptions.

The conditional forecast exercise evaluated here is distinct from an unconditional analysis of out-of-sample return predictability since it conditions the return forecast on a correct future demographic prediction. Thus, we capture the out-of-sample errors in using the demographic projection to predict stock valuations and returns, but we do not capture or evaluate the error in the demographic forecasts themselves.

We have conducted 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 OLS based forecasts perform as well or better than IVX in most cases, althoug the results were often similar. In order to conserve space, we present the OLS OOS results here and include the IVX counterpart in the extended appendix.

4.1 Conditional out of Sample Forecasts

In Table 3, we provide short-run (conditional) out-of-sample forecasting results. We run recursive forecasting regressions for the one-year ahead horizon ($k = 1$). Similar to Table 2, panels A and B correspond to the period 1901-2015 and 1947-2015, respectively. 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.

Column 1 lists the candidate models. HM is the historical mean. PR is the classical predictive regression model, using the lagged dividend-price ratio as the predictor. FGT stands for the model in the spirit of Favero et al. (2011), which provides forecasts based on (15) for $k = 1$. MO, MY and MO & MY provide the estimates of (10) using the actual future values of the demographic ratios MO_{t+1}, MY_{t+1} and (MO_{t+1}, MY_{t+1}) to replace the demographic ratio projections replacing $\widehat{dr}_{t+1|t}$ in (9). The use of the actual future demographic values is appropriate when assessing the effectiveness of the model in providing *conditional* forecasts, where the conditioning is based on the assumption of a correct demographic forecast.[7]

Within the first part of each panel, columns 2,4, and 6 of Table 3 show the out-of-sample mean square errors (MSE OOS) of each model under various training periods. Consistent with the in-sample finding, MY dominates all of the other models in panel A and performs as one of the best models in panel B. Interestingly, even though it places second to MY in most cases, FGT performs consistently better than the remaining models. Indeed, following Favero et al. (2011), our finding gives more evidence that the MY ratio does improve the short-run prediction. Columns 3, 5, and 7 present the out of sample R^2 (OOS R^2), introduced by Campbell and Thompson (2007).[8] The results support the conclusion that MY outperforms the prevailing historical mean during the full sample (Panel A), whereas none of the models show a positive OOS R^2 during the post WWII

[7]This is because the accuracy of the conditional forecast does not depend on the accuracy of the demographic forecast itself, since we condition on it. By contrast, the unconditional forecast accuracy does depends on the accuracy of the demographic forecast and cannot be accessed in this manner without the introduction of severe look-ahead bias. To assess the accuracy of the unconditional forecast, we would need to use real-time Census Bureau forecasts for $\widehat{dr}_{t+1|t}$. Unfortunately, the real-time demographic forecast history made available by the Census Bureau is far too limited for us to pursue this. Thus all models are compared here in terms of their ability to provide *conditional* out of sample forecasts.

[8]The OOS R^2 is defined as $R^2 = 1 - \frac{MSE\ OOS}{MSE\ HM}$, where $\frac{MSE\ OOS}{MSE\ HM}$ 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 ($MSE\ HM$).

period (Panel B). On the other hand, MY outperforms the predictive regression model with a higher OOS R^2 in all cases.

The second part of each panel of Table 3 reports tests of the demographic models against two benchmark forecasts: the historical mean and the predictive regression. 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.[9]

The second part of Columns 2, 4, and 6 show the CW test for out-performance relative to the historical mean under each training period. In five out of six cases, the MY model outperforms the historical mean at the 10% significant level or better. Columns 3, 5, and 7 provide the CW test with the predictive regression as the benchmark model. For the MY model, the test statistics are all significant at the 10% levels and many are also significant at the five and one percent levels. This confirms the significance of MY's out-performing OOS-MSE and OOS-R^2 from the first part of Panel A (full sample). In Panel B, the test results appear more favorable to the MY model than the corresponding OOS-MSE or OOS-R^2 results. Indeed the CW test points to out-performance of the MY even for some cases in which its MSE slightly exceed that of the benchmark. This is due to the bias correction of CW, which adjusts for the fact that the larger nested model – in this case the demographic model – is expected to have the larger OOS-MSE even under the null hypothesis in which the two models are equivalent. Therefore, if the MSE of the larger model is even close to matching that of the smaller model it nests, this may be strong enough evidence to reject the null in its favor.

Intriguingly, in many cases, the P values of FGT seem to be fairly close to the P values of MY. Intuitively, this is natural due to the fact that, for one-year ahead forecasting, the only difference between FGT and MY is the replacement of MY_t by MY_{t+1}. Considering that the demographical ratio cannot vary greatly in the short-run, it is reasonable to expect MY_{t+1} to contribute only marginally to the information already contained in MY_t. As a result, in the short-run, FGT and MY are most likely to show similar forecasting accuracy.

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

4.2 Long-horizon Out-of-Sample Results

In Table 4, we provide long-run out-of-sample forecasting results. Panel A and Panel B correspond to the full sample and post-WWII samples. 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 ahead forecasts. Due to the cumulative nature of the long horizon forecast, we may estimate the predictive return both in a recursive way as defined in (10) and (11) and in a direct way as defined in (14). 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 again the candidate models. PR is now the long-horizon version of the predictive regression model in (2) since $k > 1$. FGT again provides the estimates of (15), now with $k = 5$. Models MO, MY_R and $MY\&MO$ provide estimates of (10) and (11) by again replacing the projected demographic ratios $\widehat{dr}_{t+h|t}$ in (9) by the actual realized values MO_{t+h}, MY_{t+h}, or $(M0_{t+h}, MY_{t+h})$ respectively [10]. These extend our earlier conditional demographic-based forecasts to the long-horizon by recursive forward substitution of the one-period ahead forecasts. Alternatively, MY_D instead provides a direct long-horizon forecast conditional on demographics based on (14), where we again replace $\widehat{dr}_{t+h|t}$ by the actual future demographic MY_{t+h}. As discussed previously, the use of actual future demographics is appropriate for the evaluation of the conditional (but not unconditional) forecast performance.

Columns 2, 4, 6 and columns 3, 5, 7 of the first part of each panel show the OOS-MSE and the OOS-R^2 respectively. For the OOS-MSE of both panels, either MY_D or MY_R reports the lowest value in all six cases and both 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 middle-young demographic ratio (MY) improves the long run forecast. We report the CW test results 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. Overall, the conditional forecast improvements from the use of the MY ratio are more pronounced at the five year horizon than the one year horizon. This is not surprising since

[10]For the iterative out-of-sample forecasts, IVX estimation can be performed in three ways. 1. we can use IVX approach to estimate (9) and OLS method to estimate (10). 2. we can estimate (9) by OLS method and (10) by IVX method. 3. We can use IVX approach to estimate both (9) and (10). In the separate appendix, we report the IVX results of the first way (IVX-OLS) due to the fact we found huge biases with the other two approaches.

demographic ratios are a slow moving variables which correlate with the long-term trends in stock valuations (Favero et al. (2011)). Furthermore, unlike the short run results of Table 3, we 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 conditions on the future path of MY_t performs much better than FGT, which uses only current MY_t, for the purpose of conditional forecasting.

4.3 Rolling Sample Analysis

To complement our recursive out-of-sample forecasting, we run rolling conditional forecasting regressions for both the 1 and 5 year ahead horizon. Tables 14 and 15 provide the rolling short-run and long-run results respectively. At the one year horizon, Table 14, the MY again has the lowest OOS-MSE and a positive OOS-R^2 in five of six cases and nearly ties the historical mean for the lowest MSE in the last case. In Table 15, either MY_R or MY_D always show 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 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, which 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 Tables 14-15, show the GW test for out-performance in rolling samples relative to the historical mean predictive regression. In the short-run, contrary to the recursive results of Table 3, using this more stringent test, MY never significantly outperforms either the historical mean model or the predictive regression model at the one hear horizon, despite its out-performace in terms of OOS MSE and R^2. However, we do observe some evidence on the ability of MY to predict long-run stock returns even using the 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 15 Panel B. In some cases it also outperforms at the 5% or 1% levels.

To explore the sensitivity of our findings to the training periods and window size, we provide Figures 2-12 to show the OOS R^2 and p-values for out-performance against the historical mean (vertical axis) for a range of training periods (horizontal axis). Figures 2 and 10 show the one and five year ahead recursive forecasts, expanding on Tables 3-4 while Figures 11 and 12 show the corresponding rolling forecasts and expand on Tables 14-15. 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 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 (MY) ratio as the best predictor. Moreover, from a conditional forecast perspective, we find that the demographic ratio projection performs better than the FGT model in both the short- and long-run.

5 Very Long-Horizon Forecasts Conditional on Demographic Projections

The official projections for the demographic ratios MY_t and MO_t until 2060 are plotted together with their historical values in Figure 6. The middle age population peaked a little after 2000 relative to both young and old. It is projected to keep declining until about 2020. It is then projected to grow slowly, reaching a second smaller "echo" peak at around 2040. Currently, the MO ratio is well below its historic mean, whereas the MY ratio is slightly above its mean.

The resulting forecasts for the dividend-price ratio, together with their historical values, are shown in Figure 7. With a brief exception around the 2008 crisis, the dividend-price ratio remains below its long-term average since the early 1990s. Using (3) 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 and/or MO_t as in (6) (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. On the other hand, the MO ratio ends the sample below its historical average, implying that the dividend-price ratio is even lower (market more over-valued) relative to its mean conditional on MO. 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 demographic ratios. Since the dividend-price ratio is more severely undervalued relative to the middle-old ratio, the forecast using only MO projects a much larger correction. However, the projections from including both MO and MY are much closer to the more modest correction implied by MY alone. Putting further near-term upward pressure on the dividend-price ratio, both MO and MY are projected to fall over the next 5 years.

Starting in the early 2020s, the relative size of the middle age population is projected to increase again with both MY and MO ratios rising again. This leads to a projected decline in the dividend price ratio. Using either MY or both MY and MO, the decline in the dividend-price ratio is projected to drop below its current value by 2040, when MY and MO reach a new peak. Basing the forecast on only MO, the dividend-price ratio still has a projected decline in the 2020s and 2030s but never remains much higher than current values.

Figure 8 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 and MO ratios. 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. The quantitative prediction is again more moderate using either MY or both MY and MO, but all three forecasts agree on its direction.

Returns are subsequently projected to slowly recover over the next 15 years during the 2020s and first half of the 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 captures the prediction that returns are higher when markets are less highly valued (higher dividend-price ratio).

Overall, there are some striking differences in the forecasts depending on which demographic ratios are employed: MO, MY or both. The forecasts are essentially unanimous on the directions of change, but predict starkly different magnitudes. Judging relative to MO alone, stocks are more highly valued, are projected to undergo a much larger correction in the near term and their valuations are projected to stay lower over the long term. Based on either MY or both MY and MO, the predictions are similar in direction, but far more moderate in magnitude, with the dividend-price ratio increasing in the near term, but eventually settling back to similar levels as in 2015 over the 20-25 year horizon.

Since the elderly not only reduce savings, but also sell off assets, there might be good a priori arguments for preferring MO over MY (Liu and Spiegel (2011)). However, using both MY and MO as predictors, we obtain predictions close to those using MY, which is a strong argument in favor of the moderate forecast. Moreover, both our in-sample and pseudo out-of-sample results consistently indicated that MY provides superior predictions relative to MO.

The divergence in forecast magnitudes points to a large degree of model uncertainty, when accessing the likely impact of either valuation corrections or demographic trends on future stock valuations and returns. Quantifying this uncertainty is not straight-forward. Even quantifying 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.

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 conditional prediction. Furthermore, because the Census produces projections until 2060, this allows us to provide very long-horizon predictions for both valuation ratios and returns.

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, they can prove useful by providing an indication of the likely effect of both current valuations and future demographic trends on stock market valuations. The models indicate an end-of-sample valuation below that predicted by the current value of the middle-age population to either the young or elderly. Furthermore, these population ratios are projected to decline over the next several years. The models thus predict an increase in the dividend price ratio over the next five years. This is followed by a gradual predicted decline over the following 15 years as the relative size of the middle age population begins to grow again, reaching a second peak at around 2040. While all of our models agree on the direction of these predictions, the magnitude of the predictions vary considerably depending on whether the ratio of middle-young or middle-old is used as the primary demographic ratio. This underlines the inherent uncertainty in any forecast of future valuations or returns.

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.

Elliott, G., Rothenberg, T. 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. A., 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.

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.

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**.

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.

A Derivation of Equation 13

Plugging (12) into (10) and simplifying gives us the iterated forecast formula for the one period ahead return $h+1$ periods ahead as

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

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

$$\begin{aligned}
\widehat{r}_{t+h+1|t} &= \widehat{\beta}_0 + \widehat{\beta}_1\widehat{dp}_{t+h|t} + \widehat{\beta}_2\widehat{dp}_{t+h+1|t} \\
&= \widehat{\beta}_0 + \widehat{\beta}_2\widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2\widehat{\rho}_1\right) dp_{t+h|t} + (\beta_2\rho_2)\,\widehat{my}_{t+h+1|t} \\
&= \widehat{\beta}_0 + \widehat{\beta}_2\widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2\widehat{\rho}_1\right)\left[\widehat{\rho}_0 \sum_{j=0}^{h-1} \widehat{\rho}_1^j + \widehat{\rho}_2 \sum_{j=0}^{h-1} \widehat{\rho}_1^j \widehat{my}_{t+h-j|t} + \widehat{\rho}_1^h dp_t\right] + \left(\widehat{\beta}_2\widehat{\rho}_2\right)\widehat{my}_{t+h+1|t} \\
&= \widehat{\beta}_0 + \widehat{\beta}_2\widehat{\rho}_0 + \left(\widehat{\beta}_1 + \widehat{\beta}_2\widehat{\rho}_1\right)\widehat{\rho}_0 \sum_{j=0}^{h-1} \widehat{\rho}_1^j + \left(\widehat{\beta}_2\widehat{\rho}_2\right)\widehat{my}_{t+h+1|t} + \left(\widehat{\beta}_1 + \widehat{\beta}_2\widehat{\rho}_1\right)\widehat{\rho}_2 \sum_{j=0}^{h-1} \widehat{\rho}_1^j \widehat{my}_{t+h-j|t} \\
&+ \left(\widehat{\beta}_1 + \widehat{\beta}_2\widehat{\rho}_1\right)\widehat{\rho}_1^h dp_t.
\end{aligned}$$

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.

Table 1: **Dividend Price Ratio Model IVX Estimation c=-1, α=0.95**

Model:	AR(1)		Augmented AR(1)					
			MO and		MY and		MO, MY	
	lag dp		lag dp		lag dp		and lag dp	
	OLS	IVX	OLS	IVX	OLS	IVX	OLS	IVX
				Panel A: 1901-2015				
Const	-0.368^{**}	-0.368^{**}	-0.513^{**}	-0.513^{**}	-0.395^{***}	-0.395^{***}	-0.529^{***}	-0.529^{***}
lap dp	0.889^{***}	0.886^{***}	0.875^{***}	0.887^{***}	0.757^{***}	0.776^{***}	0.745^{***}	0.779^{***}
MO			0.053	-0.002			0.049	-0.006
MY					-0.504^{***}	-0.400^{*}	-0.500^{***}	-0.395^{*}
Model Test	395.415^{***}	374.89^{***}	198.339^{***}	384.65^{***}	222.308^{***}	409.87^{***}	148.506^{***}	425.03^{***}
p-value	0	0	0	0	0	0	0	0
				Panel B: 1947-2015				
Const.	-0.309^{**}	-0.309^{**}	-0.227	-0.227	-0.335^{***}	-0.335^{***}	-0.345^{**}	-0.345^{**}
lap dp	0.915^{***}	0.904^{***}	0.879^{***}	0.817^{***}	0.829^{***}	0.777^{***}	0.828^{***}	0.778^{***}
MO			-0.130	-0.227			0.014	-0.122
MY					-0.312^{**}	-0.408^{**}	-0.3297	-0.254
Model Test	377.506^{***}	170.61^{***}	193.103^{***}	269.69^{***}	200.331^{***}	228.78^{***}	131.526^{***}	282.94^{***}
p-value	0	0	0	0	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 provides IVX estimates of (3) & (6). Column 2 provides estimates for the pure AR(1) process in (3). Column 3-5 provide estimates for the augmented AR(1) process including demographic ratios as in (6). In columns 3, 4, and 5 respectively, MO, MY, and both MO and MY are employed respectively 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.

Table 2: **IVX Estimation of Return Regression Models (In Sample), c=-1, α=0.95**

Model:	lag dp		MO and lag dp		MY and lag dp		MO, MY, and lag dp	
	OLS	IVX	OLS	IVX	OLS	IVX	OLS	IVX
				Panel A: 1901-2015				
Const	0.257***	0.257**	-0.065	-0.065	-0.022	-0.022	-0.025	-0.025
$\widehat{dp}_{t+1\|t}$			-0.8736	-0.418	−0.755***	−0.649**	−0.765**	−0.691**
dp_t	0.062	0.062	0.839*	0.434	0.733**	0.642**	0.742***	0.681**
Model Test	2.615	2.537	1.857	2.831	5.789***	7.817**	6.343***	8.195**
p-value	0.109	0.111	0.161	0.243	0.004	0.020	0.003	0.017
				Panel B: 1947-2015				
Const	0.402**	0.402**	0.097	0.097	0.032	0.032	0.028	0.028
$\widehat{dp}_{t+1\|t}$			-0.987	-1.742	−1.198***	−1.549***	−1.210***	−1.534***
dp_t	0.098**	0.114*	1.001*	1.732	1.194***	1.550***	1.205***	1.535***
Model test	5.008**	3.561*	3.780**	3.980	6.509***	12.250***	6.610***	12.807***
p-value	0.029	0.0589	0.028	0.1367	0.003	0.0022	0.002	0.0017

$***$ 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 provides IVX estimation of equations (1), (4), and (5). The dependent variable in all cases are yearly log returns including dividends. Column 2 (lag dp), provides the estimates of (1) in which only on the past dp_t is employed as a predictor. Columns 3-5 provide estimates of (5), using three different specifications of the demographic ratio projection $dp_{t+1|t}$ in (6): only MO (Column 3), only MY (Column 4), and both MO and MY (Column 5). Results for the time period 1901-2015 are shown in Panel A and results for 1947-2015 period are shown in Panel B.

Table 3: **Results of Return Regression Models (Out of Sample, Recursive, 1 Year Ahead Conditional Forecast)**

Panel A: 1901-2015, forecasts begin tp years after sample.

training period (tp)	tp=30		tp=40		tp=60	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.0354	0	0.0266	0	0.0262	0
PR	0.0367	-0.0368	0.0269	-0.0132	0.0274	-0.0482
FGT	0.0355	-0.0020	0.0245	0.0780	0.0254	0.0280
MO	0.0402	-0.1347	0.0275	-0.0325	0.0294	-0.1237
MY	0.0347	0.0204	0.0237	0.1090	0.0250	0.0457
MY & MO	0.0435	-0.2274	0.0250	0.0596	0.0261	0.0025
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	0.2071		0.0941*		0.2090	
FGT	0.1029	0.0228**	0.0104**	0.0035***	0.0635*	0.0216**
MO	0.6797	0.6480	0.1307	0.1784	0.4325	0.5595
MY	0.0512*	0.0077***	0.0048**	0.0014***	0.0531*	0.0156**
MY & MO	0.4416	0.2981	0.0217**	0.0094***	0.1068	0.0569*

Panel B: 1947-2015: forecasts begin tp years after sample.

training period (tp)	tp=20		tp=25		tp=30	
Model	MSE OOS	R^2	MSE OOS	R^2	MSE OOS	R^2
HM	0.0282	0	0.0293	0	0.0240	0
PR	0.0292	-0.0350	0.0313	-0.0666	0.0288	-0.2010
FGT	0.0306	-0.0850	0.0309	-0.0536	0.0269	-0.1214
MO	0.0326	-0.1582	0.0336	-0.1443	0.0283	-0.1797
MY	0.0287	-0.0178	0.0295	-0.0063	0.0257	-0.0726
MY & MO	0.0326	-0.1550	0.0335	-0.1403	0.0290	-0.2100
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	0.1028		0.1654		0.4126	
FGT	0.0290**	0.0343**	0.0859*	0.0486**	0.1706	0.0179**
MO	0.3432	0.2520	0.2997	0.1826	0.3115	0.0857*
MY	0.0314**	0.0275**	0.0818*	0.0299**	0.1602	0.0075***
MY & MO	0.2204	0.5535	0.3020	0.4196	0.3833	0.1923

$***$, $**$, $*$ 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 1 year out of sample forecasting results. The dependent variable in all cases are yearly 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. Column 2-3 reports the out of sample mean square error (OOS MSE) and out of sample R^2 (OOS R^2) with 30/20 years training period for panel A/panel B. Column 4-5 shows OOS MSE and OOS R^2 with 40/25 years training period for panel A/panel B. Column 6-7 gives OOS MSE and OOS R^2 with 60/30 years 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. Favero provides the estimates of (10) in which the $\hat{x}_{t+1}$ is estimated by x_t and MY_t. MO,MY , and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+1|t}$ in (9). CW is the Clark and West (2007) test.

Table 4: **Results of Return Regression Models (Out of Sample, Recursive, 5 Year Ahead Conditional Forecast)**

Panel A: 1901-2015: Forecasts begin tp years after sample.

training period (tp)	$tp = 30$		$tp = 40$		$tp = 60$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	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
MO	0.2123	-0.4767	0.1484	-0.0723	0.1719	-0.4683
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
MY &MO	0.5542	-2.8545	0.1212	0.1242	0.0815	0.3043
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
MO	0.5195	0.6172	0.2101	0.2609	0.6015	0.8223
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*
MY &MO	0.7268	0.7354	0.1053	0.0565*	0.0991*	0.0279**

Panel B: 1947-2015: Forecasts begin tp years after sample.

training period (tp)	$tp = 20$		$tp = 25$		$tp = 30$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	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
MO	0.1717	-0.0946	0.1738	-0.3513	0.1815	-0.5541
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
MY & MO	0.2057	-0.3112	0.2057	-0.5994	0.2234	-0.9133
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
MO	0.0990*	0.1567	0.1123	0.1770	0.2059	0.2405
MY_D	0.0532*	0.0785*	0.0521*	0.0729*	0.0810*	0.0919*
MY	0.0826*	0.0415**	0.0895*	0.0570*	0.1632	0.0758*
MY & MO	0.2168	0.9267	0.3075	0.8996	0.4336	0.8900

$***$, $**$, $*$ 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 5 year out of sample forecasting results. The dependent variable in all cases are yearly 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. The dependent variable in all cases are yearly log returns including dividends. HM stands for historical mean. PR denotes the 5 year ahead forecast obtained by forward recursion from the one-year ahead predictive regression forecast. MY_D shows the results from (14) when using MY as the demographic variable with $k = 5$. MO, MY_R , and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+h+1|t}$ in (9) with $h = 4$. CW is the Clark and West (2007) test.

Table 5: **Results of Return Regression Models (Out of Sample, Rolling, 1 Year Ahead Conditional Forecast)**

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

window (w)	$w = 30$		$w = 40$		$w = 60$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.0366	0	0.0272	0	0.0268	0
PR	0.0375	-0.0266	0.0270	0.0057	0.0279	-0.0389
FGT	0.0380	-0.0402	0.0273	-0.0071	0.0271	-0.0123
MO	0.0404	-0.1056	0.0305	-0.1244	0.0333	-0.2425
MY	0.0356	0.0260	0.0268	0.0132	0.0268	-0.0005
MY & MO	0.0478	-0.3068	0.0295	-0.0852	0.0331	-0.2339
Model	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	0.6870		0.4625		0.6330	
FGT	0.6593	0.5592	0.5228	0.5472	0.5352	0.3901
MO	0.8476	0.7679	0.7838	0.8034	0.9384	0.9417
MY	0.3947	0.2801	0.4610	0.4751	0.5013	0.3610
MY & MO	0.9112	0.8826	0.7068	0.7252	0.8869	0.9187

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

window (w)	$w = 20$		$w = 25$		$w = 30$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.0296	0	0.0308	0	0.0252	0
PR	0.0287	0.0296	0.0339	-0.1013	0.0291	-0.1537
FGT	0.0283	0.0450	0.0294	0.0473	0.0263	-0.0430
MO	0.0313	-0.0587	0.0322	-0.0466	0.0271	-0.0.0731
MY	0.0278	0.0610	0.0289	0.0630	0.0235	0.0668
MY & MO	0.0410	-0.3839	0.0389	-0.2621	0.0306	-0.2130
Model	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	0.3961		0.8211		0.9610	
FGT	0.3610	0.4516	0.3665	0.1756	0.6482	0.2326
MO	0.6194	0.6666	5979	0.3973	0.6138	0.3788
MY	0.2931	0.3892	0.3060	0.1292	0.2632	0.0598*
MY & MO	0.9258	0.9424	0.9493	0.8080	0.8105	0.5942

$***$, $**$, $*$ 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 1 year out of sample forecasting results based on the rolling forecast method. The dependent variable in all cases are yearly 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. Column 2-3 reports the out of sample mean square error (OOS MSE) and out of sample R^2 (OOS R^2) with 30/20 years training period for panel A/panel B. Column 4-5 shows OOS MSE and OOS R^2 with 40/25 years training period for panel A/panel B. Column 6-7 gives OOS MSE and OOS R^2 with 60/30 years 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 (10) in which the $\hat{x}_{t+1}$ is estimated by $x_t$and MY_t. MO,MY, and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+1|t}$ in (9). GW is the adjusted one-sided Giacomini and White (2006) test.

Table 6: **Results of Return Regression Models (Out of Sample, Rolling, 5 Year Ahead Conditional Forecast)**

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

window (w)	$w = 30$		$w = 40$		$w = 60$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	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
MO	0.2394	-0.4102	0.1853	-0.2423	0.2229	-0.7597
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
MY & MO	0.7292	-3.2948	0.2054	-0.3767	0.1749	-0.3809
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
MO	0.8077	0.8171	0.7196	0.6810	0.9346	0.9345
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
MY & MO	0.8737	0.8776	0.7556	0.7426	0.6974	7346

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

window (w)	$w = 20$		$w = 25$		$w = 30$	
Model for $\widehat{dp}_{t+1\|t}$	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	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
MO	0.1967	-0.0134	0.1622	0.0424	0.1800	-0.2292
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
MY & MO	0.4498	-1.3175	0.2971	-0.7535	0.3514	-1.3994
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
MO	0.5100	0.3964	0.4652	0.5148	0.6247	0.4150
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**
MY & MO	0.9445	0.9243	0.8255	0.6207	0.8814	0.7650

$***$, $**$, $*$ 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 5 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 yearly 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. The dependent variable in all cases are yearly log returns including dividends. HM stands for historical mean. PR denotes the 5 year ahead forecast obtained by forward recursion from the one-year ahead predictive regression forecast. MY_D shows the results of (14) by using MY as demographical choice with $k = 5$. MO, MY_R , and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+h+1|t}$ in (9) with $h = 4$. GW is the adjusted one-sided Giacomini and White (2006) test.

Figure 1: Common factor in stock returns and valuation ratios, and demographics

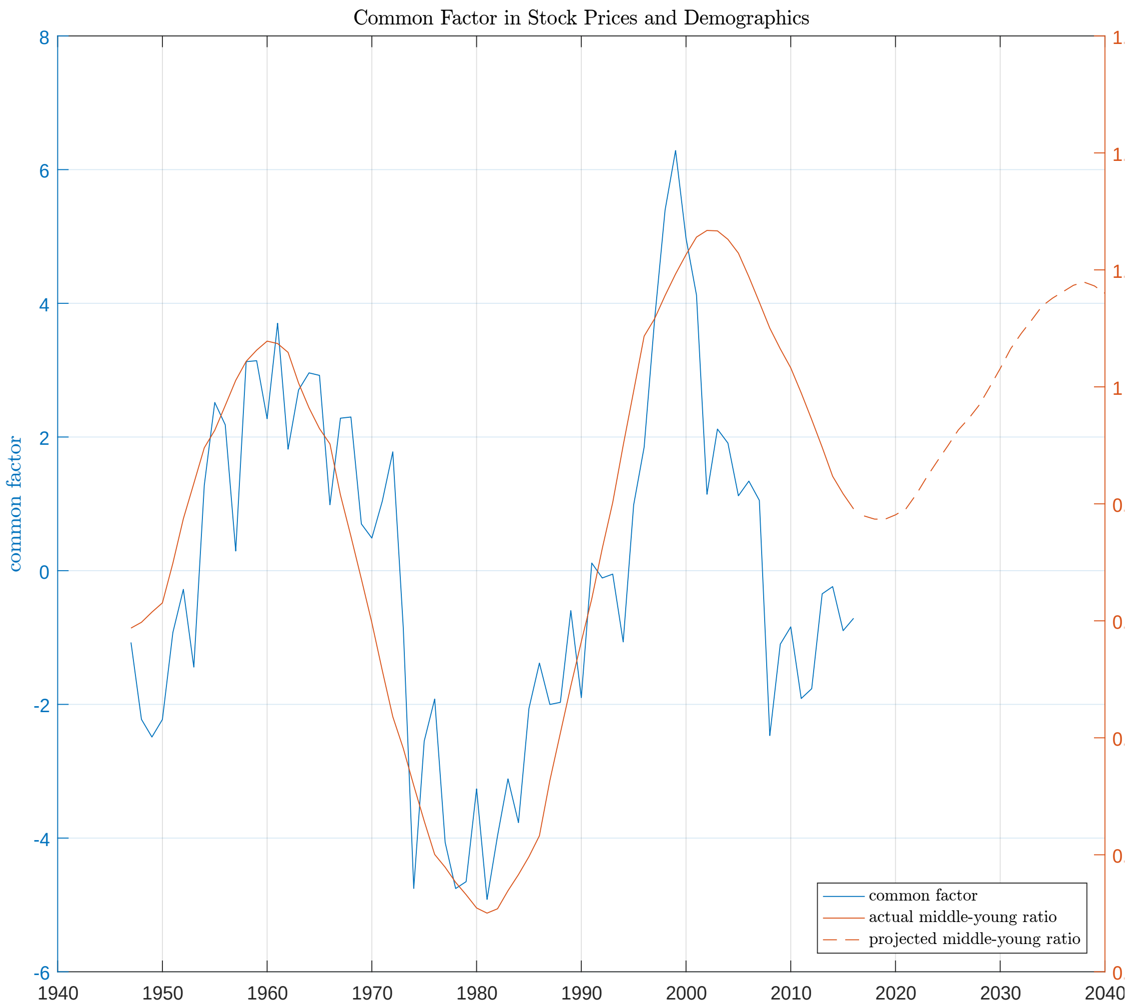
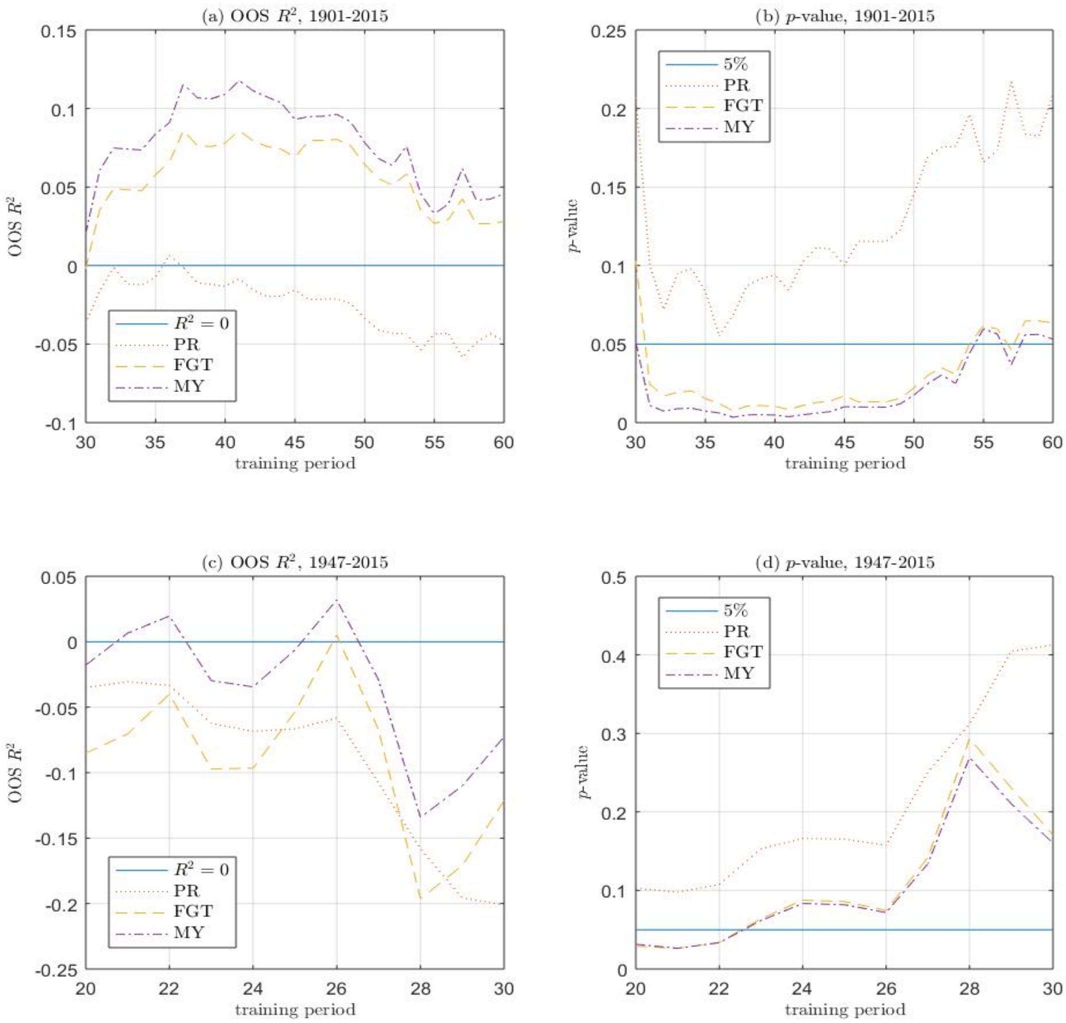
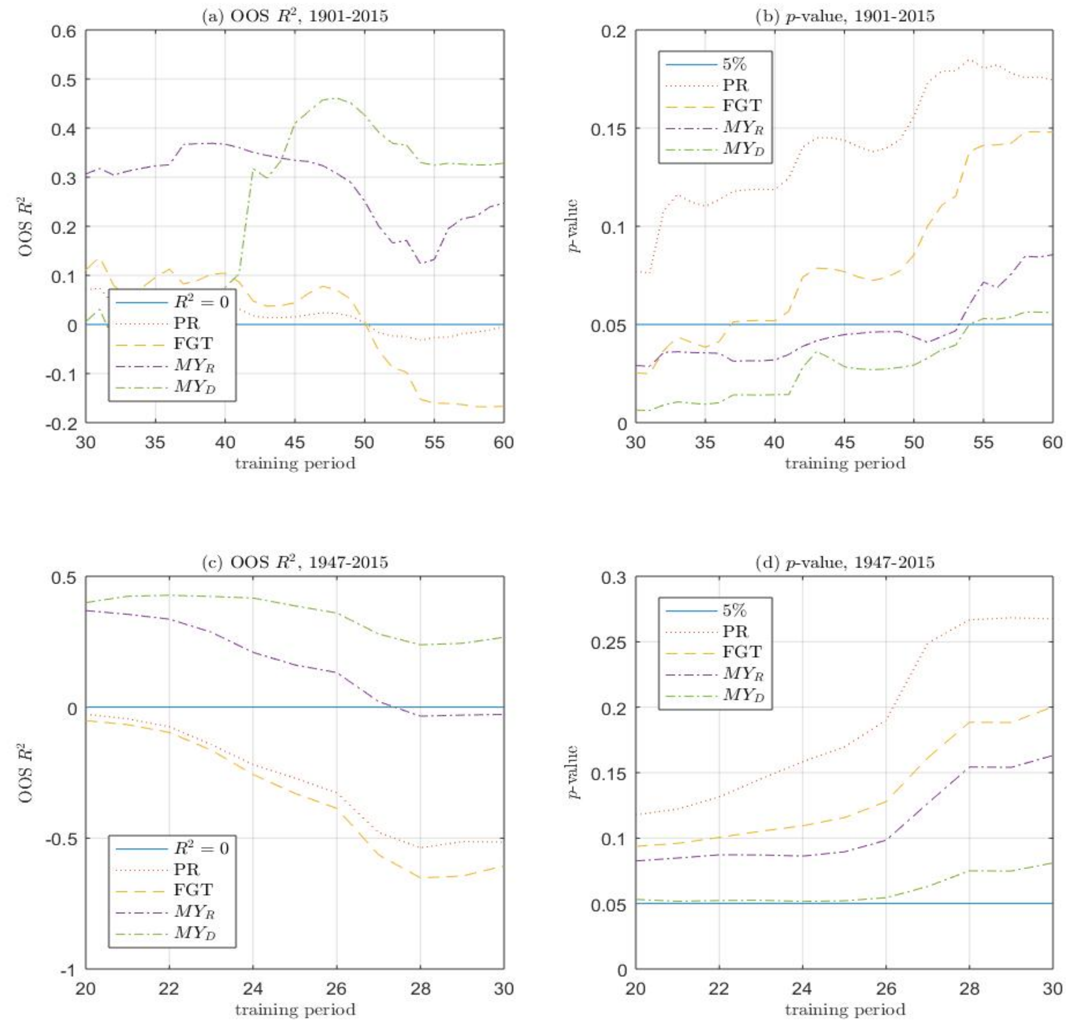


Figure 2: Conditionally Recursive Forecasts, 1 Year Ahead



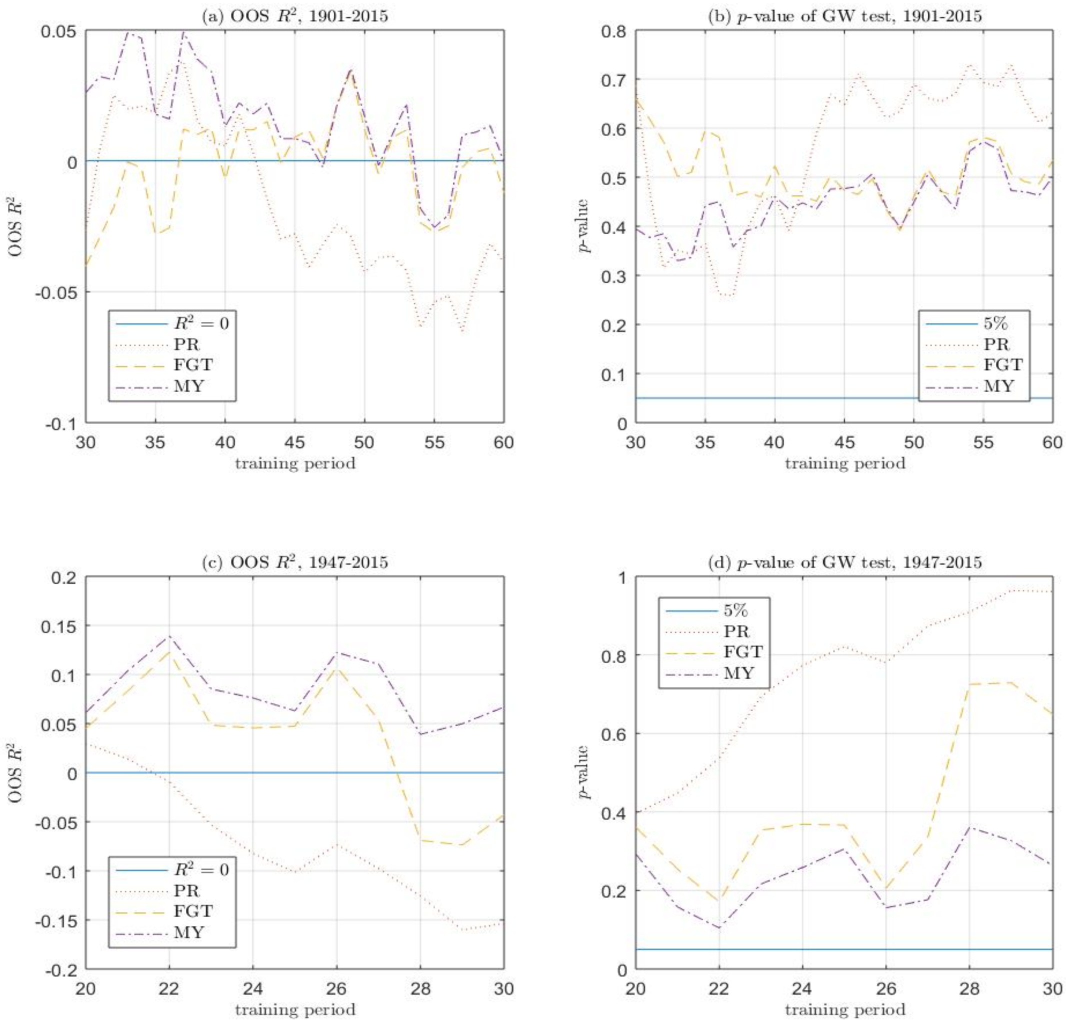
This figure provides 1 year ahead out of sample forecasting results with respect to different window sizes based on the recursive method. (a)-(b) show the OOS-R^2s and one-sided p-values for the full sample period 1901-2015 with the window size varying from 30-60 years. (c)-(d) show the analogous result for the post WWII period 1947-2015 with the window size varying from 20-30 years. The purple, red, and yellow dashed lines stand for the MY, 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).

Figure 3: Conditionally Recursive Forecasts, 5 Year Ahead



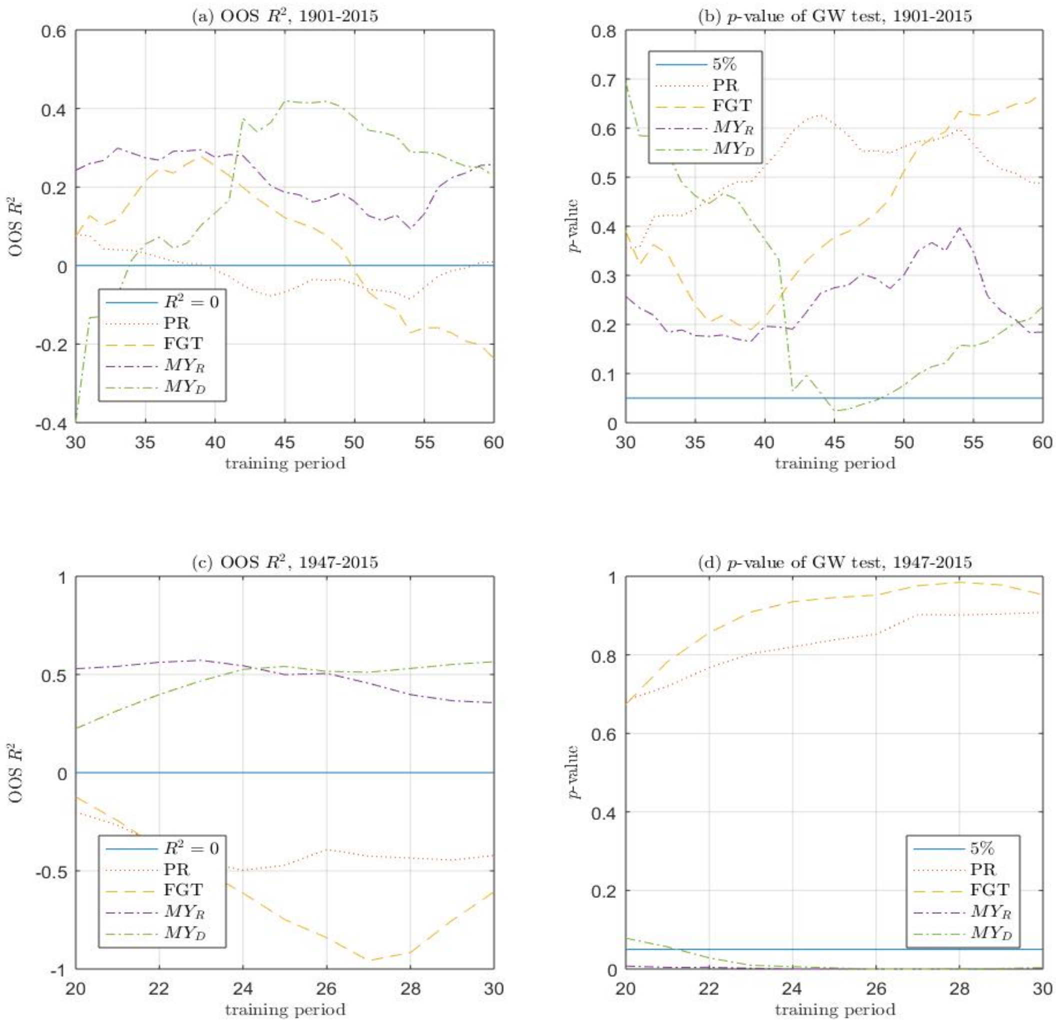
This figure provides 5 year ahead out of sample forecasting results with respect to different window size based on the recursive method. (a)-(b) show the OOS-R^2s and one-sided p-values for the full sample period 1901-2015 with the window size varying from 30-60 years. (c)-(d) show the analogous result 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).

Figure 4: Conditionally Rolling Forecasts, 1 Year Ahead



This figure provides 1 year ahead out of sample forecasting results with respect to different window size based on the rolling method. (a)-(b) show the OOS-R^2s and p-values 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 purple, red, and yellow dashed lines stand for the MY, 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).

Figure 5: Conditionally Rolling Forecasts, 5 Year Ahead



This figure provides 5 year ahead out of sample forecasting results with respect to different window size based on the rolling method. (a)-(b) show the OOS-R^2s and p-values 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).

Figure 6: Historical series and projections for MY_t and MO_t

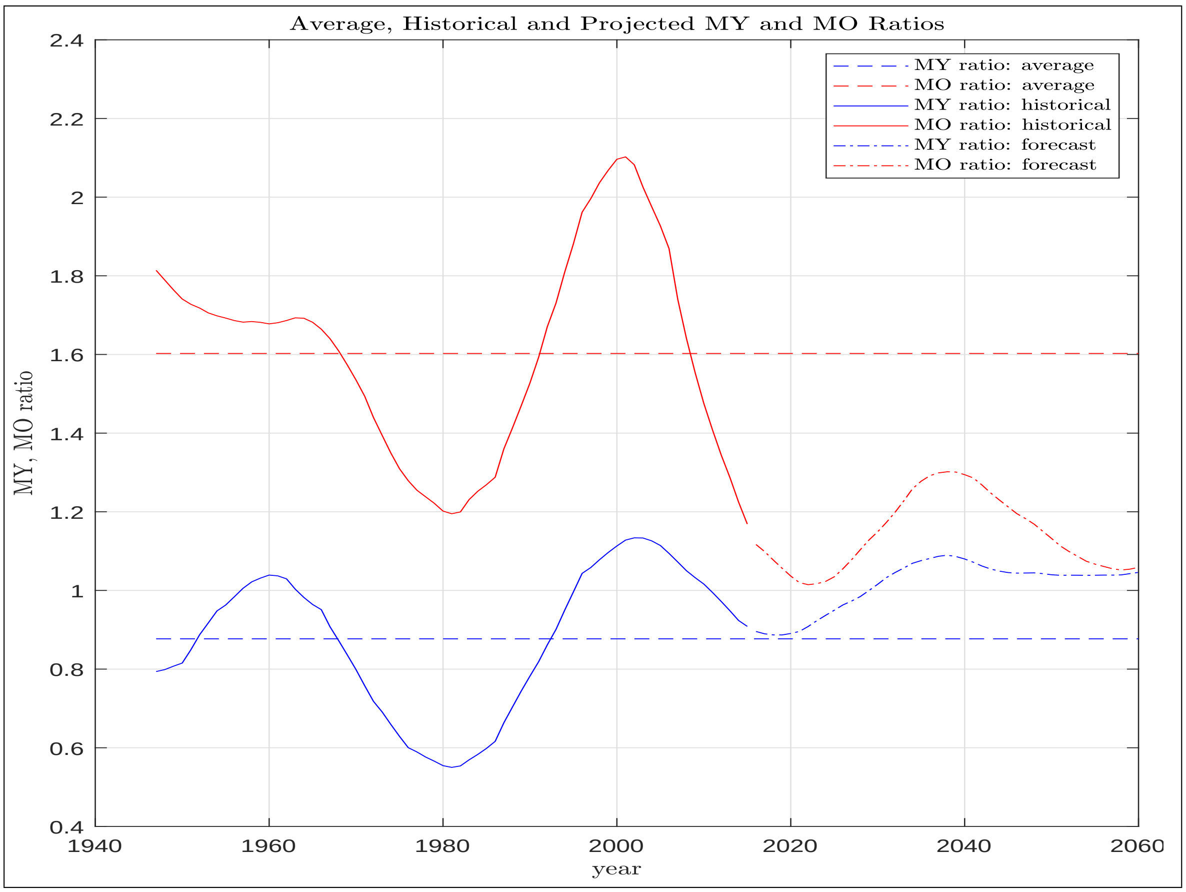


Figure 7: Historical series and projections for the dp ratio based on demographic ratios

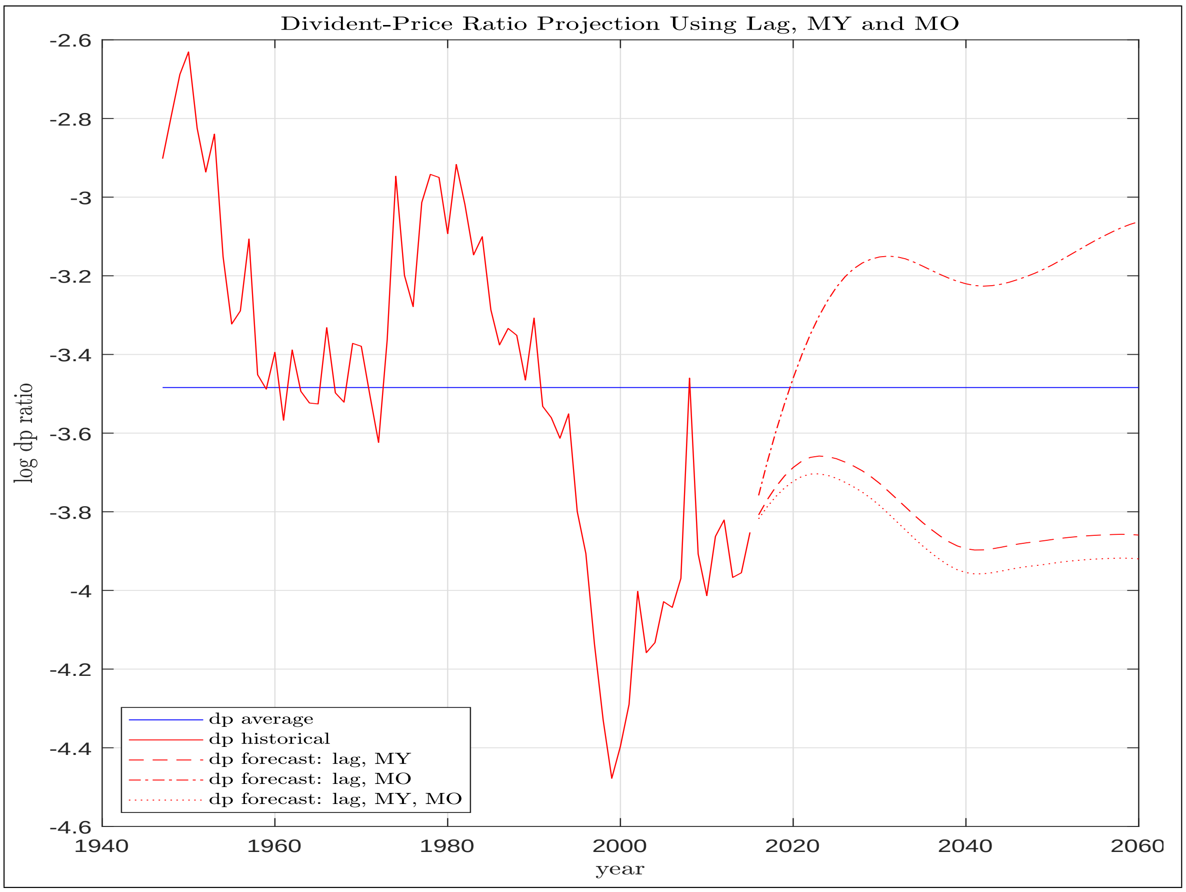
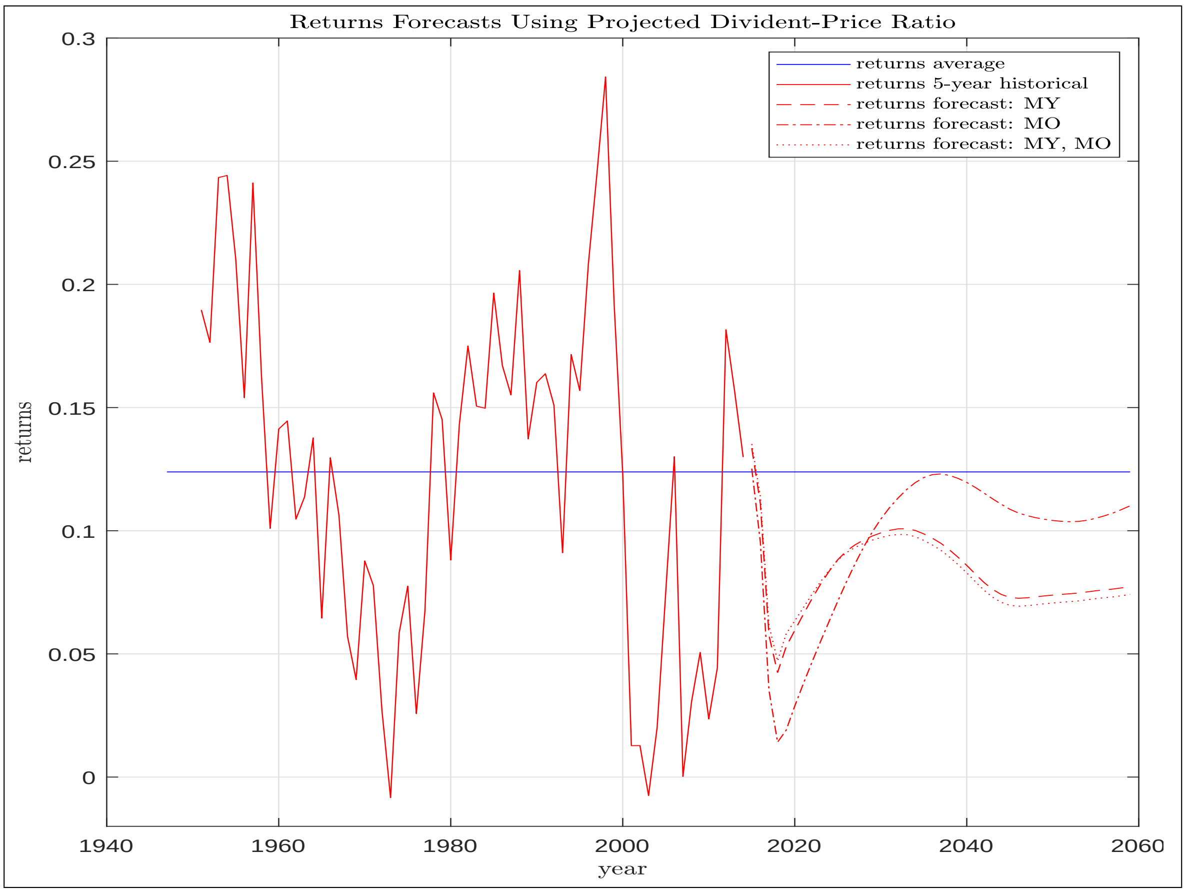


Figure 8: Historical series and projected 5 Year Rolling Average Stock Returns



B Separate Appendix for: Long-Horizon Stock Valuation and Return Forecasts Conditional on Demographic Projections

B.1 Unit Roots and Cointegration Tests

Table 7 provides unit root test results on the dividend price ratio (dp), middle-old ratio (MO), middle-young ratio (MY) and the log return with dividends (r) using the Augmented Dickey-Fuller Generalized Least Squares (ADF-GLS) test of Elliott et al. (1996). Panel A presents results for the full sample (1901-2015), whereas panel B presents results for the later part of the sample (1947-2015). A maximum of four lags was employed. The row corresponding to the lag with the minimum Modified Aikake Information Criterion (MAIC) value is in bold font and the row corresponding to the minimum Bayesian Information Criterion (BIC) is in italic. The tests allow for both an intercept and a linear trend under the alternative.

The results are somewhat dependent on the lag length and sample period. For both the dp and MY ratios we fail to reject at all but the BIC selected lag-length in the full sample. For dp we also fail to reject at all lag lengths in the later part of the sample, whereas for MY we reject for both BIC and MAIC lag length choice. The tests for MO reject using both BIC and MAIC in both samples. 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 a literal 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 three variables are highly persistent.

Favero et al. (2011) report evidence of cointegration between the dp and MY ratios. In Table 8, we next test for cointegration among dp, MO, and MY, both in pairs and all together. Panels I and II, respectively, provide results for the Engle-Granger two-step and Johansen cointegration tests. Within each panel, Sub-Panels A and B provide results for the full sample (1901-2015) and post WWII sample (1947-2015) periods, respectively.

The Engle-Granger two-step cointegration test results at the five percent significance level are presented in the first column of each sub-panel in Panel I. This tests the null hypothesis of no cointegration (unit root in the residuals), with a rejection of this null corresponding to a presence cointegration. The p-value is given in the second column and the AIC value corresponding to each lag choice is given in the third column. The rows corresponding to the minimum AIC value are in bold font. A maximum of five lags was employed. The tests allow for both an intercept and a linear trend under the alternative.

The results are again somewhat sensitive to the lag length and sample. Overall, the Engle-Granger test result does not provide strong evidence of cointegration between the dp ratio and the the two demographic variables. The only case in which we can reject the null of no cointegration at the 5 percent level at the AIC-selected lag length is for (dp, MO) in the 1947-2015 period. In all other cases, we fail to reject in favor of cointegration at the AIC-selected lag length.

The results of the Engle-Granger tests run somewhat contrary to the earlier findings of Favero et al. (2011) and raises the possibility of a spurious relationship between the financial ratios and demographic variables. However, as in the case of the unit root tests, the failure to reject does not imply acceptance of the null hypothesis, especially when the power of the test is known to be low.[11] Moreover, our specification in linking the dividend price ratio and demographic ratios in (6) includes a lag of the dividend price ratio, guaranteeing the stationarity of the residual and thus ruling out the possibility of a spurious regression. As discussed in the paragraph below, the Johansen tests are also generally more supportive of cointegration.

The Johansen test results are provided in the first column of each sub-panel in Panel II and the corresponding p-values are given in the second column. This tests the null of only one cointegrating vector against the alternative of one or more cointegrating vectors at the five percent significance level. A linear trend is allowed under the alternative hypothesis. Generally, the Johansen test provides stronger evidence of cointegration than the Engle-Granger test, with the exception of the (dp, MY) pair over the full sample. However, none of the cointegration tests can be reliable given the possibility of very persistent but not exactly unit roots variables. This is why in the main paper and below we confirm the robustness of our results by employing the IVS method of Kostakis et al.(2015).

B.2 In Sample Estimation

In this section we present the additional results we have for in sample estimation. Table 9 and Table 10 include, respectively, OLS estimation of the Dividend Price Ration Models and the Return Regression Models. Table 11 presents robustness checks for the IVX estimation with c=-0.9 and α=1.

[11] The Engle-Granger test is expected to have lower power in the presence of strong endogeneity (Pesavento (2004)).

B.3 Out of Sample Estimation

The rest of the Appendix presents results for Out of Sample Estimation when IVX is used in the first step of the estimation to take into account possible persistency of the variables. The Figures replicate what we have in the paper with IVX estimation.

Table 7: **Augmented Dickey Fuller - Generalized Least Squares Test 1901-2015**

		Panel A: 1901-2015		**Panel B: 1947-2015**	
Time Series	lag length	ADF -GLS	Unit Root	ADF -GLS	Unit Root
dp_t	0	-3.75^{***}	*Reject*	-2.48	*Fail*
	1	-2.78	Fail	-2.19	Fail
	2	-1.96	**Fail**	-1.82	**Fail**
	3	-2.07	Fail	-2.09	Fail
	4	-2.14	Fail	-2.60	Fail
$\ln(MO_t)$	0	-0.56	Fail	-0.21	Fail
	1	-4.24^{***}	**Reject**	-3.55^{**}	**Reject**
	2	-4.94^{***}	Reject	-4.22^{***}	Reject
	3	-5.04^{***}	Reject	-4.41^{***}	Reject
	4	-6.22^{***}	*Reject*	-5.72^{***}	*Reject*
$\ln(MY_t)$	0	-0.91	Fail	-0.60	Fail
	1	-1.67	**Fail**	-1.94	Fail
	2	-1.99	Fail	-3.23^{**}	**Reject**
	3	-2.45	Fail	-4.30^{***}	Reject
	4	-2.91^{*}	*Reject*	-5.40^{***}	*Reject*
r_t	0	-10.59^{***}	**Reject**	-8.45^{***}	**Reject**
	1	-9.24^{***}	*Reject*	-7.13^{***}	Reject
	2	-6.38^{***}	Reject	-5.05^{***}	Reject
	3	-5.97^{***}	Reject	-3.80^{***}	Reject
	4	-6.23^{***}	Reject	-2.98^{*}	Reject

The table shows the results from the Augmented Dickey Fuller Generalized Least Squares (ADF-GLS) unit root test of Elliott et al. (1996) for the four time series shown in the first column. Column 2 (lag length) shows the number of lagged first differences included in the ADF-GLS regression specification. Results for the time period 1901-2015 are shown in Panel A (Columns 3-4) and results for 1947-2015 period are shown in Panel B (Columns 5-6). Column 3 and 5 show the value of the ADF-GLS test statistic, with one, two and three stars(*) denoting rejection at 10%, 5% and 1% respectively. Column 4 and 6 denotes whether the unit root was rejected at least at the 10% level, with "reject" noting a rejection of the unit root hypothesis, and "Fail" noting a failure to reject. All the tests are run with an intercept and time trend. The critical values for the ADF-GLS test are -3.70 (1%), -3.13 (5%) and -2.83 (10%). Bold indicates the lag length chosen by the MAIC criteria with a maximum of 4 lags. Italics indicates the lag length chosen by BIC.

Table 8: **Cointegration Tests**

		Panel-I: Engle-Granger 2-Step						Panel -II: Johansen			
		Sub-Panel A: 1901-2015			Sub-Panel B: 1947-2015			Sub-Panel A: 1901-2015		Sub-Panel B: 1947-2015	
Time Series	lag length	No Cointegration	P Value	AIC Value	No Cointegration	P Value	AIC Value	No Cointegration	P Value	No Cointegration	P Value
dp	0	Reject	0.045	-22.88	**Reject**	**0.034**	**-56.245**	Fail	0.549	Fail	0.733
&	1	Fail	0.081	-19.949	Fail	0.077	-53.769	Reject	0.002	Reject	0.03
MO	2	Fail	0.415	-24.93	Fail	0.29	-56.207	Reject	0.004	Reject	0.007
	3	**Fail**	**0.258**	**-25.616**	Fail	0.152	-55.307	Reject	0.002	Reject	0.001
	4	Fail	0.214	-22.988	Fail	0.075	-54.974	Reject	0. 001	Reject	0.001
	5	Fail	0.284	-19.929	Fail	0.136	-50.814	Reject	0.016	Reject	0.003
dp	0	Reject	0.004	-35.663	Reject	0.023	-56.182	Reject	0.02	Fail	0.56
&	1	Reject	0.011	-33.239	Fail	0.069	-53.1	Fail	0.09	Reject	0.027
MY	2	Fail	0.22	-42.447	**Fail**	**0.404**	**-56.291**	Fail	0.419	Reject	0.01
	3	**Fail**	**0.136**	**-43.761**	Fail	0.37	-55.072	Fail	0.269	Reject	0.008
	4	Fail	0.115	-42.025	Fail	0.2	-53.517	Fail	0.175	Reject	0.001
	5	Fail	0.14	-41.23	Fail	0.302	-49.544	Fail	0.096	Reject	0.001
dp,	0	Reject	0.005	-34.52	Fail	0.056	-34.52	Reject	0.04	Reject	0.02
MO	1	Reject	0.0150	-32.1018	**Fail**	**0.151**	**-52.766**	Reject	0.03	Reject	0.02
&	2	Fail	0.254	-41.448	Fail	0.596	-49.549	Reject	0.04	Reject	0.01
MY	3	**Fail**	**0.154**	**-42.668**	Fail	0.596	-52.649	Reject	0.02	Reject	0.004
	4	Fail	0.131	-40.836	Fail	0.404	-51.027	Reject	0.001	Reject	0.001
	5	Fail	0.159	-39.9	Fail	0.5	-45.639	Reject	0.003	Reject	0.001

The table shows both the two-step Engle-Granger and the Johansen cointegration tests of the null hypothesis of no cointegration. The combination of variables tested are shown in Column 1, where dp is the log dividend price ratio, and MO and MY are the middle-to-old and middle-to-young ratios, respectively. Column 2 (lag length) shows the number of lagged first differences included in the Engle-Granger specification. Engle-Granger test results are provided in Panel I and Johansen test results are provided in Panel II. Within each panel results for the time period 1901-2015 are shown in Sub-Panel A (Columns 3-5) and results for 1947-2015 period are shown in Sub-Panel B (Columns 6-8). Columns 3 and 6 (No cointegration) the result of testing the null hypothesis of no cointegration at the five percent significance level, with "reject" noting a rejection of the unit root hypothesis, and "Fail" noting a failure to reject. The Engle-Granger test P-values are provided in Columns 4 and 7. The Aikake Information Criterion (AIC) for the number of lagged differences is shown in Columns 5 and 8. The row with the lag-length that minimizes this criteria is shown in bold. The result at the five percent level for the Johansen test of the null of no cointegrating vectors, with the alternative of one or more cointegrating vectors are shown in Columns 9 and 11. The corresponding p-values are shown in Columns 10 and 12.

Table 9: **Dividend Price Ratio Model Estimation**

Model:	AR(1)	Augmented AR(1)		
		MO and	MY and	MO, MY,
	lag dp	lag dp	lag dp	and lag dp
Panel A: 1901-2015				
Constant	-0.3679^{**}	-0.5131^{**}	-0.3952^{***}	-0.5287^{***}
	(0.1474)	(0.2040)	(0.1410)	(0.1951)
lap dp	0.889^{***}	0.8751^{***}	0.7570^{***}	0.7452^{***}
	(0.0447)	(0.0467)	(0.0576)	(0.0589)
MO		0.05327		0.04904
		(0.0518)		(0.0495)
MY			-0.5039^{***}	-0.5002^{***}
			(0.1477)	(0.1478)
Observations	114	114	114	114
Adjusted R^2	0.777	0.777	0.797	0.797
Root MSE	0.219	0.219	0.209	0.209
Panel B: 1947-2015				
Constant	-0.3085^{**}	-0.2269	-0.3347^{***}	-0.3450^{**}
	(0.1652)	(0.1730)	(0.1617)	(0.1903)
lap dp	0.9153^{***}	0.8790^{***}	0.8292^{***}	0.8281^{***}
	(0.0471)	(0.0529)	(0.0620)	(0.0633)
MO		-0.1301		0.0141
		(0.0888)		(0.1336)
MY			-0.3115^{**}	-0.3297
			(0.1503)	(0.2296)
Observations	68	68	68	68
Adjusted R^2	0.849	0.852	0.856	0.854
Root MSE	0.171	0.169	0.167	0.168

$***$ 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 provides preliminary OLS estimation of equation (3) & (6), with standard errors in parenthesis. Column 2 provides estimates for the pure AR(1) process in (3). Column 3-5 provide estimates for the augmented AR(1) process including demographic ratios as in (6). In columns 3, 4, and 5 respectively, MO, MY, and both MO and MY are employed respectively 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. Note that the OLS estimates, standard errors and significance levels do not account for the potential non-stationarity of the variables and should be treated only as preliminary findings. The standard errors, cannot for example, be used to assess the significance of the difference of the coefficient on lag dp and one. Non-HAC standard errors are reported due to the inclusion of the lagged dependent variable.

Table 10: **Estimation of Return Regression Models (In Sample)**

Model:	lag dp	MO and lag dp	MY and lag dp	MO, MY, and lag dp
Panel A: 1901-2015				
Constant	0.2566**	-0.0649	-0.02107	-0.0248
	(0.1265)	(0.3321)	(0.1541)	(0.1512)
$\widehat{dp}_{t+1\|t}$		-0.8736	-0.7546^{***}	-0.7647^{**}
		(0.8346)	(0.2547)	(0.2435)
dp_t	0.0621	0.8387*	0.7329**	0.7419***
	(0.0384)	(0.7430)	(0.2294)	(0.2196)
Observations	114	114	114	114
Adjusted R^2	0.0141	0.0149	0.0781	0.0864
F Statistic	2.62	1.86	5.79**	6.34**
MSE	0.0353	0.0353	0.0331	0.0328
Out-performance P value ($Versus$ lag dp, $1year$)		0.2640	0.0027***	0.0020***
Out-performance P value ($Versus$ lag dp, $5year$)		0.0268**	0.0000***	0.0000***
Panel B: 1947-2015				
Constant	0.4015**	0.09715	0.0319	0.0283
	(0.1530)	(0.2466)	(0.1987)	(0.1983)
$\widehat{dp}_{t+1\|t}$		-0.9865	-1.1978^{***}	-1.2096^{***}
		(0.6311)	(0.4369)	(0.4358)
dp_t	0.0977**	1.0006*	1.1941***	1.2048***
	(0.0436)	(0.5793)	(0.4021)	(0.4010)
Observations	68	68	68	68
Adjusted R^2	0.0564	0.0766	0.141	0.143
F Statistic	5.01**	3.78**	6.51***	6.61***
MSE	0.025	0.0243	0.0228	0.0228
Out-performance P value ($Versus$ lag dp, $1year$)		0.0967*	0.0246**	0.0244**
Out-performance P value ($Versus$ lag dp, $5year$)		0.0056***	0.0000***	0.0000***

$***$ 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 provides preliminary OLS estimation of equations (1), (4), and (5) with standard errors in parenthesis. The dependent variable in all cases are yearly log returns including dividends. Column 2 (lag dp), provides the estimates of (1) in which only on the past dp_t is employed as a predictor. Columns 3-5 provide estimates of (5), using three different specifications of the demographic ratio projection $dp_{t+1|t}$ in (6): only MO (Column 4), only MY (Column 5), and both MO and MY (Column 6). Results for the time period 1901-2015 are shown in Panel A and results for 1947-2015 period are shown in Panel B. Note that the standard errors and significance levels do not account for the potential non-stationarity or near-non-stationarity of the regressors and should be considered as preliminary findings only. Standard errors are employed without HAC since the one year excess returns show no significant autocorrelations for the first five lags according to Ljung-Box tests available upon request.

Table 11: **Dividend Price Ratio Model IVX Estimation c=-0.9, α=1**

Model:	AR(1)	Augmented AR(1)		
		MO and	MY and	MO, MY,
	lag dp	lag dp	lag dp	and lag dp
Panel A: 1901-2015				
Constant	-0.3679^{**}	-0.5131^{**}	-0.3952^{***}	-0.5287^{***}
lap dp	0.888^{***}	0.888^{***}	0.778^{***}	0.770^{***}
MO		0.024		0.010
MY			-0.418^{**}	-0.424^{*}
Overall test of significance	382.76^{***}	388.12^{***}	417.08^{***}	429.07^{***}
p-value	0	0	0	0
Panel B: 1947-2015				
Constant	-0.3085^{**}	-0.2269	-0.3347^{***}	-0.3450^{**}
lap dp	0.906^{***}	0.834^{***}	0.792^{***}	0.794^{***}
MO		-0.202		-0.084
MY			-0.380^{**}	-0.274
Overall test of significance	169.47^{***}	276.74^{***}	229.30^{***}	290.95^{***}
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 provides IVX estimates of (3) & (6). Column 2 provides estimates for the pure AR(1) process in (3). Column 3-5 provide estimates for the augmented AR(1) process including demographic ratios as in (6). In columns 3, 4, and 5 respectively, MO, MY, and both MO and MY are employed respectively 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.

Table 12: **Results of Return Regression Models (Out of Sample, Recursive, 1 Year Ahead Conditional Forecast, IVX Estimation)**

Panel A: 1901-2015, forecasts begin tp years after sample.

training period (tp)	tp=30		tp=40		tp=60	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.0354	0	0.0266	0	0.0262	0
PR	0.0392	-0.1063	0.0268	-0.0075	0.0284	-0.0848
FGT	0.0340	0.0395	0.0237	0.1073	0.0250	0.0464
MO	0.0482	-0.3613	0.0277	-0.0406	0.0298	-0.1401
MY	0.0337	0.0475	0.0234	0.1215	0.0252	0.0391
MY & MO	0.0560	-0.5807	0.0277	-0.0402	0.0260	0.0054
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	0.3802		0.0432**		0.4121	
FGT	0.0320	0.0000***	0.0040**	0.0018***	0.0547*	0.0040**
MO	0.7088	0.8324	0.0716*	0.3136	0.5163	0.3780
MY	0.3252	0.0005***	0.0048**	0.0014***	0.0551*	0.0039***
MY & MO	0.4416	0.2981	0.0181**	0.0094***	0.1369	0.0090***

Panel B: 1947-2015: forecasts begin tp years after sample.

training period (tp)	tp=20		tp=25		tp=30	
Model	MSE OOS	R^2	MSE OOS	R^2	MSE OOS	R^2
HM	0.0282	0	0.0293	0	0.0240	0
PR	0.0358	-0.2717	0.0390	-0.3281	0.0379	-0.5788
FGT	0.0657	-1.33270	0.0636	-1.1691	0.0599	-1.4965
MO	0.1591	-4.6462	0.0470	-0.6009	0.0405	-0.6890
MY	0.0599	-1.1266	0.0581	-0.9790	0.0553	-1.3047
MY & MO	0.1678	-4.9548	0.0758	-1.5836	0.0532	-1.2168
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	0.0535*		0.0973*		0.3030	
FGT	0.0666**	0.0054***	0.1520	0.0051***	0.2808	0.0094***
MO	0.7970	0.0382**	0.0.4716	0.0162**	0.3502	0.0131**
MY	0.0759*	0.0083**	0.1598	0.0069**	0.2837	0.0127**
MY & MO	0.8750	0.4698	0.7493	0.6488	0.4642	0.2374

$***$, $**$, $*$ 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 1 year out of sample forecasting results. The dependent variable in all cases are yearly 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. Column 2-3 reports the out of sample mean square error (OOS MSE) and out of sample R^2 (OOS R^2) with 30/20 years training period for panel A/panel B. Column 4-5 shows OOS MSE and OOS R^2 with 40/25 years training period for panel A/panel B. Column 6-7 gives OOS MSE and OOS R^2 with 60/30 years 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. Favero provides the estimates of (10) in which the $\hat{x}_{t+1}$ is estimated by x_t and MY_t. MO,MY , and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+1|t}$ in (9). CW is the Clark and West (2007) test.

Table 13: **Results of Return Regression Models (Out of Sample, Recursive, 5 Year Ahead Conditional Forecast, IVX Estimation)**

Panel A: 1901-2015, forecasts begin *tp* years after sample.

training period (*tp*)	*tp*=30		*tp*=40		*tp*=60	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.1438	0	0.1384	0	0.1171	0
PR	0.1339	0.0690	0.1349	0.0253	0.1205	-0.0291
FGT	0.1125	0.2174	0.1155	0.1650	0.1150	0.0179
MO	0.2672	-0.8588	0.1654	-0.1952	0.1929	-0.6475
MY_D	0.1344	0.0654	0.1293	0.0654	0.0755	0.3555
MY_R	0.1053	0.2677	0.0908	0.3442	0.0909	0.2233
MY & MO	0.2913	-1.0260	0.1346	0.0271	0.1175	-0.0033
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	0.0764*		0.1221		0.1808	
FGT	0.0202**	0.0173**	0.0411**	0.0325**	0.1148	0.1226
MO	0.5767	0.7727	0.1912	0.4533	0.4540	0.9266
MY_D	0.0062***	0.0071***	0.0134**	0.0149**	0.0517*	0.0665*
MY_R	0.0324**	0.0299**	0.0327**	0.0222**	0.0817*	0.0797*
MY & MO	0.4935	0.6605	0.0913*	0.0914*	0.1312	0.1250

Panel B: 1947-2015: forecasts begin *tp* years after sample.

training period (*tp*)	*tp*=20		*tp*=25		*tp*=30	
Model	MSE OOS	R^2	MSE OOS	R^2	MSE OOS	R^2
HM	0.1569	0	0.1286	0	0.1168	0
PR	0.1599	-0.0196	0.1603	-0.2465	0.1720	-0.4727
FGT	0.2298	-0.4652	0.2445	-0.9011	0.2750	-1.3548
MO	0.7401	-3.7184	0.1618	-0.2579	0.1532	-0.3117
MY_D	0.0986	0.3714	0.0866	0.3262	0.0903	0.2271
MY_R	0.1344	0.1430	0.0992	0.2288	0.0997	0.1463
MY & MO	2.2532	-13.3650	0.9047	-6.0350	0.8311	-6.1170
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	0.1345		0.2006		0.2934	
FGT	0.0847*	0.3325	0.1091	0.3773	0.1725	0.6799
MO	0.1258	0.0908*	0.1203	0.1199	0.2271	0.1961
MY_D	0.0568*	0.0751*	0.0564*	0.0743*	0.0853*	0.1014
MY_R	0.0495**	0.1437	0.0387**	0.1085	0.0792*	0.1055
MY & MO	0.1558	0.3665	0.2881	0.6965	0.5900	0.8630

$***$, $**$, $*$ 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 1 year out of sample forecasting results. The dependent variable in all cases are yearly 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. Column 2-3 reports the out of sample mean square error (OOS MSE) and out of sample R^2 (OOS R^2) with 30/20 years training period for panel A/panel B. Column 4-5 shows OOS MSE and OOS R^2 with 40/25 years training period for panel A/panel B. Column 6-7 gives OOS MSE and OOS R^2 with 60/30 years 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. Favero provides the estimates of (10) in which the $\hat{x}_{t+1}$ is estimated by x_t and MY_t. MO,MY , and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+1|t}$ in (9). CW is the Clark and West (2007) test.

Table 14: **Results of Return Regression Models (Out of Sample, Rolling, 1 Year Ahead Conditional Forecast, IVX Estimation)**

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

window (w)	$w = 30$		$w = 40$		$w = 60$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.0366	0	0.0272	0	0.0268	0
PR	0.0513	-0.4030	0.0326	-0.1992	0.0282	-0.0511
FGT	0.0538	-0.4706	0.0580	-1.1354	0.0463	-0.7265
MO	0.0636	-0.7407	0.0439	-0.6168	0.0617	-1.3026
MY	0.0617	-0.6863	0.0631	-1.3256	0.0496	-0.8492
MY & MO	0.0915	-1.5024	0.0384	-0.4142	0.0579	-1.1588
Model	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	0.9850		0.9372		0.6493	
FGT	0.9654	0.5946	0.9702	0.9375	0.9694	0.9869
MO	0.9897	0.8701	0.9956	0.9398	0.9850	0.9873
MY	0.9988	0.8654	0.9883	0.9698	0.9776	0.9899
MY & MO	0.9900	0.9484	0.9750	0.8138	0.9783	0.9791

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

window (w)	$w = 20$		$w = 25$		$w = 30$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.0296	0	0.0308	0	0.0252	0
PR	0.0558	-0.8840	0.0564	-0.8303	0.0556	-1.2031
FGT	0.1163	-2.9293	0.0457	-0.4838	0.0380	-0.5055
MO	0.1658	-4.5997	0.0438	-0.4224	0.0342	-0.3559
MY	0.0893	-2.0153	0.0348	-0.1281	0.0254	-0.0079
MY & MO	0.3405	-10.4998	0.1042	-2.3805	0.0525	-1.0798
Model	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	0.9481		0.9348		0.9937	
FGT	0.9999	0.9932	0.9221	0.2959	0.9593	0.1051
MO	0.9717	0.9346	0.8880	0.2597	0.8557	0.0640*
MY	0.9998	0.9442	0.6798	0.1216	0.5194	0.0096***
MY & MO	0.9990	0.9971	0.9973	0.9339	0.9985	0.4176

$***$, $**$, $*$ 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 1 year out of sample forecasting results based on the rolling forecast method. The dependent variable in all cases are yearly 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. Column 2-3 reports the out of sample mean square error (OOS MSE) and out of sample R^2 (OOS R^2) with 30/20 years training period for panel A/panel B. Column 4-5 shows OOS MSE and OOS R^2 with 40/25 years training period for panel A/panel B. Column 6-7 gives OOS MSE and OOS R^2 with 60/30 years 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 (10) in which the $\hat{x}_{t+1}$ is estimated by $x_t$and MY_t. MO,MY, and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+1|t}$ in (9). GW is the adjusted one-sided Giacomini and White (2006) test.

Table 15: **Results of Return Regression Models (Out of Sample, Rolling, 5 Year Ahead Conditional Forecast, IVX Estimation)**

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

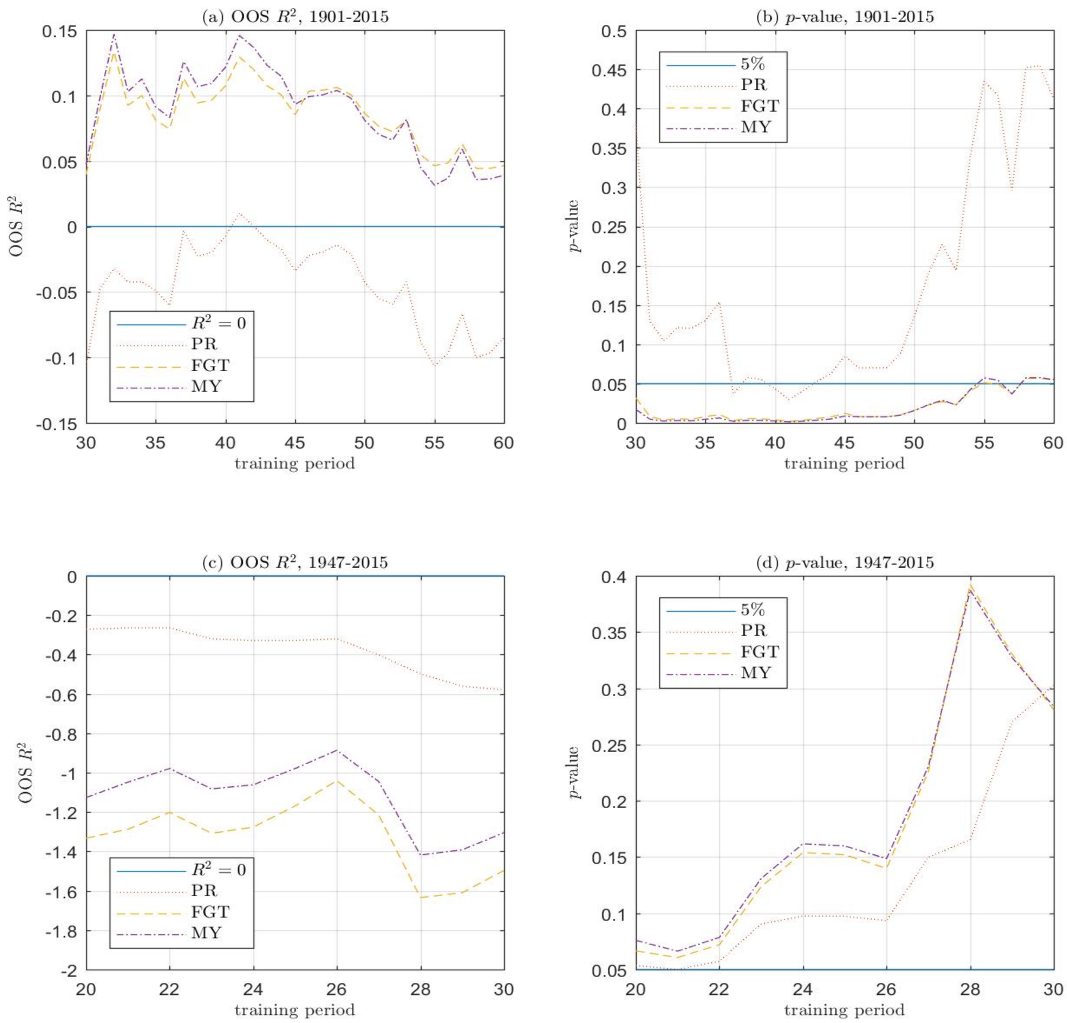
window (w)	$w = 30$		$w = 40$		$w = 60$	
Model	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.1698	0	0.1492	0	0.1266	0
PR	0.2028	-0.1942	0.1775	-0.1897	0.1273	-0.0055
FGT	0.2594	-0.5277	0.1988	-0.3327	0.2641	-1.0853
MO	0.4108	-1.4194	0.1898	-0.2721	0.2012	-0.5886
MY_D	0.2559	-0.5071	0.1353	0.0930	0.1407	-0.1107
MY_R	0.1286	0.2428	0.1080	0.2757	0.0940	0.2578
MY & MO	0.6161	-2.6287	0.3286	-1.2028	0.3054	-1.4111
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	0.8432		0.8361		0.5174	
FGT	0.8253	0.6974	0.7502	0.5893	0.9644	0.9793
MO	0.9836	0.9552	0.7137	0.5581	0.8392	0.8584
MY_D	0.7175	0.6338	0.4128	0.2917	0.5910	0.5877
MY_R	0.2571	0.1816	0.1964	0.1529	0.1842	0.1477
MY & MO	0.9635	0.9496	0.9646	0.9151	0.8941	0.9067

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

window (w)	$w = 20$		$w = 25$		$w = 30$	
Model for $\widehat{dp}_{t+1\|t}$	MSE OOS	OOS R^2	MSE OOS	OOS R^2	MSE OOS	OOS R^2
HM	0.1941	0	0.1694	0	0.1465	0
PR	1.1298	-4.8204	0.2727	-0.6100	0.1742	-0.1893
FGT	0.1599	0.1760	0.3184	-0.8795	0.3079	-1.1021
MO	1.0988	-4.6605	0.4388	-1.5904	0.3435	-1.3454
MY_D	0.1490	0.2324	0.0758	0.5527	0.0603	0.5880
MY_R	0.0914	0.5291	0.0848	0.4991	0.0943	0.3561
MY & MO	3.7911	-18.5305	0.8328	-3.9159	0.4105	-1.8024
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	0.9226		0.9678		0.9078	
FGT	0.2008	0.0728*	0.8402	0.6184	0.8949	0.8395
MO	0.9592	0.4865	0.9979	0.9756	0.9780	0.9368
MY_D	0.0691*	0.0689*	0.0029***	0.0029***	0.0040***	0.0007***
MY_R	0.0075***	0.0607*	0.0005***	0.0027***	0.0025***	0.0016***
MY & MO	0.9749	0.8994	0.9969	0.9890	0.9728	0.9429

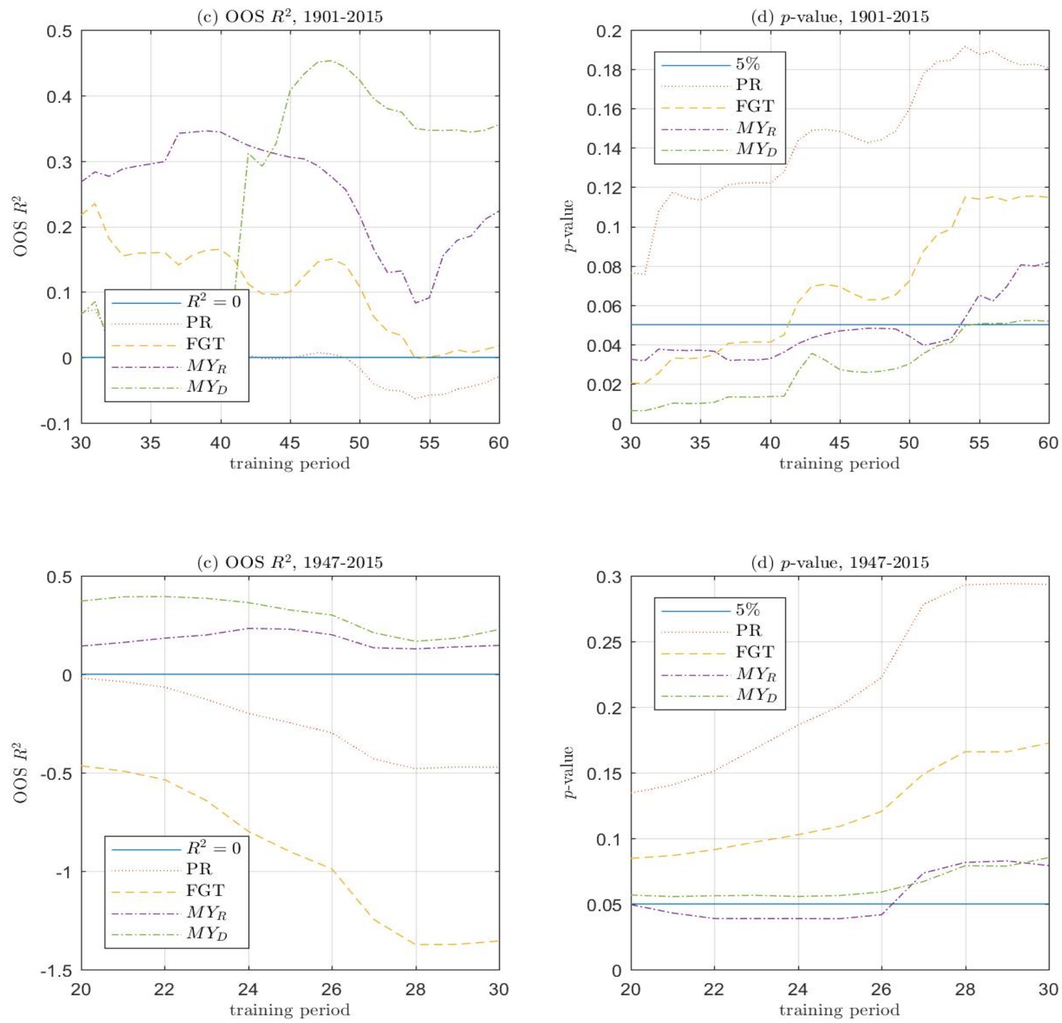
$***$, $**$, $*$ significantly out-performs the benchmark forecasts at the 1% level, 5% level, and 10 % level, respectively. Reported p-values are one-sided. This table provides 5 year out of sample forecasting results based on the rolling forecast method. It also provides the out-performance test results. The dependent variable in all cases are yearly 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. The dependent variable in all cases are yearly log returns including dividends. HM stands for historical mean. PR denotes the 5 year ahead forecast obtained by forward recursion from the one-year ahead predictive regression forecast. MY_D shows the results of (14) by using MY as demographical choice with $k = 5$. MO, MY_R , and MYMO provide estimates of (10) using three different specifications of the demographic ratio projection $dr_{t+h+1|t}$ in (9) with $h = 4$. GW is the adjusted one-sided Giacomini and White (2006) test.

Figure 9: Conditionally Recursive Forecasts, 1 Year Ahead, IVX Estimation



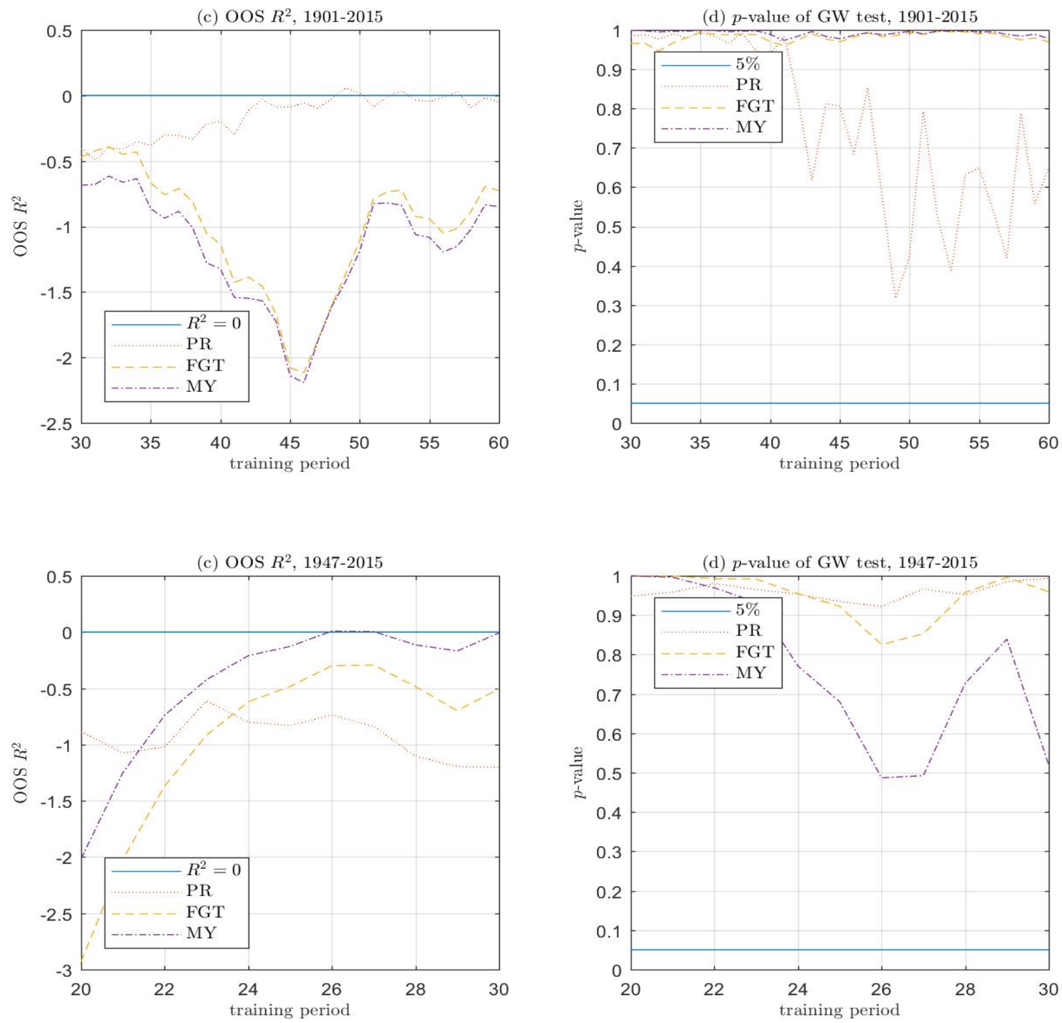
This figure provides IVX estimates of 1 year ahead out of sample forecasting results with respect to different window sizes based on the recursive method. (a)-(b) show the OOS-R^2s and one-sided p-values for the full sample period 1901-2015 with the window size varying from 30-60 years. (c)-(d) show the analogous result for the post WWII period 1947-2015 with the window size varying from 20-30 years. The purple, red, and yellow dashed lines stand for the MY, 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).

Figure 10: Conditionally Recursive Forecasts, 5 Year Ahead, IVX Estimation



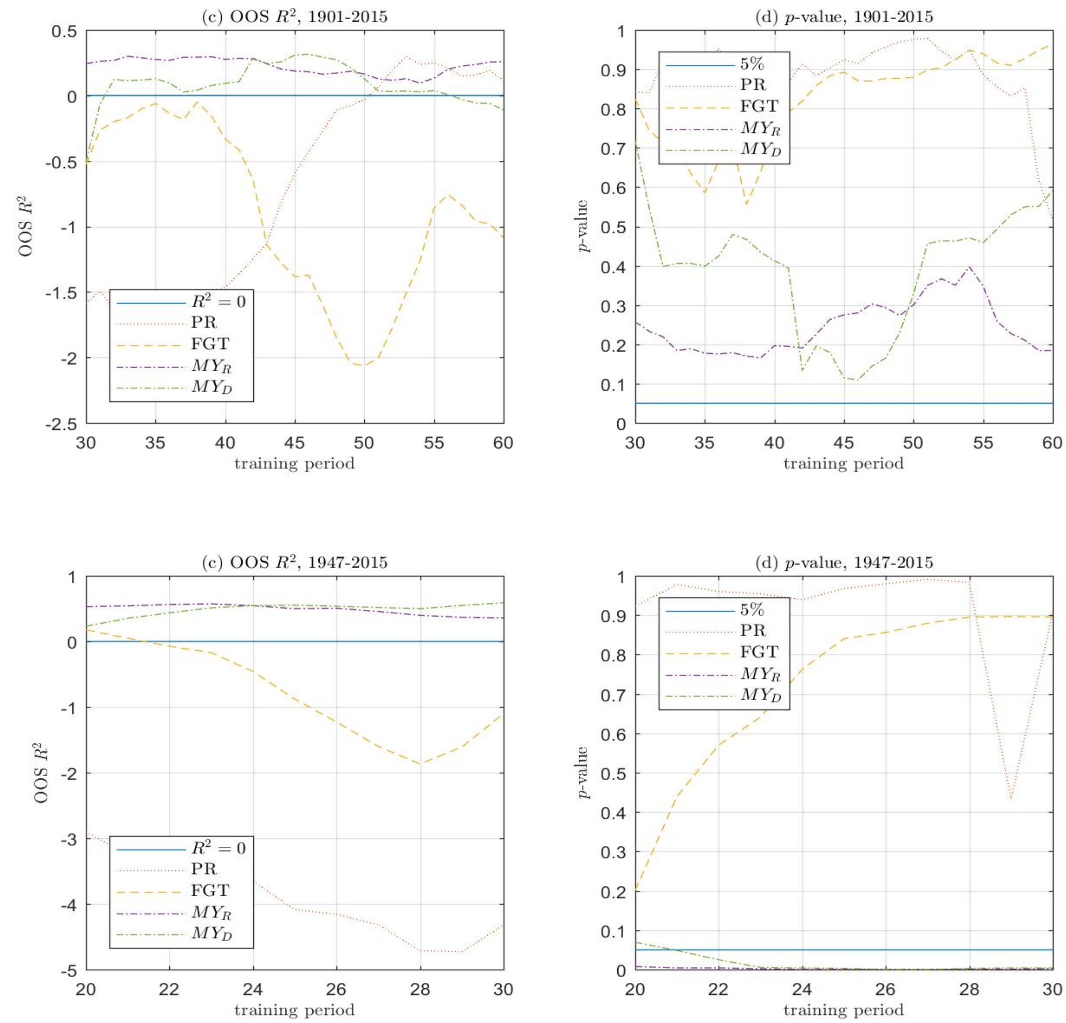
This figure provides IVX estimates of 5 year ahead out of sample forecasting results with respect to different window size based on the recursive method. (a)-(b) show the OOS-R^2s and one-sided p-values for the full sample period 1901-2015 with the window size varying from 30-60 years. (c)-(d) show the 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).

Figure 11: Conditionally Rolling Forecasts, 1 Year Ahead, IVX Estimation



This figure provides IVX estimates of 1 year ahead out of sample forecasting results with respect to different window size based on the rolling method. (a)-(b) show the OOS-R^2s and p-values 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 purple, red, and yellow dashed lines stand for the MY, 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).

Figure 12: Conditionally Rolling Forecasts, 5 Year Ahead, IVX Estimation



This figure provides IVX estimates of 5 year ahead out of sample forecasting results with respect to different window size based on the rolling method. (a)-(b) show the OOS-R^2s and p-values 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).