

Life-Cycle Consumption and Portfolio Choice with an Imperfect Predictor*

Yuxin Zhang [†]

Imperial College London

23rd December 2016

*I am indebted to Alex Michaelides for guidance and encouragement. I would also like to thank Franklin Allen, Harjoat Bhamra, Rustam Ibragimov, Alexandros Kostakis, David Miles, Lubos Pastor, Tarun Ramadorai and seminar participants at Imperial College London for many helpful comments. All remaining errors are my own.

[†]Address: Department of Finance, Imperial College Business School, South Kensington Campus, SW7 2AZ: y.zhang11@imperial.ac.uk.

Life-Cycle Consumption and Portfolio Choice with an Imperfect Predictor

Abstract

I study the effect of observable predictors that imperfectly predict conditional expected stock returns on optimal life-cycle consumption and portfolio choice in the presence of undiversifiable labor income risk. Investors filter the unobservable expected stock returns from realized predictive variables and stock returns. Young stockholders hold more conservative portfolios, better matching empirical observations, than models assuming a predictor perfectly delivering the conditional expected stock return or models assuming i.i.d. stock returns. Welfare losses from ignoring imperfect predictability can be substantial.

(JEL D14, G11, G17)

Key Words: Portfolio Choice over the Life Cycle, Stock Market Mean Reversion, Filtering, Stock Market Predictability, Imperfect Predictor.

1 Introduction

Optimal life-cycle portfolio choice is a classic problem in financial economics, encountered by every investor. Samuelson (1969) argues that the investment decision is independent of wealth and consumption-saving decisions. However, Samuelson's conclusion is confined to the assumption of independent and identically distributed (i.i.d.) stock returns and the absence of undiversifiable, risky labor income. Cocco et al. (2005) solve for optimal portfolio choice, consumption and saving decisions numerically and show that the labor income stream is a key factor for optimal life-cycle portfolio choice with mortality risk, borrowing and short-sale constraints, and time-separable power utility preferences. Their findings provide rationale for age-varying investment advice such as recommending target-date funds (TDFs) that reduce exposure to stocks as retirement approaches.¹ These authors, however, assume that the stock returns are i.i.d., a classical view meaning that the expected return is constant over time.

Nevertheless, recent empirical studies provide evidence supporting the predictability of stock returns. Many papers find that a number of variables forecast stock returns. The main method is a simple predictive regression: if we can find $|b| > 0$ in $r_{t+1} = \alpha + bq_t + z_{t+1}$, then we know that $E_t(r_{t+1}) = bq_t$. This implies that the expected stock return can be perfectly predicted by the predictor. The popular predictors (q_t) provided by the literature are the dividend/price ratio (D/P), earnings per share (EPS) or consumption-wealth ratio (CAY).² Since these predictors themselves follow a persistent auto-regressive process (AR model), the r_t essentially is a mean reversion process.³

¹Heaton and Lucas (2000), Viceira (2001), Haliassos and Michaelides (2003) and Gomes and Michaelides (2005) also study the effect of labor income risk on portfolio choice while ignoring the predictability of stock returns.

²See Lettau and Ludvigson (2001) and Lan (2015).

³For instance, Campbell (1987) and Fama and French (1988) show that dividend/price ratios predict stock returns. Campbell and Shiller (1988a) also make this point by proposing the following regressions:

$$\begin{cases} r_{t+1} = r_f + b\mu_t + z_{t+1} \\ \mu_{t+1} = a + \beta\mu_t + \varepsilon_{t+1} \end{cases}, \quad \begin{bmatrix} z_{t+1} \\ \varepsilon_{t+1} \end{bmatrix} \sim Normal(0, \Omega),$$

where r_{t+1} denotes the real stock market return from time t to $t+1$, μ_t is the predictor such as the dividend/price ratio at time t , α and β are the regression's intercept and slope coefficients of the predictor, r_f is the real risk free interest rate and z_{t+1} and ε_{t+1} are the white noises following a bi-variate normal distribution with mean of zero and

In response to the evidence on the stock market predictability, various papers have studied its implications for optimal portfolio choice and consumption.⁴ Michaelides and Zhang (Forthcoming) build a model in which an investor chooses consumption and optimal asset allocation over the life cycle to maximize an Epstein-Zin-Weil preference function assuming that the dividend yield can perfectly predict the expected stock returns (hereafter, the **perfect predictor model**). This model, however, seems restrictive because it assumes that an observable predictor such as the dividend yield can perfectly predict expected stock returns. This assumption can be criticized for data mining, non-robustness of test statistics and incorrect inference in small samples. Goyal and Welch (2008) re-examine the performance of predictors such as the dividend yield and find that these predictors are both weak in-sample, and out-of-sample, indicating that the predictability of expected stock returns is quite uncertain. It seems more likely that the predictors are noisy proxies, in that they are correlated with the time-varying expected stock returns but can not predict them perfectly.⁵

More recently, the idea that the predictive relation between the predictor and expected stock returns is quite uncertain has gained more ground. For example, Xia (2001) assumes that the predictability parameter (b) in the predictive regression is ambiguous. This uncertainty in b is just one specific example that the expected risk premium is hard to precisely observe. Pastor and Stambaugh (2009) generalize Xia (2001) by assuming that the current expected stock return is unobservable and the predictor is imperfect so that the estimation of expected stock returns using the predictive regression omits some important features. In fact, the unexpected stock returns nega-

covariance structure of Ω . When $\beta = 0$, this regression becomes the i.i.d. stock return model. Fama and French really focus on the importance of the D/P on long-time horizon. These observations show that the predictability of stock return is economically and statistically significant phenomenon that can not be dismissed. Fama and French (1989) is an excellent summary and example that documents and illustrates the time variation of expected stock returns.

⁴Kim and Omberg (1996), Brennan et al. (1997), Brandt (1999), Campbell and Viceira (1999), Balduzzia and Lynch (1999), Campbell et al. (2001, 2003), and Wachter (2002) show that stock market risk premiums change materially with respect to the predictive factor(s) and analyze the implications for optimal portfolio choice.

⁵Ang and Bekaert (2007) also examine the predictive power of the dividend yield for forecasting the excess stock returns. They find that the univariate dividend yield regression provides a rather poor proxy to the true expected stock return.

tively correlate with the innovations in the unobservable expected stock returns, when the stock returns exhibit mean reversion (Pastor and Stambaugh (2012)). Pastor and Stambaugh (2009) construct an imperfectly predictive system with noisy predictors to estimate the expected stock returns and find that this imperfection has a significant effect on the conditional expected stock returns.

How does the presence of such imperfect predictability affect optimal consumption and portfolio choice for a stockholder over the life cycle? In this paper, I solve a life-cycle portfolio choice model with an imperfect predictor, jointly modeling an imperfect predictive system, liquidity constraints and non-diversifiable background labor income risk to analyze the normative implications for life-cycle consumption and portfolio choice using Epstein-Zin (1989) preferences (hereafter, the **imperfect predictor model**). The key feature of this model is to include the imperfection in the predictive relation of stock returns model to understand how this type of uncertainty affects saving and portfolio choice over the life cycle.

When calibrated to the observed dividend yield and stock returns from 1946 to 2015, under the imperfect predictive system of stock returns, the portfolio allocation is more conservative than that in the perfect predictor model or in the i.i.d. stock returns model. This result substantially alters one of the main insights of models ignoring imperfect predictability. Specifically, such models predict that "stocks are for the young" and such advice has been popularized by Target Date Funds (TDFs) that advise a more aggressive asset allocation in stocks when young and a gradual reduction in this exposure as the investor gets older. With imperfect predictability, consistent with Pastor and Stambaugh (2012), stocks become more volatile in the long run, and therefore young households hold more conservative (balanced) portfolios.

Interestingly, this prediction of the imperfect predictor model is more consistent with empirical observation than either the i.i.d. stock returns or the perfect predictor models. When compared with the data from the U.S. Survey of Consumer Finances (hereafter, **SCF**), the imperfect predictor model matches the data better than either

the perfect predictor model or the i.i.d. stock returns model. Specifically, in the SCF data stockholder portfolios are balanced between bonds and stocks. Recently, Wachter and Yogo (2010) generate balanced portfolios through nonhomothetic utility over basic and luxury goods. In this paper, the balanced portfolio early in life arises due to the additional stock market uncertainty arising from imperfect predictability.

From all the underlying parameters studied, the main parameters that materially affect the optimal consumption and investment choice are the volatility of the unobservable expected stock return, the persistence of the unobservable expected stock returns and the correlation between the innovations to stock returns and shocks to unobserved expected stock returns. Therefore, we should pay more attention to these parameters when making investment decisions. I also experiment with respect to the correlation between permanent earnings shocks and stock market innovations, the correlation between innovations to stock returns and shocks to the dividend yield and the correlation between shocks to the dividend yield and innovations to the unobserved expected stock returns. I find that these correlations do not substantially change wealth accumulation and consumption, but they do significantly alter the portfolio allocation.

These findings influence the design of target date funds (TDFs) because market timing through the utilization of different information affects optimal portfolio choice. The presence of imperfect predictability affects tactical asset allocation and alters the prediction of models where investors expect either i.i.d. stock returns or use a model with a perfect predictor to compute expected stock returns. Therefore, the imperfection of the predictor significantly changes the asset allocation decision, with potentially significant implications for the design of optimal TDFs.

To illustrate the importance of taking imperfect predictability into account when designing TDFs, I compare the welfare across different models by computing the consumption certainty equivalent under different settings. Specifically, I simulate 10,000 individual life histories assuming that the data generating process of stock returns is an imperfect predictive system. In the imperfect predictor model, the investor chooses

the investment policy according to the expected return filtered from the observed data. On the contrary, investors using the perfect predictor model or the i.i.d. stock returns model make investment decisions without caring about any observed stock returns. As to the investors using the Vanguard TDFs investment rules (hereafter, **Vanguard TDF model**), they adjust their portfolio allocation only depending on age. I can then calculate the ratio of value functions from the imperfect predictor model to the ones from the other models and report the consumption certainty equivalent based on this ratio. In this way, I can compare the change in investor welfare between the imperfect predictor model and the other three models: the perfect predictor model, the i.i.d. stock returns model, and the Vanguard TDF model.

The perfect predictor model has the smallest welfare loss, and the i.i.d. stock returns model generates the largest welfare loss. The Vanguard TDF model obtains the second largest welfare loss. All of these welfare losses vary with the correlation between unexpected returns and shocks to the predictors, and increase as this correlation approaches 1. These losses are maximized at around age 50 when the increase in average wealth accumulation slows down and the net saving rate (the difference between labor income and consumption) turns negative.

Where do these welfare rankings come from? I show that these substantial welfare losses relative to the baseline can be explained by the differences in the first two moments of household consumption. The imperfect predictor model has the highest mean consumption and volatility of consumption, and the i.i.d. stock returns model generates the lowest mean consumption and volatility of consumption over the working life. In the middle is the perfect predictor model.

The paper is organized as follows. Section 2 explains the theoretical model in the paper and a rough description of the numerical solution. Section 3 illustrates the estimation method and discusses the calibration. Section 4 builds a baseline model with the risky labor income and Epstein–Zin preferences to study the effect of the imperfect predictive system on the portfolio choice over the life cycle, Section 5 contains the

welfare analysis across different models including the TDFs and Section 6 concludes.

2 The Model

2.1 Model Specification

2.1.1 Preference Model

I denote adult age by t ($t \in [20, 100]$). The investor chooses the portfolio and consumption policies to maximize the following Epstein-Zin preferences:

$$\begin{cases} V_t = \max_{(c_t, \alpha_t)} \left\{ (1 - \beta) C_t^{1-1/\psi} + \beta [\mathfrak{R}_t(V_{t+1})]^{1-1/\psi} \right\}^{1/(1-1/\psi)} \\ \mathfrak{R}_t(V_{t+1}) = [E_t(p_{t+1} V_{t+1}^{1-\gamma} + b(1-p_{t+1}) X_{t+1}^{1-\gamma})]^{1/(1-\gamma)} \end{cases} \quad (1)$$

where V_t is the continuation value at age t , \mathfrak{R}_t is the uncertainty aggregator, X_{t+1} is the terminal wealth if the investor is dead at age $t+1$, β is the discount factor, ψ is the elasticity of inter-temporal substitution (hereafter, **EIS**), γ is the risk aversion parameter, b is the strength of the bequest motive and p_{t+1} is the conditional probability of surviving next period conditional on having survived until age t .

2.1.2 Labor Income Process

Following the same method as Cocco et al. (2005) and Carroll (1997), I build the labor income process before retirement as follows:

$$Y_{it} = Y_{it}^p U_{it} \quad (2)$$

$$Y_{it}^p = \exp[g(t, Z_{it})] Y_{it-1}^p N_{it} \quad (3)$$

where $g(t, Z_{it})$ is a deterministic function of age and household i 's characteristics Z_{it} , Y_{it}^p is a permanent component with innovation N_{it} of household i 's age t labor income, and U_{it} is a transitory component of household i 's age t labor income.

In equations (2) - (3), I assume that $\ln(U_{it})$ and $\ln(N_{it})$ are independent and identically distributed with mean $\{-0.5\sigma_u^2, -0.5\sigma_n^2\}$, and variances σ_u^2 and σ_n^2 , respectively. As to Y_{it}^p , $\ln(Y_{it}^p)$ evolves as a random walk with a deterministic drift, $g(t, Z_{it})$. For simplicity, retirement is assumed to be exogenous and deterministic, with all households retiring in time period K , corresponding to age 65 ($K = 46$). Earnings in retirement ($t > K$) are given by $Y_{it} = \lambda Y_{iK}^p$, where λ is the replacement ratio ($\lambda = 0.68$) of the last working period permanent component of labor income.

Durable goods, and in particular housing, can provide an incentive for higher spending early in life. We exogenously subtract a fraction of labor income every year allocated to durables (housing), and this fraction includes both rental and mortgage expenditures. This empirical process is taken from Gomes and Michaelides (2005) and is based on Panel Study Income Dynamics (hereafter, **PSID**) data. We choose not to model explicitly the returns from housing following the empirical evidence (e.g., Cocco and Lopes (2015) and references therein) that households tend not to decumulate housing as fast as life-cycle models predict. A prominent explanation tends to be a psychological benefit from continuing to own one's house, an explanation that is consistent with the low observed demand for home equity conversion mortgages (Davidoff (2015)). For these reasons we do not explicitly model the potential effects of housing returns, and focus instead only on investments of liquid financial wealth for rich households (that empirically tend to be both stockholders and homeowners).

For convenience, I will take logarithms on both sides of (2) and (3) while solving the investor's problem. Hence, $\log(Y_{it}^p) = g(t, Z_{it}) + \log(Y_{it-1}^p) + \log(N_{it})$ and $\log(Y_{it}) = \log(Y_{it}^p) + \log(U_{it})$.

2.1.3 Stock Return Predictability Model

I assume that there are two assets in which the investor can invest, a risk-free asset, such as T-bills, and a risky asset, such as stocks. The risk free asset has a constant gross real return of r_f , and the risky asset has a gross real return r_t . As to modeling the gross

real return of risky asset, I follow the idea of Pastor and Stambaugh (2009) that the expected stock returns are unobservable and that investor must filter these expected stock returns from the other observable information. Denote (r_t, q_t, μ_t) as the stock return, the predictor and the unobservable expected stock return, respectively. Then, an imperfect predictive system can be defined as follows:

$$\mu_{t+1} = \alpha_\mu + \phi_\mu \mu_t + \varepsilon_{t+1} \quad (4)$$

$$r_{t+1} = r_f + \mu_t + z_{t+1} \quad (5)$$

$$q_{t+1} = \alpha_q + \phi_q q_t + v_{t+1} \quad (6)$$

where $\begin{bmatrix} \varepsilon_{t+1} & z_{t+1} & v_{t+1} \end{bmatrix} \sim Normal(0, \Omega)$ and $\Omega = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{z\varepsilon} & \sigma_{v\varepsilon} \\ & \sigma_z^2 & \sigma_{zv} \\ & & \sigma_v^2 \end{bmatrix}$.

This imperfect predictive system is a generalization of the classical predictive regression. The unobservable expected stock return (μ_t) follows a simple AR(1) process described by equation (4). Equation (5) defines the next period's stock return (r_{t+1}) as a sum of the risk free rate (r_f) , the unobservable expected stock return (μ_t) and an innovation term (unexpected stock return, z_t). Equation (6) assumes that the predictor (q_t) evolves in a manner of a persistent AR(1) process, which is a standard assumption in the literature about the predictability of stock returns. This model is consistent with a variety of economic models in which the expected return not only varies over time but also exhibits mean reversion.

Based on this imperfect predictive system, the investor must filter out μ_t from the other observable variables (r_t, q_t) . Applying the simplest filtering algorithm (see theorem 7.1 in the appendix, the conditional distribution of a multivariate normal distribution), the first two conditional expected moments of μ_t can be rewritten as

$$E(\mu_t|d_t) = E_r + \Sigma_{\mu d} \Sigma_d^{-1} \left(\begin{bmatrix} r_t - r_f \\ q_t \end{bmatrix} - \begin{bmatrix} E_r \\ E_q \end{bmatrix} \right) \quad (7)$$

$$Var(\mu_t|d_t) = \sigma_\mu^2 - \Sigma_{\mu d} \Sigma_d^{-1} \Sigma'_{\mu d} \quad (8)$$

where $d_t = [r_t, q_t]$, $\Sigma_{\mu d} = [\sigma_{\mu r}, \sigma_{\mu q}]$ and $\Sigma_d = \begin{bmatrix} \sigma_r^2 & \sigma_{rq} \\ \sigma_{rq} & \sigma_q^2 \end{bmatrix}$.

(7) and (8) can be further simplified as:

$$E(\mu_t|[r_t, q_t]) = \hat{\mu}_{t|t} = E_r + \kappa_r [r_t - r_f - E_r] + \kappa_q [q_t - E_q] \quad (9)$$

$$Var(\mu_t|[r_t, q_t]) = \sigma_\mu^2 - \kappa_r \sigma_{\mu r} - \kappa_q \sigma_{\mu q} \quad (10)$$

where $\kappa_r = \frac{\sigma_{\mu r} \sigma_q^2 - \sigma_{rq} \sigma_{\mu q}}{\sigma_r^2 \sigma_q^2 - \sigma_{rq}^2} < 0$, $\kappa_q = \frac{\sigma_r^2 \sigma_{\mu q} - \sigma_{rq} \sigma_{\mu r}}{\sigma_r^2 \sigma_q^2 - \sigma_{rq}^2} > 0$, $E_r = \frac{\alpha_\mu}{1 - \phi_\mu}$, $E_q = \frac{\alpha_q}{1 - \phi_q}$, $\sigma_r^2 = \sigma_\mu^2 + \sigma_z^2$, $\sigma_\mu^2 = \frac{\sigma_\varepsilon^2}{(1 - \phi_\mu^2)}$, $\sigma_q^2 = \frac{\sigma_v^2}{(1 - \phi_q^2)}$, $\sigma_{\mu r} = \rho_{z\varepsilon} \sigma_z \sigma_\varepsilon + \frac{\sigma_\varepsilon^2}{(1 - \phi_\mu^2)}$, $\sigma_{\mu q} = \frac{\rho_{v\varepsilon} \sigma_v \sigma_\varepsilon}{(1 - \phi_\mu \phi_q)}$ and $\sigma_{rq} = \rho_{zv} \sigma_z \sigma_v + \frac{\phi_q \sigma_v \sigma_\varepsilon}{(1 - \phi_\mu \phi_q)}$.

(9) and (10) say that the conditional moments of μ_t consist of three information sources. The first source is the unconditional mean of risk premium (E_r). The second source is the current stock return (r_t), and the last one is the current dividend yield (q_t). Similarly, the conditional variance of μ_t can be decomposed into three parts: the variance of unobservable expected stock returns (σ_μ^2), the covariance between the unobservable expected stock returns and the realized stock returns ($\sigma_{\mu r}$) and the covariance between the unobservable expected stock returns and the dividend yield ($\sigma_{\mu q}$).

Several important conclusions can be drawn from (9) and (10). First, κ_r is negative, which implies that an unexpected increase in the stock return leads to the decrease in the next period's expected stock return. κ_r , therefore, measures the mean reversion effect. In contrast, the positive κ_q measures the predictability effect because a positive shock to the dividend yield predicts an increase in the next period's expected stock

return and vice versa.

Second, when $\rho_{\mu q} = 1$, $E_r = E_q$, $\sigma_\mu = \sigma_q$ and $\rho_{\mu r} = \rho_{rq}$, $\kappa_r = 0$ and $\kappa_q = 1$. (9) and (10), therefore, become

$$\hat{\mu}_{t|t} = E(\mu_t | [r_t, q_t]) = q_t \quad (11)$$

$$Var(\mu_t | [r_t, q_t]) = 0 \quad (12)$$

(11) and (12) implies that $E_t(r_{t+1}) = q_t$, namely, the predictor perfectly predicts the expected stock return. The imperfect predictive system ((4) - (6)) degenerates into the classical predictive regression used in Campbell and Shiller (1988a), Campbell and Viceira (1999), Michaelides and Zhang (Forthcoming) etc. Similarly, the i.i.d. stock returns model is also a special case of this imperfect predictive system. In contrast, if $|\rho_{v\varepsilon}| < 1$ and $\rho_{v\varepsilon} \neq 0$, the predictor (q_t), is not a perfect proxy of μ_t , and the information from r_t and q_t enters the conditional expected μ_t according to (9) - (10). Hence, the expected stock return of the next period is not completely determined by the observed predictor so that uniquely relying on the this predictor can deliver an inaccurate estimation.

Third, the conditional moments of the unobservable expected stock return depend on both the observed data (r_t, q_t) and the correlations among the unobservable expected stock return, the observed predictor and the current stock return ($\rho_{\mu r}, \rho_{rq}, \rho_{\mu q}$). This also explains why the correlation between the innovations to observable predictor and the shocks to current stock return does not play a key role in the perfect predictor model solved by Michaelides and Zhang (Forthcoming)⁶. The perfect predictor model rule out the effect of these correlations from calculating the conditional expected stock return of the next period ($E_t[r_{t+1}] = q_t$) and conditional variance ($Var_t[r_{t+1}] = \sigma_z^2$),

⁶Michaelides and Zhang (Forthcoming) use the perfect predictor model/classical predictive regression to solve the life-cycle portfolio choice problem and find that only the correlation between the innovations of stock returns and the permanent earning shocks of labor income (ρ_{zn}) materially affects the optimal portfolio choice.

which means that these correlations only have a small effect on the optimal investment and consumption decision.

2.2 The Investor's Optimization Problem

At the beginning of period t , investor i has a wealth $W_{i,t}$. Then, during this period, labor income $Y_{i,t}$ is realized. Following Deaton (1991), cash on hand $X_{i,t}$ can be defined as $X_{i,t} = W_{i,t} + Y_{i,t}$. Then, the investor must determine how much to consume, $C_{i,t}$ and how to invest the remaining savings in stocks $S_{i,t}$ and the risk free asset $B_{i,t}$. In the next period, before earning period $t + 1$'s labor income, the wealth at $t + 1$ is given by $W_{i,t+1} = S_{i,t} (1 + r_{t+1}) + B_{i,t} (1 + r_f) = \alpha_{it} (1 + r_{t+1}) + (1 - \alpha_{it}) (1 + r_f)$, where $S_{i,t}$ is the investment in the stock market in the previous period, $B_{i,t}$ is the investment in risk-free asset in the previous period and $\alpha_{i,t}$ is the share of wealth in stocks in the previous period and defined as $\alpha_{i,t} = \frac{S_{i,t}}{B_{i,t} + S_{i,t}}$. The budget constraint of investor i at time t is $S_{i,t} + B_{i,t} = W_{i,t} + Y_{i,t} - C_{i,t}$.

The investor maximizes the household's utility subject to the budget constraint and the constraints (2) through (6) with the non-negativity restrictions on $C_{i,t}$, $B_{i,t}$ and $S_{i,t}$. These non-negativity constraints on $B_{i,t}$ and $S_{i,t}$ guarantee the investor not to borrow against his/her future labor income or retirement wealth.

In this optimization problem, μ_t is unobservable and the investor has to estimate it through (9) - (10) conditional on the observed information (r_t, q_t) available at time t . The state variables of the investor's problem are t , $X_{i,t}$, $\hat{\mu}_{t|t}$ and $Y_{i,t}^p$, the control variables are $C_{i,t}$ and $\alpha_{i,t}$, and the policy functions are defined as $C_{i,t}(X_{i,t}, Y_{i,t}^p, \hat{\mu}_{t|t})$ and $\alpha_{i,t}(X_{i,t}, Y_{i,t}^p, \hat{\mu}_{t|t})$.

Since, the problem uses the Epstein-Zin utility, the value function is homogeneous with respect to the current permanent part of labor income. This property allows us to normalize the investor's cash on hand ($X_{i,t}$) by dividing $Y_{i,t}^p$, which means the number of state variables is reduced by one. The policy functions, therefore, become $c_{i,t}(x_{i,t}, \hat{\mu}_{t|t})$ and $\alpha_{i,t}(x_{i,t}, \hat{\mu}_{t|t})$, where $x_{i,t} = \frac{X_{i,t}}{Y_{i,t}^p}$.

2.3 Numerical Solution

The optimization problem faced by the investor can be rewritten as the following optimization model:

$$\begin{aligned}
 V_t(x_{i,t}, \hat{\mu}_{t|t}) = \underset{(c_{i,t}, \alpha_{i,t})}{Max} & \left\{ (1 - \beta) c_{i,t}^{1-\frac{1}{\psi}} + \beta \left[\{ E_t(p_{t+1} V_{t+1}^{1-\gamma}(x_{i,t+1}, \hat{\mu}_{t+1|t+1}) \right. \right. \\
 & \left. \left. + b(1 - p_{t+1}) x_{i,t+1}^{1-\gamma} \}^{\frac{1}{\gamma}} \right]^{1-\frac{1}{\psi}} \right\}^{\frac{1}{1-\frac{1}{\psi}}} \\
 s.t. & \begin{cases} \mu_{t+1} = \alpha_\mu + \phi_\mu \mu_t + \varepsilon_{t+1} \\ r_{t+1} = r_f + \mu_t + z_{t+1} \\ q_{t+1} = \alpha_q + \phi_q q_t + v_{t+1} \\ \ln(N_{t+1}) = \mu_n + n_{t+1} \\ \ln(U_{t+1}) = \mu_u + u_{t+1} \\ x_{i,t+1} = \frac{Y_{i,t}^p}{Y_{i,t+1}^p} (r_{t+1} \alpha_{i,t} + r_f [1 - \alpha_{i,t}]) (x_{i,t} - c_{i,t}) + U_{i,t+1} \end{cases}
 \end{aligned} \tag{13}$$

where $\hat{\mu}_{t|t}$ is a linear function of (r_t, q_t) and updated through formula (9), c_{it} is the normalized consumption of household i at time t , x_{it} is the normalized cash on hand of household i at time t and $\alpha_{i,t}$ is the risky asset allocation of household i at time t .

This problem has no analytical solution. I, therefore, solve this problem numerically by using backward induction. In the last period (hereafter, T), the optimal policy functions are easy to solve because the investor does not invest any more and consumes all wealth except for the saving bequeathed to heirs. Then, I can now replace the value function in the Bellman equation (13) with the optimal policy function solved at time T and calculate the optimal policies for T-1. Repeating this procedure up to age 20, I can obtain the policy functions at each age.

In the backward induction algorithm, grid search is used to find the optimal policy functions of the problem (13) based on a fine discrete approximation of the following VAR model:

$$\begin{cases} \mu_{t+1} = \alpha_\mu + \phi_\mu \mu_t + \varepsilon_{t+1} \\ r_{t+1} = r_f + \mu_t + z_{t+1} \\ q_{t+1} = \alpha_q + \phi_q q_t + v_{t+1} \\ \ln [N]_{t+1} = \mu_n + n_{t+1} \end{cases} \quad (14)$$

I use Tauchen and Hussey (1991) method to discretize the state space of the VAR model (14) and calculate the transition probabilities among these grid points assuming that they follow a Markov Chain. Then, using the grid points from the discretization of (14)⁷, I can construct the next period's return by:

$$\begin{cases} r_{t+1|t} = r_f + \hat{\mu}_{t|t} + z_{t+1} + w_{t+1} \\ \hat{\mu}_{t|t} = E_r + \kappa_r [r_t - r_f - E_r] + \kappa_q [q_t - E_q] \end{cases} \quad (15)$$

where w_{t+1} is an independent innovation term introduced by the filtering algorithm and follows $N(0, Var\{\mu_t|[r_t, q_t]\})$.

Finally, I iteratively apply the backward induction algorithm to solve the consumption and investment policy functions of the optimization problem (13) based on $r_{t+1|t}$ from age T to age 20. The details of numerical implements are the same as the Online Appendix of Michaelides and Zhang (Forthcoming).

I implement this numerical algorithm using Fortran 2003 on a Windows workstation⁸. For accelerating the time of computation, I parallelize this algorithm according to the state variables using OpenMP⁹, which makes the problem can be solved in twenty four hours.

⁷The temporary part of labor income ($\ln(U_t)$) is not correlated with the other variables. Its grid points are, therefore, generated independently.

⁸Intel Xeon E5-2699 v3 2.3GHz RAM 256GB

⁹OpenMP is a set of compiler directives, library routines, and environment variables to enable programmers to develop parallel applications for shared memory multiprocessor computer.

3 Empirical Analysis

3.1 Data

The stock market data used in this paper comes from the Center for Research in Securities Prices (CRSP). I screen out the annual one year bond return, annual CPI growth rate, monthly value-weighted cumulative return of S&P 500 and monthly value-weighted ex-return of S&P 500 from Dec. 31st, 1946 to Dec. 31st, 2015. Next, I construct annual cumulative and ex-dividend S&P 500 price index based on the monthly data with the initial cumulative price of 1.00. Using the difference between annual cumulative and ex-dividend price index, I can easily obtain the annual cumulative return and annual ex-dividend return. The annual dividend is calculated by multiplying the lagged total annual price index by the difference between the annual cumulative return and ex-dividend return. Finally, I compute the real return as the difference between the annual cumulative return and annual CPI growth rate. Table 1 shows the summary of stock market data.

The empirical portfolio holding data are based on the SCF 2007. The empirical asset holding is defined as either $\alpha = equity/(equity + bond)$ or $\alpha = equity/(equity + bond + liquidity)$, where *liquidity* is the financial wealth with high liquidity such as the cash.

3.2 Parameter Estimation

The first step of solving the investor's optimization problem is to estimate the parameters of the equation (4) - (6). For estimating this VAR model through the observed data, I transform it into the following VARMA(1,1) model:

$$\begin{cases} r_{t+1} - r_f = (1 - \phi_\mu) E_r + \phi_\mu(r_t - r_f) + nv_t - (\phi_\mu - m)\omega_t + \omega_{t+1} \\ q_{t+1} = (1 - \phi_q)E_q + \phi_q q_t + v_{t+1} \end{cases} \quad (16)$$

where m and n are constant parameters derived based on the equations (4) - (6) and ω_t is forecast error ($\omega_t = r_t - \hat{r}_{t|t-1}$) and serially uncorrelated.

TABLE 1
Descriptive Statistics

Table 1 presents descriptive statistics of the annual data from CRSP. The real risk free is defined as the mean of the difference between the 1-Year bond return and annual CPI growth rate. Real adjusted return (r_t) is defined as the difference between the annual value weighted adjusted returns and annual CPI growth rate. SD is the standard deviation.

1946/12/31~2015/12/31	Mean(%)	SD(%)	Skewness	Kurtosis
1-Year Bond Return	5.07	3.9	1.00	4.23
Annual CPI Growth Rate	3.79	3.4	1.81	7.14
Value Weighted Adjusted Returns	12	17	-0.40	3.02
Value Weighted Ex>Returns	8.28	17	-0.41	2.93
Dividend/Price	3.39	1.5	0.47	3.05
Real Adjusted Return	8.22	18	-0.43	3.04
Real Risk Premium ($r_t - r_f$)	6.93	18	-0.43	3.04
Real Risk Free Rate	1.29	-	-	-

TABLE 2
The Results of Parameter Estimation

Table 2 shows the parameter estimation of the equations (4) - (6). E_r is the unconditional expectation of the risk premium, E_q is the unconditional expectation of the predictor, ϕ_q is the persistence parameter of the predictor, ϕ_μ is the persistence parameter of the unobserved expected stock return process, σ_v is the standard deviation of the predictor's innovations, σ_ω is the standard deviation of the forecast error specified in (16), m and n are the parameters in (16) which are derived from equations (4) - (6), $\rho_{\omega v}$ is the correlation between the innovations of the predictor process and the forecast errors, σ_r is the standard deviation of stock returns, $\rho_{r,q}$ is the correlation between the stock returns and the predictors and σ_q is the standard deviation of the predictor.

E_q	0.0326	ϕ_q	0.9553	m	-0.2242	σ_r	0.1921
E_r	0.069	σ_v	0.004968	n	5.7292	$\rho_{r,q}$	-0.0932
ϕ_μ	0.1022	σ_ω	0.1813	$\rho_{\omega v}$	-0.6562	σ_q	0.01681

The appendix describes how to derive this VARMA(1,1) model and estimate it using MLE.¹⁰ Table 2 summarizes the results of the parameter estimation.

¹⁰I thank Lubos Pastor for kindly providing matlab codes to perform this estimation.

Some parameters in the covariance matrix of equations (4) - (6) remain unidentified because the covariance matrix consisting of three variables can not be exactly estimated through only two observed variables (see appendix 7.3.4). I, therefore, describe the solution space of the covariance matrix (Ω) with respect to a specific variable. As $\sigma_{z\varepsilon}$ play a critical role in determining the conditional expected return, I solve the solution space of the covariance matrix (Ω) with respect to $\sigma_{z\varepsilon}$. The details about how to derive the solution space of the covariance matrix are explained in the appendix. In short, the solution space of the covariance matrix with respect to $\sigma_{z\varepsilon}$ is simplified into the following linear system:

$$\begin{cases} \sigma_z^2 = \sigma_r^2 - (Cov(r_t, r_{t-1}) - \sigma_{z\varepsilon}) / \phi_\mu \\ \sigma_\varepsilon^2 = (Cov(r_t, r_{t-1}) - \sigma_{z\varepsilon}) (1 - \phi_\mu^2) / \phi_\mu \end{cases} \quad (17)$$

s.t. $|\rho_{zv}| < 1$ and $|\rho_{v\varepsilon}| < 1$

As $\sigma_{z\varepsilon} = \rho_{z\varepsilon}\sigma_\varepsilon\sigma_z$, I can only discuss the correlation between the shocks to unobservable expected stock returns and the innovation of stock returns ($\rho_{z\varepsilon}$) instead of the covariance, $\sigma_{z\varepsilon}$. Various studies provide empirical evidence that $\rho_{z\varepsilon} < 0$. Pastor and Stambaugh (2009) find that this correlation is negative if the stock returns exhibit mean reversion. Figure 1, Panel A, plots the solution space of $(\rho_{v\varepsilon}, \rho_{zv}, \rho_{z\varepsilon})$ while changes $\rho_{z\varepsilon}$ from -1 to 0. Panel B projects the solution space onto the plane consisting of ρ_{zv} and $\rho_{z\varepsilon}$, and panel C describes the relationship between $\rho_{z\varepsilon}$ and $\rho_{v\varepsilon}$.

Several conclusions can be drawn from the Figure 1. First, based on the data, the ranges of $\rho_{z\varepsilon}$, ρ_{zv} and $\rho_{v\varepsilon}$ are approximately $[-0.66, -0.99]$, $[-0.65, -0.99]$ and $[0.37, 0.94]$, respectively. Second, panel B shows that the correlation between the innovations of stock returns and the shocks to the dividend yield (ρ_{zv}) has approximately a negative relation with the correlation between the innovations of stock returns and the shocks to the unobservable expected stock returns ($\rho_{z\varepsilon}$). When $\rho_{z\varepsilon}$ tends to be a perfect negative correlation, $|\rho_{zv}|$ decreases from 0.99 to 0.65. In contrast, the correlation between the shocks to the unobservable expected stock returns and the innovations of

the dividend yield ($\rho_{v\varepsilon}$) positively relates with $\rho_{z\varepsilon}$. When $\rho_{z\varepsilon}$ tends to be a perfect negative correlation, $\rho_{v\varepsilon}$ is close to a perfect positive correlation.

For better understanding the effect of the imperfect predictive system of stock returns on the life-cycle consumption and portfolio choice, I set up a baseline model, where $\rho_{z\varepsilon} = -0.7$, $\sigma_\varepsilon = 0.09852$, $\sigma_z = 0.1646$, $\rho_{zv} = -0.723$, $\rho_{v\varepsilon} = 0.56$ and $\rho_{zn} = 0.15$.

4 Optimal Consumption and Portfolio Choice

4.1 The Baseline Model

4.1.1 Parameter Choice

Even though empirical predictability studies are typically done on a monthly or quarterly frequency, I solve the model at an annual frequency to maintain comparability with the existing life-cycle portfolio literature. Carroll (1997) estimates the variances of the idiosyncratic shocks using data from the PSID, and the baseline simulations use values close to those: 0.1 for σ_u and 0.1 for σ_n . The deterministic component of labor income is identical to the values used by most life cycle papers, for example, Cocco et al. (2005), and this setting also facilitates comparisons between this model and its counterparts such as perfect predictor model and i.i.d. stock returns model. The relatively large estimate for the replacement ratio during retirement (68% of last working period labor income) arises from using both social security and private pension accounts to estimate the benefits in the PSID data and is consistent with not explicitly modeling tax-deferred saving through retirement accounts.

The baseline preference specification is taken to capture the observed behavior of stockholders. Gomes and Michaelides (2005) argue that this is well achieved, when using a coefficient of relative risk aversion (γ) equal to 5. The elasticity of inter-temporal substitution (ψ) is set to be 0.5. These choices are close to the empirical estimates for the EIS in Vissing-Jorgensen (2002) and the empirical preference parameter estimates in Gomes et al. (2009). The bequest parameter is set to 2.5 to capture the empirical

observation that few rich stockholders die with zero financial assets. As to the discount rate, much macroeconomic research estimates this rate to be 1% per quarter or approximate 4% per year. In order to emphasize that the results in this paper does not stem from extreme assumptions about discount factor, β in the baseline model is 0.96, which means the discount rate is assumed to be 4% per year.

The parameters used in the imperfect predictive system of the stock market are listed in Table 1 and 2. In addition, I set a trading cost of 2.9% to reflect transaction cost, tax and other implicit trading costs, which implies a risk premium of 4% the same as in the most of the life-cycle portfolio literature.

There is no estimate of the correlation between the innovations of the unobservable expected stock returns and the permanent, idiosyncratic earnings shocks to the labor income ($\rho_{n\varepsilon}$) in the literature. I therefore set this correlation equal to zero. Angerer and Lam (2009) note that the correlation between the innovations of stock returns and transitory part of labor income (ρ_{zu}) does not empirically affect portfolios and this is consistent with the simulation results in life cycle models (Cocco et al. (2005)). I set this correlation at zero. Similarly, I also set ρ_{nv} to zero. The correlation between the permanent earning shocks to the labor income and the innovations of stock returns (ρ_{zn}) is set equal to 0.15 in the baseline model, which follows the same setting as Michaelides and Zhang (Forthcoming). Table 3 summarizes the parameter values used in the baseline model.

4.1.2 Consumption and Portfolio Choice in the Baseline Model

Figure 2 plots the life-cycle profiles of wealth accumulation, consumption, labor income and share of wealth in stocks by simulating 10,000 individual life histories and reports the average.

Panel A shows the mean wealth accumulation and consumption over the life cycle in the presence of a bequest motive and labor income. The wealth accumulation increases as the investor approaches retirement and reaches the peak at the retirement age. After

the retirement, the wealth accumulation starts to decrease as agent ages.

TABLE 3

Summary of Parameter Choice

Table 3 presents the parameter choice used in the baseline model. The σ_z is the standard error of the stock returns, σ_ε is the standard deviation of the shocks to the unobservable expected stock return, σ_v is the standard error of the predictor, σ_n is the standard error of the permanent part of labor income, σ_u is the standard deviation of the transitory component of labor income, E_r is the unconditional expected risk premium, E_q is the unconditional expected dividend yield, γ is the risk aversion, ϕ_q is the persistence parameter of the dividend yield process, ϕ_u is the persistence of the unobservable expected stock returns, r_f is the real risk free rate, $\rho_{z\varepsilon}$ is the correlation between the innovations of stock returns and the shocks to the dividend yield, ρ_{vz} is the correlation between the shocks to the dividend yield and the innovations of stock returns, $\rho_{v\varepsilon}$ is the correlation between the innovations of the dividend yield and the shocks to the unobservable stock returns, ψ is the elasticity of inter-temporal substitution, b is the bequest motive, ρ_{zu} is the correlation between the innovations of stock returns and the transitory component of labor income, ρ_{vn} is the correlation between the innovations of dividend yield process and the shocks to the permanent part of labor income, ρ_{vu} is the correlation between the innovations of dividend yield process and the transitory component of labor income, $\rho_{\varepsilon n}$ is the correlation between the innovations of unobservable expected stock returns and the shocks to the permanent part of labor income. $\rho_{\varepsilon u}$ is the correlation between the innovations of unobservable expected stock returns and the transitory component of labor income, $E[\ln(N_t)]$ is the expectation of logarithm of the permanent earning shocks to the labor income, $E[\ln(U_t)]$ is the expectation of logarithm of the transitory earning shocks to the labor income, and β is the discount factor of the utility function.

Parameter	Value	Parameter	Value	Parameter	Value
σ_z	0.1646	ϕ_q	0.955	ρ_{zu}	0.0
σ_ε	0.0985	ϕ_μ	0.1022	ρ_{vn}	0.0
σ_v	0.00497	r_f	0.0129	ρ_{vu}	0.0
σ_n	0.1	$\rho_{z\varepsilon}$	-0.7	$\rho_{\varepsilon n}$	0.0
σ_u	0.1	ρ_{vz}	-0.723	$\rho_{\varepsilon u}$	0.0
E_r	0.069	$\rho_{v\varepsilon}$	0.56	$E[\ln(N_t)]$	-0.005
E_q	0.0326	ρ_{zn}	0.15	$E[\ln(U_t)]$	-0.005
γ	5	ψ	0.5	β	0.96
<i>Trading Cost</i>	0.029	b	2.5		

Panel A also shows that the consumption tracks labor income very closely before retirement and the gap between consumption and labor income gets larger as the wealth deaccumulates, reflecting that the liquidity constraint becomes less binding. When the agent approaches death, the consumption path decreases. Panel B graphs the mean

share of wealth in stocks over the life cycle. Early in life, a higher proportion of wealth is invested in the risky asset except for at the very beginning of life. As the agent approaches retirement, the share of wealth in stocks slopes down. After retirement, the mean stock allocation bounces up a little then keeps highly stable until the agent reaches the end of life. During the whole life cycle, the mean stock allocation is clearly less than 1 and fluctuates between 40% and 65%. These findings (Panel A and B) are consistent with Cocco et al. (2005), Gomes and Michaelides (2005) and Michaelides and Zhang (Forthcoming).

Figure 3 compares the life-cycle profiles between the baseline model (imperfect predictor model), perfect predictor model, i.i.d. stock returns model and the Vanguard TDF model. The Vanguard TDF model's basic recommendation is to invest 90% of a household's financial wealth in stocks until age 40, and start decreasing that share as retirement approaches reach 50% at age 65. After retirement, the Vanguard TDF model recommends the investor to continuously reduce the stock market exposure to approximate 30% and keep this proportion until death. To simulate wealth profiles for this case, I take the portfolio rule as exogenous but the household still makes optimal consumption-saving decisions, taking this portfolio decision into account.

The mean wealth accumulation and consumption shows a notable difference between the baseline model and the other three models. Panel A shows that the wealth accumulation and the consumption in the baseline model are the highest in all of these models. This arises here because the imperfection of the predictive system leads the investor to increase the precautionary saving in the baseline model. Panel B describes the difference in simulated average consumption over the life cycle. The mean consumption of the baseline and perfect predictor model are the highest and the second highest respectively because the investor takes advantage of predictability. Panel C depicts the mean share of wealth in stocks over the life cycle. The i.i.d. stock returns model maintains the highest proportion of wealth in the stock market, except between ages 45 and 65, and the baseline model has the lowest mean share allocation. On the other

hand, the perfect predictor model falls in between. The glide path of the Vanguard TDF model is exogenous as it is fixed at each age without considering any information.

The remarkable difference of mean portfolio allocation across these models can be explained by the investment policy functions. Figure 4 shows the share of wealth in stocks with respect to low, medium and high estimations of the expected stock return for age 25, 55 and 75 (Panel A, B and C show the share of wealth in stocks for age 25, Panel D, E and F for age 55, and Panel G, H and I for age 75). The investment policy functions of the i.i.d. stock returns model vary with age besides the cash on hand, and does not depend on the other factors. In the baseline and perfect predictor models, the portfolio allocation can drastically shift up or down depending on the estimation of the expected stock return besides age and cash on hand. When focusing on the baseline model and the perfect predictor model, I find that the investment policy functions in the baseline model are always less than that of the perfect predictor model. This result arises because the imperfection of the predictive system increases the conditional variance of the next period's return given the same estimation of the expected return.

An empirical puzzle arises that the predictions of portfolio allocation from the i.i.d. and perfect predictor model have a large gap during the working age over the life cycle. Figure 5 compares the mean share of wealth in stocks from the perfect predictor model and the imperfect predictor model with the data of SCF 2007. Panel A compares the mean share of wealth in stocks with the empirical portfolio allocation without considering liquidity. Panel B, however, includes the asset with high liquidity in the calculation of empirical portfolio allocation. The smoothed empirical portfolio allocation is calculated by the linear regression method. From Figure 5, we can find that the prediction from the imperfect predictor model matches the SCF data better than the perfect predictor model, which shows that the imperfection of the predictive system possibly make an important contribution to explain the observed pattern of household portfolio choice.

4.1.3 The Analysis of Model Parameter Uncertainty

Even though the baseline model has considered the imperfection of predictability in the stock returns, the estimation of parameter such as the unconditional expected risk premium (E_r), persistence of the unobservable expected stock returns (ϕ_μ) and standard error of the unobservable expected stock returns (σ_ε) possibly still have an estimation error. These parameters materially affect the mean wealth accumulation, consumption and asset allocation when the preference parameters such as risk aversion(γ) and EIS(ψ) remain unchanged. Therefore, this section measures the sensitivity of the baseline model to these parameters.

Figure 6 shows the effect of a higher unconditional expected risk premium (E_r) on the mean wealth accumulation (Panel A), consumption (Panel B) and share of wealth in stocks (Panel C) over the life cycle of the baseline model, perfect predictor model and i.i.d. stock returns model, respectively. For obtaining a higher unconditional expected risk premium ($E_r = 7\%$), I set up a 0% of the Trading Cost. Under the scenario in which the risk premium is perceived to be higher, the mean wealth accumulation, consumption and portfolio allocation all shift up. A higher unconditional expected risk premium makes investor lean to holding stocks, which leads to a higher wealth accumulation and, then, a higher consumption.

Figure 7 plots how a lower standard error of the unobserved expected stock return (σ_ε) affects the life-cycle profiles of the baseline model. When the volatility of the unobserved expected stock return (σ_ε) decreases to 0.005 from 0.0985, the mean wealth accumulation and consumption decrease, and portfolio allocation shifts up except for the 45 - 65 age group. A lower σ_ε leads to the unobservable expected stock returns fluctuating around the unconditional expectation of risk premium within a narrow band, which makes the imperfect predictive system act as a i.i.d. stock returns. The life-cycle profiles are, therefore, closer to that of the i.i.d. stock returns model.

The parameter ϕ_μ measures the persistence of the unobservable expected stock returns. This parameter is of our interest because the predictor used in the predictive

regression is often a highly persistent process in the classical literature such as Campbell and Shiller (1988b), Fama and French (1988), Xia (2001) and Cochrane (2005). Figure 8 depicts the life-cycle profile given a higher persistence of the unobservable expected stock returns (Panel A shows the mean wealth accumulation and consumption, and Panel B describes the mean share of wealth in stocks). From Panel A and B, a higher persistence makes the agent take advantage of predictability so that the mean share of wealth in stocks shifts up and seems close to that of the perfect predictor model. This is reasonable because the unobservable expected stock return is close to the high persistent predictor process when its persistence is high. On the other hand, the high persistence of the unobservable expected stock returns makes the investor more willing to consume in the earlier stage of life, attaining a lower mean wealth accumulation at retirement. The conclusions drawn from Figure 7 and 8 remind us that it is dangerous to depend entirely on an imperfect predictor such as dividend yield. The characteristics of high persistence and low volatility in the dividend yield process can lead to more aggressive investment policies and inappropriate consumption decisions.

Admittedly, the unconditional expected risk premium and standard error and persistence of the unobservable expected stock return are not the whole story. The variations due to correlations such as ρ_{zn} and $\rho_{z\varepsilon}$ are also crucial in the household financial decisions. I analyze these effects in the next subsection.

4.2 Hedging Demands

How does the correlations among the different innovations change the results of baseline model? In the i.i.d. stock returns model and perfect predictor model, the most important correlation generating quantitatively substantial hedging demands is the correlation between the permanent earnings shocks and the innovations to stock returns (ρ_{zn}), and the other correlations such as $\rho_{z\varepsilon}$ do not materially affect the results. Does this conclusion change when I introduce the imperfection to the predictive regression?

4.2.1 Correlation between the Shocks to the Unobservable Expected Stock Returns and the Innovations of Stock Returns

To investigate the importance of the correlation between the shocks to the unobservable expected stock returns and the innovations of stock returns ($\rho_{z\varepsilon}$), I vary $\rho_{z\varepsilon}$ from -0.9 to -0.5 and use the baseline model ($\rho_{z\varepsilon} = -0.7$) and perfect predictor model as benchmarks for comparison. Figure 9 plots the mean wealth accumulation, consumption (Panel A) and the mean share of wealth in stocks (Panel B) over the life cycle due to the variation of $\rho_{z\varepsilon}$. When $\rho_{z\varepsilon}$ tends to be 0 from a perfect negative correlation, the investor views the dividend yield as a better predictor of the unobserved expected stock return. From Table 4, we know that a smaller $|\rho_{z\varepsilon}|$ decreases the mean reversion effect and increases the predictability effect. This implies that results are close to that from the perfect predictor model. The investor, therefore, decreases the wealth accumulation and consumption (Panel A) and increases the stock holding. On the contrary, when this correlation is close to perfect negative, the mean asset allocation in risky stocks shifts down and the mean wealth accumulation and consumption move up.

4.2.2 Correlation between the Permanent Earnings Shocks and the Innovations of Stock Returns

I also measure the sensitivity of the correlation between the permanent earnings shocks and the innovations of stock returns (ρ_{zn}). In the baseline model, this correlation is 0.15, a value that reflects the substantial idiosyncratic risk that exists in labor income data. I vary this correlation from -0.15 to 0.3.

Figure 10 plots its effect of ρ_{zn} on the results from the baseline model. From Panel A, when ρ_{zn} changes, I find that the mean wealth accumulation and consumption rarely change. However, in Panel B, I find that the investor is more willing to invest risky stocks when this correlation decreases. The labor income acts more as a risk less asset when ρ_{zn} is small, which leads investors to taking more risk exposure in the stock market. On the contrary, when this correlation increase, it crowds out the risky

investment because the labor income acts more like a risky stock. Hence, the portfolio allocation negatively correlates with ρ_{zn} , which is consistent with the results found in Cocco et al. (2005) and Michaelides and Zhang (Forthcoming).

TABLE 4

Conditional Expectation and Standard Deviation due to Variation of $\rho_{z\varepsilon}$

Table 4 presents the conditional moments of the unobservable expected stock return (μ_t) and the next period's stock return (r_{t+1}) with different correlations between the shocks to the stock market and the innovations of unobservable expected stock returns ($\rho_{z\varepsilon}$). $E(\mu_t|r_t, q_t)$ is the conditional expectation of the current unobserved expected stock return based on the current observed return (r_t) and dividend yield (q_t), $\sigma_{\mu_t|r_t, q_t}$ is the conditional standard deviation of the current unobserved expected stock return based on the current observed return (r_t) and dividend yield (q_t), and $\sigma_{r_{t+1}|r_t, q_t}$ is the conditional standard deviation of next period's stock return based on the current observed return (r_t) and dividend yield (q_t).

$\rho_{z\varepsilon}$	$E(\mu_t r_t, q_t)$	$\sigma_{\mu_t r_t, q_t}$	$\sigma_{r_{t+1} r_t, q_t}$
-0.5	$E_r - 0.185[r_t - r_f - E_r] + 0.88[q_t - E_q]$	0.091	0.188
-0.7(benchmark)	$E_r - 0.274[r_t - r_f - E_r] + 0.79[q_t - E_q]$	0.082	0.184
-0.9	$E_r - 0.363[r_t - r_f - E_r] + 0.69[q_t - E_q]$	0.068	0.178

4.2.3 Correlation between the Innovations of Stock Returns and the Shocks to Dividend Yield

Changing the correlation between the innovations of stock returns and the shocks to dividend yield (ρ_{zv}) does not materially affect the mean wealth accumulation and consumption (Figure 11, Panel A), but does significantly change the portfolio allocation (Figure 11, Panel B). According to Table 5, when $|\rho_{zv}|$ close to 0, the predictability effect from dividend yield becomes stronger, while the mean reversion effect from the current return does not change. This makes the investment behavior from the imperfect predictor model more like that from the perfect predictor model. This conclusion is different from the result of Michaelides and Zhang (Forthcoming) because ρ_{zv} has little effect on determining the conditional moments of the next period's return in the perfect predictor model.

TABLE 5

Conditional Expectation and Standard Deviation due to Variation of ρ_{zv}

Table 5 presents the conditional moments of unobservable expected stock return (μ_t) and the next period's stock return (r_{t+1}) with different correlations between the shocks to the stock market and the innovations of dividend yield (ρ_{zv}). $E(\mu_t|r_t, q_t)$ is the conditional expectation of the current unobservable expected stock return based on the current observed return (r_t) and the dividend yield (q_t), $\sigma_{\mu_t|r_t, q_t}$ is the conditional standard deviation of the current unobservable expected stock return based on the current observed return (r_t) and dividend yield (q_t), and $\sigma_{r_{t+1}|r_t, q_t}$ is the conditional standard deviation of the following period's stock return based on the current observed return (r_t) and dividend yield (q_t).

ρ_{zv}	$E(\mu_t r_t, q_t)$	$\sigma_{\mu_t r_t, q_t}$	$\sigma_{r_{t+1} r_t, q_t}$
-0.5	$E_r - 0.277[r_t - r_f - E_r] + 0.96[q_t - E_q]$	0.082	0.184
-0.723(benchmark)	$E_r - 0.274[r_t - r_f - E_r] + 0.79[q_t - E_q]$	0.082	0.184
-0.9	$E_r - 0.273[r_t - r_f - E_r] + 0.65[q_t - E_q]$	0.082	0.184

4.2.4 Correlation between the Innovations of the Dividend Yield and the Shocks to the Unobservable Expected Stock Returns

What happens when the correlation between the innovations of the dividend yield and the shocks to the unobservable expected stock returns ($\rho_{v\varepsilon}$) varies? Figure 12, Panel A plots the mean wealth accumulation and consumption over the life cycle, and Figure 12, Panel B plots the mean share of wealth in stocks. When $\rho_{v\varepsilon}$ increases from 0.2 to 0.8, the mean wealth accumulation and consumption rarely change. The mean share of wealth in stocks, however, shows a positive correlation with $\rho_{v\varepsilon}$ before retirement. Table 6 shows that the predictability effect from the dividend yield becomes strong, while the mean reversion effect from the stock returns only slightly decreases. Therefore, as $\rho_{v\varepsilon}$ increases, the portfolio choices from the imperfect predictor model tend toward the predictions from the perfect predictor model.

TABLE 6

Conditional Expectation and Standard Deviation due to Variation of $\rho_{v\varepsilon}$

Table 6 presents the conditional moments of the unobservable expected stock return (μ_t) and the next period's stock return (r_{t+1}) for different correlations between the shocks to the unobservable expected stock returns and the innovations of dividend yield ($\rho_{v\varepsilon}$). $E(\mu_t|r_t, q_t)$ is the conditional expectation of the current unobservable expected stock return based on the current observed return (r_t) and the dividend yield (q_t), $\sigma_{\mu_t|r_t, q_t}$ is the conditional standard deviation of the current unobservable expected stock return based on the current observed stock return (r_t) and the dividend yield (q_t), and $\sigma_{r_{t+1}|r_t, q_t}$ is the conditional standard deviation of next period's return based on the current observed return (r_t) and the dividend yield (q_t).

$\rho_{v\varepsilon}$	$E(\mu_t r_t, q_t)$	$\sigma_{\mu_t r_t, q_t}$	$\sigma_{r_{t+1} r_t, q_t}$
0.2	$E_r - 0.28 [r_t - r_f - E_r] - 0.10 [q_t - E_q]$	0.0831	0.1844
0.4	$E_r - 0.28 [r_t - r_f - E_r] + 0.39 [q_t - E_q]$	0.0828	0.1843
0.56(benchmark)	$E_r - 0.27 [r_t - r_f - E_r] + 0.79 [q_t - E_q]$	0.0820	0.1839
0.8	$E_r - 0.27 [r_t - r_f - E_r] + 1.37 [q_t - E_q]$	0.0798	0.1830

5 Optimal TDFs

Financial advisors argue that the share of wealth in stocks should decrease as the investor approaches retirement and also quantify this as what the i.i.d. stock returns model predicts. The target date fund (TDF) using the results from the i.i.d. stock returns model has therefore become quite a popular financial advice, commonly recommended by large financial advisors like Vanguard TDFs. When the stock returns are predictable, however, the share of wealth in stocks should change according to market timing. This is what the perfect predictor model predicts. Retrospecting the financial crisis in 1929, 1997, 2001 and 2008, blindly following the rules suggested by life style funds for households entering retirement would not have been sound investment advice. Hence, the enhanced TDF (eTDF) has been proposed to take advantage of changing market conditions and expectations. This paper, however, shows that it is not easy to take advantage of changing market conditions and expectations. When the predictor is imperfect, the investment decision from the perfect predictor model seems over optimistic. But, what is the quantitative magnitude of investor welfare from investment

rules given by the different models? In this section I evaluate the welfare loss of the investor with respect to the perfect predictor model, the i.i.d. stock returns model and the Vanguard TDF model when the stock market is modeled as the imperfect predictive system.

5.1 Welfare Evaluation

To measure welfare changes I use the value functions across different models. Given that I have solved for saving, portfolio choices and value functions for all periods in the life cycle, I know that the value functions at a particular age are a sufficient statistic for welfare effects. Let $V_0(x_{i,t}, \hat{\mu}_{t|t})$ be the value function from the baseline model, and $V_n(x_{i,t}, f_t)$ be the value function from the alternative model such as the perfect predictor model or the i.i.d. stock returns model or the Vanguard TDF model. In these notations, $\hat{\mu}_{t|t}$ is the conditional expectation of unobservable expected stock return and f_t is the observed state factor. The f_t in perfect predictor model or Vanguard TDF model is the dividend yield. In contrast, f_t in i.i.d. stock returns model is a null variable because the policy functions are all the same for the different dividend yield.

I compute consumption certainty equivalent for a particular age as the follows:

$$E \left\{ \left[\frac{V_n(x_{i,t}, f_t)}{V_0(x_{i,t}, \hat{\mu}_{t|t})} \right]^{1/(1-\gamma)} - 1 \right\} \quad (18)$$

where $i \in I_{age}$ and $x_{i,t}$ is the same in both V_0 and V_n . This definition is the consumption certainty equivalent because I convert the change of the value into the dimension of expenditure before taking the unconditional expectation. Moreover, this consumption certainty equivalent is computed when stock returns are simulated based on the imperfect predictive system.

Figure 13 plots the consumption certainty equivalent of the different models relative to the baseline mode over the life cycle when changing $\rho_{v\varepsilon}$ from 0.2 to 0.8. Panel A illustrates substantial welfare loss from following the strategy predicted by the perfect

predictor model relative to using the optimal investment policy given by the imperfect predictor model. Panel B shows that the welfare losses are even more substantial from following the i.i.d. stock returns model, and Panel C reports that welfare loss from taking the investment rules from the Vanguard TDF model is in the middle.

Several observations can be drawn based on Figure 13. First, the welfare losses are economically significant: they vary from 2% to 4% of consumption equivalent when the investor follows the strategy from the perfect predictor model, from 5% to 11% when the investor follows the strategy from the i.i.d. stock returns model and from 2% to 6% when investor follows the strategy from the Vanguard TDF model. Second, these welfare losses positively correlated with $\rho_{v\varepsilon}$. When the predictor is close to perfect positive ($\rho_{v\varepsilon} \approx 1$), the welfare loss becomes even larger. Third, the welfare losses from the i.i.d. stock returns model and perfect predictor model tend to get maximized at around age 50, whereas average wealth accumulation is maximized at the exogenous retirement age of 65. On the contrary, the welfare losses from the Vanguard TDF model has a peak at around age 70.

To better understand these welfare shapes and magnitudes, it is helpful to recall that given the preference for consumption smoothing, welfare is increasing in average consumption and decreasing in the volatility of consumption. I can therefore obtain an insight on where the welfare differences are coming from by comparing the mean change of consumption and the change of standard deviation of consumption over the life cycle. To do so, I define the average change of consumption for a particular age as $E_t \left(\frac{C_1 - C_2}{C_2} \right)$, where C_1 is the consumption stream from the first model and C_2 is from the second model, and the change of standard deviation of consumption as $\frac{SD(C_1) - SD(C_2)}{SD(C_2)}$.

Figure 14 plots the mean change of consumption and the change of standard deviation of consumption for the baseline model relative to the perfect predictor model (Panels A and B), the i.i.d. stock returns model (Panels C and D) and the Vanguard TDF model (Panels E and F). The I.I.D model produces the largest change in consumption volatility over the working life. Given the preferences for smoother consumption,

this increased consumption inequality translates into a welfare loss that essentially gets maximized at mid life (around age 50), justifying the peak in welfare losses depicted in Figure 14.

The perfect predictor model on the other hand generates a lower mean consumption change over the life cycle. Since the portfolio rules are more stable than i.i.d. stock returns model, consumption variability is actually lower with the perfect predictor model relative to the i.i.d. stock returns model but higher relative to the baseline model.

When compared to the Vanguard TDF model, the welfare loss approaches the peak at about age 70 (see Figure 14, Panel E) because the mean change of consumption reaches the summit at age 70 (Figure 14, Panel E and F).

6 Conclusions

In this paper, I jointly analyze the implications of an imperfect predictive system, undiversifiable labor income risk and exogenously imposed liquidity constraints on optimal consumption and portfolio decisions over the life cycle. In the presence of an imperfect predictor of the unobservable expected stock returns, the optimal portfolio choice is shown to be more conservative than that predicted by an i.i.d. and perfect predictor model when calibrated to the observed data from 1946 to 2015. Different from Wachter and Yogo (2010) which use the nonhomothetic utility over basic and luxury goods to generate balanced portfolios, this paper generates the balanced portfolios through introducing the imperfection to the predictor of stock returns. Compared with the SCF 2007, the imperfect predictor model matches the data better than the i.i.d. stock returns model and the perfect predictor model. Moreover, when the imperfection is introduced into the perfect predictor model, almost all correlations ($\rho_{z\varepsilon}, \rho_{zv}, \rho_{zn}, \rho_{v\varepsilon}$) become important, which is different from one of the conclusion of Michaelides and Zhang (Forthcoming). Hence, a financial advisor should pay more attention to these correlations when giving investment advice.

To measure the benefits of taking the imperfect predictor into account, I compare

the welfare loss of the perfect predictor, i.i.d. and Vanguard TDF model relative to baseline model. The largest welfare loss is obtained from following the rules predicted by the i.i.d. stock returns model. The perfect predictor model has the smallest welfare loss, and the Vanguard TDF model is in the middle. Hence, an investment strategy uniquely relying on a single information source or the unconditional expected stock return leads to an incorrect investment decision and substantial welfare loss.

Future directions of research include the explicit introduction of ambiguity aversion in preferences, ambiguity in the parameters such as the risk premium and persistence of the unobservable expected return process, Bayesian posterior distributions for the parameters (Barberis (2000), Xia (2001) and Pastor and Stambaugh (2009)), a stochastic volatility in stock returns and an explicit model of housing. All these extensions will require additional computational power to achieve the desired required solution accuracy, but will further improve our understanding of life cycle portfolio choice under uncertainty and provide reasonable advice to billions of households increasingly making their own financial decisions.

7 Appendix

7.1 A Theorem of Multivariate Normal Distribution

Theorem 7.1 (Tsay (2010), Ch11): Suppose that x and y are two random vectors such that their joint distribution is multivariate normal. In addition, assume that the diagonal block covariance matrix Σ_{ww} is non-singular for $w = x, y$. Then,

$$1. E(x|y) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$$

$$2. Var(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{yx}$$

This theorem provides us with an algorithm of filtering the unobservable state, x , from the observable variables y .

7.2 Definitions and Notations

- μ_t is the unobservable expected stock return
- q_t is the observable dividend yield.
- E_r is the unconditional expectation of μ_t ;
- E_q is the unconditional expectation of q_t ;
- $d_t = \begin{pmatrix} r_t \\ q_t \end{pmatrix}$, $D_t = (d_0, d_1, \dots, d_t)$, where D_t is the full history of d_t ;
- $e_t = E(d_t|\mu_t, D_{t-1})$ is the expectation of observable d_t conditional on the current unobservable μ_t and historical observed values of d_t up to time $t - 1$;
- $\hat{\mu}_{t|t-1} = E(\mu_t|D_{t-1})$ is the expectation of unobservable μ_t conditional on the historical observed values of d_t up to time $t - 1$;
- $\hat{d}_{t|t-1} = E(d_t|D_{t-1})$ is the expectation of observable d_t conditional on the historical observed values of d_t up to time $t - 1$;

- $\hat{\mu}_{t|t} = E(\mu_t|D_t)$ is the estimation of unobservable μ_t conditional on the historical observed values of d_t up to time t ;
- $G_t = Cov(d_t, \mu_t|D_{t-1})$ is the covariance between the observable variables d_t and the unobservable μ_t conditional on the historical observed data up to time $t - 1$.
- $P_t = Var(\mu_t|D_{t-1})$ is the variance of unobservable μ_t conditional on the historical observed data up to $t - 1$.
- $R_t = Var(d_t|\mu_t, D_{t-1})$ is the variance of d_t conditional on the current unobservable expected stock return (μ_t) and the historical observed data up to time $t - 1$.
- $Q_t = Var(\mu_t|D_t)$ is the variance of the unobservable μ_t conditional on the historical observed data up to t .
- $S_t = Var(d_t|D_{t-1})$ is the variance of d_t conditional on the historical observed data up to time $t - 1$.

7.3 Maximum Likelihood Estimation

Since the expected stock return (μ_t) is unobservable, the classical maximum likelihood estimation needs a modification that rewrites the VAR (24) to make it only involve the observable variables, the stock returns (r_t) and the predictor (q_t). For doing this modification, I follow the same idea of Pastor and Stambaugh (2009) and use this method to estimate the parameters of the imperfect predictive system conditional on all the observed data. The first step is to set up the recursive formula to update conditional moments of the unobservable expected stock return using the observed data.

7.3.1 Starting the Recursion

The recursion begins with $\mu_0|D_0$, where D_0 is the empty set. At time 0, since there are no observations of d_t , I use the unconditional mean of μ_t as the estimation of

unobservable expected stock returns, $\hat{\mu}_{0|0} = E(\mu_0|D_0) = E_r$. Similarly, the $Q_0 = Var(\mu_0|D_0) = V_{\mu\mu}$ is the unconditional variance of μ_t .

At time 0, I can predict the moments of d_1 conditional on D_0 . Assuming that the process of μ_t is stationary, I obtain $E(\mu_1|D_0) = \alpha_q + \phi_\mu E(\mu_0|D_0)$. This implies:

$$\hat{\mu}_{1|0} = E(\mu_1|D_0) = E_r;$$

$$P_1 = Var(\mu_1|D_0) = \Sigma_{\mu\mu};$$

and

$$\hat{d}_{1|0} = E(d_1|D_0) = \begin{pmatrix} 0 \\ \alpha_q \end{pmatrix} + E\left(\begin{bmatrix} \mu_0 \\ q_0 \end{bmatrix} | D_0\right) \implies$$

$$\hat{d}_{1|0} = \begin{pmatrix} E_r \\ E_q \end{pmatrix} \text{ and } S_1 = Var(d_1|D_0) = \begin{bmatrix} \Sigma_{rr} & \Sigma_{rq} \\ \Sigma_{qr} & \Sigma_{qq} \end{bmatrix}.$$

As $(d_1|D_0)$ and $(\mu_1|D_0)$ follow bi-variate normal distribution, I can obtain the following through applying theorem 7.1;

$$e_1 = E(d_1|\mu_1, D_0) = \begin{pmatrix} E_r \\ E_q \end{pmatrix} + \begin{bmatrix} \Sigma_{r\mu} \\ \Sigma_{q\mu} \end{bmatrix} \Sigma_{\mu\mu}^{-1} (\mu_1 - \hat{\mu}_{1|0})$$

and

$$R_1 = Var(d_1|\mu_1, D_0) = \begin{bmatrix} \Sigma_{rr} & \Sigma_{rq} \\ \Sigma_{qr} & \Sigma_{qq} \end{bmatrix} - \begin{bmatrix} \Sigma_{r\mu} \\ \Sigma_{q\mu} \end{bmatrix} \Sigma_{\mu\mu}^{-1} \begin{bmatrix} \Sigma_{r\mu} & \Sigma_{q\mu} \end{bmatrix}$$

Hence, I find the formulae for e_1 and R_1 as follows:

$$e_1 = \hat{d}_{1|0} + G_1 P_1^{-1} (\mu_1 - \hat{\mu}_{1|0}) \tag{19}$$

$$\text{where } G_1 = Cov(d_1, \mu_1|D_0) = \begin{bmatrix} \Sigma_{r\mu} \\ \Sigma_{q\mu} \end{bmatrix}$$

and

$$R_1 = S_1 - G_1 P_1^{-1} G_1' \tag{20}$$

where $S_1 = \begin{bmatrix} \Sigma_{rr} & \Sigma_{rq} \\ \Sigma_{qr} & \Sigma_{qq} \end{bmatrix}$ and $P_1^{-1} = \Sigma_{\mu\mu}^{-1}$

7.3.2 Updating via Bayes Theorem

Recalling the conditional probability formula, I can obtain the following equation, $f(\mu_1|d_1, D_0) = \frac{f(d_1|\mu_1, D_0)f(\mu_1|D_0)}{f(d_1|D_0)}$. Since $D_1 = (d_1, D_0)$, I have a probability density function as follows.

$$f(\mu_1|d_1, D_0) \propto f(d_1|\mu_1, D_0) f(\mu_1|D_0) \quad (21)$$

Now, I expand the right hand side of equation (21) as follows

$$\begin{aligned} & f\left(\begin{bmatrix} r_1 & q_1 \end{bmatrix} | \mu_1, D_0\right) f(\mu_1|D_0) = \\ & \left(\frac{1}{\sqrt{2\pi}}\right)^2 |P_1|^{-1} |R_1|^{-1} \exp\left\{-\frac{1}{2} [(d_1 - e_1)' R^{-1} (d_1 - e_1) + (\mu_1 - \hat{\mu}_{1|0})' P^{-1} (\mu_1 - \hat{\mu}_{1|0})]\right\} \\ & \propto \exp\left\{-\frac{1}{2} \left(d_1 - \hat{d}_{1|0} + G_1 P_1^{-1} (\mu_1 - \hat{\mu}_{1|0})\right)' R^{-1} \left(d_1 - \hat{d}_{1|0} + G_1 P_1^{-1} (\mu_1 - \hat{\mu}_{1|0})\right) \right. \\ & \quad \left. + (\mu_1 - \hat{\mu}_{1|0})' P^{-1} (\mu_1 - \hat{\mu}_{1|0})\right\} \end{aligned}$$

After rearranging the terms and ignoring the irrelevant quadratic terms of d_1 , I can obtain the follows:

$$\begin{aligned} & f\left(\begin{bmatrix} r_1 & q_1 \end{bmatrix} | \mu_1, D_0\right) f(\mu_1|D_0) \propto \\ & \exp\left\{-\frac{1}{2} \left[\left(d_1 - \hat{d}_{1|0}\right)' G_1 P_1^{-1} R^{-1} (\mu_1 - \hat{\mu}_{1|0}) + (\mu_1 - \hat{\mu}_{1|0})' (P_1^{-1})' G_1' R^{-1} \left(d_1 - \hat{d}_{1|0}\right) \right. \right. \\ & \quad \left. \left. + (\mu_1 - \hat{\mu}_{1|0})' (P_1^{-1})' G_1' R^{-1} G_1 P_1^{-1} (\mu_1 - \hat{\mu}_{1|0}) + (\mu_1 - \hat{\mu}_{1|0})' P^{-1} (\mu_1 - \hat{\mu}_{1|0}) \right] \right\}. \end{aligned}$$

Since μ_t , P_t and $\left(d_1 - \hat{d}_{1|0}\right)' G_1 P_1^{-1} R^{-1}$ are scalars, I can rewrite this formula as

$$\begin{aligned} & f\left(\begin{bmatrix} r_1 & q_1 \end{bmatrix} | \mu_1, D_0\right) f(\mu_1|D_0) \propto \\ & \exp\left\{-\frac{1}{2} \left[2P_1^{-1} G_1' R^{-1} \left(d_1 - \hat{d}_{1|0}\right) (\mu_1 - \hat{\mu}_{1|0}) + \right. \right. \\ & \quad \left. \left. \left(P_1^{-1} + (P_1^{-1})' G_1' R^{-1} G_1 P_1^{-1}\right) (\mu_1 - \hat{\mu}_{1|0})^2 \right] \right\} \\ & = \exp\left\{-\frac{1}{2} \left[2P_1^{-1} G_1' R^{-1} \left(d_1 - \hat{d}_{1|0}\right) (\mu_1 - \hat{\mu}_{1|0}) \right. \right. \\ & \quad \left. \left. + P_1^{-1} P_1 \left(P_1^{-1} + (P_1^{-1})' G_1' R^{-1} G_1 P_1^{-1}\right) P_1 P_1^{-1} (\mu_1 - \hat{\mu}_{1|0})^2 \right] \right\} \\ & = \exp\left\{-\frac{1}{2} \left[2P_1^{-1} G_1' R^{-1} \left(d_1 - \hat{d}_{1|0}\right) (\mu_1 - \hat{\mu}_{1|0}) + \right. \right. \\ & \quad \left. \left. P_1^{-1} (P_1 + G_1' R^{-1} G_1) P_1^{-1} (\mu_1 - \hat{\mu}_{1|0})^2 \right] \right\} \end{aligned}$$

$$\propto \exp \left\{ - \frac{[\mu_1 - \hat{\mu}_{1|0} - P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1} (d_1 - \hat{d}_{1|0})]^2}{2 P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} P_1} \right\}.$$

This is the kernel of the normal distribution again. Hence, after the Bayesian update, conditional probability density of μ_t is still a normal distribution, and its conditional moments are:

$$\hat{\mu}_{1|1} = E(\mu_1 | D_1) = \hat{\mu}_{1|0} + P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} G_1' R_1^{-1} (d_1 - \hat{d}_{1|0}) \quad (22)$$

and

$$Var(\mu_1 | D_1) = Q_1 = P_1 (P_1 + G_1' R_1^{-1} G_1)^{-1} P_1 \quad (23)$$

For finding all conditional probability densities and moments of μ_t for $t = 2, \dots, T$, I rewrite the equations (4) - (6) as follows:

$$\begin{bmatrix} r_t - r_f - E_r \\ q_t - E_q \\ \mu_t - E_r \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \phi_q & 0 \\ 0 & 0 & \phi_\mu \end{bmatrix} \begin{bmatrix} r_{t-1} - r_f - E_r \\ q_{t-1} - E_q \\ \mu_{t-1} - E_r \end{bmatrix} + \begin{bmatrix} z_t \\ v_t \\ \varepsilon_t \end{bmatrix} \quad (24)$$

Assuming the VAR represented by (24) is stationary, I obtain the following:

$$E(\mu_t - E_r | D_{t-1}) = \phi_\mu E(\mu_{t-1} - E_r | D_{t-1}) \implies$$

$$E(\mu_t | D_{t-1}) = (1 - \phi_\mu) E_r + \phi_\mu E(\mu_{t-1} | D_{t-1}).$$

Since the μ_t is unobservable, I can not delete the $E(\cdot)$ in $E(\mu_{t-1} | D_{t-1})$. Simplifying the formula above, I rewrite it as:

$$\hat{\mu}_{t|t-1} = (1 - \phi_\mu) E_r + \phi_\mu \hat{\mu}_{t-1|t-1} \quad (25)$$

Similarly, I can rewrite $\hat{d}_{t|t-1}$ as the following:

$$\hat{d}_{t|t-1} = E(d_t | D_{t-1}) = \begin{bmatrix} \hat{\mu}_{t-1|t-1} \\ (1 - \phi_q) E_q + \phi_q q_{t-1} \end{bmatrix} \quad (26)$$

and

$$P_t = Var(\mu_t|D_{t-1}) = \phi_\mu^2 Var(\mu_{t-1}|D_{t-1}) + \sigma_\varepsilon^2 = \phi_\mu^2 Q_{t-1} + \sigma_\varepsilon^2 \quad (27)$$

Next, for getting the other update formulae, I take the variance on both sides of the VAR (24).

$$\begin{aligned} & Var \left(\begin{bmatrix} r_t \\ q_t \\ \mu_t \end{bmatrix} \middle| D_{t-1} \right) = \\ & \begin{bmatrix} 0 & 0 & 1 \\ 0 & \phi_q & 0 \\ 0 & 0 & \phi_\mu \end{bmatrix} Var \left(\begin{bmatrix} r_{t-1} - E_r \\ q_{t-1} - E_q \\ \mu_{t-1} - E_r \end{bmatrix} \middle| D_{t-1} \right) \begin{bmatrix} 0 & 0 & 1 \\ 0 & \phi_q & 0 \\ 0 & 0 & \phi_\mu \end{bmatrix} + Var \left(\begin{bmatrix} z_t \\ v_t \\ \varepsilon_t \end{bmatrix} \middle| D_{t-1} \right) \\ & = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \phi_x & 0 \\ 0 & 0 & \phi_\mu \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q_{t-1} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & \phi_q & 0 \\ 0 & 0 & \phi_\mu \end{bmatrix} + \begin{bmatrix} \sigma_z^2 & \rho_{zv}\sigma_z\sigma_v & \rho_{z\varepsilon}\sigma_z\sigma_\varepsilon \\ & \sigma_v^2 & \rho_{v\varepsilon}\sigma_v\sigma_\varepsilon \\ & & \sigma_\varepsilon^2 \end{bmatrix} \\ & = \begin{bmatrix} S_t & G_t \\ G'_t & P_t \end{bmatrix} \\ & \text{Hence, I obtain} \end{aligned}$$

$$e_t = E(d_t|\mu_t, D_{t-1}) = \hat{d}_{t|t-1} + G_t P_t^{-1} (\mu_t - \hat{\mu}_{t|t-1}) \quad (28)$$

$$R_t = Var(d_t|\mu_t, D_{t-1}) = S_t - G_t P_t^{-1} G'_t \quad (29)$$

and, the conditional moments of μ_t

$$\hat{\mu}_{t|t} = E(\mu_t|D_t) = \hat{\mu}_{t|t-1} + P_t (P_t + G'_t R_t^{-1} G_t)^{-1} G'_t R_t^{-1} (d_t - \hat{d}_{t|t-1}) \quad (30)$$

$$Q_t = Var(\mu_t|D_t) = P_t (P_t + G'_t R_t^{-1} G_t)^{-1} P_t \quad (31)$$

7.3.3 Maximum Likelihood Estimation of Parameter

Denote $[m_t, n_t] = P_t (P_t + G_t' R_t^{-1} G_t)^{-1} G_t' R_t^{-1}$. Then, according to (30), I have

$$\begin{aligned} \hat{\mu}_{t|t} &= \hat{\mu}_{t|t-1} + [m_t, n_t] \left[\begin{array}{c} r_t - r_f - \hat{\mu}_{t-1|t-1} \\ x_t - E_{t-1}(q_t) \end{array} \right] \\ &= (1 - \phi_\mu) E_r + \phi_\mu \hat{\mu}_{t-1|t-1} + m_t (r_t - r_f - \hat{\mu}_{t-1|t-1}) + n_t v_t \end{aligned}$$

The last equality hold because $\hat{\mu}_{t|t-1} = (1 - \phi_\mu) E_r + \phi_\mu \hat{\mu}_{t-1|t-1}$ (equation 25).

Then, I rewrite this formula as follows.

$$\hat{\mu}_{t|t} = (1 - \phi_\mu) E_r + (\phi_\mu - m_t) \hat{\mu}_{t-1|t-1} + m_t (r_t - r_f) + n_t v_t \quad (32)$$

Next, I define the forecast error of $r_{t+1} - r_f$ conditional on information at time t as

$$\omega_{t+1} = (r_t - r_f) - E_t (r_t - r_f) \quad (33)$$

Since $r_{t+1} - r_f = \mu_t + z_{t+1}$ (equation (5)), I have $E_t (r_{t+1} - r_f) = E_t (\mu_t) = \hat{\mu}_{t|t}$.

Now, replacing $E_t (r_t - r_f)$ with $\hat{\mu}_{t|t}$ in (33), I obtain $\omega_{t+1} = r_{t+1} - r_f - \hat{\mu}_{t|t}$.

Rearrange this equation, I obtain as the following:

$$r_{t+1} - r_f = (1 - \phi_\mu) E_r + \phi_\mu (r_t - r_f) + n_t v_t - (\phi_\mu - m_t) \omega_t + \omega_{t+1} \quad (34)$$

Combining (34) with (6), I obtain a new equation system consisting of only observable data as follows:

$$\begin{cases} r_{t+1} - r_f = (1 - \phi_\mu) E_r + \phi_\mu (r_t - r_f) + n_t v_t - (\phi_\mu - m_t) \omega_t + \omega_{t+1} \\ q_{t+1} = \alpha_q + \phi_q q_t + v_{t+1} \end{cases} \quad (35)$$

In steady state, I can delete the time index of m_t and n_t and rewrite the equation system (35) as the following VARMA(1,1) model.

$$\begin{aligned} \begin{bmatrix} r_t - r_f \\ q_t \end{bmatrix} - \begin{bmatrix} E_r \\ E_q \end{bmatrix} &= \begin{bmatrix} \phi_\mu & 0 \\ 0 & \phi_q \end{bmatrix} \left(\begin{bmatrix} r_{t-1} - r_f \\ q_{t-1} \end{bmatrix} - \begin{bmatrix} E_r \\ E_q \end{bmatrix} \right) + \\ & \begin{bmatrix} -(\phi_\mu - m) & n \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \omega_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \omega_t \\ v_t \end{bmatrix} \end{aligned} \quad (36)$$

Now, define the state variable as $\xi_t = (\omega_t, v_t, \omega_{t-1}, v_{t-1})'$. This VARMA(1,1) model can be rewritten as a state space model consisting of a observation equation and a state equation and estimated by the Kalman Filter algorithm.

$$\begin{cases} d_t^* = Ad_{t-1}^* + H\xi_t \\ \xi_t = F\xi_{t-1} + \begin{bmatrix} e_t \\ 0 \end{bmatrix} \end{cases} \quad (37)$$

$$\text{where } d_t^* = d_t - [E_r, E_q]', \quad A = \begin{bmatrix} \phi_\mu & 0 \\ 0 & \phi_q \end{bmatrix}, \quad H_{12} = \begin{bmatrix} -(\phi_\mu - m) & n \\ 0 & 0 \end{bmatrix},$$

$$F = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times 2} \\ I_{2 \times 2} & 0_{2 \times 2} \end{bmatrix}, \quad H = \begin{bmatrix} I_{2 \times 2} & H_{12} \end{bmatrix}, \quad e_t = [\omega_t, v_t]', \quad \text{and } \text{Var}(e_t) = \Sigma^*.$$

Given a sample of d_t , the joint likelihood function of the state space model (37) is $L = \prod_{t=1}^T f(d_t^* | d_{t-1}^*)$, where $f(d_t^* | d_{t-1}^*)$ is the conditional probability density of d_t^* . I, therefore, can estimate the parameters by maximizing the following log-likelihood function¹¹:

$$-2\ln(L) = \sum_{t=1}^T \left(\ln|V_{t|t-1}| + [d_t^* - \hat{d}_{t|t-1}^*]' V_{t|t-1}^{-1} [d_t^* - \hat{d}_{t|t-1}^*] \right) \quad (38)$$

The terms in equation (38) are defined as the following:

- $\text{Var}(r_t) = \sigma_r^2 = (1 - \phi_\mu)^{-1} [n\sigma_v^2 + (1 - \phi_\mu^2 + m^2)\sigma_\omega^2 + 2m\sigma_{v\omega}n]$ ¹²;

¹¹This log-likelihood function (38) is based on the logarithm of the equation (13.4.1) from Hamilton (1994)

¹²Taking variance on both sides of (34), I obtain $\sigma_r^2 = \phi_\mu^2 \sigma_r^2 - 2\phi_\mu(\phi_\mu - m)\sigma_\omega^2 + (\phi_\mu - m)^2 \sigma_\omega^2 +$

- $Cov(r_t, q_t) = \sigma_{qr} = (1 - \phi_\mu \phi_q)^{-1} [\phi_q \sigma_v^2 n + (1 - (\phi_\mu - m)\phi_q) \sigma_{v\omega}]$ ¹³;
- $Var(q_t) = \sigma_q^2 = (1 - \phi_q^2)^{-1} \sigma_v^2$;
- $\hat{d}_{t|t-1}^* = E_{t-1}(d_t^*) = Ad_{t-1}^* + H_{12} \Sigma^* V_{t-1|t-2}^{-1} [d_{t-1}^* - \hat{d}_{t-1|t-2}^*]$; and
- $V_{t|t-1} = V_{t-1}(d_t^*) = H_{12} \left(\Sigma^* - \Sigma^* V_{t-1|t-2}^{-1} \Sigma^* \right) H_{12}' + \Sigma^*$

The formula of updating $\hat{d}_{t|t-1}^*$ is from taking $E_{t-1}(\cdot)$ on the both sides of the observation equation of (37).

$$\begin{aligned}
\hat{d}_{t|t-1}^* &= E_{t-1}(d_t^*) = Ad_{t-1}^* + H \hat{\xi}_{t|t-1} \\
&= Ad_{t-1}^* + HF \hat{\xi}_{t-1|t-1} \\
&= Ad_{t-1}^* + \begin{bmatrix} H_{12} & 0_{2 \times 2} \end{bmatrix} \left(\hat{\xi}_{t-1|t-2} + \Sigma_{\xi d^*} V_{t-1|t-2}^{-1} [d_{t-1}^* - \hat{d}_{t-1|t-2}^*] \right) \\
&= Ad_{t-1}^* + \begin{bmatrix} H_{12} & 0_{2 \times 2} \end{bmatrix} \begin{pmatrix} \hat{\omega}_{t-1|t-2} \\ \hat{v}_{t-1|t-2} \\ \hat{\omega}_{t-2|t-2} \\ \hat{v}_{t-2|t-2} \end{pmatrix} + \begin{bmatrix} H_{12} & 0_{2 \times 2} \end{bmatrix} \Sigma_{\xi d^*} V_{t-1|t-2}^{-1} [d_{t-1}^* - \hat{d}_{t-1|t-2}^*].
\end{aligned}$$

As $\begin{bmatrix} \hat{\omega}_{t-1|t-2} \\ \hat{v}_{t-1|t-2} \end{bmatrix} = 0$ and $\begin{bmatrix} H_{12} & 0_{2 \times 2} \end{bmatrix} \Sigma_{\xi d^*} = H_{12} \Sigma^*$, this equation can be simplified as:

$$\hat{d}_{t|t-1}^* = Ad_{t-1}^* + H_{12} \Sigma^* V_{t-1|t-2}^{-1} [d_{t-1}^* - \hat{d}_{t-1|t-2}^*].$$

As to $V_{t|t-1}$, the update formula is from taking $Var(\cdot)$ on both sides of state equation of (37).

$$\begin{aligned}
V_{t|t-1} &= V_{t-1}(d_t^*) = HF Var_{t-1}(\xi_t) F' H' + HF Var \begin{pmatrix} e_t \\ 0 \end{pmatrix} F' H' \\
&= \begin{bmatrix} I_{2 \times 2} & H_{12} \end{bmatrix} \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times 2} \\ I_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} \left(\Sigma_{\xi \xi} - \Sigma_{\xi d} V_{t-1|t-2}^{-1} \Sigma_{d \xi}' \right) \begin{bmatrix} 0_{2 \times 2} & I_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} \begin{bmatrix} I_{2 \times 2} \\ H_{12} \end{bmatrix} \\
&+ \begin{bmatrix} I_{2 \times 2} & H_{12} \end{bmatrix} \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times 2} \\ I_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} \Sigma^* \begin{bmatrix} 0_{2 \times 2} & I_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} \begin{bmatrix} I_{2 \times 2} \\ H_{12} \end{bmatrix}.
\end{aligned}$$

¹³ $n^2 \sigma_v^2 + \sigma_\omega^2 - 2(\phi_\mu - m)n\sigma_{v\omega} + 2\phi_\mu n\sigma_{v\omega}$. Solve this equation for σ_r^2

¹³Taking covariance between r_t and q_t based on (35), I obtain the equation $\sigma_{qr} = \phi_q \phi_\mu \sigma_{xr} + n\phi_q \sigma_v^2 - (\phi_\mu - m)\phi_q \sigma_{v\omega} + \sigma_{v\omega}$. Then, σ_{qr} is from solving this equation.

Since $\begin{bmatrix} I_{2 \times 2} & H_{12} \end{bmatrix} \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times 2} \\ I_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} = H_{12}$, I have:

$$V_{t|t-1} = H_{12} \left(\Sigma^* - \Sigma^* V_{t-1|t-2}^{-1} \Sigma^* \right) H'_{12} + \Sigma^*.$$

Last, the necessary initial values for solving the log-likelihood function (38) are:

- $\hat{d}_{1|0}^* = 0$; and

- $V_{1|0} = Var(d_t) = \begin{bmatrix} Var(r_t) & Cov(r_t, q_t) \\ & Var(q_t) \end{bmatrix}.$

7.3.4 Unidentified Covariance Matrix

Through the MLE, I can obtain the estimation of parameters in VARMA model (16).

Our goal, however, is to find the covariance matrix of equations (4) - (6), $\Omega = \begin{bmatrix} \sigma_z^2 & \sigma_{zv} & \sigma_{z\varepsilon} \\ & \sigma_v^2 & \sigma_{v\varepsilon} \\ & & \sigma_\varepsilon^2 \end{bmatrix}.$

Unfortunately, in Ω , σ_{vv} is the only term that can be identified. As to estimate the other terms in Ω , obviously, a good starting point is the moments because the moments of VARMA model (16) should equal to that of the original VAR model (24). The moments we need in VARMA model can be computed as follows:

- $\sigma_r^2 = (1 - \phi_\mu)^{-1} [n\sigma_v^2 + (1 - \phi_\mu^2 + m^2)\sigma_\omega^2 + 2m\sigma_{v\omega}n]$
- $\sigma_{qr} = (1 - \phi_\mu\phi_q)^{-1} [\phi_q\sigma_v^2n + (1 - (\phi_\mu - m)\phi_q)\sigma_{v\omega}]$
- $Cov(r_t, r_{t-1})$

$$= Cov(\phi_\mu(r_{t-1} - r_f) + nv_{t-1} - (\phi_\mu - m)\omega_{t-1} + \omega_t, r_{t-1} - r_f)$$

$$= \phi_\mu\sigma_r^2 + nCov(v_{t-1}, r_{t-1} - r_f) - (\phi_\mu - m)Cov(\omega_{t-1}, r_{t-1} - r_f)$$
- $Cov(r_{t-1}, q_t)$

$$= Cov(r_{t-1}, \phi_q q_{t-1} + v_t)$$

$$= \phi_q Cov(r_{t-1}, q_{t-1})$$

- $Cov(r_t, q_{t-1})$

$$= Cov(\phi_\mu r_{t-1} + nv_{t-1} - (\phi_\mu - m)\omega_{t-1} + \omega_t, q_{t-1})$$

$$= \phi_\mu \sigma_{rq} + nCov(v_{t-1}, q_{t-1}) - (\phi_\mu - m)Cov(\omega_{t-1}, q_{t-1})$$

On the other hand, based on the VAR model (24), I can write down the following linear equation system about the moments as follows:

$$\begin{bmatrix} \sigma_r^2 \\ \sigma_{qr} \\ Cov(r_t, r_{t-1}) \\ Cov(r_{t-1}, q_t) \\ Cov(r_t, q_{t-1}) \end{bmatrix} = \begin{bmatrix} 1/(1-\phi_\mu^2) & 1 & 0 & 0 & 0 \\ 0 & 0 & \phi_q/(1-\phi_\mu\phi_q) & 1 & 0 \\ \phi_\mu/(1-\phi_\mu^2) & 0 & 0 & 0 & 1 \\ 0 & 0 & \phi_q^2/(1-\phi_\mu\phi_q) & \phi_q & 0 \\ 0 & 0 & 1/(1-\phi_\mu\phi_q) & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_\varepsilon^2 \\ \sigma_z^2 \\ \sigma_{v\varepsilon} \\ \sigma_{zv} \\ \sigma_{z\varepsilon} \end{bmatrix}$$

Since both the VARMA (16) and VAR (24) models describe the same thing, their moments must be the same as each other. This linear equation system, therefore, becomes solvable, when I replace the right hand side of this equation with the corresponding calculation from the VARMA model (16).

The rank of this linear equation system is, however, four because the second row can be eliminated by the fourth row, which makes this linear equation system be reduced into:

$$\begin{cases} \sigma_r^2 = \sigma_\varepsilon^2 / (1 - \phi_\mu^2) + \sigma_z^2 \\ \sigma_{qr} = \phi_q \sigma_{v\varepsilon} / (1 - \phi_\mu \phi_q) + \sigma_{zv} \\ Cov(r_t, r_{t-1}) = \sigma_{z\varepsilon} + \phi_\mu \sigma_\varepsilon^2 / (1 - \phi_\mu^2) \\ Cov(r_t, q_{t-1}) = \sigma_{v\varepsilon} / (1 - \phi_\mu \phi_q) \end{cases} \quad (39)$$

(39) shows that this linear equation system has not unique solution but a solution space. For obtaining the solution space, I represent this linear equation system (39) with respect to $\sigma_{z\varepsilon}$, because this parameter is important in the estimation. Then, it becomes as follows:

$$\begin{cases} \sigma_z^2 = \sigma_r^2 - (\text{Cov}(r_t, r_{t-1}) - \sigma_{z\varepsilon}) / \phi_\mu \\ \sigma_\varepsilon^2 = (\text{Cov}(r_t, r_{t-1}) - \sigma_{z\varepsilon}) (1 - \phi_\mu^2) / \phi_\mu \end{cases} \cdot$$

s.t. $\rho_{zv} < 1$ and $\rho_{v\varepsilon} < 1$

References

- Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies* 20, 651–707.
- Angerer, X. and P. Lam (2009). Income risk and portfolio choice: An empirical study. *Journal of Finance* LXIV, 1037–1055.
- Balduzzia, P. and A. W. Lynch (1999). Transaction costs and predictability: some utility cost calculations. *Journal of Financial Economics* 52, 47–78.
- Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance* LV, 225–264.
- Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional euler equations approach. *Journal of Finance* 54, 1609–1646.
- Brennan, M. J., E. S. Schwartz, and R. Lagnado (1997). Strategic asset allocation. *Journal of Economic Dynamics and Control* 21, 1377–1403.
- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics* 18(2), 373–399.
- Campbell, J. Y., J. Cocco, F. Gomes, P. J. Maenhout, and L. M. Viceira (2001). Stock market mean reversion and the optimal allocation of a long lived investor. *European Finance Review* 5, 269–292.
- Campbell, J. Y. and R. J. Shiller (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1(3), 195–228.
- Campbell, J. Y. and R. J. Shiller (1988b). Stock-prices, earnings, and expected dividends. *Journal of Finance* 43(3), 661–676.
- Campbell, J. Y. and L. M. Viceira (1999). Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics* 114(2), 433–495.

- Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics* 112(1), 1–55.
- Cocco, J. and P. Lopes (2015). Reverse mortgage designs. *LBS/LSE mimeo*.
- Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005). Consumption and portfolio choice over the life cycle. *Review of Financial Studies* 18(2), 491–533.
- Cochrane, J. H. (2005). *Asset Pricing (Revised Edition)*. Princeton University Press.
- Davidoff, T. (2015). Can "high costs" justify weak demand for the home equity conversion mortgage? *Review of Financial Studies* 28(8), 2364–2398.
- Deaton, A. (1991). Saving and liquidity constraints. *Econometrica* 59, 1221–1248.
- Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics* 22(1), 3–25.
- Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25(1), 23–49.
- Gomes, F. and A. Michaelides (2005). Optimal life-cycle asset allocation: Understanding the empirical evidence. *Journal of Finance* 60(2), 869–904.
- Gomes, F., A. Michaelides, and V. Polkovnichenko (2009). Optimal savings with taxable and tax-deferred accounts. *Review of Economic Dynamics* 12, 718–735.
- Goyal, A. and I. Welch (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455–1508.
- Haliassos, M. and A. Michaelides (2003). Portfolio choice and liquidity constraints. *International Economic Review* 44, 143–177.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Heaton, J. and D. Lucas (2000). Portfolio choice in the presence of background risk. *The Economic Journal* 110, 1–26.

- Kim, T. S. and E. Omberg (1996). Dynamic nonmyopic portfolio behavior. *Review of Financial Studies* 9, 141–161.
- Lan, C. (2015). An out-of-sample evaluation of dynamic portfolio strategies. *Review of Finance* 19, 2359–2399.
- Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* LVI, 815–849.
- Michaelides, A. and Y. Zhang (Forthcoming). Stock market mean reversion and portfolio choice over the life cycle. *Journal of Financial and Quantitative Analysis*.
- Pastor, L. and R. F. Stambaugh (2009). Predictive systems: Living with imperfect predictors. *Journal of Finance* 64(4), 1583–1628.
- Pastor, L. and R. F. Stambaugh (2012). Are stocks really less volatile in the long run? *Journal of Finance* LXVII, 431–477.
- Samuelson, P. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics* 51, 239–246.
- Tauchen, G. and R. Hussey (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing-models. *Econometrica* 59(2), 371–396.
- Tsay, R. S. (2010). *Analysis of Financial Time Series, 3rd Edition*. Wiley.
- Viceira, L. M. (2001). Optimal portfolio choice for long-horizon investors with nontradable labor income. *Journal of Finance* 56, 433–470.
- Vissing-Jorgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy* 110, 825–853.
- Wacher, J. (2002). Optimal consumption and portfolio allocation with mean-reverting returns: An exact solution for complete markets. *Journal of Financial and Quantitative Analysis* 37, 63–91.

Wachter, J. A. and M. Yogo (2010). Why do household portfolio shares rise in wealth?
Review of Financial Studies 23, 3929–3965.

Xia, Y. H. (2001). Learning about predictability: The effects of parameter uncertainty
on dynamic asset allocation. *Journal of Finance* 56(1), 205–246.

FIGURE 1

Solution Space of MLE

Figure 1 presents the solution space from MLE method. Panel A shows a 3D graph of the relationship between ρ_{zv} , $\rho_{z\varepsilon}$ and $\rho_{v\varepsilon}$, where ρ_{zv} is the correlation between shocks to stock returns and innovations of predictor, $\rho_{v\varepsilon}$ is the correlation between innovations of predictor and unobservable expected return, and $\rho_{z\varepsilon}$ is the correlation between innovations of stock returns and unobservable expected return. Panel B is the projection of the 3-D graph in panel A on the plane of ρ_{zv} and $\rho_{z\varepsilon}$. Similarly, panel C project the solution space on the plane of $\rho_{z\varepsilon}$ and $\rho_{v\varepsilon}$.

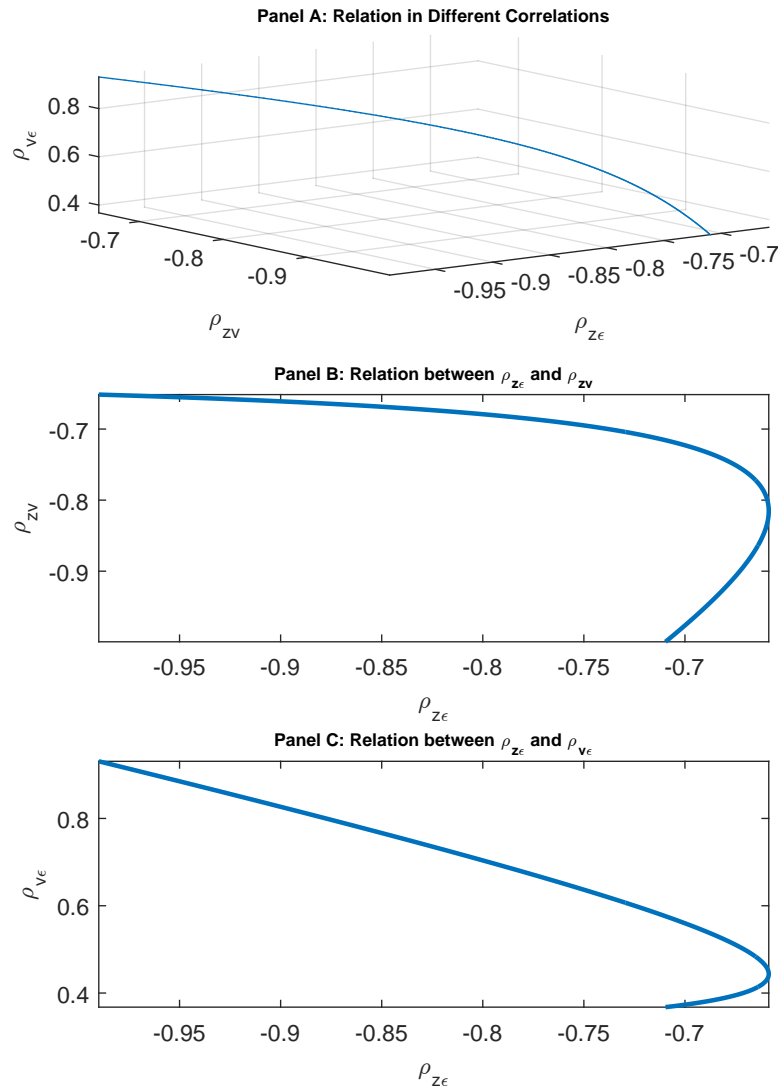


FIGURE 2

Mean Life-Cycle Profiles for Benchmark Model

Figure 2, panel A presents the mean wealth, consumption and labor income over the life cycle by simulating 6000 individual life histories. Panel B shows the mean share of wealth in stocks over the life cycle.

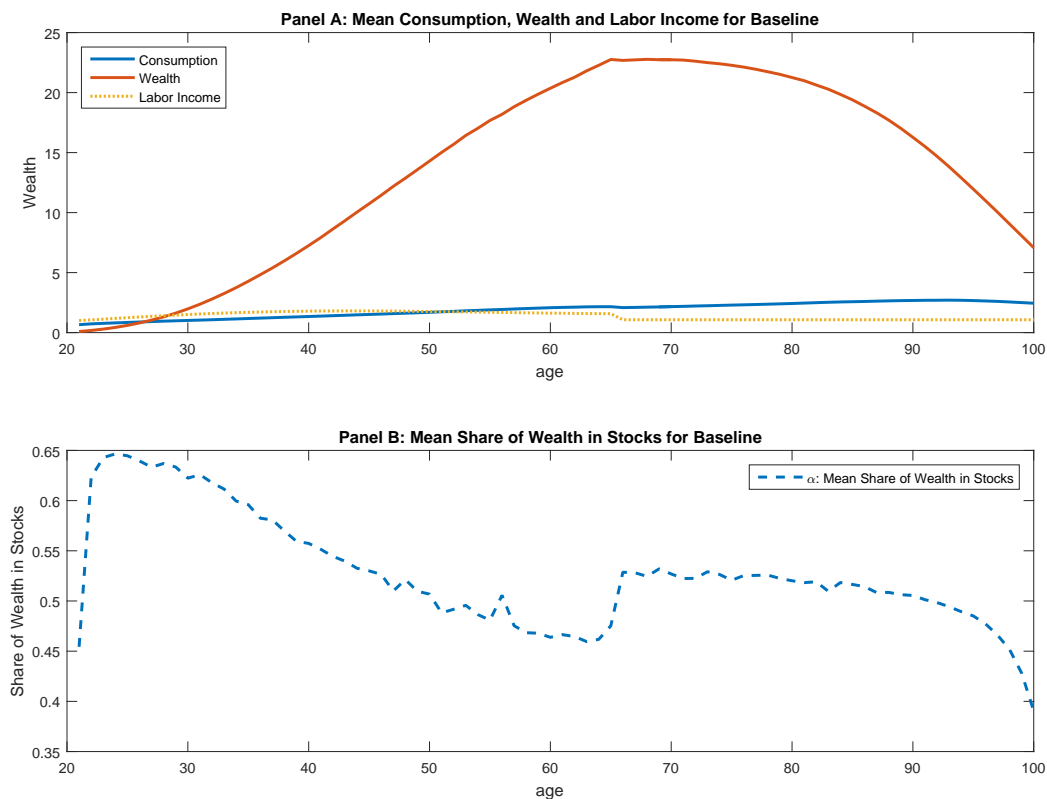


FIGURE 3

Comparison of Life-Cycle Profiles among Different Models

Figure 3, panel A presents the mean wealth over the life cycle for four different models, where PP is the perfect predictor model, i.i.d. is the i.i.d. stock returns model and Vanguard represents the Vanguard TDF model. Panel B and C describe the mean consumption and share of wealth in stocks for the corresponding models.

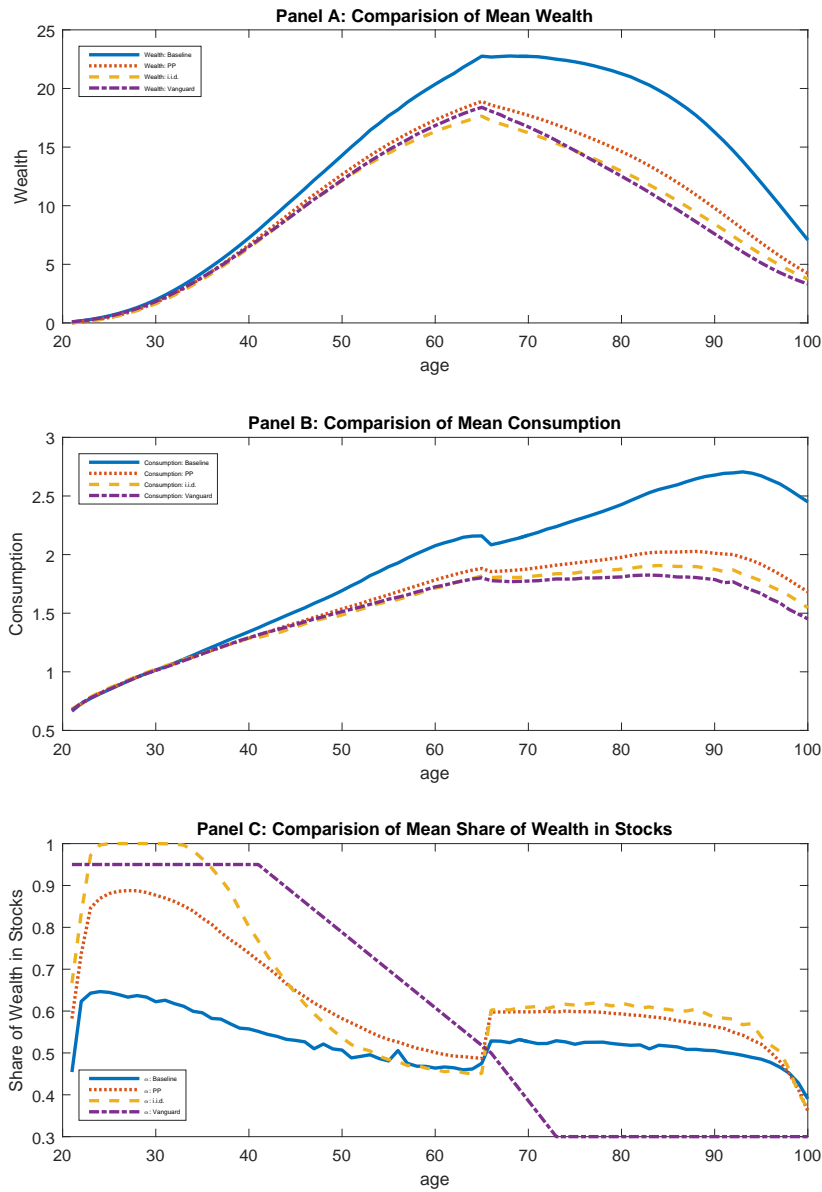


FIGURE 4

Investment Policy Function for Different Expected Stock Returns

Figure 4 presents the share allocation policy functions of different states, ages and models. The first row (Panel A, B and C) describes the share allocation policy functions between the baseline model (imperfect predictor model), perfect predictor model (PP model) and i.i.d. stock returns model for a 25-year-old investor when the estimations of the expected stock return is low, median and high respectively. The second row (Panel D, E and F) displays the share allocation policy functions for a 55-year-old agent, and the last row (Panel G, H and I) shows the share allocation policy functions for a 75-year-old agent.

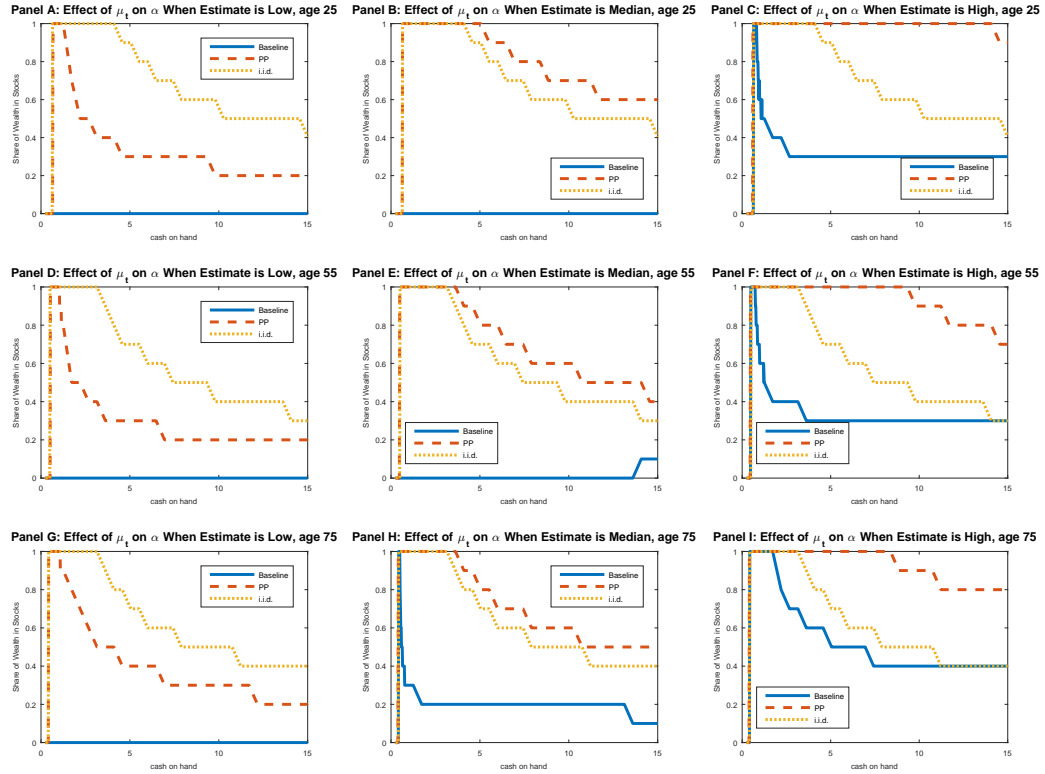


FIGURE 5

The Mean Share of Wealth in Stocks from the Baseline Model, Perfect Predictor Model and SCF Data

Figure 5, panel A and B present the mean share of wealth in stocks (α) between the baseline model, perfect predictor model (PP model) and SCF data. The only difference between panel A and B is the definition of the empirical α from SCF data. The empirical α in panel A rules out the asset with high liquidity and is defined as equity/(equity+bonds). In contrast, in panel B, the empirical α includes the asset with high liquidity and is calculated by equity/(equity+bonds+liquidity). I compute smoothed empirical α through running weighted linear least squares and a 2nd degree polynomial model with a span of 20% at each age.

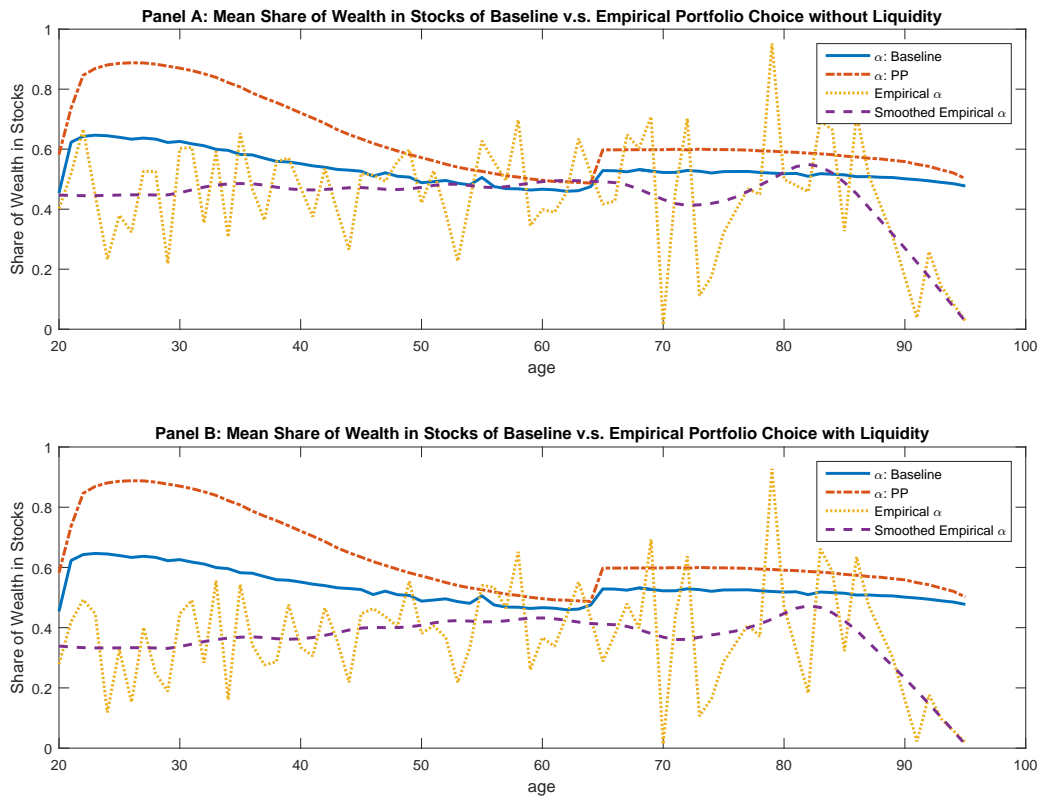


FIGURE 6

The Life-Cycle Profiles with High Risk Premium

Figure 6, panel A, presents the comparison of mean wealth between the baseline model (imperfect predictor model), perfect predictor model (PP model) and i.i.d. stock returns model over the life cycle. Panel B reports the difference of mean consumptions between these models. Panel C compares mean share of wealth (α) in stocks.

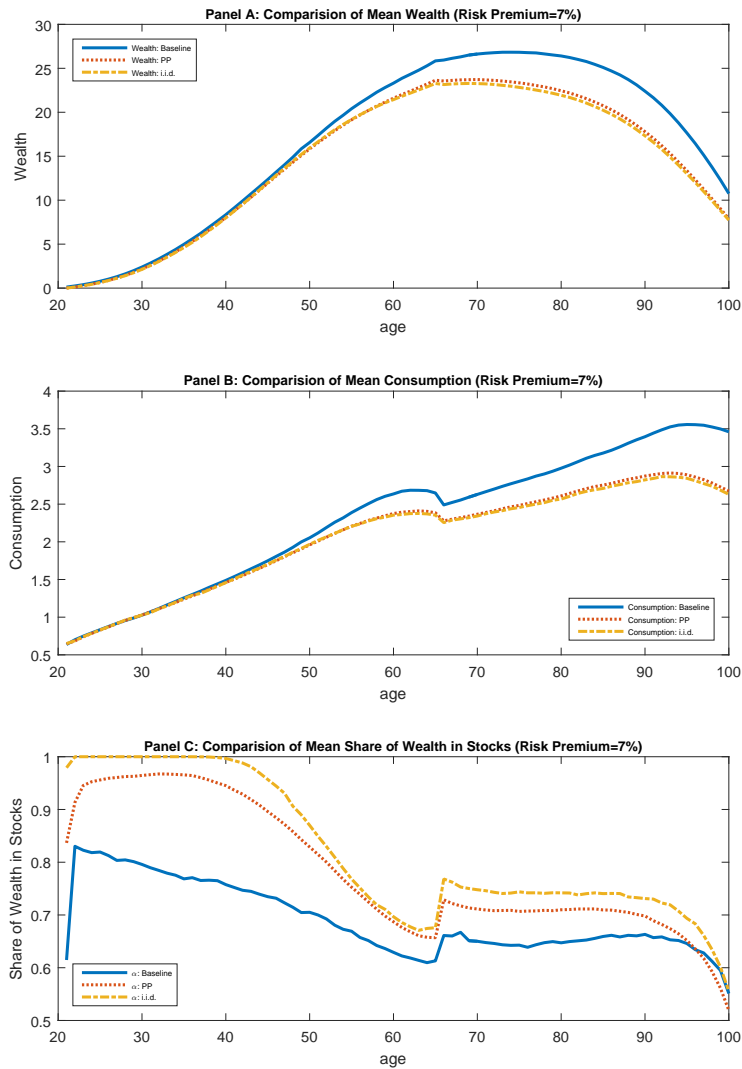


FIGURE 7

Life-Cycle Profile Comparison under the Low Volatility of Unobserved Expected Stock Returns

Figure 7, panel A, presents how the low volatility of unobserved expected stock returns σ_ε affects the mean wealth accumulation and consumptions over the life cycle. Panel B shows the change of mean share of wealth in stocks from the baseline model when the volatility of unobserved expected stock returns is low.

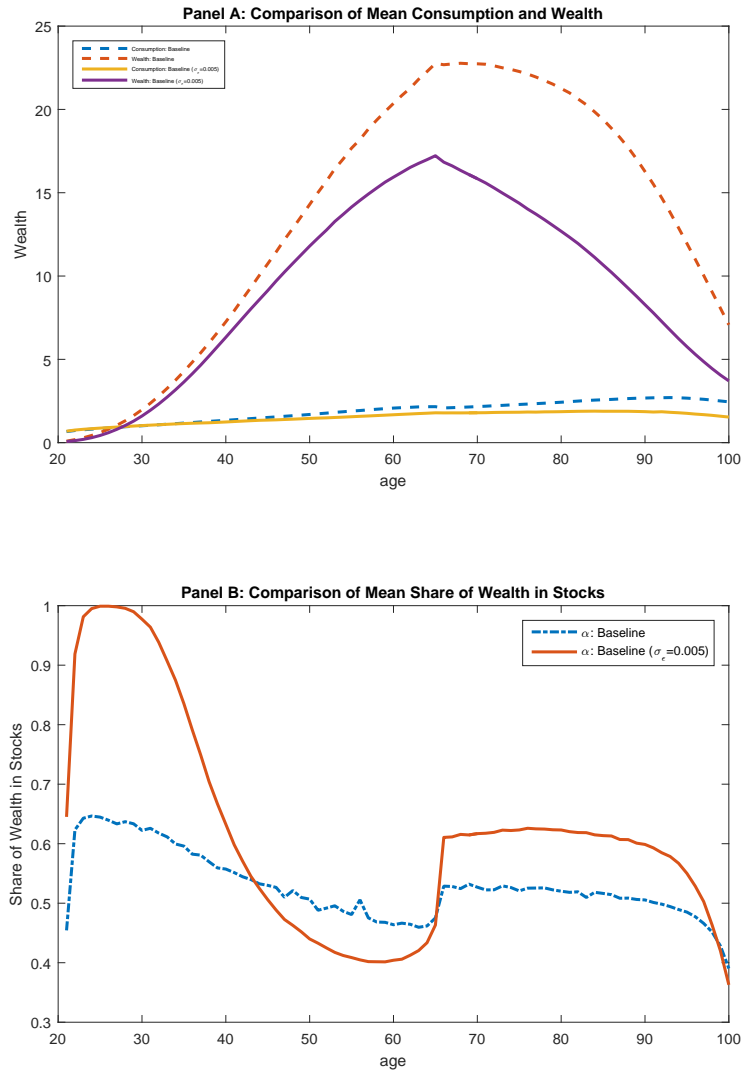


FIGURE 8

The Effect of High Persistence of Unobservable Expected Stock Return
 (ϕ_μ) on the Life Cycle Profiles

Figure 8 presents how the persistence of unobservable expected stock returns (ϕ_μ) affects mean wealth accumulation, consumption and portfolio choice. Panel A shows the mean wealth accumulation and consumption by varying ϕ_μ from 0.01 to 0.9. Panel B shows the mean share of wealth in stocks due to changing ϕ_μ .

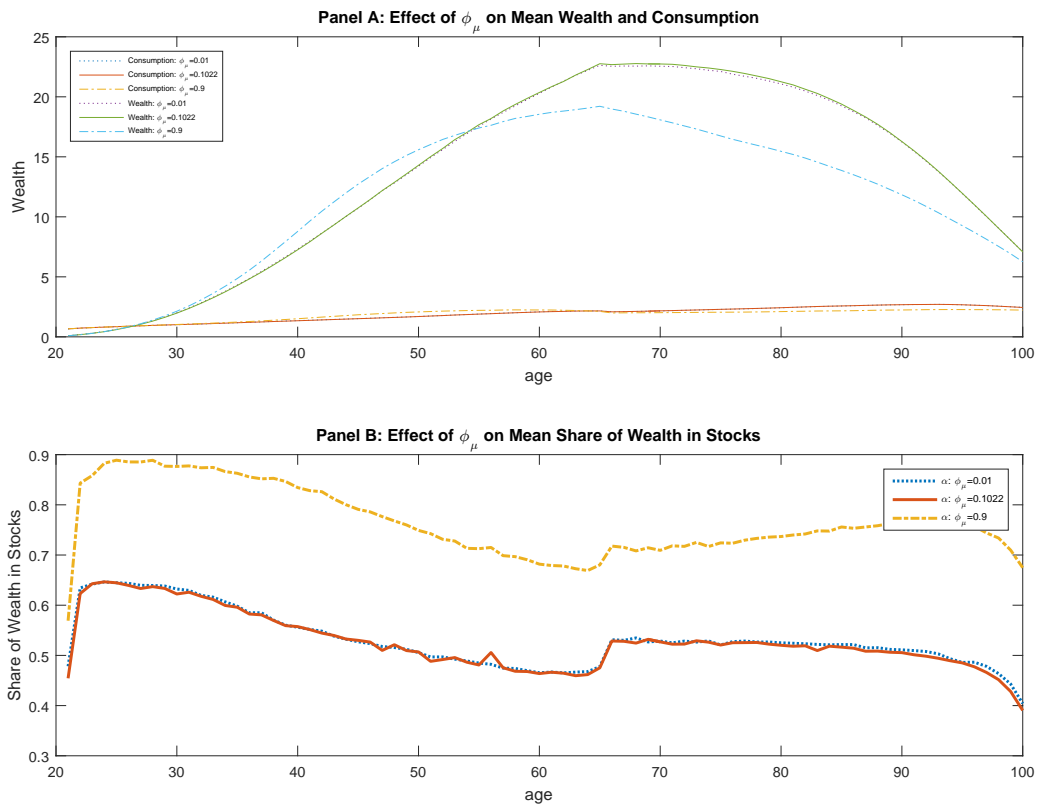


FIGURE 9

The Effect of Correlation between the Shocks to Stock Returns and Innovations of Unobservable Expected Stock Returns ($\rho_{z\varepsilon}$)

Figure 9, panel A presents the effect of correlation between the shocks to stock returns and the innovations of unobserved expected stock returns ($\rho_{z\varepsilon}$) on the mean wealth accumulation and consumption, and compares that with the perfect predictor model (PP model). Panel B shows its effect on mean share of wealth in stocks.

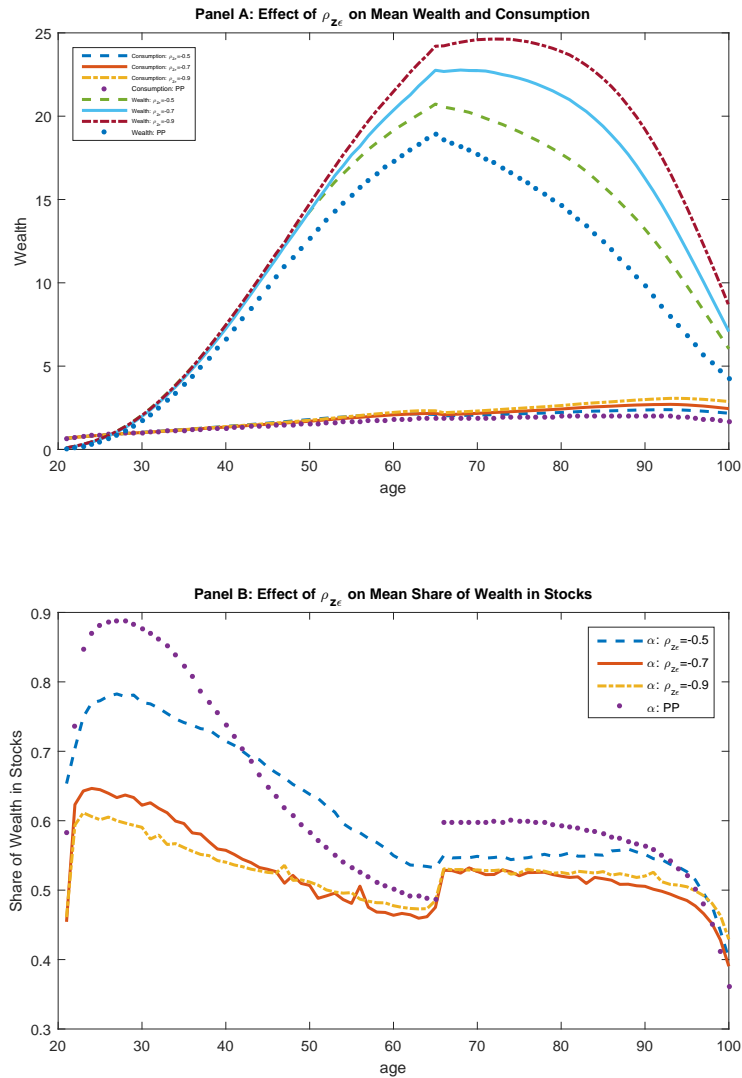


FIGURE 10

The Effect of Correlation between the Shocks to Stock Returns and the Permanent Part of Labor Income (ρ_{zn})

Figure 10, panel A presents the effect of correlation between the shocks to stock returns and the permanent part of labor income (ρ_{zn}) on the mean wealth and consumption. Panel B depicts the change in the mean share of wealth in stocks due to the variation of ρ_{zn} .

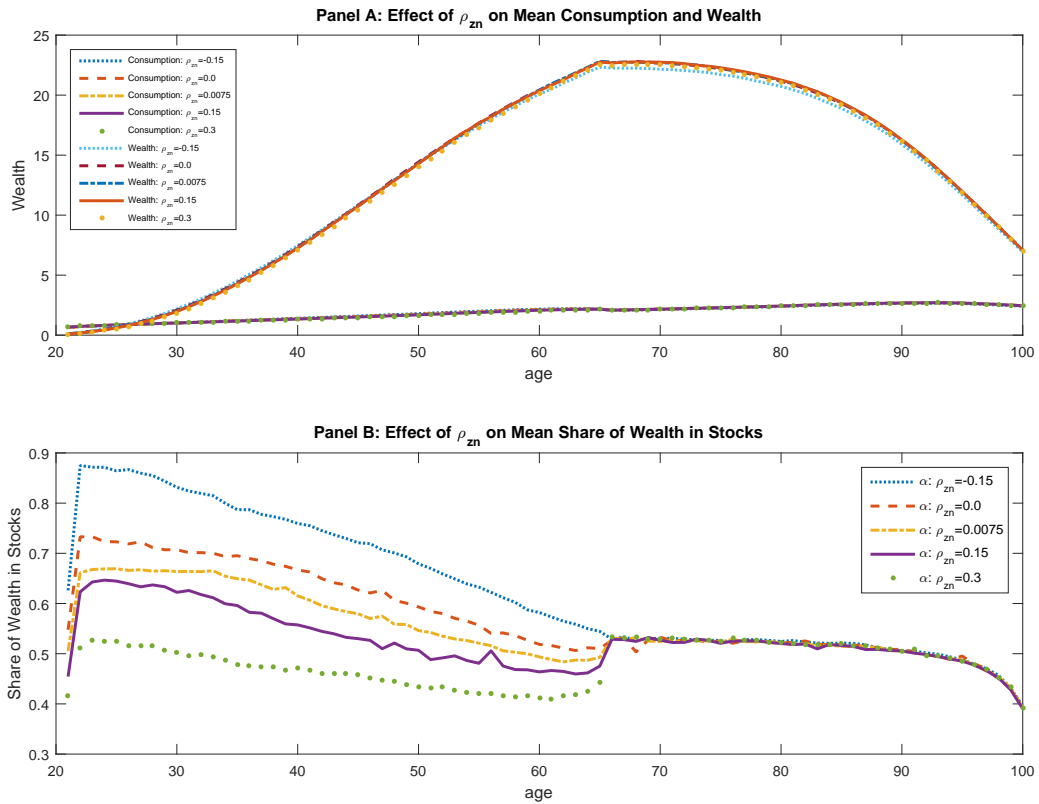


FIGURE 11

The Effect of Correlation between the Shocks to Stock Returns and the Innovations of Predictor (ρ_{zv})

Figure 11, panel A presents the effect of correlation between the shocks to stock returns and the innovations of predictor (ρ_{zv}) on the mean wealth accumulation and consumption. Panel B exhibits its effect on the mean share of wealth in stocks.

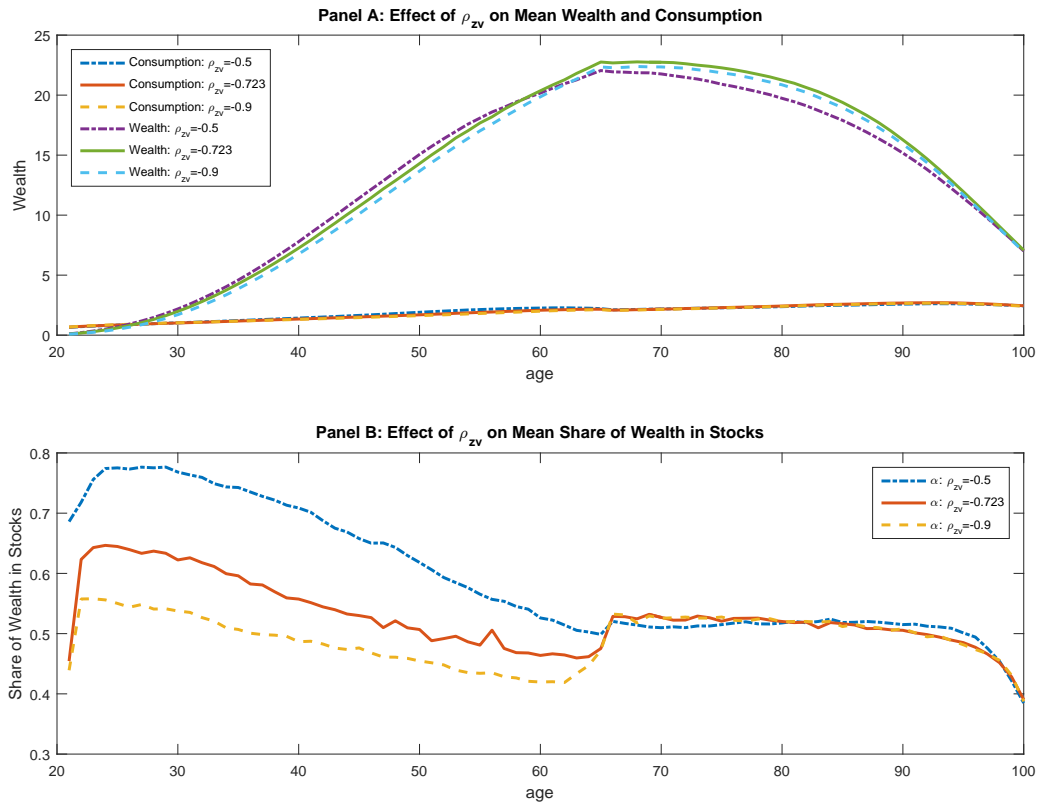


FIGURE 12

The Effect of Correlation between the Innovations of Predictor and the Shocks to the Unobservable Expected Stock Returns ($\rho_{v\varepsilon}$)

Figure 12, panel A presents the effect of correlation between the innovations of predictor and the shocks to the unobservable expected stock returns ($\rho_{v\varepsilon}$) on the mean wealth accumulation and consumption. Panel B describes the change of the mean share of wealth in stocks by varying $\rho_{v\varepsilon}$.

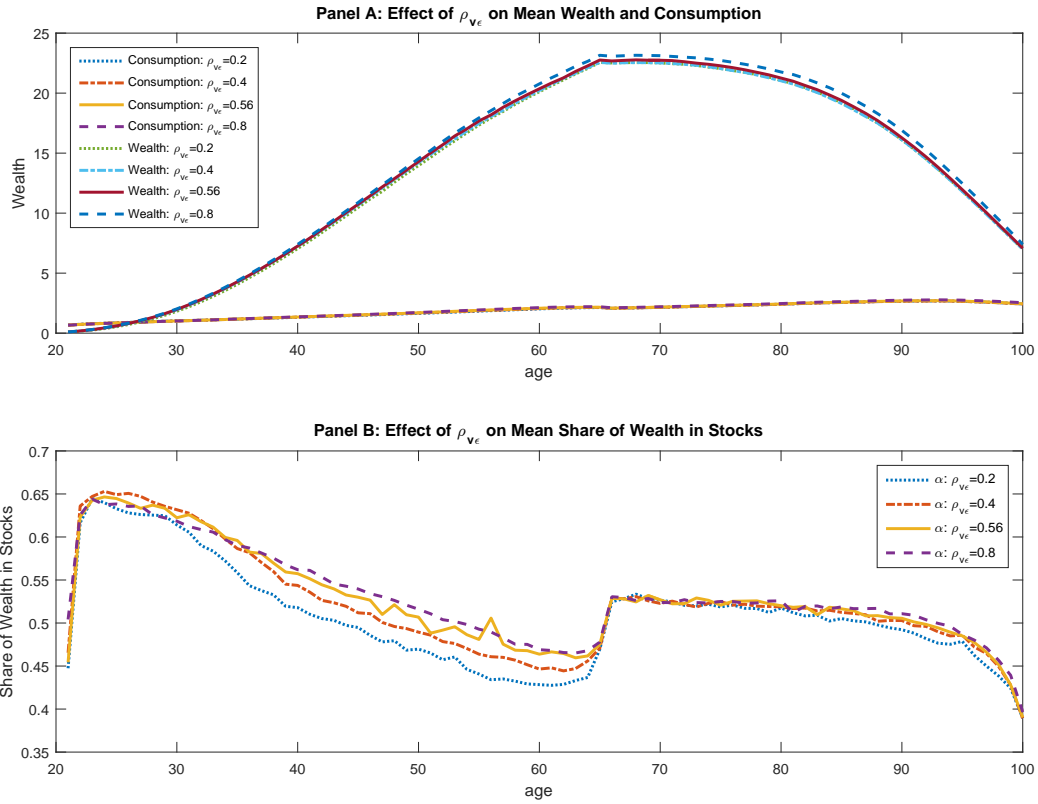


FIGURE 13

Welfare Evaluation

Figure 13 presents the change of consumption certainty equivalent of the baseline model with respect to the perfect predictor model (PP model), i.i.d. stock returns model and the Vanguard TDF model (Vanguard model) when changing the correlation between the shocks to the unobservable expected stock returns and the innovations of predictor ($\rho_{v\epsilon}$). Panel A shows the welfare loss from the perfect predictor model when $\rho_{v\epsilon}$ varies. Panel B plots the welfare loss from the i.i.d. stock returns model when changing $\rho_{v\epsilon}$. Panel C gives the welfare loss from the Vanguard TDF model for different $\rho_{v\epsilon}$.

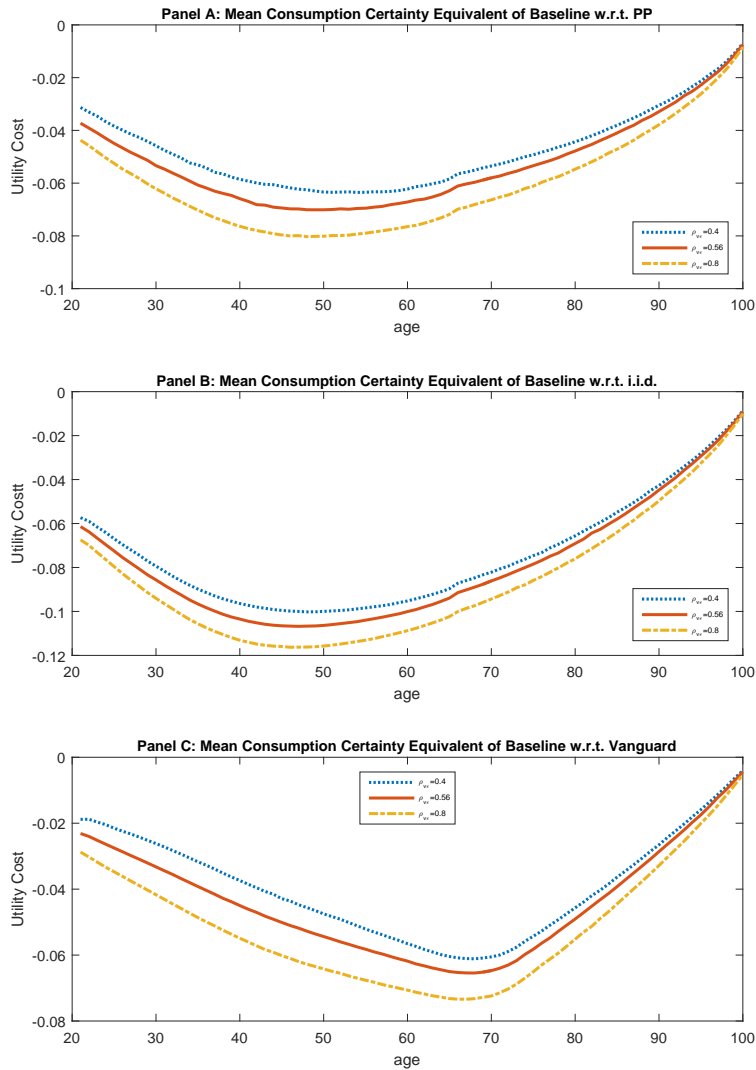


FIGURE 14

Consumption Evaluation

Figure 14 presents the average change of consumption and standard deviation of consumption over the life cycle. Panel A, C and E describe the mean change of consumption in the baseline model compared to the perfect predictor model (Panel A), the i.i.d. stock returns model (Panel C) and the Vanguard TDF model (Panel E). The mean change of consumption is defined as $E_t \left(\frac{C_1 - C_2}{C_2} \right)$. Panel B, D and F show the change of standard deviation of consumption for the corresponding models. The change of standard deviation is defined as $\left(\frac{sd[C_1] - sd[C_2]}{sd[C_2]} \right)$.

