Demographics, Stock Market Flows, and Stock Returns

Amit Goyal*

Abstract

This paper studies the link between population age structure, net outflows (dividends plus repurchases less net issues) from the stock market, and stock market returns in an overlapping generations framework. I find support for the traditional lifecycle models—the outflows from the stock market are positively correlated with the changes in the fraction of old people (65 and over) and negatively correlated with the changes in the fraction of middle-aged people (45 to 64). Changes in population age structure also add significant explanatory power to equity premium regressions. The population structure adds to the predictive power of regressions involving the investment/savings rate for the U.S. economy. Finally, international demographic changes have some power in explaining international capital flows.

I. Introduction

The importance of demographic changes to a country’s economic health needs no exaggeration. An aging population’s potential effects on social security, labor supply and employment, levels of aggregate consumption, and saving are of interest not only to academics but also to policy analysts. While actuaries have long recognized the need for factoring demographic characteristics into payment schedules, less is known, or widely accepted, about the effect of demographics on asset markets. This research takes on an added dimension in the U.S. because of the importance of the Baby Boomer generation. The aging of the U.S. population could cause significant changes in pension plans, social security, and Medicare programs.

Given the fundamental importance of the relation between savings, portfolio allocation, and lifecycle, it is not surprising that there is a rich literature on this topic. Friedman (1957) introduced the idea of permanent income, while

---

*amit.goyal@bus.emory.edu, Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, GA 30322. This paper is a revised version of the second chapter of my dissertation at the Anderson School of Management at UCLA. I thank Zhiwu Chen (associate editor and referee), Matthias Kahl, Jonathan Karpoff (the editor), Richard Roll, and Ivo Welch for feedback, support, and encouragement. Antonio Bernardo, Tyrone Callahan, Bhagwan Chowdhry, Mark Grinblatt, Jan Jindra, Robert Michaud, Pedro Santa-Clara, Avanidhar Subrahmanyan, Walter Torous, and seminar participants at the 2001 WFA conference also gave helpful suggestions. I retain the credit for any remaining errors.
the lifecycle theory of savings was pioneered by Modigliani (see Modigliani and Brumberg (1954) and Ando and Modigliani (1963)). The empirical evidence (see Modigliani (1986)) supports the essential idea that consumers want to smooth consumption over time so as to maximize their lifetime utility and, thus, have a hump-shaped saving pattern.\footnote{See also Deaton (1992) for a comprehensive review of the literature analyzing various theories of consumption using aggregate macroeconomic data.}

The intuition of the model I wish to test is derived from basic lifecycle theories. Households invest different amounts of money at different life stages: young households do not invest much because they do not own much financial wealth; middle-aged households are the most prominent investors; and old households again invest less because they have less time to enjoy the fruits of their investments. The increase in savings is a consequence of both the increasing wealth of the middle-aged consumer and the desire to save for old age when income-generating possibilities are few. I then take this solution of the optimization problem for a single consumer and aggregate it through an overlapping generations (OLG) framework to an economy consisting of four generations: Infants, Youngs, Middles, and Olds. I consider more than two generations to analyze the differences in portfolio allocations across various age groups. This aggregation leads to the natural conclusion that as the proportion of the middle-aged population increases, the aggregate savings in the economy increase. In a similar vein, an increase in proportion of the old-aged population is accompanied by a decline in aggregate savings. The inflows (net new investment) into the stock market are, therefore, directly related to \textit{changes} in the demographic structure of the economy. This aspect of changes vs. levels is often ignored in the empirical literature, which sometimes finds insignificant relations between demographics and stock market variables.\footnote{For instance, Poterba (2001) finds an insignificant relation between the fraction of population in different age groups and stock returns. Ang and Maddaloni (2001) find that average age does not have much predictive power in explaining international excess stock returns.}

Motivated by the anticipated impact of the retirement of the baby boomers, researchers have conducted numerous studies on investors' portfolio allocation based on the lifecycle model.\footnote{The National Bureau of Economic Research has a separate program on the economics of aging. See various volumes edited by David Wise (2001) consisting of papers presented as part this project.} Bergantino (1998) and Poterba and Samwick (1997) use the Federal Reserve Board's Surveys of Consumer Finances data, while Ameriks and Zeldes (2000) use data from TIAA-CREF. Schieber and Shoven (1994) examine the impact of the aging U.S. population on the funded private pension system. However, it is surprising that no study has looked specifically at aggregate investment flows and demographics. In this paper, I directly model the flows into the stock market and then test the predictions of the model using U.S. data.

The effect of demographic changes on prices in my setup also derives from the basic lifecycle models. An increase in the middle-aged population leads to an increase in demand, driving up the prices and returns in the short term. At the same time, an absence of change in fundamentals leads to a decline in longer horizon returns. The effect of an increase in the old-aged population is analogous in the opposite direction. Similar to studies on portfolio allocation, there has been...
a lively debate about the impact of baby boomers on stock prices. For instance, the popular press has often cited the baby boomers’ entry into the saving years as a possible factor contributing to the 1990s rise of the stock market. A natural corollary to this argument is the prospect of falling asset prices when the baby boomers reach the retirement years. The impact of demographics on stock returns received renewed attention after Bakshi and Chen’s (1994) influential study. Bakshi and Chen analyze two hypotheses: the lifecycle investment hypothesis and the lifecycle risk aversion hypothesis in an Euler equation framework. They find that a rise in average age predicts a rise in the equity premium. One of the issues raised by this study is whether average age adequately summarizes all the age structure information. Poterba (2001) does not find any significant historical relation between demographic structure and real returns on Treasury bills, long-term government bonds, or corporate stocks and posits that this is due to the limited power of the statistical tests. On the other hand, in an international context, Ang and Maddaloni (2001) and Erb, Harvey, and Viskanta (1997) find a positive relation between population structure and real stock returns. While these studies implicitly have the intuition that changing demand for financial assets affects stock returns, this intuition is not directly tested in these studies. In contrast, I explicitly study the relation between the demand for financial assets and demographics.

The empirical results of this paper are summarized as follows. I find support for the traditional lifecycle models that predict a link between age and investments. Specifically, I find that outflows from the stock market are positively correlated with changes in the fraction of old people (65 and over) and negatively correlated with changes in the fraction of middle-aged people (45 to 64). The population age structure also adds significant explanatory power to excess stock return predictability regressions with the $R^2$ rising to over 18%. Moreover, empirical data lend support to the intuition that an increase in the number of middle-aged people leads to an increase in prices, and a decrease in the number of old-aged people leads to a decline in prices. Forecasts using a vector autoregression (VAR) approach predict an increase in outflows from the stock market over the next 25 years to provide cash returns to retiring baby boomers. Not only are the outflows projected to increase over the next 25 years, but they also remain at high levels for almost a decade. At the same time, the average predicted outflows from the market for the next 52 years are almost the same as they have been for the past 70 years. The reason for this is twofold. First, the outflows were historically at very low levels in the 1990s (in fact, the 1990s are characterized by inflows of funds) and it takes a decade for the outflows to come back to mean levels. Second, the outflows are predicted to decline around the middle of this century because the fraction of middle-aged people (ages 45 to 64) is projected to increase by 2040. On an aggregate macroeconomic level, preliminary results indicate that the age structure has some explanatory power in predicting the investment/saving rate for the U.S. economy.

There is also reason to expect that international capital flows should be dependent on international demographics. For instance, an increase in the middle-

---

4See, for example, *Economist* (Nov. 1, 2001), Farrell and Dunkin (1998), and Targett (2000). A similar argument was made by Mankiw and Weil (1989) to explain the rise of real estate prices in the early 1980s.
aged population of a country should not only increase the investment in that country, but also should increase international investment by the residents of that country. The literature on international capital flows has approached the issue from either the perspective of portfolio diversification (see Brennan and Cao (1997), Froot, O’Connell, and Seasholes (2001), Kang and Stulz (1997), and Tesar and Werner (1995)) or the real economy aspects (see Helpman (1998), Leamer and Levinsohn (1995), and Obstfeld and Rogoff (1996)). Based on these two approaches, Portes and Rey (1999) use either stock returns or macroeconomic variables to explain international capital flows. In this paper, I add demographic changes to these tests and find that changes in the population structure of a country add to the explanatory power of regressions predicting international capital flows.

The rest of the article is organized as follows. Section II discusses the basic lifecycle model in an OLG framework and outlines the hypothesis I wish to test. Section III discusses data calculations and includes a brief discussion of some of the more noteworthy features of the data. The main results are presented in Section IV. Concluding remarks are in Section V.

II. Model

In this section, I use an OLG model to analyze the effect of changing demographic structure on the investment flows in and out of asset markets, and on asset prices. I first analyze the intertemporal consumption-investment problem of a single consumer in the economy. The solution to the dynamic programming problem of a single consumer is then aggregated in an OLG framework. The analysis in Section II.A is a partial equilibrium analysis—the returns distribution is specified exogenously. The objective is to study how the changes in population structure affect the net flows in and out of the asset market. Section II.B then endogenizes the price of the risky asset in the economy. The objective there is to see how the changing population (and the expectation of changing population) affects current stock prices.

The population consists of four generations: Infants, Youngs, Middles, and Olds. All except Infants are active participants in the economy. The Youngs have three, Middles have two, and Olds have just one more period to live. The Youngs receive incomes of \( Y_c, Y_m, Y_o \) sequentially during their lifetime, the Middles receive \( Y_m, Y_o \), while the Olds receive \( Y_o \). The income of the Olds is not restricted to zero and can be thought of as their pension income. The initial wealth of the three active generations is \( W_c(\text{Youngs}), W_m(\text{Middles}), \) and \( W_o(\text{Olds}) \). The initial wealth of the Youngs is not restricted to zero and can be thought of as part of the leftover wealth that the Infants did not consume. The wealth of the different generations is, of course, related. For instance, the wealth of the Olds at time \( t \) is the return on

---

5Abel (1999) considers an OLG model with two generations and log utility to analyze the effects of the baby boom on stock prices in the presence of social security. See also Brooks (2000), Constantinides, Donaldson, and Mehra (2002), Gourinchas and Parker (2002), Luo (2000), and Storesletten, Telmer, and Yaron (2001). However, OLG models with more than one generation and more than one period to live quickly become intractable. Although numerical techniques are available for solutions to these models (see Auerbach and Kotlikoff (1987)), these routines rely on further simplifying assumptions such as no risky asset or log utility or only two generations.
the investments made by the Middles at time \( t - 1 \). Section II.C presents the exact dynamics of wealth transfer. There is no growth in total population and the total population is normalized to one. The fraction of people belonging to the different generations changes over time. The fraction of the Infants, the Youngs, the Middles, and the Olds at time \( t \) is assumed to be \( p_{i,t}, p_{y,t}, p_{m,t}, \) and \( p_{o,t} \), respectively. There are two assets in the economy: one risk-free asset with gross return denoted by \( R_f \) and another risky stock with gross return denoted by \( R_m \).

### A. Asset Market Flows

Consider an investor living for \( T \) periods, who maximizes power utility over wealth \( W_T \) at time \( T \) and consumption \( C_t \) at time \( t \) given by

\[
U(C_0, C_1, C_2, \ldots, C_{T-1}, W_T) = \sum_{t=0}^{T-1} \delta^t \frac{C_t^{1-\gamma}}{1-\gamma} + \delta^T \frac{W_T^{1-\gamma}}{1-\gamma},
\]

where \( \delta \) is the rate of time preference and \( \gamma \) is the coefficient of relative risk aversion.

Let \( W_t \) be the wealth at the beginning of the period \( t \) and \( Y_t \) be the income received at the beginning of period \( t \). It should be noted at the outset that analytical solutions to the general consumption-investment problem in the presence of labor income are unknown.\(^6\) I present an approximate solution to this problem with non-stochastic labor income in this section. Let \( \alpha \) be the fraction of wealth invested in the risky asset. Then Proposition 1 follows (all proofs are given in Appendix A).

**Proposition 1.** The approximate solution to problem (1) is given by

\[
C_t = k_t \left( W_t + \sum_{i=1}^{T-1} \frac{Y_i}{R_f^{t-i}} \right), \quad 0 = E \left[ R_\alpha^{-\gamma} (R_m - R_f) \right] \text{ gives } \alpha
\]

where \( R_\alpha = \alpha(R_m - R_f) + R_f \), \( \rho = (\delta E[R_f^{1-\gamma}])^{1/\gamma} \), and \( k_t = (1 - \rho)/(1 - \rho^{T-t+1}) \).

The investor consumes a constant fraction of her total wealth, which includes the financial wealth and the discounted value of future labor income. This conclusion is reminiscent of the results in Bodie, Merton, and Samuelson (1992). The constant of proportionality in the consumption-wealth ratio increases over time. The fraction of wealth invested in the risky asset is given by the standard first-order conditions. Aggregating this solution to individual optimization in an OLG framework to obtain the demand for the risky asset yields the table,

<table>
<thead>
<tr>
<th>Periods to Live</th>
<th>Income</th>
<th>Current Wealth</th>
<th>Amount Invested in Risky Asset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Youngs</td>
<td>3</td>
<td>( Y_y, Y_y, Y_o )</td>
<td>( W_y ) ( \alpha ) ((W_y + Y_y)(1 - k_y) - k_y(Y_m/R_f + Y_o/R_f^2))</td>
</tr>
<tr>
<td>Middles</td>
<td>2</td>
<td>( Y_m, Y_o )</td>
<td>( W_m ) ( \alpha ) ((W_m + Y_m)(1 - k_m) - k_m Y_o/R_f )</td>
</tr>
<tr>
<td>Olds</td>
<td>1</td>
<td>( Y_o )</td>
<td>( W_o ) ( \alpha ) ((W_o + Y_o)(1 - k_o))</td>
</tr>
</tbody>
</table>

where \( k_y = (1 - \rho)/(1 - \rho^3) \), \( k_m = (1 - \rho)/(1 - \rho^3) \) and \( k_o = (1 - \rho)/(1 - \rho^2) \).

\(^6\)However, see Bodie, Merton, and Samuelson (1992), Heaton and Lucas (2000), and Viceira (2001) for analysis of this problem with labor income uncertainty under other restrictions.
I can now study how the changes in population structure affect the net flows in and out of the asset market. Intuitively, an increase in the size of the middle-aged population \((p_{m,i})\) leads to an increase in the overall investment into the stock market because the middle-aged investor has higher wealth (including capitalized labor income) than either the young-aged investor, who has little financial wealth, or the old-aged investor, who has little labor income wealth. Similarly, an increase in the size of the old-aged population \((p_{o,i})\) leads to a decrease in the overall investment as these people start to draw upon their earlier investments to finance their consumption. This is summarized in the following implication (see Appendix A for proofs).

**Implication 1.** The effect of an increase in \(p_{o,i}\) is to increase the leakage (net investment out of the stock market) at time \(t\). The effect of an increase in \(p_{m,i}\) is to decrease the leakage at time \(t\).

**B. Endogenous Prices and Changing Population**

In this subsection, the risky asset is assumed to have an exogenous dividend stream. The prices are then endogenously determined by market clearing. On an intuitive level, market clearing in this subsection takes place by adjustment of prices while in the previous subsection, markets cleared through the adjustment of supply to changing demand. I model the firms using an approach developed by Leland (1992).\(^7\)

I assume that there is a firm that supplies shares to the market. These shares promise an identical future value of \(P_{t+1}\) per share, where \(P_{t+1} = \bar{P} + \epsilon_{t+1}\). In supplying shares to the market, this firm faces costs, which can be thought of as real investment costs required to provide the return on the shares, given by \(C(q) = cq^2\), where \(C(q)\) is the cost of providing \(q\) number of shares and \(c\) is a constant. I assume, for simplicity, that the firm issues shares to optimize profits over the next period only.\(^8\) It issues shares to maximize \(\pi_t(q_t) = P_tq_t - C(q_t)\) implying an optimal share supply of

\[
q_t = zP_t,
\]

where \(z\) is a constant. Given these prices, the return over the interval \([t, t + 1]\) is given by \(r_{t+1} = P_{t+1}/P_t - 1\). My assumption of a one-period-lived firm implies that the returns over \([t, t + 1]\) are dependent only on the current price \(P_t\). In other words,

\[
E_t[r_{t+1}] = \bar{P}/P_t - 1; \quad \text{var}_t(r_{t+1}) = \text{var}(\epsilon)/P_t^2.
\]

The return-generating process derived above determines the optimal consumption and investment decisions of various generations. I have derived the process for returns only over the next period (since my firm is short-lived). I assume

\(^7\)The central issue in the complete characterization of the solution is the expectations in equation (2). Since the prices are endogenously determined, the return distribution is also determined endogenously through market clearing. The absence of a functional form for prices makes it impossible to obtain an analytical solution and leads to non-trivial issues in numerical solutions.

\(^8\)The introduction of a long-lived firm would lead me to solve the intertemporal problem of maximizing profits over the remaining horizon. This would distract from my primary focus on consumers rather than producers.
that all subsequent returns are exogenously given. In other words, the returns over 
\([t, t+1]\) are given by equation (4) and returns over subsequent periods are exoge-
nously specified to be normally distributed with constant mean and variance. The 
effect of this change in process for next period returns is to make the parameters 
for the fraction (k) of wealth consumed and the fraction (\(\alpha\)) of wealth invested in 
the risky asset functions of the current price \(P_t\). In short, the demand for financial 
assets is given by

\[
(5) \quad \text{Dollar Demand}_t = \]
\[+p_{y,t} (W_{y,t} + Y_{x,t})(1 - k_{y}(P_t)) - k_{y}(P_t) \left( Y_m/R + Y_o/R^2 \right) \alpha(P_t)\]
\[+ p_{m,t} (W_{m,t} + Y_{m,t})(1 - k_{m}(P_t)) - k_{m}(P_t) Y_o/R \alpha(P_t)\]
\[+ p_{o,t} (W_{o,t} + Y_{o,t})(1 - k_{o}(P_t)) \alpha(P_t).\]

Clearing the market for the risky asset by equating equations (3) and (5),

\[
(6) \quad q_t P_t = \text{Dollar Demand}_t,
\]

which is the non-linear equation for \(P_t\), that can be solved numerically. Consider 
the effect of increasing the fraction of the middle-aged population. There are 
two opposing effects. On the one hand, an increase in the demand for financial 
assets due to a larger number of Middles should lead to an increase in price as the 
firm supplies more shares in response to an increasing price. On the other hand, 
an increase in price today results in lower expected returns in the future. This 
adverse change in the return distribution of the risky asset lowers demand for it. 
The net effect is ambiguous. I undertake a full empirical exploration in a later 
section. However, benchmark calculations suggest that:

Implication 2. Returns fall due to an increase in the number of Olds and rise due 
to an increase in the number of Middles.

Note that this exercise only studies the impact of changing population on 
next period returns. In the absence of any change in fundamentals, it is also rea-
sonable to expect that an increase in prices today is accompanied by a fall in future 
returns and vice versa. It is also interesting to explore whether there is any relation 
between macroeconomic variables and the changing structure of population, 
even though I did not formally model the real economy. Finally, there is reason 
to expect that the international capital flows should be dependent on the interna-
tional demographics. For instance, an increase in the middle-aged population of 
a country should not only increase the investment in that country but also should 
increase in international investment by that country’s residents. To summarize, I 
wish to test the following hypotheses in the empirical data.

i) Does an increase in the fraction of middle-aged people lead to an increase in 
the inflows into the stock market, and does an increase in the fraction of old-aged 
people lead to an increase in the outflows from the stock market?

ii) Does an increase in the fraction of middle-aged people lead to an increase in 
prices today implying higher stock returns today, and does an increase in the 
fraction of old-aged people lead to a decrease in prices today implying lower stock 
returns today?
iii) Does an increase in the fraction of middle-aged people lead to a decrease in longer horizon stock returns, and does an increase in the fraction of old-aged people lead to an increase in longer horizon stock returns?

iv) Does a changing population structure have an impact on the savings rate, investment rate, and growth rate of the aggregate economy?

v) Does the changing demographic structure impact cross-country equity capital flows?

III. Data

The stock return data used in the paper is obtained from S&P. SPR is the stock return (including dividends) on the S&P500 index. SPDP is the dividend price ratio of the S&P500 index. T-bill rates, TBILL, are obtained from Ibbotson. The excess stock return, ESRET, is the difference between (logged) stock return and the (logged) T-bill rate.

The total outflows from the stock market are calculated as follows using CRSP data. For each stock \( i \), net outflow during month \( t \) is calculated as

\[
\text{Net Outflow}_{it} = -\text{Market Cap}_{it} + \text{Market Cap}_{it-1} \times (1 + \text{ret}_{it}),
\]

where Market Cap\(_i\) is the market capitalization at the end of month \( t \) and \( \text{ret}_{it} \) is the return during month \( t \). For new issues in month \( t \), I have only the first term in the equation while for delisting I have just the second term in the equation without \( \text{ret}_{it} \). Total outflows in month \( t \) are the sum of outflows of all the stocks that have data on CRSP for month \( t \),

\[
\text{Net Outflow}_t = \sum_{i=1}^{N_t} \text{Net Outflow}_{it},
\]

where \( N_t \) is the number of stocks in month \( t \) on CRSP. An equivalent way of representing the above equation in terms of aggregates is

\[
\text{Net Outflow}_t = -\text{Totval}_t + \text{Totval}_{t-1} \times (1 + \text{vwret}_t) - \text{Del}_t \times \text{vwret}_t,
\]

where \( \text{vwret}_t \) is the value-weighted return on the CRSP index, Totval is the total stock market capitalization of the CRSP index, and \( \text{Del}_t \) is the total dollar value of delistings. The LEAK variable used in the regressions is calculated as total outflows during the year deflated by last year’s market capitalization, i.e., \( \text{LEAK}_t = \frac{\text{Net Outflow}_t}{\text{Market Cap}_{t-1}} \). Total (dollar) dividends paid during the month \( t \) are calculated by using the equation,

\[
\text{Dividends}_t = \sum_{i=1}^{N_t} \text{Dividends}_{it},
\]

Dividends (net outflows) paid during the year are calculated by adding the dividends (net outflows) paid during the past 12 months. The difference between total outflows and dividends gives the net repurchases during the year.

---

9Note that the new issues calculated in the above way have to account for the fact that AMEX was added to the CRSP tapes in July 1962 and NASDAQ was added in December 1972.
Table 1 offers descriptive statistics on the main variables. The average ES-RET was 5.9% per annum with a standard deviation of about 19.1%. The median ESRET is substantially higher at 7.7%. The primary new variable introduced in this paper is the net outflows from the market (LEAK). The average LEAK was 1.4% with a standard deviation of 3.2%. Since the average dividends during this period are 4.2%, this implies that there is a net issuing activity of approximately 2.8%. Table 1, panel B is the contemporaneous correlation matrix of the variables. LEAK is only weakly correlated with the ESRET (correlation 0.02 on annual frequency) but has a correlation of 0.28 with SPDP.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel A. Descriptive Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>Std. Dev.</strong></td>
</tr>
<tr>
<td>SPR</td>
<td>12.93</td>
</tr>
<tr>
<td>SPDP</td>
<td>-3.21</td>
</tr>
<tr>
<td>TBILL</td>
<td>4.73</td>
</tr>
<tr>
<td>ESRET</td>
<td>5.93</td>
</tr>
<tr>
<td>LEAK</td>
<td>1.38</td>
</tr>
<tr>
<td>ΔAGNP</td>
<td>3.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Cross-Correlations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPDP</strong></td>
<td><strong>TBILL</strong></td>
</tr>
<tr>
<td>SPR</td>
<td>-0.426</td>
</tr>
<tr>
<td>SPDP</td>
<td>-0.193</td>
</tr>
<tr>
<td>TBILL</td>
<td>0.019</td>
</tr>
<tr>
<td>ESRET</td>
<td>-0.019</td>
</tr>
<tr>
<td>LEAK</td>
<td>0.362</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Autocorrelations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lag</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td>SPR</td>
<td>0.007</td>
</tr>
<tr>
<td>SPDP</td>
<td>0.741</td>
</tr>
<tr>
<td>TBILL</td>
<td>0.866</td>
</tr>
<tr>
<td>ESRET</td>
<td>0.067</td>
</tr>
<tr>
<td>LEAK</td>
<td>0.569</td>
</tr>
<tr>
<td>ΔAGNP</td>
<td>0.568</td>
</tr>
</tbody>
</table>

All series are measured in percent, except for the SPDP series, which is the logarithm dividend price ratio of the S&P500 index. SPR is the stock return (including dividends) on the S&P500 index. TBILL is the risk-free rate. ESRET is the log difference between SPR and TBILL. LEAK is the total outflows from the stock market deflated by the previous year's market capitalization. ΔAGNP is the growth rate in real GNP. Each mean and median is significantly different from zero at the 1% level. JqBr is the Jarque-Bera test for normality. For 73 observations, the critical level to reject normality is 5.99 at the 95% level, 9.21 at the 99% level. The correlations in panel B are contemporaneous correlations. There are 73 annual observations in the sample from 1926 to 1998.

Table 2 presents a breakup of outflows from the stock market into dividends and net issues. A negative number for net issues implies that there was a net repurchasing activity during the year. A negative number for net outflows implies that the net issuing activity exceeded dividends paid out, or in other words, there was a net inflow into the stock market. The years 1929 and 1930 were characterized by flow of funds into the stock market of sizeable proportions (11.5% in 1929 and 4.5% in 1930), although the net issuing activity did drop off in later years. The second big 20th century crash in 1987 was followed by a huge fall in the net issuing activity (the years 1988 to 1990 were characterized by a net repurchasing activity). Table 2 also shows that the 1990s were characterized by a huge inflow of funds into the stock market. Average inflow in the 1990s was $19 billion and
there was a record inflow of $138 billion in 1993. In fact, the 1990s has been the only decade with an average inflow into the stock market.

<table>
<thead>
<tr>
<th>Year</th>
<th>Market Cap.</th>
<th>Divid.</th>
<th>Net Issues</th>
<th>Total Outflows</th>
</tr>
</thead>
<tbody>
<tr>
<td>1926</td>
<td>266.1</td>
<td>12.7</td>
<td>20.9</td>
<td>-8.2</td>
</tr>
<tr>
<td>1927</td>
<td>373.6</td>
<td>15.5</td>
<td>33.6</td>
<td>-18.1</td>
</tr>
<tr>
<td>1928</td>
<td>504.7</td>
<td>16.6</td>
<td>28.5</td>
<td>-11.9</td>
</tr>
<tr>
<td>1929</td>
<td>471.2</td>
<td>18.1</td>
<td>78.7</td>
<td>-60.6</td>
</tr>
<tr>
<td>1930</td>
<td>322.6</td>
<td>19.8</td>
<td>49.4</td>
<td>-29.6</td>
</tr>
<tr>
<td>1931</td>
<td>186.7</td>
<td>17.8</td>
<td>5.6</td>
<td>12.2</td>
</tr>
<tr>
<td>1932</td>
<td>182.8</td>
<td>12.2</td>
<td>6.4</td>
<td>5.8</td>
</tr>
<tr>
<td>1933</td>
<td>287.0</td>
<td>10.0</td>
<td>6.2</td>
<td>3.9</td>
</tr>
<tr>
<td>1934</td>
<td>277.7</td>
<td>11.4</td>
<td>7.9</td>
<td>3.5</td>
</tr>
<tr>
<td>1935</td>
<td>382.7</td>
<td>12.6</td>
<td>8.0</td>
<td>4.6</td>
</tr>
<tr>
<td>1936</td>
<td>489.3</td>
<td>19.5</td>
<td>14.4</td>
<td>5.2</td>
</tr>
<tr>
<td>1937</td>
<td>294.6</td>
<td>21.0</td>
<td>10.7</td>
<td>10.3</td>
</tr>
<tr>
<td>1938</td>
<td>377.4</td>
<td>14.0</td>
<td>7.5</td>
<td>6.5</td>
</tr>
<tr>
<td>1939</td>
<td>373.0</td>
<td>16.9</td>
<td>5.1</td>
<td>11.8</td>
</tr>
<tr>
<td>1940</td>
<td>324.3</td>
<td>19.0</td>
<td>3.9</td>
<td>15.0</td>
</tr>
<tr>
<td>1941</td>
<td>259.0</td>
<td>19.7</td>
<td>3.3</td>
<td>16.4</td>
</tr>
<tr>
<td>1942</td>
<td>262.2</td>
<td>16.1</td>
<td>0.6</td>
<td>15.4</td>
</tr>
<tr>
<td>1943</td>
<td>306.6</td>
<td>15.5</td>
<td>6.2</td>
<td>9.3</td>
</tr>
<tr>
<td>1944</td>
<td>350.0</td>
<td>16.3</td>
<td>3.2</td>
<td>13.1</td>
</tr>
<tr>
<td>1945</td>
<td>458.7</td>
<td>16.4</td>
<td>5.9</td>
<td>10.5</td>
</tr>
<tr>
<td>1946</td>
<td>390.4</td>
<td>16.8</td>
<td>23.3</td>
<td>-6.5</td>
</tr>
<tr>
<td>1947</td>
<td>354.0</td>
<td>19.2</td>
<td>10.6</td>
<td>8.6</td>
</tr>
<tr>
<td>1948</td>
<td>326.2</td>
<td>21.1</td>
<td>4.9</td>
<td>16.2</td>
</tr>
<tr>
<td>1949</td>
<td>377.0</td>
<td>23.1</td>
<td>13.5</td>
<td>9.6</td>
</tr>
<tr>
<td>1950</td>
<td>467.7</td>
<td>29.1</td>
<td>14.7</td>
<td>14.4</td>
</tr>
<tr>
<td>1951</td>
<td>517.8</td>
<td>27.4</td>
<td>18.6</td>
<td>8.8</td>
</tr>
<tr>
<td>1952</td>
<td>561.6</td>
<td>28.2</td>
<td>16.2</td>
<td>12.9</td>
</tr>
<tr>
<td>1953</td>
<td>539.8</td>
<td>29.2</td>
<td>13.4</td>
<td>15.9</td>
</tr>
<tr>
<td>1954</td>
<td>785.4</td>
<td>31.3</td>
<td>20.6</td>
<td>10.7</td>
</tr>
<tr>
<td>1955</td>
<td>954.3</td>
<td>35.5</td>
<td>25.3</td>
<td>10.2</td>
</tr>
<tr>
<td>1956</td>
<td>966.7</td>
<td>38.1</td>
<td>36.7</td>
<td>1.4</td>
</tr>
<tr>
<td>1957</td>
<td>654.9</td>
<td>39.5</td>
<td>29.6</td>
<td>9.9</td>
</tr>
<tr>
<td>1958</td>
<td>1190.4</td>
<td>38.1</td>
<td>29.0</td>
<td>9.1</td>
</tr>
<tr>
<td>1959</td>
<td>1319.5</td>
<td>40.7</td>
<td>31.2</td>
<td>9.5</td>
</tr>
<tr>
<td>1960</td>
<td>2582.4</td>
<td>42.3</td>
<td>30.4</td>
<td>11.7</td>
</tr>
<tr>
<td>1961</td>
<td>1622.4</td>
<td>44.1</td>
<td>32.0</td>
<td>8.0</td>
</tr>
<tr>
<td>1962</td>
<td>1509.9</td>
<td>48.4</td>
<td>20.8</td>
<td>27.7</td>
</tr>
</tbody>
</table>

Data are from the CRSP tapes. Market Cap is the total stock market capitalization at the end of the year. Divid is the dividends paid during the year. Net Issues is the net issue activity during the year (a negative value indicates net repurchasing activity during the year). Total outflow is the difference between Divid and Net Issues (a negative value implies that net issuing activity exceeded the dividends paid out, or in other words, there was a net inflow into the stock market). In particular, the net outflow in year t is calculated using the following formula:

\[
\text{Net Outflow}_t = -\sum_{i=1}^{t-1} (\text{Total}_{i,\text{Cap}} + \text{Total}_{i-1,\text{Divid}} + (1 + \text{vwt}_{i}) - \text{Del}_{i} \cdot \text{vwt}_{i+1})
\]

where vwt is the value-weighted return on the CRSP index. Totalv is the total stock market capitalization of the CRSP index, and Del is the dollar value of delistings. The nominal data are converted to real dollars by using the implicit price deflator for GNP obtained from the National Income and Product Accounts of the United States. All figures are in billions of 1992 dollars.

The population data are collected from two sources. Data from 1900 to 1945 are collected from Historical Statistics of the United States, Colonial Times to 1970: A Statistical Abstract (U.S. Bureau of the Census (1975)). Data in the Historical Statistical Abstract are, however, available only in age groups over 10-year intervals (age groups 0-4, 5-14, . . ., 55-64, 65 and over). Fortunately, various issues of the Statistical Abstract (U.S. Bureau of the Census (various years)) give population data every decade at more refined age group levels. A simple interpolation is then used in the intervening years to calculate population numbers in the refined age groups for all years from 1900 to 1945. Citibase provides population
data (again derived from the Statistical Abstract) from 1946 onward in age groups of five-year intervals (0–4, 5–9, 10–14, . . . , 70–74, 75 and over). Population estimates from 1999 onward are based on Census Bureau projections.

Table 3 presents a few summary statistics on the U.S. population demographic structure from 1900 to 2050. A few remarkable, albeit well-known, features of the demographic structure are evident from this table. The average adult age (all people above age 25) is also expected to increase by four years over the current value of almost 49 years. The fraction of people 45–64 shows a substantial variation. It showed an almost secular increase till the 1970s when it started to decline a little. Currently, almost 22% of the population is in this “savings age group,” called the Middles in the model in Section II. Current forecasts project this fraction to increase to 27% around 2010 before falling again to 22%. The most interesting feature of these changes is the extraordinary increase in the number of old-aged people (over 65). The fraction of people belonging to this age group increases secularly till 2040 when it begins to level off. The magnitude of this increase can be gauged by the fact that the projected number of old people (65 and over) in the year 2050 is more than the entire U.S. population at the beginning of the 20th century.

### TABLE 3
Demographic Statistics Summary (1900–2050)

#### Panel A. End of Decade Numbers

<table>
<thead>
<tr>
<th>Year</th>
<th>25–44</th>
<th>45–64</th>
<th>65+</th>
<th>Avg. Age of Persons over 25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P25–44</td>
<td>P45–64</td>
<td>P65+</td>
<td>A25–64</td>
</tr>
<tr>
<td>1900</td>
<td>26.2</td>
<td>13.6</td>
<td>4.1</td>
<td>42.9</td>
</tr>
<tr>
<td>1910</td>
<td>29.3</td>
<td>14.7</td>
<td>4.3</td>
<td>43.0</td>
</tr>
<tr>
<td>1920</td>
<td>29.9</td>
<td>16.1</td>
<td>4.6</td>
<td>43.5</td>
</tr>
<tr>
<td>1930</td>
<td>29.5</td>
<td>17.5</td>
<td>5.4</td>
<td>44.5</td>
</tr>
<tr>
<td>1940</td>
<td>30.2</td>
<td>19.9</td>
<td>6.8</td>
<td>45.4</td>
</tr>
<tr>
<td>1950</td>
<td>30.0</td>
<td>20.3</td>
<td>8.1</td>
<td>46.3</td>
</tr>
<tr>
<td>1960</td>
<td>26.1</td>
<td>20.0</td>
<td>9.2</td>
<td>47.9</td>
</tr>
<tr>
<td>1970</td>
<td>23.6</td>
<td>20.5</td>
<td>9.8</td>
<td>48.6</td>
</tr>
<tr>
<td>1980</td>
<td>27.5</td>
<td>19.5</td>
<td>11.3</td>
<td>48.1</td>
</tr>
<tr>
<td>1990</td>
<td>32.4</td>
<td>18.6</td>
<td>12.5</td>
<td>47.7</td>
</tr>
<tr>
<td>2000</td>
<td>29.6</td>
<td>22.2</td>
<td>12.6</td>
<td>49.0</td>
</tr>
<tr>
<td>2010</td>
<td>25.8</td>
<td>26.5</td>
<td>13.2</td>
<td>50.4</td>
</tr>
<tr>
<td>2020</td>
<td>25.6</td>
<td>24.6</td>
<td>16.5</td>
<td>51.6</td>
</tr>
<tr>
<td>2030</td>
<td>25.1</td>
<td>21.7</td>
<td>20.0</td>
<td>52.9</td>
</tr>
<tr>
<td>2040</td>
<td>24.4</td>
<td>22.0</td>
<td>20.3</td>
<td>52.4</td>
</tr>
<tr>
<td>2050</td>
<td>24.6</td>
<td>21.8</td>
<td>20.0</td>
<td>53.3</td>
</tr>
</tbody>
</table>

#### Panel B. Cross-Correlations (1926–1998)

<table>
<thead>
<tr>
<th></th>
<th>P45–64</th>
<th>P65+</th>
<th>ΔP25–44</th>
<th>ΔP45–64</th>
<th>ΔP65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>P25–44</td>
<td>-0.391</td>
<td>0.047</td>
<td>0.065</td>
<td>0.338</td>
<td>0.130</td>
</tr>
<tr>
<td>P45–64</td>
<td>0.290</td>
<td>0.023</td>
<td>-0.233</td>
<td>-0.230</td>
<td>-0.200</td>
</tr>
<tr>
<td>P65+</td>
<td>0.221</td>
<td>-0.989</td>
<td>-0.272</td>
<td>-0.754</td>
<td>0.280</td>
</tr>
<tr>
<td>ΔP25–64</td>
<td>-0.014</td>
<td>0.280</td>
<td>0.280</td>
<td>0.014</td>
<td>0.280</td>
</tr>
</tbody>
</table>

P25–44, P45–64, and P65+ are the fractions of people in the age groups 25–44, 45–64, and 65+, respectively. Δ denotes the percentage change. Average age is calculated using the midpoint in five-year age intervals and assuming the average age of persons over 75 is 80. Source of the historical data is the Statistical Abstract of the United States. All figures from 1999 onward are based on projections provided by the Census Bureau.
IV. Empirical Results

A. Outflows from the Stock Market

I first explore the relation between the changes in the proportion of population in different age groups and the outflows from the stock market. Panel A of Table 4 reports the result of regressing LEAK on variables other than the demographic variables. The first regression is just an AR(1) model on the LEAK variable. The autoregressive root of 0.593 is not high enough to justify a unit root on LEAK (augmented Dickey-Fuller statistic (with two lags) of −3.29 against a critical value of −2.90 at the 95% level). The second regression adds the lagged excess stock returns to the regression. The negative coefficient of the LEAK on ESRET might partly result from the fact that the good performance of the stock market in the previous year prompts more people to invest (recall that a negative LEAK implies a net inflow into the stock market). Table 2 shows that of the 15 times there has been a net inflow into the stock market in the last 73 years, six occurred in the 1990s, a decade characterized by exceptionally good market performance. The third regression shows that the data lend no support to the intuition that the stock market inflows increase in response to a growing economy—ΔGNP has an insignificant effect on LEAK.

Panel B of Table 4 reports the results of including the demographic variables. The relation between demographics and the stock market flows is determined by the interaction of the optimal decisions of generations that differ in terms of their income-generating possibilities, accumulated wealth, risk aversion, and the horizon. The use of a single variable such as average age of the population seems unlikely to capture these dynamics. Previous studies (Bakshi and Chen (1994), Erb, Harvey, and Viskanta (1997), and Poterba (2001)) have looked at this variable in the context of stock return regressions (which I discuss in the next subsection) with mixed results. As expected, I do not get a substantial improvement in fit with the addition of ΔA_{25+}. Moreover, the coefficient on this variable does not have the anticipated positive sign (older population withdraws more money from the stock market than younger population). The introduction of three variables measuring the percentage change in the fraction of population in the three age groups, viz. 25–44, 45–64, and 65+, not only leads to an improved fit but also all the coefficients have the sign anticipated from my model. The Middles (45–64) have a significantly negative coefficient while the Olds (65+) have an insignificantly positive coefficient. The Youngs have a positive coefficient but I am unable to conclude that the coefficient is statistically significant. Panel B also reports the results of including the levels of demographic variables as opposed to the changes thereof. Even though my model suggests that these regressions are likely to be

---

10 LEAK is the total outflows from the stock market and is inclusive of the dividend payments and net repurchases (net of issues). Since economic theory treats both these forms of payouts as essentially the same, one could make a case for using the LEAK variable instead of the SPDP as a predictor of the next period equity premium (see Fama and French (1988), (1989) for predictability evidence of dividend yield). A huge literature documents the problems inherent in using the SPDP because of its (near) non-stationarity (see, for example, Goetzmann and Jorion (1993), Hodrick (1992), and Nelson and Kim (1993)). This would strengthen the case for using LEAK as it displays no unit root properties. Regression of one period ahead ESRET on LEAK shows a strongly significant statistical relation with a t-statistic of almost 2.5 (contrast this with a t-statistic of 0.78 on the SPDP in Table 5).
## TABLE 4
### Predicting Net Outflows (1926–1998)

<table>
<thead>
<tr>
<th>Panel A. Without Demographic Variables</th>
<th>CNST</th>
<th>LEAK</th>
<th>ESRET</th>
<th>ΔGNP</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.006</td>
<td>0.593</td>
<td>(1.57)</td>
<td>(5.35)</td>
<td>36.05</td>
<td>35.13</td>
<td>2.55</td>
</tr>
<tr>
<td>2.</td>
<td>0.008</td>
<td>0.599</td>
<td>(2.48)</td>
<td>(5.92)</td>
<td>41.18</td>
<td>39.48</td>
<td>2.46</td>
</tr>
<tr>
<td>3.</td>
<td>0.006</td>
<td>0.584</td>
<td>(1.46)</td>
<td>(5.00)</td>
<td>36.01</td>
<td>35.24</td>
<td>2.57</td>
</tr>
<tr>
<td>4.</td>
<td>0.007</td>
<td>0.573</td>
<td>(2.12)</td>
<td>(5.17)</td>
<td>41.62</td>
<td>39.05</td>
<td>2.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. With Demographic Variables</th>
<th>CNST</th>
<th>LEAK</th>
<th>ESRET</th>
<th>A25+</th>
<th>P25–44</th>
<th>P45–64</th>
<th>P65+</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>S.E.</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>0.011</td>
<td>0.571</td>
<td>(3.20)</td>
<td>(5.39)</td>
<td>(1.97)</td>
<td>(2.21)</td>
<td>42.49</td>
<td>39.96</td>
<td>2.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>0.004</td>
<td>0.516</td>
<td>(0.87)</td>
<td>(4.61)</td>
<td>(1.57)</td>
<td>0.140</td>
<td>0.064</td>
<td>0.520</td>
<td>45.32</td>
<td>41.17</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.90)</td>
<td>(1.99)</td>
<td>(1.46)</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>-0.016</td>
<td>0.594</td>
<td>(-0.14)</td>
<td>(6.37)</td>
<td>(-2.21)</td>
<td>0.011</td>
<td>0.001</td>
<td>41.24</td>
<td>38.65</td>
<td>2.48</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>-0.218</td>
<td>0.486</td>
<td>(-1.75)</td>
<td>(5.42)</td>
<td>(-2.29)</td>
<td>0.212</td>
<td>0.061</td>
<td>-0.010</td>
<td>45.87</td>
<td>41.76</td>
<td>2.42</td>
</tr>
</tbody>
</table>

Variables are of annual frequency. The dependent variable, the net outflows (LEAK) from the stock market, leads the independent variables by one year. ESRET is the (logged) difference between stock returns on the S&P500 index and the risk-free rate. ΔGNP is the percentage change in real GNP. \( P_{25–44} \), \( P_{45–64} \), and \( P_{65+} \), are the fractions of people in the age groups 25–44, 45–64, and 65+, respectively. \( A_{25+} \) is the average age of population above age 25. The fifth and sixth regressions use the growth rates of the demographic variables while the last two regressions use the levels of the demographic variables. The first row of each regression output lists the coefficient and the second row lists its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. \( R^2 \), \( R^2 \), and S.E. are in %. Wald is the Wald test for the joint significance of the three demographic variables and is calculated using the Newey-West corrected variance-covariance matrix. The number in parentheses below the Wald statistic is the p-value for this statistic based on a \( \chi^2 \) (3) distribution.

LEAK has an autoregressive process. Percent changes in demographic variables add additional explanatory power to the LEAK prediction while levels of demographic variables do not explain the next period outflows. Further, the signs of the coefficients on demographic variables are in accord with the lifecycle model in Section II.

---

misspecified, the reason for conducting this test is to examine the (non-) findings of the previous literature that document no relation between these variables and the stock market variables.\(^\text{11}\) The last two regressions of panel B make it abundantly clear that the levels of demographic variables have almost zero explanatory power—a Wald test for the joint significance of the three demographic variables has a p-value of 0.19 even though the \( R^2 \) is slightly higher.

### B. Stock Market Returns

The lifecycle investment hypothesis and the lifecycle risk aversion hypothesis (see Bakshi and Chen (1994)) seem to make opposite predictions for the effect of increasing age on asset prices. The lifecycle investment hypothesis predicts that demand for financial assets increases with age (only up to retirement), which

---

\(^\text{11}\)As mentioned earlier, previous studies, such as Bakshi and Chen (1994), Erb, Harvey, and Viskanta (1997), and Poterba (2001), have looked at demographic variables like the average age or the fraction of population in different age groups in the context of stock return regressions. On the other hand, Ameriks and Zeldes (2000) and Schieber and Shoven (1994) use pension data to construct portfolio allocation profiles to examine the impact of baby boomers. I am aware of no studies that have analyzed aggregate stock market flows from the perspective of population structure.
would drive up financial prices. The lifecycle risk aversion hypothesis, on the other hand, argues that increasing age is associated with an increasing risk aversion parameter. This implies that as one grows older, the demand for financial assets declines and one requires higher returns to invest, thereby depressing financial prices. The net effect is unclear in the absence of a fully specified model. However, my numerical simulations of a reduced form model in Section II.B suggest that an increase in the number of Middles leads to an increase in prices today and a fall in long horizon returns. Here, I undertake an empirical investigation of this issue.

Table 5 presents the results of regressing the next year's excess stock returns on dividend price ratio and demographic variables. Dividend yield is included as a regressor because previous studies have found it to have some explanatory power (see, for example, Fama and French (1988), (1989)). Although the $R^2$ improves a bit by using the lagged outflows from the stock market as an additional explanatory variable in the second regression, the coefficient on the dividend price ratio is still statistically insignificant as in the first regression. Regressions 3 to 6 use growth rates of demographic variables while the last four regressions have levels of demographic variables as regressors. As expected, the levels of demographic variables fail to increase the fit (measured by $R^2$) of the regression. Except for the variable $P_{25-44}$, all the demographic variables are insignificant. This confirms the results of Poterba (2001) who finds that the levels of the demographic variables have limited explanatory power for explaining the aggregate stock returns. It is with the introduction of the changes of demographic variables that the regressions show a significant increase in $R^2$. Similarly to Bakshi and Chen (1994), I find that the stock returns and changes in average age are positively correlated. When I include the growth rate of the variables $P_{25-44}$, $P_{45-64}$, and $P_{65+}$, I find that not only does the $R^2$ jump to over 18%, but also that all the coefficients are statistically significant.

Multi-Year Forecasts

Changing population structure affects not only the next year's stock returns but also has a longer term effect on stock prices. For instance, an increase in the number of Middles leads to an increase in this year's stock prices, but in the absence of fundamental changes also leads to a decline in the expectation of future

---

12Bakshi and Chen (1994) find that average age has significant explanatory power in equity premium regressions, while Poterba (2001) finds no significant relation between demographics and returns on stocks, bonds, and T-bills. Erb, Harvey, and Viskanta (1997) find significant correlation between population age structure and real stock returns in an international context, while Ang and Maddaloni (2001), using pooled regressions, find links between demographic structure and international excess stock returns.

13The insignificance of dividend price ratio is easily explained by their poor performance in predicting the excess stock returns in the 1990s.
TABLE 5
Predicting Excess Stock Returns (1926–1998)

<table>
<thead>
<tr>
<th>CNST</th>
<th>SPDP</th>
<th>LEAK</th>
<th>A_{25+}</th>
<th>P_{25-44}</th>
<th>P_{45-64}</th>
<th>P_{65+}</th>
<th>R²</th>
<th>R²</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>1.24</td>
<td>-0.17</td>
<td>19.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>11.59</td>
<td>9.02</td>
<td>18.54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Percentage Changes**

| 3.00 | 2.44 | 29.56 | 18.67 | 16.31 | 17.78 |
| 0.04 | 0.08 | 0.08 | 18.68 | 15.09 | 17.01 |
| 0.55 | 2.67 | -3.19 | 3.02 | 2.75 | 19.52 |
| 1.22 | 3.27 | -2.26 | 1.42 | -0.82 | 24.67 |
| 0.26 | 0.22 | 2.36 | -2.87 | 4.80 | -11.38 |
| 2.13 | 2.65 | -2.42 | 2.03 | -2.49 | 24.67 |

| 7.00 | 2.14 | -0.01 | 12.18 | 9.64 | 18.47 |
| 0.04 | 0.04 | 0.04 | 12.38 | 8.52 | 18.59 |
| 6.00 | -0.04 | -0.02 | 12.38 | 8.52 | 18.59 |
| 0.04 | -0.06 | -0.03 | 12.38 | 8.52 | 18.59 |
| 0.04 | 1.85 | 1.85 | 14.09 | 8.96 | 18.54 |
| 0.04 | -0.71 | 0.00 | 14.09 | 8.96 | 18.54 |
| 0.04 | -0.56 | 0.01 | 14.11 | 7.60 | 18.68 |
| 0.04 | -0.68 | -0.10 | 14.11 | 7.60 | 18.68 |

Levels

| 3.00 | 2.44 | 29.56 | 18.67 | 16.31 | 17.78 |
| 0.04 | 0.08 | 0.08 | 18.68 | 15.09 | 17.01 |
| 0.55 | 2.67 | -3.19 | 3.02 | 2.75 | 19.52 |
| 1.22 | 3.27 | -2.26 | 1.42 | -0.82 | 24.67 |
| 0.26 | 0.22 | 2.36 | -2.87 | 4.80 | -11.38 |
| 2.13 | 2.65 | -2.42 | 2.03 | -2.49 | 24.67 |

| 7.00 | 2.14 | -0.01 | 12.18 | 9.64 | 18.47 |
| 0.04 | 0.04 | 0.04 | 12.38 | 8.52 | 18.59 |
| 6.00 | -0.04 | -0.02 | 12.38 | 8.52 | 18.59 |
| 0.04 | -0.06 | -0.03 | 12.38 | 8.52 | 18.59 |
| 0.04 | 1.85 | 1.85 | 14.09 | 8.96 | 18.54 |
| 0.04 | -0.71 | 0.00 | 14.09 | 8.96 | 18.54 |
| 0.04 | -0.56 | 0.01 | 14.11 | 7.60 | 18.68 |
| 0.04 | -0.68 | -0.10 | 14.11 | 7.60 | 18.68 |

Variables are of annual frequency. The dependent variable, the excess stock returns (ESRET), leads the independent variables by one year. LEAK is the net outflows from the stock market deflated by last year’s market capitalization. SPDP is the log-deduct dividend price ratio of the S&P500 index. P_{25-44}, P_{45-64}, and P_{65+} are the fractions of people in the age groups 25-44, 45-64, and 65+, respectively. A_{25+} is the average age of the population above age 25. The third to sixth regressions use the growth rates of the demographic variables while the last four regressions use the levels of the demographic variables. The first row of each regression output lists the coefficient and the second row lists its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic. R², R², and S.E. are in %.

Both the dividend yield and the past outflows together help in predicting the excess stock returns but not in isolation. The percent changes in demographic variables add additional explanatory power taking the R² to almost 19%. The signs of the coefficient support the intuition that an increase in the number of Middles (Olds) leads to an increase (decrease) in prices.

stock returns. I test the impact of changing demographics on multi-year returns in this subsection.\textsuperscript{14} The basic structure of the regression is

\[
(11) \quad \sum_{u=1}^{K} ESRET_{t+u} = a_0 + a_1 SPDP_t + a_2 P_{t,25-44}^K + a_3 P_{t,45-64}^K + a_4 P_{t,65+}^K
\]

where \(P_{t,25-44}^K = (P_{t,25-44}/P_{t-K+1,25-44})^{1/K} - 1\) is the K year (annualized) percentage change in the fraction of population aged 25 to 44 and the other demographic variables are defined similarly. In other words, I forecast the next K-period returns based on the last K-period percentage changes in the population structure.

Panel A of Table 6 regresses three-year excess stock returns on three-year percentage changes in demographic variables while panel B performs the same exercise for the five-year horizon. The data set used in these regressions is slightly longer than in the previous regressions and goes back as far as possible to the 1900s. I mitigate the empirical difficulties associated with the use of overlapping observations by using the Newey-West correction for t-statistics. Table 6 shows

\textsuperscript{14}Peterba (2001) finds demographic variables have an insignificant impact on non-overlapping returns, similar to his non-finding for one-year returns. Ang and Maddaloni (2001) emphasize the role of robust standard errors for multi-year forecasts using demographic variables in an international context.
that the demographic variables are jointly highly significant and the signs of the coefficients are in accord with the hypothesis. However, individually the coefficients in the three-year regressions are not significant while only the coefficient for the Middles is significant in the five-year regressions.

### TABLE 6

**Multi-Year Forecasts of Excess Stock Returns**

<table>
<thead>
<tr>
<th>Panel A</th>
<th>Three-Year Forecasts (1900-1996)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CST</td>
</tr>
<tr>
<td>1.</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
</tr>
<tr>
<td>2.</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
</tr>
<tr>
<td>3.</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(2.00)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>Five-Year Forecasts (1910-1998)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CST</td>
</tr>
<tr>
<td>1.</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>(3.26)</td>
</tr>
<tr>
<td>2.</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
</tr>
</tbody>
</table>

Variables are of annual frequency. ESRET is the (logged) excess stock return (including dividends) on the S&P500 index. SPDP is the logged dividend-price ratio of the S&P500 index. Table 6 reports the results of the following regression:

$$\sum_{i=1}^{K} ESRET_{t-i} = a_0 + a_1 SPDP_{t-i} + a_2 P_{25-44}^{K} + a_3 P_{25-64}^{K} + a_4 P_{65+}^{K},$$

where $P_{25-44}^{K} = (P_{25-44}/P_{25-44}^{K-1})^{1/K}$ - 1 is the K-year (annualized) percentage change in the fraction of population aged 25 to 44 and the other demographic variables are defined similarly. Panel A reports the result for three-year forecasts of excess stock returns while panel B reports the result for five-year forecasts of excess stock returns. The first row of each regression output lists the coefficient and the second row lists its Newey-West heteroskedasticity and autocorrelation adjusted t-statistic, $R^2$, $R^2$, and S.E. are in %. Wald is the Wald test for the joint significance of the three demographic variables and is calculated using the Newey-West corrected variance-covariance matrix. The number in parentheses below the Wald statistic is the p-value for this statistic based on a $\chi^2(3)$ distribution.

The significance of the dividend price ratio in predicting multi-year stock returns increases with horizon. The percent changes in demographic variables add additional explanatory power to forecasts. The signs of the coefficient support the intuition that an increase in the number of Middles (Olds) leads to a(n) decrease (increase) in future expected stock returns, although the evidence is not overwhelmingly strong.

### C. VAR and Forecasts

One reason for the interest in demographics is to analyze the effect of baby boomers in the next few years. The baby boomers, now largely in the 45 to 55 age group, are going to start retiring over the next 10 to 20 years. The effect of this on Social Security and Medicare programs has been analyzed extensively. In this subsection, I forecast the outflows from the stock market over the next 52 years. Since the stock returns, dividends, and leakages are part of a system (refer to Tables 4 and 5 for evidence on predictability), I employ a VAR system for the purpose of forecasting. My VAR is a little different from traditional VARs in that...

---

15See various volumes edited by David Wise (2001) consisting of papers presented as part of the National Bureau of Economic Research's ongoing project on the economics of aging. Also see Cutler, Poterba, Sheiner, and Summers (1990) for an analysis of the effect of changing demographics on national savings.
I have more than one exogenous variable. As Appendix B explains, this creates little problem in estimation or inference.

The results of the estimation are presented in Table 7. Standard errors are corrected for heteroskedasticity and autocorrelation using the Newey-West weighting scheme. The coefficients on the demographic variables retain their signs in predicting the LEAK but are somewhat less significant. The regression using SPDP has a high $R^2$ because of the close-to-unit root properties of the dividend price ratio.\(^\text{16}\)

<table>
<thead>
<tr>
<th>TABLE 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VAR with Stock Market Variables (1926–1998)</strong></td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ESRET(+1)</td>
</tr>
<tr>
<td>(2.95)</td>
</tr>
<tr>
<td>SPDP(+1)</td>
</tr>
<tr>
<td>(-3.30)</td>
</tr>
<tr>
<td>LEAK(+1)</td>
</tr>
<tr>
<td>(-0.09)</td>
</tr>
</tbody>
</table>

Variables are of annual frequency. ESRET is the logged excess stock returns on the S&P500 index. LEAK is the net outflows from the stock market deflated by last year's market capitalization. SPDP is the logged dividend price ratio of the S&P500 index. $P_{25-44}$, $P_{45-64}$, and $P_{65+}$ are the fractions of people in the age groups 25–44, 45–64, and 65+, respectively. The demographic variables are measured as growth rates of fractions of people in different age groups (\(\Delta\) denotes percentage change). The dependent variables lead the independent variables by one year (the VAR order is 1). The first row of each regression output lists the coefficient and the second row its Newey-West heteroskedasticity and autocorrelation adjusted $t$-statistic. Number of observations is 72. $R^2$, $R^2$, and S.E. are in %. The interpretation of coefficients on regressions of LEAK and ESRET is in Tables 4 and 5, respectively. SPDP has a high $R^2$ because of its close-to-unit root properties.

The estimated coefficients (and their standard errors) are then utilized for forecasting. Appendix B outlines a simulation technique for obtaining the standard errors of the forecasts, due to the non-availability of analytical formulas for the same. Using the projections for the demographic variables (provided by the Census Bureau), I calculate the outflows (and the other endogenous variables in the system) for the next 52 years. This process is repeated 10,000 times and Figure 1 shows summary statistics on the outflows. The figure shows that as baby boomers continue to retire over the next 10 years, increasing demands will be placed on the stock market to provide cash flows to them in the form of dividends or repurchases. This trend continues until 2025 before the outflows start to drop. A quick look at the population data reveals that 2025 is characterized by a leveling of the number of people over 65 ($P_{65+}$) and a slight increase in the number of people in the savings group $P_{45-64}$. This is in accord with my lifecycle model that predicts increasing outflows with an increase in the number of Olds.

Note that not only are the outflows projected to increase over the next 20 years, but they also remain at high levels for almost a decade. At the same time, the levels of outflows are not unprecedented and the latter half of the 1980s were

\(^{16}\)Results of the first-order autoregressive process on SPDP are (Newey-West $t$-statistics are in parentheses),

\[
SPDP_{t+1} = -0.473 + 0.858 \times SPDP_t; \quad R^2 = 63.5\%.
\]

\((-1.49) \quad (8.49)\)
characterized by almost the same level of outflows. Finally, note that the average predicted outflows from the market for the next 52 years are almost the same as they have been in the last 70 years. The reason for this is twofold. First, the outflows were at historically very low levels in the 1990s (in fact, the 1990s are characterized by inflows of funds) and it takes a decade for the outflows to come back to mean levels. Second, the outflows are predicted to decline in the mid-21st century because the fraction of middle-aged people (ages 45 to 64) is projected to increase around 2040. Thus, the popular perception that the retirement of the baby boomers would result in huge outflows as retirees start to withdraw money from their investments seems overstated. In fact, Cutler, Poterba, Sheiner, and Summers (1990) conclude that demographic changes will improve American standards of living in the near future and suggest that an optimal policy response is a reduction in the national savings rate.

D. Demographics and Macroeconomy

Even though I did not formally model the real economy, it needs no emphasizing that demographics impact the real sector as much as (if not more than) the financial sector. There is an extensive literature studying the determinants of the aggregate savings rate/investment rate in a country and analyzing cross-country differences in these variables.¹⁷ In this subsection, I explore whether there is any

¹⁷See the excellent survey by Deaton (1992). For a more recent analysis at an international level, see Culhane (2001).
relation between macroeconomic variables and the changing structure of population. The macroeconomic indicators chosen are: i) ratio of real gross private domestic investment to the real gross national product, INVST; ii) ratio of personal savings to personal disposable income, SAVING; and iii) gross national product, GNP. Since real GNP is a trending variable, I use the Baxter and King (1999) band-pass filter to calculate business cycle components of real GNP. This filter passes frequencies between years two and eight and uses three lags.

Panel A of Table 8 regresses INVST on lagged macroeconomic and demographic variables. INVST has a high autoregressive coefficient and lagged stock returns add explanatory power to the regression. A possible explanation comes from the forward looking nature of stock prices: prices might rise in anticipation of an increase in the economy’s investment rate. The demographic variables are also jointly significant (p-value of 0.03) although the coefficient on Middles is not high and does not have the anticipated positive sign (an increase in the fraction of Middles raises the level of real activity in the economy). Panel B presents the same results for SAVING. Lagged stock returns have less success in predicting the next period’s personal savings rate but the demographic variables are again jointly highly significant. However, the coefficient on Middles is again negative (Deaton (1992), p. 48), suggests that the presence of children, by placing an additional burden on young workers, may force young households to borrow more than would be suggested by simple lifecycle models. Panel C shows that the demographic variables have limited explanatory power in predicting the business cycle components of real GNP.

E. Demographics and International Capital Flows

The two broad perspectives on the determinants of international capital flows deal with portfolio rebalancing and international macroeconomics and trade. The finance perspective treats capital flows as rising from diversification motives although the empirical literature documents a significant home bias (see, for instance, French and Poterba (1991) and Kang and Stulz (1997)). Tesar and Werner (1995) show an abnormal turnover of foreign equity holdings by U.S. residents while Brennan and Cao (1997) model capital flows as rising from informational asymmetry between domestic and foreign residents. Froot, O’Connell, and Seasholes (2001) find a contemporaneous correlation between flows and returns. Using the international economics literature as a motivation, Portes and Rey (1999) use additional macroeconomic variables such as GNP growth and measures of geographical distance in explaining equity flows. I add the demographic variables to the regressions explaining equity flows. The hypothesis I wish to test is again derived from the basic lifecycle effects. An increase in the proportion of the middle-aged population of an economy leads to an increase in the economic investment, implying net investment by this economy’s residents in a foreign country (U.S.) should increase.

I consider capital flows to the U.S. from six developed countries, namely, Canada, France, Germany, Italy, Japan, and the U.K. The primary sources for

---

TABLE 8
Demographic Structure and Macroeconomy


<table>
<thead>
<tr>
<th>CNST</th>
<th>INVS</th>
<th>ESRET</th>
<th>ΔGNP</th>
<th>ΔP_{25-44}</th>
<th>ΔP_{45-64}</th>
<th>ΔP_{65+}</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>S.E.</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.02</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68.47</td>
<td>67.99</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.57)</td>
<td>(9.51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>0.01</td>
<td>0.87</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td>77.79</td>
<td>77.12</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.42)</td>
<td>(13.05)</td>
<td>(6.56)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>0.01</td>
<td>0.89</td>
<td>0.06</td>
<td>0.09</td>
<td></td>
<td></td>
<td>79.59</td>
<td>78.65</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.54)</td>
<td>(13.31)</td>
<td>(5.28)</td>
<td>(1.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>0.04</td>
<td>0.78</td>
<td>0.06</td>
<td>0.09</td>
<td>0.25</td>
<td>0.01</td>
<td>80.63</td>
<td>78.75</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.90)</td>
<td>(9.47)</td>
<td>(5.54)</td>
<td>(1.15)</td>
<td>(2.25)</td>
<td>(0.06)</td>
<td></td>
<td></td>
<td></td>
<td>(0.03)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>CNST</th>
<th>SAVING</th>
<th>ESRET</th>
<th>ΔGNP</th>
<th>ΔP_{25-44}</th>
<th>ΔP_{45-64}</th>
<th>ΔP_{65+}</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>S.E.</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.01</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71.58</td>
<td>71.15</td>
<td>2.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(7.63)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>0.01</td>
<td>0.85</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td>71.66</td>
<td>70.80</td>
<td>2.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.61)</td>
<td>(7.40)</td>
<td>(0.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>0.01</td>
<td>0.71</td>
<td>0.01</td>
<td>0.31</td>
<td></td>
<td></td>
<td>81.07</td>
<td>80.20</td>
<td>2.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.20)</td>
<td>(9.02)</td>
<td>(0.35)</td>
<td>(2.64)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>0.01</td>
<td>0.67</td>
<td>0.01</td>
<td>0.07</td>
<td>0.32</td>
<td>0.01</td>
<td>82.25</td>
<td>80.53</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.01)</td>
<td>(9.16)</td>
<td>(0.32)</td>
<td>(2.70)</td>
<td>(0.48)</td>
<td>(1.95)</td>
<td></td>
<td></td>
<td></td>
<td>(1.07)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>CNST</th>
<th>GNPBP</th>
<th>ESRET</th>
<th>ΔP_{25-44}</th>
<th>ΔP_{45-64}</th>
<th>ΔP_{65+}</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>S.E.</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-2.05</td>
<td>0.32</td>
<td>-0.45</td>
<td>29.88</td>
<td>28.30</td>
<td>446.46</td>
<td>4.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.27)</td>
<td>(4.27)</td>
<td>(4.27)</td>
<td>(4.27)</td>
<td>(4.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>-2.31</td>
<td>0.37</td>
<td>-1.64</td>
<td>29.88</td>
<td>28.30</td>
<td>446.46</td>
<td>4.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.59)</td>
<td>(4.59)</td>
<td>(4.59)</td>
<td>(4.59)</td>
<td>(4.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>-2.13</td>
<td>0.96</td>
<td>-0.66</td>
<td>511.09</td>
<td>512.47</td>
<td>-244.18</td>
<td>30.86</td>
<td>26.84</td>
<td>449.75</td>
</tr>
</tbody>
</table>

Variables are of annual frequency. The dependent variable leads the independent variables by one year. INVS is the ratio of real gross private domestic investment to the real GNP. SAVING is the ratio of personal savings to personal disposable income. GNPBP is the result of the application of Baxter and King’s (1999) band-pass filter to real GNP. This filter passes frequencies between years two and eight, and uses three lags, and is intended to give business cycle components of real GNP. ESRET is the logged excess stock return on the S&P500 index. \( P_{25-44} \), \( P_{45-64} \), and \( P_{65+} \) are the fractions of people in the age groups 25–44, 45–64, and 65+, respectively. \( \Delta \) denotes percentage change. The first row of each regression output lists the coefficient and the second row its Newey-West heteroskedasticity and autocorrelation adjusted \( t \)-statistic. \( R^2 \), \( R^2 \), and S.E. are in %. Wald is the Wald test for the joint significance of the three demographic variables and is calculated using the Newey-West corrected variance-covariance matrix. The number in parentheses below the Wald statistic is the p-value for this statistic based on a \( \chi^2 \) distribution.

Demographics do not have a significant impact on the business cycle components of GNP, but have some explanatory power in predicting the aggregate investment/savings rate.

data on demographics are Demographic Statistics (Eurostat), Japan Statistical Yearbook, Canada Year Book, and Demographic Yearbook (United Nations). The international capital flows are collated from various issues of the “Treasury Bulletin” published by the U.S. Department of the Treasury. For my selected countries, the population data in different age groups and the capital flows data are readily available starting in 1975. The stock returns on country indices are taken from Datastream. All capital flows are measured in U.S. dollars and deflated by the U.S. inflation rate. The following fixed effects panel regression is estimated,

\[
Ni_{i,t+1} = \alpha_0 + \alpha_1 Ni_{i,t} + \alpha_2 r_{i,t} + \alpha_3 \gamma_{US} + \beta_1 \Delta P_{25-44,t}^i + \beta_2 \Delta P_{45-64,t} + \beta_3 \Delta P_{65+,t}^i
\]

where \( Ni_{i,t} \) is the net investment (purchases less sales) in the U.S. by residents of country \( i \), \( r_{i,t} \) is the return on stocks in country \( i \), \( \Delta \) denotes the percentage change, and the demographic variables are as defined earlier. The return chasing hypothesis of Bohn and Tesar (1996) predicts a positive \( \alpha_3 \) (higher U.S. returns prompt
foreign investors to increase their U.S. investment) and a negative $\alpha_2$ (higher domestic returns prompt foreign investors to decrease their U.S. investment). The lifecycle hypothesis predicts a positive $\beta_2$ (an increase in Middles leads to an increase in foreign investment) and a negative $\beta_3$ (increase in Olds leads to a decrease in foreign investment). The results of the above regression are presented in Table 9.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff.</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{t-1}^{f}$</td>
<td>0.64</td>
<td>5.10</td>
</tr>
<tr>
<td>$r_{t-1}^{f}$</td>
<td>-596.89</td>
<td>-1.39</td>
</tr>
<tr>
<td>$r_{t-1}^{d}$</td>
<td>29.36</td>
<td>0.02</td>
</tr>
<tr>
<td>$\Delta P_{25-44,t-1}^{f}$</td>
<td>22170.30</td>
<td>1.02</td>
</tr>
<tr>
<td>$\Delta P_{25-64,t-1}^{f}$</td>
<td>56000.74</td>
<td>2.06</td>
</tr>
<tr>
<td>$\Delta P_{65+,t-1}^{f}$</td>
<td>4391.77</td>
<td>0.33</td>
</tr>
</tbody>
</table>

The following fixed-effects panel regression is estimated,

$$N_{t}^{f} = \alpha_0 + \alpha_1 N_{t-1}^{f} + \alpha_2 r_{t-1}^{f} + \alpha_3 r_{t-1}^{d} + \alpha_4 \Delta P_{25-44,t-1}^{f} + \beta_1 \Delta P_{25-64,t-1}^{f} + \beta_2 \Delta P_{65-,t-1}^{f},$$

where $N_{t}^{f}$ is the net investment (purchases less sales) in the U.S. by residents of country $i$ in period $t$, $r_{t}^{f}$ is the return on stocks in country $i$. $\Delta$ denotes the percentage change, and the demographic variables are as defined earlier. The countries included in this regression are Canada, France, Germany, Italy, Japan, and the U.K. See the main text for sources of data. The sample period is 1977–1997 giving 126 observations. The constants, $\alpha_0$, are not reported to conserve space. t-Statistics are White adjusted for heteroskedasticity.

International capital flows react to the past returns in predictable fashion although the evidence is not statistically strong. Demographic variables also impact the capital flows with an increase in the number of Middles ($P_{25-64}$) leading to an increase in foreign investment by residents of other countries.

The evidence in Table 9 does not lend support to the return chasing hypothesis. The coefficients on lagged returns are statistically insignificant. There is some evidence, though, for the lifecycle hypothesis. When the number of Middles in a country increases, there is a significant ($t$-statistic of 2.06) increase in the net investment by the residents of that country. Unfortunately, the evidence also seems to suggest an increase in the number of Olds would lead to an (insignificant) increase in net investment, which is inconsistent with lifecycle effects.

V. Conclusion

This paper attempts to model the investments in the stock market in an OLG model with wealth effects in which each generation has more than two periods to live. The conclusions are in accord with the basic lifecycle models. The implications of the model are tested empirically using U.S. data. I find that outflows are positively correlated with changes in the fraction of old people and negatively correlated with changes in the fraction of middle-aged people. I also find that outflows over the next 50 years are not expected to rise to levels that cause concern even with the retirement of baby boomers.

In addition to a decrease in outflows from the stock market with an increase in the middle-aged population, the stock prices are posited to rise. Moreover, in absence of fundamental changes, the long horizon returns are predicted to fall. The empirical regressions confirm these results. Stock returns increase following
an increase in the middle-aged population and decrease following an increase in the old-aged population. The $\tilde{R}^3$ of excess stock return predictability regressions rises to over 18%. Long horizon returns are also negatively correlated with an increase in the middle-aged population, and positively correlated with an increase in the middle-aged population.

I also empirically assess whether demographics have an appreciable impact on the real economy. I find no significant impact of demographics on the business cycle components of the GNP, but do find some explanatory power in predicting the aggregate investment/savings rate. Finally, demographic changes have some power in explaining international capital flows.

Appendix A

Proof of Proposition 1. The solution to the optimal consumption-investment problem follows from the dynamic programming approach using the value function. Refer to Ingersoll ((1987), Ch. 11) for details on this approach.\(^{19}\)

The investor consumes $C_t$ at time $t$ and has $I_t = W_t + Y_t - C_t$ left over for investment in the risky and risk-free assets. Assume that fraction $\alpha_t$ of this investment $I_t$ is invested in the risky asset and the rest is invested in the risk-free asset. For convenience, define $R_{t+i,t+i} = \alpha_t (R_{m,t+i} - R_f) + R_f$. Then the wealth $W_{t+i}$ is given by $W_{t+i} = (W_t + Y_t - C_t)R_{t+i,t+i}$. The value function is defined as usual by

$$
(A-1) \quad J_t \left( W_t; \{Y_t\}_{t=T-1}^{T-1} \right) = \max_{C_t,\alpha_t} \left\{ u_t(C_t) + E_t \left[ J_{t+1} \left( W_{t+1}; \{Y_t\}_{t=T-1}^{T-1} \right) \right] \right\}.
$$

I guess that the functional form of $J_t(\cdot)$ is given by

$$
(A-2) \quad J_t(\cdot) = \frac{k_t^\gamma}{1 - \gamma} \left( W_t + \sum_{t=1}^{T-1} Y_t \frac{R_f^{T-t}}{R_f^{T-t}} \right)^{1-\gamma},
$$

\(^{19}\)The choice of utility function is an important part of the assumptions. The most popular choices are: log utility, negative exponential (CARA) utility, and power utility. Log utility investors are myopic in their consumption and investment choices. Thus, log utility is inadequate for analyzing the long-run effects of changing population. CARA utility displays wealth independence, which means that the dollar demand for risky assets is independent of wealth, and depends only of the level of risk aversion. Moreover, risk aversion of a CARA utility investor with a finite horizon increases with age. The solution to the optimization problem for CARA investors living for $T$-periods is

$$
\theta_t = \frac{\tilde{r}_{t+1} - r_f}{a_t \sigma_{t+1}^2}, \quad \alpha_t = \frac{a_{t+1}(1 + \tilde{r}_t)^{T-t}}{(1 + r_f)^{T-t} - 1},
$$

$$
C_t = \frac{a_t}{a} (W_t + Y_t) + \frac{\gamma - \alpha_t}{a^2} \left[ \rho + \frac{1}{2} \left( \tilde{r}_{t+1} - r_f \right)^2 - \log \left( \frac{a_{t+1}(t+1)}{a_{t+1}} \right) \right],
$$

where $\theta_t$ is the dollar investment in the risky asset at time $t$, $C_t$ is the consumption in period $t$, $W_t$ is the wealth at time $t$, $Y_t$ is the income received in time $t$, $r_t$ is the risk-free rate, $\tilde{r}_{t+1} \sim N(\tilde{r}_{t+1}, \sigma_{t+1}^2)$ is the risky asset distribution for period $[t, t+1]$, $\rho$ is the rate of time preference, $f(t)$ is a constant that depends only on time $t$, and $a$ is the coefficient of risk aversion. Thus, the investor reduces her demand for risky assets as she grows older. Therefore, as the economy grows older, outflows from the market will increase regardless of whether the aging economy is due to an increase in the number of middle-aged people or of old-aged people, which is counterfactual. Finally, the demand for risky assets for CARA investors depends on the next period’s return distribution. In this sense, investors are myopic (although their consumption decisions do depend on the whole lifecycle return distribution). Both of the above effects make CARA utility an inadequate choice in this setting.
where the constant $k_t$ is to be determined later. The first-order conditions for maximization are given by

\begin{equation}
C_t^{-\gamma} = k_{t+1}^{-\gamma} E_t \left[ \left( W_t + Y_t - C_t \right) R_{\alpha,t+1} + \sum_{s=t+1}^{T-1} \frac{Y_s}{R_{s-1}^{t-1}} \right] R_{\alpha,t+1}^{-\gamma},
\end{equation}

\begin{equation}
0 = E_t \left[ \left( W_t + Y_t - C_t \right) R_{\alpha,t+1} + \sum_{s=t+1}^{T-1} \frac{Y_s}{R_{s-1}^{t-1}} \right] \left( R_{m,t+1} - R_f \right).
\end{equation}

Equation (A-3) illustrates one of the difficulties in obtaining an analytical solution to this problem: I cannot take $C_t$ out of the expectation on the right-hand side of the equation. However, the following approximate first-order conditions help in obtaining an analytical solution,

\begin{equation}
C_t^{-\gamma} \approx k_{t+1}^{-\gamma} E_t \left[ \left( W_t + \sum_{s=t}^{T-1} \frac{Y_s}{R_{s-1}^{t-1}} - C_t \right) R_{\alpha,t+1}^{-\gamma} \right],
\end{equation}

\begin{equation}
0 \approx E_t \left[ \left( W_t + \sum_{s=t}^{T-1} \frac{Y_s}{R_{s-1}^{t-1}} - C_t \right) R_{\alpha,t+1}^{-\gamma} \left( R_{m,t+1} - R_f \right) \right].
\end{equation}

It is then easily verified that the complete solution to the problem is given by

\begin{equation}
C_t = k_t \left( W_t + \sum_{s=t}^{T-1} \frac{Y_s}{R_{s-1}^{t-1}} \right),
\end{equation}

\begin{equation}
0 = E \left[ k_{t+1}^{-\gamma} R_{\alpha,t+1}^{-\gamma} (r_{m,t+1} - r_f) \right] \text{ gives } \alpha_t,
\end{equation}

\begin{equation}
k_t^{-1} = 1 + \left[ \delta E_t \left[ k_{t+1}^{-\gamma} R_{\alpha,t+1}^{-\gamma} \right] \right]^{1/\gamma}.
\end{equation}

The results in the main text for Proposition 1 are just a special case of the above with iid stock returns.

**Proof of Implication 1.** Consider first the Olds at time period $t$. Their wealth at the beginning of period $t$ is the return on the amount invested by the Middles at time period $t-1$. Money invested by each Middle at time period $t-1$ was $W_{m,t-1} Y_n (1-k_m) - k_m Y_n / R_f$, and has grown to $\left( W_{m,t-1} + Y_n \right) (1-k_m) - k_m Y_n / R_f$ at time period $t$. Note that of the total wealth passed on to the Olds, $p_{m,t-1} \left( W_{m,t-1} + Y_n \right) (1-k_m) - k_m Y_n / R_f \alpha R_m$ is already invested in the stock market. Now, since the number of the Olds at $t$ is $p_{0,t}$, wealth of each old person is given by

\begin{equation}
W_{o,t} = \frac{p_{m,t-1}}{p_{o,t}} \left[ \left( W_{m,t-1} + Y_n \right) (1-k_m) - k_m Y_n / R_f \right] R_{\alpha}.
\end{equation}

Investment in the stock market by each Old is $\alpha \left( W_o + Y_o \right) (1-k_o)$. Substituting the expression for $W_{o,t}$ from equation (A-6) and using the fact that the number of Olds at time $t$ is $p_{0,t}$,

\begin{equation}
\text{New Investment}_{o,t} = \alpha p_{m,t-1} \left[ \left( W_{m,t-1} + Y_n \right) (1-k_m) - k_m Y_n / R_f \right] \times \left[ -R_m + (1-k_o) R_{\alpha} \right] + \alpha p_{m,t-1} Y_o (1-k_o).
\end{equation}
A similar calculation for the Middles gives

\[(A-8) \quad \text{New Investment}_{m,t} = \alpha p_{m,t-1} \left[ (W_{y,t-1} + Y_y)(1 - k_y) - k_y \left( \frac{Y_m}{R_f} + \frac{Y_o}{R_f^2} \right) \right] \times \left[ -R_m + (1 - k_m)R_o \right] + \alpha p_{m,t} \left[ Y_m(1 - k_m) - k_m \frac{Y_o}{R} \right].\]

The Youngs are new investors (since whatever wealth they have at the beginning is the result of leftovers from the Infants). Thus,

\[(A-9) \quad \text{New Investment}_{y,t} = p_{y,t} \left[ (W_{y,t} + Y_y)(1 - k_y) - k_y \left( \frac{Y_m}{R_f} + \frac{Y_o}{R_f^2} \right) \right] \alpha.\]

Since Infants just consume the wealth bequeathed to them by the Olds,

\[(A-10) \quad \text{New Investment}_{i,t} = -p_{o,t-1}(W_{o,t-1} + Y_o)(1 - k_o)\alpha R_m.\]

Combining equations (A-7)–(A-10),

\[(A-11) \quad \text{Leakage}_{t} = -\text{New Investment}_{t} + \alpha p_{o,t-1}(W_{o,t-1} + Y_o)(1 - k_o)R_m
\times \alpha p_{m,t-1} \left[ (W_{m,t-1} + Y_m)(1 - k_m) - k_m \frac{Y_o}{R_f} \right] \left[ R_m - (1 - k_m)R_o \right]
\times \alpha p_{y,t-1} \left[ (W_{y,t-1} + Y_y)(1 - k_y) - k_y \left( \frac{Y_m}{R_f} + \frac{Y_o}{R_f^2} \right) \right] \left[ R_m - (1 - k_m)R_o \right]
- \alpha p_{o,t} Y_o(1 - k_o) - \alpha p_{m,t} \left[ Y_m(1 - k_m) - k_m \frac{Y_o}{R} \right]
- \alpha p_{y,t} \left[ (W_{y,t} + Y_y)(1 - k_y) - k_y \left( \frac{Y_m}{R_f} + \frac{Y_o}{R_f^2} \right) \right].\]

Consider first the effect of an increase in the fraction of the Olds, \(p_{o,t} \). Assuming (for simplicity) that the number of the Middles (\(p_{m,t-1} \)) in period \(t - 1 \) is approximately equal to the number of the Olds (\(p_{o,t} \)) in period \(t \), the sign of a change in \(p_{o,t} \) depends on the sign of the expression,

\[(A-12) \quad \left[ (W_{m,t-1} + Y_m)(1 - k_m) - k_m \frac{Y_o}{R_f} \right] \left[ R_m - (1 - k_m)R_o \right] - Y_o(1 - k_o).\]

Similarly, the sign of a change in \(p_{m,t} \) depends on the sign of the expression,

\[(A-13) \quad \left[ (W_{y,t-1} + Y_y)(1 - k_y) - k_y \left( \frac{Y_m}{R_f} + \frac{Y_o}{R_f^2} \right) \right] \left[ R_m - (1 - k_m)R_o \right]
- Y_m(1 - k_m) + k_m \frac{Y_o}{R}.\]

Finally, consider the effect of an increase in the fraction of Youngs, \(p_{y,t} \). Equation (A-11) predicts that an increase in \(p_{y,t} \) would lead to a decline in the leakage at time \(t \) (the sign of \(p_{y,t} \) in equation (A-11) is negative) because the Youngs are first-time investors and, therefore, an increase in their proportion leads to an increase in net inflows into the asset market. However, in my model an increase in Youngs comes only due to a previous increase in Infants. Thus, it is possible that the total wealth passed on to the Youngs is negligible and, therefore, I do not expect any strong relation between asset market flows and the proportion of the Youngs.
Calibrating the parameter values to match U.S. data wherever possible, I assume that the risky return is normally distributed with a mean of 9% and standard deviation of 20%. The risk-free rate is assumed to be 3%. The risk aversion parameter $\gamma$ is assumed to be five and the rate of time preference $\delta$ is assumed to be 0.95. These parameters imply that $\alpha = 30.79\%$, $k_0 = 0.5447$, $k_m = 0.3945$, and $k_s = 0.3206$. Substituting these values in expression (A-12), I find that the effect of an increase in the number of Olds on net leakage depends on the sign of $0.37(W_m + Y_m) - 0.69Y_s$. For all reasonable values of $W_m$ and $Y_m > Y_s$ (Middles are expected to have a higher income than Olds), this expression is positive. In short, the effect of an increase in $p_{m,s}$ is to increase the leakage at time $t$. Similarly, expression (A-13) evaluates to $0.31(W_s + Y_s) - 0.75Y_s + 0.24$. Again, the initial wealth of the Youngs is expected to be negligible, while their income $Y_s$ is expected to be lower than the income of the Middles $Y_m$. Thus, the effect of an increase in $p_{m,s}$ is to decrease the leakage at time $t$.

Proof of Implication 2. The parameters for the comparative statics exercise are as before, namely: $r_f = 3\%$, $\gamma = 5$, $\delta = 0.95$, and $r_m \sim N(9\%, 20\%)^2$. I initially assume that there are equal fractions of people in all generations, i.e., $p_t = p_m = p_s = 0.25$. Assuming an exogenous return-generating process gives $\alpha = 0.31$ and $r_m = 4.85\%$. I assume that the income of Youngs, Middles, and Olds is $S_6$, $S_7$, and $S_3$, and their wealth is $S_1$, $S_4$, and $S_3$, respectively, based on the information in the latest Statistical Abstract. Note that in this equilibrium state, the price is equal to $1/1.09 = 0.9174$. The last parameter I need is $z$, the constant in the supply function. Using the above parameter values and substituting in equation (6) to solve for $z$, I get $z = 0.7664$.

To find out the effects of increasing the fraction of Olds in period $t$, I give a shock to the number of Middles in period $t - 1$. Similarly, to calculate the effect of increasing the fraction of Middles in period $t$, I give a shock to the number of Youngs in period $t - 1$. The complete set of experiments is summarized in the following table.

<table>
<thead>
<tr>
<th>Time</th>
<th>$p_t$</th>
<th>$p_m$</th>
<th>$p_s$</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t - 1$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>$0.9202$</td>
</tr>
<tr>
<td>$t$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>$0.9194$</td>
</tr>
<tr>
<td>$t - 1$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>$0.9186$</td>
</tr>
<tr>
<td>$t$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>$0.9202$</td>
</tr>
</tbody>
</table>

This table shows that increasing the proportion of Olds at time $t$ leads to a drop in prices (rows 1 and 2), while increasing the proportion of Middles at time $t$ leads to a rise in prices (rows 3 and 4). These results translate into a fall in returns due to an increase in the number of Olds and a rise in returns due to an increase in the number of Middles.

Appendix B

Consider a VAR(1) in the $(k \times 1)$ vector $y_t$. The explanatory variables are one lag of $y_t$ and one lag of an additional $(m \times 1)$ vector $z_t$. The system is

$$
    y_t = \begin{bmatrix} A & C \end{bmatrix} \begin{bmatrix} y_{t-1} \ z_{t-1} \end{bmatrix} + u_t
$$

(B-1)

Suppose I have $T$ observations of the variables of interest. Stack the variables as follows,

$$
    Y = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{bmatrix}, \quad x_t = \begin{bmatrix} y_{t-1}' \\ z_{t-1}' \end{bmatrix}, \quad X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{T-1} \end{bmatrix}, \quad U = \begin{bmatrix} u_1' \\ u_2' \\ \vdots \\ u_T' \end{bmatrix}
$$

(B-2)
Let $B = [A \quad C]'$. Then the system can be summarized as $Y = XB + U$. Define $\bar{y} = \text{vec}(Y)$, $\bar{b} = \text{vec}(B)$, and $\bar{u} = \text{vec}(U)$. Then an equivalent representation of the system is

$$\bar{y} = (I_k \otimes X) \bar{b} + \bar{u}. \tag{B-3}$$

Assuming the errors $u$ are normally distributed, I can define the log likelihood function as

$$\mathcal{L} = -\frac{KT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \text{tr} \left( U' U \Sigma_u^{-1} \right) \tag{B-4}$$

$$= -\frac{KT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \left( \bar{y} - (I_k \otimes X) \bar{b} \right)' \left( \Sigma_u^{-1} \otimes I_T \right) \left( \bar{y} - (I_k \otimes X) \bar{b} \right).$$

The first and second derivatives are given by

$$\frac{\partial \mathcal{L}}{\partial \bar{b}} = \left( \bar{y} - (I_k \otimes X) \bar{b} \right)' \left( \Sigma_u^{-1} \otimes I_T \right) (I_k \otimes X) \tag{B-5}$$

$$= \left[ \left( \Sigma_u^{-1} \otimes X' \right) \bar{y} - \left( \Sigma_u^{-1} \otimes XX' \right) \bar{b} \right]'.$$

$$\frac{\partial^2 \mathcal{L}}{\partial \bar{b} \partial \bar{b}'} = -(I_k \otimes X)' \left( \Sigma_u^{-1} \otimes I_T \right) (I_k \otimes X)$$

$$= -\left( \Sigma_u^{-1} \otimes XX' \right). \tag{B-6}$$

From these, the estimate $\hat{b}$ and its variance are given by

$$\hat{b} = (I_k \otimes (X'X)^{-1}X') \bar{y} = \text{vec} \left( (X'X)^{-1}X'y \right). \tag{B-7}$$

$$\sqrt{T} (\hat{b} - b) \rightarrow \mathcal{N} \left( 0, \Sigma_u \otimes Q^{-1} \right) \text{ where } Q = \text{plim} \left( \frac{XX'}{T} \right). \tag{B-8}$$

From equation (B-1), it is clear that an $h$ period ahead value of $y_{t+h}$ is given by

$$y_{t+h} = A^h y_t + \sum_{i=0}^{h} A^i C \bar{z}_{t+h-i-1} + \sum_{i=0}^{h} A^i u_{t+h-i}. \tag{B-9}$$

The optimal (in the sense of minimizing MSE) forecast is thus

$$y_t(h) = A^h y_t + \sum_{i=0}^{h} A^i C \bar{z}_{t+h-i-1}. \tag{B-10}$$

The error in the forecast is

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h} A^i u_{t+h-i} \sim \mathcal{N} \left( 0, \Sigma_t(h) \right), \tag{B-11}$$

where $\Sigma_t(h)$ is given by

$$\Sigma_t(h) = \Sigma_t(h-1) + A^{h-1} \Sigma_u \left( A^{h-1} \right)' \quad ; \quad \Sigma_t(1) = \Sigma_u. \tag{B-12}$$

It is easy to verify that $\Sigma_t(h)$ as $h \rightarrow \infty$ is given by $I_y$, where

$$\text{vec} \left( I_y \right) = (I_{K^2} - A \otimes A)^{-1} \text{vec} \left( \Sigma_u \right). \tag{B-13}$$
The above, of course, assumes that the coefficients are known without error. In the presence of estimation error, the forecast error variance is given by (see Lütkepohl (1993) for details)

\[(B.14) \quad \Sigma_y(h) = \Sigma_y(h) + \frac{1}{T} \Omega(h) \quad \text{where} \quad \Omega(h) = E \left[ \left( \frac{\partial y_i(h)}{\partial \beta} \right) \left( \sum_k \frac{\partial y_i(h)}{\partial \beta_k} \right) \right].\]

The computation of an analytical expression for the expectation in \(\Omega(h)\) is complicated by the fact that \(y_i(h)\) depends on the exogenous variables \(\{z_{it}\}_{t=0}^{h-1}\). Hence the paper relies on numerical alternative.

To obtain the standard errors, I resort to simulation. In each iteration, I draw the \((21 \times 1)\) simulation coefficient vector from a normal distribution with mean equal to the estimated coefficients and covariance matrix equal to the covariance matrix of the estimated coefficients. These simulated coefficients are then used as the "true" coefficients for the next 52 years. In each iteration, I also draw a \((52 \times 1)\) error vector for the errors using the estimated covariance matrix of the errors.\(^{20}\) Finally, using the projections for the demographic variables (provided by the Census Bureau) I am able to arrive at the outflows (and the other endogenous variables in the system) for the next 52 years. This process is repeated 10,000 times and Figure 1 shows summary statistics on the outflows.

References


\(^{20}\)Note that by construction the estimated coefficients and the errors are uncorrelated.