

UCLA

Recent Work

Title

Learning About Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation

Permalink

<https://escholarship.org/uc/item/3167f8mz>

Author

Xia, Yihong

Publication Date

2000-05-10

Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation*

Yihong Xia[†]

University of California, Los Angeles

First Draft: July 1, 1998

This Revision: May 10, 2000

Abstract

This paper examines the effects of uncertainty about the predictability of stock returns on optimal dynamic portfolio choice in a continuous time setting with a long horizon. Uncertainty about the predictive relation affects the optimal portfolio choice through dynamic learning, and leads to a rich set of relations between the optimal portfolio choice and the investment horizon. There are also substantial market timing elements in the optimal hedge demands, which are caused by stochastic covariance and variance terms arising from dynamic learning. The opportunity cost of ignoring predictability or learning is found to be quite substantial.

*The author thanks Michael Brennan for his encouragement, insightful discussions, and constructive suggestions. The author also thanks Mathias Kahl, Anthony Lynch, Richard Roll, Pedro Santa-Clara, Robert Stambaugh, Rene Stulz (the editor), Luis Viceira, David Wessels, Fernando Zapatero and two anonymous referees for their helpful comments. All remaining errors are mine.

[†]Corresponding Address: Department of Finance, The Anderson Graduate School of Management; University of California, Los Angeles; 110 Westwood Plaza; Los Angeles; CA 90095-1481. Phone: (310) 825-8155. Fax: (310) 206-5455. E-mail: yihong.xia@anderson.ucla.edu.

Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation

Abstract

This paper examines the effects of uncertainty about the predictability of stock returns on optimal dynamic portfolio choice in a continuous time setting with a long horizon. Uncertainty about the predictive relation affects the optimal portfolio choice through dynamic learning, and leads to a rich set of relations between the optimal portfolio choice and the investment horizon. There are also substantial market timing elements in the optimal hedge demands, which are caused by stochastic covariance and variance terms arising from dynamic learning. The opportunity cost of ignoring predictability or learning is found to be quite substantial.

Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation

How much should a “long horizon” investor allocate to equity? The conventional wisdom says that a long horizon investor should invest more in equity because, over long horizons, above-average returns tend to offset below-average returns. This is the notion of “time diversification”. Samuelson (1989, 1990), among others, has argued that the notion of “time diversification” is spurious: when stock returns are i.i.d., for example, the optimal portfolio is independent of the horizon for an investor with isoelastic utility function. When stock returns are predictable, however, the optimal stock allocation does depend on the investment horizon, even if the investor has an isoelastic utility. In this setting, the investment opportunity set is stochastic and the intertemporal hedge demand introduced by Merton (1971) becomes central to the dynamics of asset allocation. In this paper, we study how learning about stock return predictability affects the intertemporal hedge demand and the optimal dynamic portfolio rules, and re-examine the validity of the prediction of “time diversification” in the context of uncertain predictability.

Although there is a growing body of evidence that stock returns are predictable, the existence of predictability is still subject to considerable debate. On the one hand, many studies have identified variables that predict future stock returns in a statistically significant way¹. The most powerful predictive variables in the U.S. have been found to be past market returns, the market dividend yield, the market earnings/price ratio, and term structure variables². On the other hand, critics of these studies point to the possibility of data-mining, the non-robustness of test statistics and incorrect inferences in small samples. For example, most of the test statistics are not robust to non-normality of the return distributions and studies with long horizon returns usually lack power and are subject to small sample problems³.

The controversy surrounding stock return predictability is symptomatic of the fact that the predictive relation is quite uncertain. An investor must take into account this uncertainty in choosing the optimal rule for consumption and asset allocation. Suppose that empirical work finds that the future stock return, r_{t+1} , is predicted by the current dividend, d/p_t , so that $r_{t+1} = a_t + b_t (d/p_t) + \epsilon$, where b_t is statistically significant different from zero in a regression using data

up to time t . Knowing the criticisms of such empirical studies, an investor must be concerned that the regression may be mis-specified and the t -statistic misleading. A rational investor will neither completely ignore the empirical results and treat stock return as i.i.d. series, nor will he use the estimate b_t as if it were the true parameter. Instead, he will derive the optimal dynamic asset allocation strategy allowing for the possibility that the true value of the predictive coefficient may be different from his current estimate b_t .

Early work by Brown, Klein and Bawa⁴ among others has considered the effect of parameter uncertainty on portfolio selection in a single period context and has shown that the “predictive distribution” of returns that is obtained by integrating the conditional distribution over the distribution of the uncertain parameters is different from the distribution that is obtained when the parameters are treated as known. The additional risk that is introduced by parameter uncertainty in this context has been labeled “estimation risk”.

More recently, Kandel and Stambaugh (1996) have explored the *economic* importance of stock return predictability and the effect of estimation risk when asset returns are partially predictable and the coefficients of the predictive relation are estimated rather than known. Given that the exact predictive relation is unknown and that the investor is assumed to trade only at a discrete one month interval, uncertainty about the parameters of the conditional return distribution - estimation risk - affects the investor’s optimal portfolio decision. While this simple framework highlights the economic importance of stock return predictability with estimation risk, the one period investment horizon assumption precludes considerations of the dynamic effects induced both by full information hedging and by learning through time, which may be important for investors with longer horizons.

Gennotte (1986) and Feldman (1992) have shown that in a continuous time setting in which security prices follow diffusion processes, the effects of parameter uncertainty are rather different from those found in the discrete time single period model. In particular, the Brown *et al* “estimation risk” effect disappears, so that an investor with logarithmic utility *ignores parameter uncertainty entirely* in his portfolio decision⁵. An investor with non-logarithmic utility must take account of parameter uncertainty, not because it affects the optimal instantaneous mean variance efficient portfolio, but because he will *learn* more about the parameters as time passes. His

estimates of the unknown parameter values are “state variables” in his dynamic optimization problem and it is the need to hedge against unanticipated changes in these state variables that affects the optimal portfolio choice.

The effect of stock return predictability for an investor with a long investment horizon in the absence of parameter uncertainty has been studied in several recent papers, including Brennan, Schwartz and Lagnado (1997), Campbell and Viceira (1998), Brandt (1998), and Lynch and Balduzzi (1998). However, none of these papers take into account the fact that the underlying predictive relation is uncertain. Barberis (1999) studies a problem that is closely related to the one analyzed in this paper. He derives a dynamic strategy in a discrete time setting with estimation risk; however in doing so, he simplifies the problem by ignoring the possibility that the investor will learn more about the predictive relation as time passes. As a result of this simplification, the investor’s opportunity set is governed by the predictive variable (the dividend yield) alone, and the investor ignores the hedge demand for stock that is induced by parameter uncertainty. In this paper, the investor’s opportunity set depends on not only the current value of the predictive variable but also the current estimate of the parameter and the variance of the estimate, and we show that the hedge demand induced by parameter uncertainty is an important component of a long horizon investor’s optimal portfolio.

This paper examines the optimal dynamic portfolio strategy for a long horizon investor who takes account of the uncertain evidence of stock return predictability. We study the effect of parameter uncertainty and its associated hedge component, in a dynamic continuous time context with a potentially time varying investment opportunity set induced by possible return predictability. The drift of the single risky asset price process is assumed to be known to be a linear function of the value of a signal which follows a known Markov process, but the parameters of the linear predictive relation are unknown and must be estimated. We derive the evolution of the investor’s beliefs about the unknown parameters when the asset prices are possibly predictable and the predictive variables themselves are stochastic, and show that this introduces a stochastic covariance between the current estimate of the parameter and the stock returns. It also introduces a stochastic variance for the estimated parameters. This feature of the model distinguishes learning effects with return predictability from learning effects without return predictability analyzed by

Gennotte (1986), Brennan (1998), and Brennan and Xia (1998a).

Since returns are *potentially* predictable and time varying, we are able to consider the interaction of investment horizon and parameter uncertainty. The hedge demand associated with the uncertain parameters plays a predominant role in the optimal strategy, and is the major element of the horizon effect. When the stock return is **possibly** predictable (even if the true but unknown return process is i.i.d.), the optimal allocation is horizon-dependent: the optimal stock allocation can increase, decrease or vary non-monotonically with the horizon⁶, because parameter uncertainty induces a state dependent hedge demand that may increase or decrease with horizon. Therefore, the conventional advice that young investors should allocate more wealth in equity does not hold for everyone. The introduction of interim consumption reduces the magnitude of the horizon effect, because it reduces the effective horizon or duration of the consumption.

Market timing, the dependence of the portfolio allocation on the predictive variable, is a natural consequence of return predictability. The relation between the optimal portfolio allocation and the predictive variable depends crucially on future learning and investment horizon. Without learning, the optimal allocation increases monotonically with the current predictive variable⁷, because the investor takes advantage of return predictability by investing significantly more in stock when the expected return is high. When there is learning, however, the optimal allocation is less sensitive to and no longer monotone in the predictive variable, because as the expected return becomes higher, the negative amount of stock the investor uses to hedge parameter uncertainty eventually dominates. In addition, the horizon effect of market timing depends on whether the investor faces parameter uncertainty. When the predictive relation is known with certainty, the portfolio allocation is more sensitive to the predictive variable for a long horizon investor than that for a short horizon investor⁸. When there is uncertainty about the predictive relation, the allocation becomes less sensitive to the predictive variable for a long horizon investor. Simulated results using both historical as well as artificially generated data show that investors who ignore market timing can incur very large opportunity costs.

The paper is organized as follows. The model is developed in section 2. Section 3 describes the data and the calibrated parameter values used in the numerical calculations. Section 4 contains three subsections. The first subsection discusses the effect of parameter uncertainty, interim

consumption and investment horizon. The second subsection examines the economic importance of market timing in the context of uncertain predictability. While the first two subsections are based on results with a constant predictive parameter, the third subsection briefly comments on the results with stochastic predictive parameter. Section 5 discusses implications for future research and concludes the paper.

I. The Model

A. The Basic Setting

Consider an investor with a long horizon who can trade continuously in a risk free asset and a single risky stock. The real return on the risk free asset is assumed to be a constant r , so that the price of the risk free asset is described by

$$dB = rBdt. \tag{1}$$

The stock price with dividend reinvestment, P , is assumed to follow a simple stochastic process of the following type:

$$\frac{dP}{P} = \mu(t)dt + \sigma_P dz, \tag{2}$$

where dz is a $(k \times 1)$ vector of increments to standard Wiener processes, dz^i is independent of dz^j for all $i, j = 1, \dots, n$ and $i \neq j$, and σ_P is a known constant $(1 \times k)$ vector. The instantaneous proportional drift, $\mu(t)$, however, is not known to the investor, but is related to an n -vector of predictive variables, S , by a functional relation $\mu(t) = \mu(S, t)$. The true relation between the drift and the predictive variables is not known to the investor, and it is possible that the predictive variables have no power to predict μ at all. For tractability, it is assumed that the investor knows that this relation is linear, but that the coefficients are possibly stochastic and are unobservable to the investor:

$$\mu(t) = \alpha + \beta' S(t), \tag{3}$$

where α is an unknown scalar and β is an $n \times 1$ vector of unknown (unobservable) predictive coefficients. The coefficients β are assumed to evolve according to the following known diffusion

process⁹:

$$d\beta = (a_0(P, S, t) + a_1(P, S, t)\beta)dt + \eta(P, S, t)dz. \quad (4)$$

Here $a_0(P, S, t)$ is an $(n \times 1)$ vector, $a_1(P, S, t)$ is an $(n \times n)$ matrix and $\eta(P, S, t)$ is an $(n \times k)$ matrix, all of which are assumed to be known. We can view equation (4), the evolution of the vector of the unobservable predictive coefficients, as the “transition equation” in a continuous time analog to the Kalman filtering procedure. When all the coefficients in equation (4) are zero, β is a vector of unobservable constants. If, in addition, $\beta = 0$ in equation (3), the model corresponds to that of iid returns with unknown drift, which is analyzed by Brennan (1998).

For simplicity, we assume that both the long run mean of μ , $\bar{\mu}$, and the mean of the predictive variables, \bar{S} , are known constants so that $\alpha \equiv \bar{\mu} - \beta'\bar{S}$ is known whenever the value of β is determined. To complete the model, we assume that the vector of predictive variables, S , follows a known joint Markov process:

$$dS = (A_0(P, S, t) + A_1(P, S, t)\beta)dt + \sigma_S(P, S, t)dz. \quad (5)$$

where $A_0(P, S, t)$ is $(n \times 1)$, $A_1(P, S, t)$ is $(n \times n)$ and $\sigma_S(P, S, t)$ is $(n \times k)$. We refer to the processes for the predictive variables, equations (2) and (5), as “measurement (or observation) equations”, following the Kalman filtering literature.

The assumption that the parameters in equations (4) and (5) are known is a strong one because an investor will not generally know for certain either the processes for the predictive variables or the evolution of the coefficients. A more realistic model would allow for unobservability in the processes (4) and (5) as well. It is also possible that the processes of the coefficients and the signals are subject to regime shifts and jumps¹⁰. While these are potentially important considerations, we simplify by modelling the predictive variables as Markovian diffusions with known parameters. Even this small source of imperfect information has major implications for optimal portfolio choice and investor welfare.

Following Merton (1971), the investor is assumed to maximize the expected value of a von-Neumann-Morgenstern utility function defined over consumption, $c(t)$, and wealth at the horizon $W(T)$ by choosing an optimal consumption and portfolio strategy, $c(t)$ and $x(t)$, $\forall t \in [0, T]$,

where $x(t)$ is the proportion of wealth allocated to the risky asset:

$$\max_{\{x(t), c(t)\}} E \left\{ \int_0^T U(c(t), t) dt + B(W(T), T) | \mathcal{F}_0^{\mathcal{I}} \right\}, \quad (6)$$

$$\text{s.t. } dW = (rW + xW(\bar{\mu} + \beta(S - \bar{S}) - r) - c)dt + xW\sigma_P dz, \quad (7)$$

where β (or μ) is unknown to the investor and must be inferred, and $\mathcal{F}_0^{\mathcal{I}}$ is the investor's information set at time 0. We assume that the investor has iso-elastic utility so that the utility of instantaneous consumption is

$$U(c(t)) = \begin{cases} e^{-\rho t} \frac{c_t^{1-\gamma}}{1-\gamma} & \text{if } \gamma > 0 \text{ and } \gamma \neq 1 \\ e^{-\rho t} \log c_t & \text{if } \gamma = 1 \end{cases}, \quad (8)$$

where ρ is the constant time preference parameter, and the bequest function, $B(W(T), T)$, is defined analogously. We assume implicitly that all necessary and sufficient technical conditions are satisfied for the investor's problem to have a well-defined solution.

The investor forms his assessment of the predictive relation coefficient β from the observed stock return and the predictive variable processes. As Detemple (1986), Dothan and Feldman (1986) and Genotte (1986) have shown in a related setting, the investor's decision problem may be decomposed into two separate problems¹¹: an inference problem in which the investor updates his estimate of the current value of the unobservable state variable, β ; and an optimization problem in which he uses his current estimate of β to choose an optimal portfolio, taking account of possible future learning. Although the value of the underlying state variable β is unknown, the separation theorem implies that the investor's optimization problem can be solved in terms of its estimated value.

B. The Investor's Inference Problem: A Learning Process

The standard Brownian motions z are defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with a standard filtration $\mathcal{F} = \{\mathcal{F}_t : t \leq T\}$. The investor's information structure is summarized by the filtration $\mathcal{F}^{\mathcal{I}}$ generated by the joint processes of signals $\mathcal{I}(t) = (P(t), S(t))$, and $\mathcal{F}_t^{\mathcal{I}} \subset \mathcal{F}_t$. Processes of dz and β are adapted to \mathcal{F}_t but not to $\mathcal{F}_t^{\mathcal{I}}$, since the investor does not directly observe β . Let

$b_t \equiv E(\beta | \mathcal{F}_t^I)$ and $\nu_t \equiv E((\beta - b_t)(\beta - b_t)' | \mathcal{F}_t^I)$ denote the conditional mean and variance of the investor's estimate. The investor is assumed to have a Gaussian prior probability distribution over β , with mean b_0 and variance ν_0 ¹².

Following Liptser and Shirayev (1978), the distribution of β conditional on $\mathcal{I}(t) = (P(t), S(t))$ is also Gaussian with mean b_t and variance ν_t (subscript t is dropped later). In order to gain more intuition for the continuous time Bayesian updating rule, we concentrate on a specific simplification of the model and leave the general case and details of derivation to Appendix A. Suppose that there is one predictive variable which follows a Ornstein-Uhlenbeck (O-U) process:

$$ds = \kappa(\bar{s} - s)dt + \sigma_s dz_s. \quad (9)$$

Then, we can rewrite the stock return process as

$$\frac{dP}{P} = (\bar{\mu} + \beta(s - \bar{s}))dt + \sigma_P dz_P. \quad (10)$$

Let us assume that the coefficient β also follows an O-U process¹³,

$$d\beta = \lambda(\bar{\beta} - \beta)dt + \sigma_\beta dz_\beta, \quad (11)$$

and let $E(dz_P dz_s) = \rho_{sP} dt$, $E(dz_P dz_\beta) = \rho_{\beta P} dt$ etc.

Conditional on the investor's filtration \mathcal{F}_t^I , the stock price follows the stochastic process

$$\frac{dP}{P} = (\bar{\mu} + b(s - \bar{s}))dt + \sigma_P d\hat{z}_P, \quad (12)$$

and the updating rules for the conditional mean and variance are

$$db = \lambda(\bar{b} - b)dt + [\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})] \begin{pmatrix} \sigma_P^2 & \sigma_{sP} \\ \sigma_{sP} & \sigma_s^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_P d\hat{z}_P \\ \sigma_s d\hat{z}_s \end{pmatrix}, \quad (13)$$

$$= \lambda(\bar{b} - b)dt + \nu_1 d\hat{z}_P + \nu_2 d\hat{z}_s, \quad (14)$$

$$\frac{d\nu}{dt} = -2\lambda\nu + \sigma_\beta^2 - [\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})] \begin{pmatrix} \sigma_P^2 & \sigma_{sP} \\ \sigma_{sP} & \sigma_s^2 \end{pmatrix}^{-1} [\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})]', \quad (15)$$

$$= -2\lambda\nu + \sigma_\beta^2 - [\nu_1(s)^2 + \nu_2(s)^2 + 2\nu_1(s)\nu_2(s)\rho_{sP}], \quad (16)$$

where \bar{b} , $\nu_1(s)$, $\nu_2(s)$, $d\hat{z}_s$, and $d\hat{z}_P$ are given in equations (A12) to (A15). Note that the innovations in both the stock price and the parameter estimate b are governed by the Brownian increment $d\hat{z}_P$, which is adapted to the investor's filtration \mathcal{F}^I , while dz_P is not. Conditional on the measure innovated by \mathcal{F}^I , the budget constraint (7) becomes

$$dW = (rW + xW(\bar{\mu} + b(s - \bar{s}) - r) - c)dt + xW\sigma_P d\hat{z}_P. \quad (17)$$

As shown in equation (A18) in Appendix A, the state variable, β , and the observation of stock return and predictive variable are jointly normally distributed. The “best” current estimate of β is then its conditional mean, b . b in this model is the slope coefficient of regressing stock return on dividend yield. The updating process is a recursive procedure for computing the optimal estimate b and its variance at time t based *only* on the information available at t , and it enables the estimate to be continually updated as new observations become available.

The change in the best estimate, db in equation (13), consists of two terms: the first term is a deterministic drift equal to the expected change in b given the current estimate, $E(db|q)$. The second term is given by a vector of regression coefficients times a vector of innovations, which are derived from the signals. The vector of covariances between the innovations in b and the innovations in the stock price and the predictive variable, $[\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})]$, determines how much of the new information is incorporated in the updating of b . It depends on several important parameters of the model. When the investor is less confident of his current estimate (higher ν) or when the signal is more highly correlated with the unobservable parameter (larger values of $\rho_{\beta P}$ and $\rho_{\beta s}$), he will put more weight on the new information, so that the second term contributes more to the updating.

The variance updating rule for ν in equation (16) is the usual Riccati equation. The first two terms correspond to the incremental uncertainty induced by the stochastic variation in β itself, while the term in the bracket denotes the reduction in estimation error when additional information becomes available. The rate of learning depends positively on the mean-reverting parameter and negatively on the variance of the β process: intuitively, the less variable is β the more rapidly the investor learns about its current value. Equation (16) differs from the standard

Gaussian conditional variance updating rule in that the rate of learning is a function of the current value of the predictive variable, s . The investor is trying to learn about a regression coefficient between stock return and $s - \bar{s}$. When $s \approx \bar{s}$, the investor does not learn much about β , and his uncertainty reduces at a slower rate.

The steady state value for ν , $\bar{\nu}$, is obtained by setting the right hand side of equation (16) to zero, which means that the investor cannot improve his estimate of the parameter after he reaches the steady state. When β is stochastic, $\bar{\nu} > 0$ in general, so that the investor is always “one-step behind”. In this case, the steady state estimation risk, $\bar{\nu}$, is a function of s , because the path of s determines how informative the data is as compared to the variation in β . When β is a constant (i.e., $\lambda = 0$ and $\sigma_\beta = 0$), although the learning path is still state-dependent, the investor eventually learns about the true value of β and $\bar{\nu} = 0$.

As detailed in Appendix A, the covariance between the observation and the state variable depends on the coefficient matrix in the observation equation, which is a function of the predictive variable s . Therefore, an interaction of return predictability and learning means that $cov\left(db, \frac{dP}{P}\right)$ depends on both the current level of uncertainty ν and the value of the predictive variable s :

$$\begin{aligned}\sigma_{bP}(s) &= cov\left(db, \frac{dP}{P}\right) = \sigma_P(\nu_1 + \rho_{sP}\nu_2), \\ &= \nu(s - \bar{s}) + \rho_{\beta P}\sigma_\beta\sigma_P.\end{aligned}\tag{18}$$

When there is no learning, the first term is zero so that σ_{bP} is constant. When there is learning, the sign of the first term depends on whether s is above or below \bar{s} . In the case of $\sigma_\beta = 0$, the sign of the covariance solely depends on $s - \bar{s}$. Suppose that $s > \bar{s}$, then an *unexpectedly* high return ($d\hat{z}_P > 0$) means that the current estimate b_t is too low, and the investor adjusts b upwards (i.e., $db > 0$). Thus, in the case of $s > \bar{s}$, the covariance is positive. Suppose that $s < \bar{s}$, then b_t is multiplying a negative number, and an *unexpectedly* high stock return ($d\hat{z}_P > 0$) means that the current estimate of b is too high, so the investor adjusts b downwards ($db < 0$). Since the sign of this covariance affects the role of the stock as an instrument for hedging against changes in b , the stochastic variation in σ_{bP} introduces a “timing element” into the hedge demand for the stock.

C. The Investor's Optimization Problem

Given the investor's information set, \mathcal{F}_t^I , his decision problem at time t is completely characterized by his wealth level W_t , his current assessment of the predictive coefficient, as summarized by its conditional mean, b_t , and variance, ν_t , and the current observation of the predictive variable, s_t . Thus, the investor's lifetime expected utility under the optimal policy can be written as:

$$J(W, b, \nu, s, t) = \max_{\{(c(\tau), x(\tau)): t \leq \tau \leq T\}} E \left\{ \int_t^T e^{-\rho\tau} \frac{c^{1-\gamma}}{1-\gamma} d\tau + \frac{e^{-\rho T} W_T^{1-\gamma}}{1-\gamma} \middle| \mathcal{F}_t^I \right\}, \quad (19)$$

with a terminal condition

$$J(W, b, \nu, s, T) = e^{-\rho T} \frac{W_T^{1-\gamma}}{1-\gamma}. \quad (20)$$

Imposing the budget constraint (17), using the stochastic processes (14) and (16), and noting that the correlation between $d\hat{z}_P$ and $d\hat{z}_s$ is also ρ_{sP} , we can derive the condition for optimality in the investor's decision problem. The first order conditions for the optimal consumption-portfolio choice $c^*(t)$ and $x^*(t)$ are derived according to Merton (1971). Under the iso-elastic utility assumption in equation (8), $J(W, b, s, \nu, t)$ is separable in wealth, and can be written as:

$$J(W, b, \nu, s, t) = e^{-\rho t} \frac{W^{1-\gamma}}{1-\gamma} \phi(b, \nu, s, t). \quad (21)$$

Calculating the derivatives of J using Equation (21) and then substituting into the original optimality condition yields a nonlinear second order partial differential equations for ϕ in four state variables¹⁴ (b, s, ν, t).

The optimal portfolio strategy and consumption-wealth ratio are given by

$$x^* = \frac{\bar{\mu} + b(s - \bar{s}) - r}{\gamma\sigma_P^2} + \left[\frac{\phi_b}{\gamma\sigma_P^2\phi} \nu(s - \bar{s}) + \frac{\phi_s}{\gamma\sigma_P^2\phi} \sigma_P \sigma_s \rho_{sP} + \frac{\phi_b}{\gamma\sigma_P^2\phi} \sigma_P \sigma_b \rho_{\beta P} \right], \quad (22)$$

and

$$\left(\frac{c}{W} \right)^* = \phi^{-\frac{1}{\gamma}}. \quad (23)$$

The first term in equation (22) is the familiar myopic portfolio allocation to stock, while the second term represents the investor's hedge demand for stock. This hedge demand has three components: the first term in the bracket is the hedge demand due to learning, which represents

the *direct* effect of parameter uncertainty; this term disappears when there is no uncertainty about β (i.e., $\nu = 0$) or s happens to equal \bar{s} . The second term arises from the need to hedge against changes in the investment opportunity set due to changes in the predictive variable s . This term drops out if the stock return is uncorrelated with the change in the predictive variable s , i.e., if ρ_{sP} is zero. The third term arises from the need to hedge against changes in the estimate b . This term disappears if the correlation between β and stock return is zero, or if the true value of β is a constant (i.e., $\sigma_\beta = 0$).

The first term is instantaneously zero when s happens to equal \bar{s} , because the covariance between the estimation error and stock return, $\nu(s - \bar{s})$, is zero in this special case. Even in this case, however, the optimal portfolio strategy with and without parameter uncertainty is different from each other. The magnitude of the second and the third terms in the bracket is different from that without parameter uncertainty, because the function ϕ depends on a different set of state variables due to the presence of parameter uncertainty. As a result, there is an *indirect* effect of parameter uncertainty represented by the difference between the hedge demands associated with s and β in the two cases. The total effect of parameter uncertainty is then given by the summation of the *direct* and *indirect* effects.

In a discrete time setting, estimation risk as discussed by Brown et al. (1979) affects the investor's portfolio choice through the variance-covariance matrix of the returns, which is increased by the estimation risk associated with uncertainty about β ¹⁵. In a continuous time long horizon model, the instantaneous variance-covariance matrix of the returns is not increased by the static estimation risk, but the dynamic effect of learning introduces an additional hedge demand represented by the first term in the bracket and changes the magnitude of the hedge demands associated with s and β in the second and third terms in the bracket. Barberis (1999) uses the Brown *et. al.* (1979) model to study return predictability with static estimation risk in a multi-period discrete time context. This introduces a hedging demand associated with the predictive variable represented by the second term in the bracket, but he did not treat the parameter estimate as a state variable, and thus ignores both the hedge demand associated directly with dynamic learning and the change in the hedge demand of s caused indirectly by it¹⁶. While we emphasize both, we are able to ignore the effect of static estimation risk on the optimal portfolio since it

disappears in a continuous time setting.

II. Data, Model Calibration and Numerical Approximation

Two different optimization problems are studied. In the first, the investor maximizes the expected utility of lifetime consumption and bequest. In the second problem, the investor maximizes the expected utility of terminal wealth. The risk aversion coefficient, γ , is set to be 5.0. For computational reasons we consider only a single predictive variable. While several important predictive variables¹⁷ have been identified, we take the dividend yield as the single predictive variable, $d/p \equiv s$, because this variable plays a prominent role in studies of return predictability and enables us to compare our results with existing work. For simplicity, we first assume that β is constant so that λ and σ_β in equation (11) are zero¹⁸.

The parameters of the joint stochastic process for the real stock return and dividend yield were calibrated to the moments of historical observations of U.S. stock market real returns and dividend yield for the period from January 1950 through December 1997. The stock return is from the CRSP monthly returns file VWRETD, and the dividend yield series is constructed by factoring out the dividend from VWRETD and VWRETX in CRSP. The return on the value weighted CRSP index adjusted by the monthly inflation rate¹⁹ was taken as the return of the stock. The historical annual average inflation rate and the annualized long run mean of the treasury bill rates were used to derive an estimate for the real interest rate, r , of approximately 3.4%²⁰. The historical mean and volatility of the real return on the CRSP value weighted index for the period January 1950 to December 1997 are used as estimates of $\bar{\mu}$ and σ_p . Values for κ , \bar{s} and σ_s were obtained by estimating the exact discrete equivalent of (9) by nonlinear ordinary least squares regression using monthly data on the dividend yield,

$$s_t = \bar{s}(1 - e^{-\frac{\kappa}{12}}) + e^{-\frac{\kappa}{12}}s_{t-\Delta t} + e_t, \quad (24)$$

where s_t is the dividend yield at time t , $\Delta t = 1/12$ (year), and the standard deviation of the regression residual σ_e is related to the volatility σ_s by $\sigma_s = \sigma_e \sqrt{\frac{2\kappa}{1 - e^{-\kappa/6}}}$.

Table I reports the parameter notations and calibrated values used in the numerical study. In order to obtain a reasonable range for the value of prior mean b_0 and variance ν_0 , a VAR regression using stock return and dividend yield with one lag is carried out using data from the whole sample and one sub-samples (January 1950 to December 1977) and reported in the last four rows of Table I. The regression results are close to those of Barberis (1998, Table II)¹ when the same sub-sample is used. We use the VAR regression results for the values of the prior mean and variance.

The nonlinear partial differential equation in ϕ is solved using an implicit finite difference approximation on a $(40 \times 200 \times 60)$ grid for the state variables $(s \times b \times \nu)$. The second-order partial derivatives with respect to the state variables are discretized using second-order accurate central difference approximations. The first-order partial derivatives with respect to the state variables are discretized using first-order accurate difference approximations. The dividend yield is allowed to range from 0% to 20%. The range of b is set to $[-10, 10]$, and the value of ν_0 varies from zero to 12.0 or the standard deviation $\sqrt{\nu_0}$ varies from 0 to 3.46. Thus, the step sizes in the three state variables are 0.5%, 0.1 and 0.2, respectively. The time step is set to be one month ($\frac{1}{12}$ year). For each time step, initial trial values of the optimal control, x^* , are computed using values of the partial derivatives from the previous time step. Current values of $\phi(b, s, \nu, t)$ are then calculated from the partial differential equation using the method of successive over relaxation, and the optimal control is then re-computed using values of the partial derivatives from the currently computed values of ϕ . The new control is then used to compute new values of ϕ . This procedure is repeated until it yields satisfactory convergence.

In solving the partial differential equation, we assume that ϕ is linear at both the upper and lower boundary of b . The upper boundary for s is also imposed to be linear. To preclude the possibility of a negative dividend yield, a reflecting barrier at $s = 0$ is imposed². The upper boundary for ν is inaccessible in theory since ν declines monotonically as the investor learns more through time. We assume that ϕ has zero second derivative at the upper boundary of ν . When ν reaches zero, it is absorbed, and the investor no longer has estimation risk. When β is constant, the partial differential equation reduces to a one-state variable case, and the values of ϕ can be calculated according to the closed form formulas given by Kim and Omberg (1996)

and summarized in Appendix B, which provides a natural lower boundary for ν . When β is stochastic, the pde along the boundary at $\nu = 0$ has two state variables. We numerically evaluate this pde and use the result as the natural boundary for the three-state variable PDE.

III. Effect of Learning about Predictability

Numerical results in the first two subsections are based on the assumption of a constant true value for β and a risk aversion coefficient of 5.0. The predictive variable is the dividend yield, so that $d/p \equiv s$ in all the results. In subsection A, the effect of the horizon on optimal portfolio choice is analyzed. Subsection B is devoted to a discussion of the dependence of the portfolio decision on the current dividend yield and of the economic value of market timing. In subsection C, we briefly comment on the results when β follows an O-U process.

A. Parameter Uncertainty, Interim Consumption and Horizon

Table II summarizes the solution to an investor's optimization problems with and without interim consumption. The results are obtained by assuming $b_0 = 4.5$ and $\nu_0 = 4.0$, which corresponds to the VAR estimate b and its standard error using the whole sample of 1950-1997. Within each table, the optimal proportion of wealth allocated to stock, the myopic stock allocation, the two components of hedge demands associated with learning and the stochastic dividend yield, are reported for five values of d/p and five investment horizons T . Consider the last five columns of Panel A, where optimal stock allocations are reported for an investor maximizing the expected utility of terminal wealth. We observe roughly two patterns of horizon effect. When $d/p < 4\%$, the optimal allocation generally increases with the horizon. When $d/p \geq 4\%$, it first increases and then decreases with the horizon. This contrasts with Barberis (1998), who finds that parameter uncertainty reduces the impact of horizon, but that the long horizon investor still holds more stock as compared to short horizon ones²³.

The driving force for the difference is the two hedge components. The first hedge component directly reflects the effect of parameter uncertainty in a dynamic setting. As shown in Panel C,

the magnitude of this hedge demand increases with the investment horizon and the absolute value of $(s - \bar{s})$. In the case of $d/p = 2\%$, for example, this first hedge component is 0% of the wealth when $T = 1m$, and it increases to 25% for $T = 20y$. As d/p is closer to its long run of 4%, the size of the hedge demand decreases for all horizons. This hedge component is positive when $s < \bar{s}$ and negative when $s > \bar{s}$. Suppose that $b_0 > 0$ and the investor observes $s > \bar{s}$, then an unexpectedly high stock return (a piece of good news) means that the current estimate is too low and the investor revises the estimate upward (another piece of good news) so that the revision of b , which represents a change in the predictability and the future investment opportunity set, is positively correlated with current stock return. To hedge for the revision of b or the change in the future opportunity set, the investor wants to sell (hold) something positively (negatively) correlated with it. Therefore, the hedge demand for stock associated with state variable b is negative. Conversely, if $s < \bar{s}$, an unexpectedly high stock return (a piece of good news) means that the current estimate of b is too high and the investor revises the estimate downward (a piece of bad news) so that the covariance between db and the stock is negative. To hedge for the revision of b , the investor's hedge demand for stock is positive. The second hedge component indirectly reflects the impact of parameter uncertainty. When $\nu_0 = 0$, both the optimal allocation and the hedge demand, as shown in Appendix B, increase with horizon when the current estimate of excess stock return is positive. However, when $\nu_0 = 4.0$, as shown in Panel D of Table II, the second hedge demand can first increase and then decrease with horizon. A possible reason is that the investor's indirect utility function with parameter uncertainty becomes less and less sensitive to the predictive variable as the horizon becomes longer, so that $\frac{\partial \ln \phi(b, s, \nu, t)}{\partial s}$ first increases and then decreases with T . This is reasonable because the investor is uncertain about the predictive power of dividend yield, and the uncertainty becomes more important as T increases.

Comparing results in the first five columns with those in the last five columns, we find that the general horizon effect with and without interim consumption is similar qualitatively, but that the magnitude of the effect is smaller. Interim consumption reduces the effective horizon relative to the stated horizon, and so the horizon effect is accordingly smaller. This is also reflected in the difference in hedge demands between long and short horizon investors. The two hedge components with interim consumption are sometimes only 50% of those with terminal wealth.

This indicates that the presence of interim consumption mainly reduces the impact of parameter uncertainty for any given investment horizon.

The total impact of parameter uncertainty is further highlighted in Table IV., where the optimal demand for stock is compared for the myopic strategy (KS), approximates the strategy adopted by an investor in Kandel and Stambaugh (1996), the dynamic strategy without parameter uncertainty (BSL), which approximates the strategy adopted by the investor in Brennan, Schwartz and Lagnado (1997), and the optimal strategy (O). Panel A compares the portfolio allocation under the three strategies for $b_0 = 0$. Under both KS and BSL strategies, $b_0 = 0$ implies that the stock return is not predictable and follows an i.i.d. distribution. The optimal allocation for every horizon is a constant myopic 55%. In contrast, the optimal allocation under strategy O is horizon-dependent. Even though the current estimate indicates no predictability, the investor optimally allows for the possibility that β is not equal to zero and he will learn more about it in the future. Panel B reports results for $b_0 = 4.5$. A certain return predictability plus a negative correlation between dividend yield and stock return introduces a mean-reverting pattern into the return, and the allocation under BSL strategy increases with the horizon, because the hedge demand for stock is positive and increases with horizon. In contrast, when the return predictability is uncertain, a high current return could imply a better future investment opportunity set as mentioned earlier, which results in positive serial correlation in stock return and a negative horizon effect. Not surprisingly, the allocation under the O strategy could decrease with the horizon. The importance of the indirect impact of parameter uncertainty is indicated in the case of $d/p = 4\%$. Even though the direct impact given in the first hedge component is zero, the optimal allocation under BSL and O are quite different, because the indirect utility function, ϕ , and its sensitivity to a change in d/p , $\frac{\partial \ln \phi(b,s,\nu,t)}{\partial s}$, differ in the two strategies.

The comparison of the BSL and O strategies is also carried out in Figure 1. We note that the horizon effect is state-dependent. It could be increasing, decreasing or non-monotone in horizon. Thus, although horizon matters, the conventional wisdom that the young investor holds more proportion of wealth in equity does not hold for every state and for every investor.

Tables IV, V and VI examine the impact of parameter uncertainty for a long horizon investor when he has different levels of prior uncertainty, ν_0 . Table IV shows that the size

of the first hedge component increases with the level of parameter uncertainty: the investor wants to hold more stock to hedge the revision of b when he is more uncertain about it. The sign of this hedge component changes when d/p changes from below to above \bar{s} , because the covariance between db and stock return changes sign. Table V shows that, while an increase in parameter uncertainty increases the importance of the first hedge demand, an increase in the level of parameter uncertainty reduce the importance of the second hedge component as the size of the hedge demand for stock is smaller. When parameter uncertainty is high, the optimal strategy and the optimal indirect utility function are less responsive to a change in the dividend yield, because the investor optimally choose to act more conservatively in return predictability. This conservatism makes the indirect expected utility of wealth and thus the hedge component against the change in dividend yield small. As a result, the first (second) hedge component dominates when ν_0 is high (low). When the first hedge component decreases with parameter uncertainty, then the optimal demand for stock also decreases with it. When the first hedge component increases with parameter uncertainty, the optimal demand for stock also increases with it as observed in Table VI.

From these tables, we also observe that the magnitude of the hedge demand increases with b_0 for any fixed d/p and ν_0 . As b_0 gets larger, return predictability is more important. A change in b_0 not only affects the size of the short run expected stock return but also the importance of stock return predictability in the long run. Thus, the need to hedge increases with the size of b_0 .

We can gain a little intuition about the horizon effect on optimal portfolio choice by looking at the moments of cumulative stock returns for a buy-and-hold mean-variance optimizing investor. Appendix C contains the derivation of the mean, variance and covariance for stock returns over horizon $(T - t)$ when the return predictability is certain. From equations (C4) and (C5), we know that both the cumulative mean, E , and the cumulative variance, V , increase with the horizon. For long horizon investors and with commonly estimated parameter values, E generally increases faster than V , so that the mean variance ratio, $\frac{E}{V}$, increases with the horizon. A buy-and-hold mean-variance optimizing investor would hold more wealth in stock when the investment horizon is longer. For an investor engaging in optimal dynamic rebalancing, we consider the first order autocovariance, C_1 , of the stock return as given in equation (C8). A large negative ρ_{bP} implies a

negative C_1 , which in turn implies mean-reversion in the stock return. Not surprisingly, both the current analysis and that of Barberis (1999) find that the optimal stock allocation is increasing in the horizon when there is no learning.

B. Market Timing and the Value of Uncertain Predictability

A natural consequence of return predictability is market timing: the optimal stock allocation depends on the current value of the predictive variable. The intensity of market timing depends crucially on future learning and investment horizon.

Figure 2 compares the optimal market timing under the KS, BSL and the optimal strategies. The first two panels assume that $b_0 = 0$, i.e., the current estimate indicates no return predictability. Both KS and BSL treat stock return as if it was truly i.i.d.. Not surprisingly, both stock allocation rules do not depend on the current observation of d/p . The optimal strategy, instead, correctly allows for the possibility of future learning and accordingly hedges for future updating in b . There are two competing effects in the optimal portfolio allocation when d/p increases. On the one hand, an increase in the dividend yield implies a higher expected stock return, which calls for an increase in stock allocation. On the other hand, when the dividend yield becomes much larger than its long run mean and the expected stock return becomes too high, the investor becomes more cautious and the hedge demand associated with parameter uncertainty becomes more negative, so the optimal allocation should decrease with the dividend yield. Whether the optimal stock demand increases or decreases with dividend yield depends on which term is more important. The negative hedge component eventually dominates, and a further increase in dividend yield implies that the investor reduces the optimal demand for stock.

The bottom two panels report the comparison when $b_0 = 1.5$, i.e., the current estimate indicates return predictability. Both KS and BSL strategies lead to a linear dependence of optimal allocation on the predictive variable as also given in Campbell and Viceira (1998). The stock allocation under the KS strategy is $x_{KS}^* = \frac{\bar{\mu} - r + b_0(s - \bar{s})}{\gamma\sigma_P^2}$, and x_{KS}^* increases with s with a slope of $\frac{b_0}{\gamma\sigma_P^2}$. The stock allocation under the BSL strategy, x_{BSL}^* , is given in equation (B4), and is linearly dependent on s , with a slope of $\frac{b_0}{\gamma\sigma_P^2} + \frac{b_0^2\rho_{sP}\sigma_s C(\tau)}{\gamma\sigma_P}$, where $C(\tau) < 0$ is given in (B11) and ρ_{sP} is negative.

Thus, the slope of the BSL strategy is always larger than that of the KS strategy by the amount of $\frac{b_0^2 \rho_s P \sigma_s C(\tau)}{\gamma \sigma_P}$. As dividend yield becomes large, the investor with KS or BSL strategy holds a very large position in stock. The investor with the optimal strategy instead holds very reasonable positions even when dividend yield becomes extremely large, because the negative hedge demand for parameter uncertainty is more than enough to offset the increase in the allocation caused by a greater expected return. In contrast with the linear dependence of KS and BSL strategy, the optimal strategy again leads to a hump-shaped relation between optimal allocation and the predictive variable.

This non-monotone dependence of the optimal strategy on the predictive variable is also given in Stambaugh (1999). The pattern in Stambaugh (1999) arises from the conditional skewness in the predictive distribution of long-horizon returns in the case of large dividend yield. Our observation is driven mainly by the increasingly large negative hedge demand associated with parameter uncertainty as the predictive variable becomes larger.

Figure 3 plots the proportion of wealth invested in stock given historical data of stock return and dividend yield from January 1978 to December 1997. The investment horizon is fixed at twenty years, so the investor starts with a targeted horizon date at December 1997 and faces a horizon date moving through time. We assume that the investor's risk aversion parameter is equal to 5.0. The investor uses the VAR regression results, $\hat{\beta}$ and $var(\hat{\beta})$, from the sample of January 1950 - December 1977, as his prior mean and variance of the estimate, $b_0 = 6.76$ and $\nu_0 = 6.38$. He updates the estimate b and its variance ν as he observes the stock return and dividend yield through time. The portfolio choice is calculated under the optimal, the "KS", the "BSL" and the "iid" strategies. Under the "iid" strategy the investor ignores predictability and assumes that the expected stock return is constant and equals $\bar{\mu}$, so that there is no market timing and the proportion of wealth allocated to stock is a constant 0.55. The three other strategies take strong advantage of return predictability. Compared to the optimal strategy, "BSL" strategy, which does not account for parameter uncertainty, shows great variability through time when the stock return and dividend yield change. In a short period, the "BSL" strategy could adopt very large positive or negative positions, indicating a much too aggressive market timing. Although "KS" strategy is less aggressive, the magnitude and direction of market timing differ quite substantially from those

under the optimal strategy in certain periods. In 1990's, the large stock price run-up resulted in a historically low dividend yield. The investor continuously revised the estimate b downward, but it was still positive at the end of 1997. A positive b with a very small dividend yield implies that the expected excess stock return, $b(s - \bar{s})$, is negative. Not surprisingly, all market timing strategies call for large short positions in the stock market in recent years. The short position for the optimal strategy, however, is the smallest, because conditional on the current state of $s < \bar{s}$ the hedge demand associated with parameter uncertainty is positive, which offsets the negative stock demand associated with a negative excess stock return. As the estimate of β becomes more precise, the hedge demand becomes less important, and the optimal portfolio strategy converges toward that of BSL as time passes.

To assess the economic value of uncertain stock return predictability, we run a horse race among the four strategies. Define the certainty equivalent wealth (CEW) of a specific strategy δ for an investor as the amount of wealth that makes the investor indifferent between receiving CEW for sure at the horizon T and having \$1 today to invest up to the horizon using strategy δ :

$$e^{-\rho T} \frac{CEW_{\delta}(b, s, \nu, T)^{1-\gamma}}{1-\gamma} = \frac{\$1^{1-\gamma}}{1-\gamma} \phi(b, s, \nu, T|\delta), \quad (25)$$

which simplifies to:

$$CEW_{\delta}(b, s, \nu, T) = e^{\rho T} \phi(b, s, \nu, T|\delta)^{\frac{1}{1-\gamma}},$$

where T is the remaining time to the horizon. Define the present value of the $CEW_{\delta}(b, s, \nu, T)$ as $PVCEW_{\delta}(b, s, \nu, T) \equiv e^{-rT} CEW_{\delta}(b, s, \nu, T)$, where r is the constant real risk free rate. $\phi(b, s, \nu, T|\delta)$ is the value of the investment opportunities remaining till the horizon T for an investor with \$1 of wealth. It is computed numerically by solving the optimization equation with the portfolio weight given by strategy δ .

Figure 4 highlights the economic value of the optimal vs. the myopic ‘‘KS’’ strategy using the historical stock returns and dividend yield. The investor's prior mean and variance of β , b_0 and ν_0 , are set to equal to 6.76 and 6.38, the VAR regression results, $\hat{\beta}$ and $var(\hat{\beta})$, with a sample of January 1950 to December 1977. b_t and ν_t are updated each month from January 1978 to December 1997 as the investor observes the historical values of stock return and the

dividend yield. The optimal strategy, which correctly hedges for the updating of b , significantly improves the investor’s welfare at every point of time. In certain historical periods, the optimal strategy can double the investor’s present value of certainty equivalent wealth when compared to the myopic strategy. Even in recent years, when “KS” strategy is close to the optimal strategy, the improvement is still about 15%. This indicates that dynamic strategy allowing for learning has significant economic value to the investor.

C. Stochastic β

When β is stochastic and follows an O-U process, we estimate λ , \bar{b} and σ_β together with parameters in Table I. We discretize equations (12) and (11) with a one-month time step and write them in a state space representation. The parameters of this state space model are estimated from monthly data on the real stock return and dividend yield via the Kalman filter algorithm by assuming that the state variable β is uncorrelated with the observations of stock return and dividend yield.

The estimates for λ , \bar{b} and σ_β are, respectively, 0.115, -1.0 and 1.226. The estimates indicate that the β process is very persistent with high volatility. Using this estimates and the assumption that $\rho_{s\beta} = 0$ and $\rho_{P\beta} = 0$, we can solve the partial differential equation for the optimal portfolio weights. Because of the assumption $\rho_{P\beta} = 0$, the third hedge component in equation (22) remains zero, but the importance of the other two hedge components changed. Even when the investor treats the current estimate of β as its true value (i.e., $\nu_0 = 0$), there are two state variables, β and d/p , governing the investment opportunity set. Because β varies stochastically over time, the investor’s hedge behavior associated with d/p changes as well. When β is constant, the mean reverting of d/p and the negative correlation between d/p and stock return translate into a mean reverting pattern in stock return, and thus the investor’s hedge demand for stock as well as optimal stock allocation increases with horizon. In contrast, the optimal stock allocation with stochastic β first increases and then decreases with horizon. The effect of learning in this case, however, is similar to the case with constant β . We find that the optimal stock allocation generally decreases with horizon when $d/p > 4\%$ and increases with horizon when $d/p \leq 4\%$.

The sensitivity of optimal stock allocation to dividend yield is generally smaller when there is parameter uncertainty. At a specific horizon with the same parameter values, the investor with parameter uncertainty ($\nu_0 > 0$) usually holds less stock as compared with the investor without uncertainty ($\nu_0 = 0$) when $d/p > 4\%$ and the opposite is true for $d/p \leq 4\%$. This indicates that the hedge demand for stock associated with parameter uncertainty is generally negative for $d/p > 4\%$ and positive for $d/p \leq 4\%$. All these qualitative effects of parameter uncertainty with stochastic β are consistent with those with constant β .

IV. Conclusions and Future Work

We have analyzed the effect of learning on optimal consumption and portfolio choice for an investor with a long investment horizon when there is uncertain evidence of return predictability. The uncertainty about the predictive parameter introduces dynamic learning into the model. Our learning model is characterized by a stochastic variance of the estimate and a stochastic covariance between the stock return and the current estimate of the predictive parameter. The optimal stock allocation with learning can increase, decrease or vary non-monotonically with the horizon, which is quite different from when there is no learning or predictability. The prospect of learning affects not only the magnitude but also the sign of the horizon effect. In addition, interim consumption reduces or even reverses the relation between the optimal stock allocation and the horizon. We examine the dependence of the optimal allocation on the current value of the predictive variable. When there is no learning, the long-horizon investor times the market more aggressively than a myopic investor. In contrast, a myopic investor times the market more aggressively than a long-horizon investor when there is learning. The consideration of interim consumption reduces this sensitivity. Simulation results using historical data show that investors who ignore the opportunity of market timing can incur very large opportunity costs, so that return predictability, even if quite uncertain, is economically valuable. The hedge demands associated with learning and stochastic predictive variable depend on the current estimate of the parameter, the level of uncertainty about the estimate, the value of the predictive variable, and the investment horizon. The optimal hedge demands are non-zero even if the current estimate indicates no return predictability.

This study is only an initial attempt to assess the economic importance of the empirical evidence on the stock return predictability. The dynamic effect of learning has substantial implications for the optimal portfolio choice and investor welfare. Although we have considered uncertainty about the predictive parameter β , we have assumed that the investor knows everything else: the nature of the predictive relation and the process for the predictive variable. We have also ignored model uncertainty. For example, the investor could have probability distributions over two possible stock return processes: one with i.i.d. returns and the other with predictability. Brennan and Xia (1999) assess the importance of this type of model uncertainty on the optimal portfolio decision. A challenging future task is to combine model uncertainty with parameter uncertainty in the context of predictable stock returns.

Table I:
Model Calibration: Parameter Values Used in the Numerical Analysis

This table lists parameter notations and calibrated values used in the numerical analysis of the nonlinear second order differential equations in terms of ϕ with four state variables (b, s, ν, ℓ) given in footnote 16. The parameter values are estimated from U.S. historical monthly data of real stock returns and dividend yield from January 1950 to December 1997. When β is constant, parameter values are estimated using (nonlinear) OLS. When β is stochastic, parameter values are estimated using Kalman filter.

Parameter Descriptions	Notation	Parameter Values
		β is constant
1. volatility (standard deviation) of the dividend yield	σ_s	0.6%
2. historical long run mean of the stock return	$\bar{\mu}$	9.1%
3. volatility (standard deviation) of the stock return	σ_P	14.4%
4. correlation coefficient between the dividend yield and the stock return processes	ρ_{sP}	-0.93
5. mean reversion coefficient for the dividend yield process	κ	0.19
6. historical long run mean of the dividend yield	\bar{s}	4.0%
7. coefficient of risk aversion	γ	5.0
8. subjective discount factor	ρ	1%
9. real interest rate	r	3.4%
10. current estimate of β (January 1950 - December 1997)	b_0	4.54
11. current estimate variance	ν_0	4.33
12. current estimate of β (January 1950 - December 1977)	b_0	6.76
13. current estimate variance	ν_0	6.38

Table II:
Optimal Stock Allocation: The Horizon Effect (1)
 $(\gamma = 5.0, b_0 = 4.5 \text{ and } \nu_0 = 4.0)$

This table presents of the optimal stock allocation, the myopic allocation and the two hedge demands for stock at different investment horizons for different values of the investment horizon and the predictive variable (dividend yield) $s \equiv d/p$ when both predictability and learning are present. We assume that the true value of beta is a constant. The myopic stock allocation is defined by $x_1^* = \frac{\bar{\mu} + b(s - \bar{s}) - r}{\gamma \sigma_P^2}$, and the first hedge component is defined by $x_2^* = \frac{\phi_b}{\gamma \sigma_P^2 \phi} \nu (s - \bar{s})$, while the second hedge demand is defined by $x_3^* = \frac{\phi_s}{\gamma \sigma_P^2 \phi} \sigma_P \sigma_s \rho_{sP}$. The optimal stock allocation is then given by $x_1^* + x_2^* + x_3^*$. All results are obtained by numerically solving the nonlinear second order differential equations in terms of ϕ with three state variables (b, s, ν) . The first five columns report the optimal stock allocation when the investor maximizes the expected utility of lifetime consumption, and the second five columns report the allocation when the investor maximizes the expected utility of terminal wealth. In this table, the investor's risk aversion coefficient is $\gamma = 5.0$. The uncertainty about the investor estimate is given by $\sqrt{\nu_0} = 2$, which corresponds to the standard deviation of $\hat{\beta}$ in the VAR regression if the investor uses the sample period of 1950-1997 to form his prior. The allocation is conditional on a prior of $b_0 = 4.5$, which is the VAR estimate of β for the same sample period.

d/p	Interim consumption					Terminal Wealth				
	$1m$	$1y$	$5y$	$10y$	$20y$	$1m$	$1y$	$5y$	$10y$	$20y$
Panel A: Optimal Stock Allocation										
2.0%	-0.34	-0.34	-0.32	-0.27	-0.18	-0.32	-0.34	-0.29	-0.14	0.09
3.0	0.12	0.13	0.20	0.27	0.36	0.12	0.13	0.27	0.43	0.54
4.0	0.56	0.61	0.75	0.83	0.86	0.56	0.63	0.87	0.95	0.88
5.0	0.99	1.07	1.19	1.22	1.19	0.99	1.09	1.27	1.23	1.04
6.0	1.43	1.49	1.49	1.42	1.32	1.42	1.48	1.46	1.30	1.05
Panel B: Myopic Allocation of Stock x_1^*										
2.0%	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32
3.0	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
4.0	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
5.0	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
6.0	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42
Panel C: Allocation to Stock to Hedge Parameter Uncertainty x_2^*										
2.0%	0.00	0.01	0.04	0.06	0.11	0.00	0.02	0.06	0.11	0.25
3.0	0.00	0.00	0.01	0.02	0.05	0.00	0.00	0.01	0.05	0.15
4.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.0	0.00	-0.01	-0.06	-0.09	-0.14	0.00	-0.02	-0.10	-0.17	-0.26
6.0	-0.01	-0.08	-0.25	-0.33	-0.41	-0.01	-0.08	-0.33	-0.47	-0.62
Panel D: Allocation to Stock to Hedge the Stochastic Predictive Variable x_3^*										
2.0%	-0.02	-0.03	-0.05	-0.02	-0.03	0.00	-0.03	-0.04	0.07	0.16
3.0	0.00	0.02	0.08	0.14	0.19	0.00	0.02	0.14	0.26	0.28
4.0	0.01	0.06	0.20	0.28	0.31	0.01	0.08	0.32	0.40	0.33
5.0	0.01	0.10	0.27	0.33	0.35	0.01	0.12	0.39	0.41	0.31
6.0	0.02	0.15	0.31	0.33	0.32	0.01	0.15	0.37	0.35	0.25

Table III:
Stock Allocation Under Different Investment Strategies
 $(\nu_0 = 4.0, T = 20 \text{ years and } \gamma = 5.0)$

This table compares the stock allocation decision under three different strategies with a **twenty-year** investment horizon for different values of the predictive variable d/p and the investment horizon T . All the three strategies give rise to almost same optimal stock allocations as the myopic one when the horizon is one-month, so the comparisons for one-month horizon is omitted. The results are obtained by solving the nonlinear second order differential equations in terms of ϕ with three state variables (b, s, ν) . In this table, the first five columns report comparisons when the investor maximizes the expected utility of lifetime consumption and the second five columns report results when he maximizes the expected utility of terminal wealth. The level of parameter uncertainty is summarized by $\sqrt{\nu_0} = 2.0$, which corresponds to the standard deviation of $\hat{\beta}$ in the VAR regression for the sample period 1950-1997. Investment strategy (i) is the “myopic portfolio strategy” (KS), strategy (ii) is the “dynamic portfolio strategy ignoring learning” (BSL), and strategy (iii) is the “dynamic portfolio strategy considering learning” (O). The iid strategy implies a constant stock allocation for a given γ , which equals 0.55 when $\gamma = 5.0$. Panel A reports the optimal portfolio under the three strategies when prior estimate is $b_0 = 0.0$ (current estimate indicates no predictability), and Panel B reports the results by assuming $b_0 = 4.5$ (VAR estimate of the predictive coefficient in the whole sample of 1950-1997).

d/p Strategy		Investment Horizon: T									
		1m	1y	5y	10y	20y	1m	1y	5y	10y	20y
		Interim Consumption					Terminal Wealth				
Panel A: $b_0 = 0.0\%$											
2%	(KS)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(BSL)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(O)	0.53	0.53	0.50	0.49	0.50	0.55	0.52	0.48	0.50	0.56
4%	(KS)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(BSL)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(O)	0.55	0.55	0.55	0.55	0.54	0.55	0.55	0.55	0.54	0.52
6%	(KS)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(BSL)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(O)	0.55	0.53	0.49	0.46	0.43	0.55	0.52	0.45	0.40	0.32
Panel B: $b_0 = 4.5\%$											
2%	(KS)	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32	-0.32
	(BSL)	-0.32	-0.35	-0.36	-0.32	-0.25	-0.32	-0.35	-0.35	-0.20	0.08
	(O)	-0.34	-0.36	-0.32	-0.27	-0.18	-0.32	-0.34	-0.29	-0.14	0.09
4%	(KS)	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(BSL)	0.55	0.61	0.77	0.91	1.04	0.56	0.63	0.97	1.32	1.71
	(O)	0.56	0.61	0.75	0.83	0.86	0.56	0.63	0.87	0.95	0.88
6%	(KS)	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42
	(BSL)	1.42	1.56	1.85	2.02	2.15	1.43	1.61	2.29	2.85	3.34
	(O)	1.42	1.48	1.49	1.44	1.35	1.42	1.48	1.46	1.30	1.05

Table IV:
Optimal Stock Demand to Hedge Parameter Uncertainty
($T = 20$ years and $\gamma = 5.0$)

This table summarizes the effect of parameter uncertainty on the optimal stock allocation for different values of the investor's prior b and predictive variable $s \equiv d/p$. The results are obtained by solving the nonlinear second order differential equations in terms of ϕ with three state variables (b, s, ν) , and the investment strategy is the "dynamic portfolio strategy considering parameter uncertainty". When the investment horizon is one month, the optimal stock allocation is the same as the myopic one, so the comparison is trivial and omitted. Only the comparisons with a twenty year horizon is reported. Panel A reports results with $d/p = 2\%$, panel B has results with $d/p = 4\%$, and Panel C with $d/p = 6\%$. An investor using the whole sample estimation has a $\mu = 4.0$; an investor using Barberis' sample or the sample of first twenty years' data has roughly a $\mu = 6.0$; if the investor ignores parameter uncertainty, then $\nu_0 = 0.0$. The remaining values of ν_0 are randomly chosen.

Levels of Parameter Uncertainty ν_0						
Prior b_0	$\nu_0 = 0.0$	$\nu_0 = 1.0$	$\nu_0 = 2.0$	$\nu_0 = 3.0$	$\nu_0 = 4.0$	$\nu_0 = 6.0$
Panel A: $d/p = 2\%$						
0.0	0.00	0.01	0.02	0.03	0.03	0.05
1.5	0.00	0.03	0.06	0.08	0.10	0.13
2.5	0.00	0.05	0.08	0.12	0.15	0.20
3.5	0.00	0.06	0.11	0.16	0.20	0.26
4.5	0.00	0.08	0.15	0.20	0.25	0.33
Panel B: $d/p = 4\%$						
0.0	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.00	0.00	0.00	0.00	0.00
2.5	0.00	0.00	0.00	0.00	0.00	0.00
3.5	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: $d/p = 6\%$						
0.0	0.00	-0.08	-0.14	-0.18	-0.22	-0.28
1.5	0.00	-0.14	-0.23	-0.29	-0.35	-0.43
2.5	0.00	-0.18	-0.29	-0.37	-0.44	-0.53
3.5	0.00	-0.22	-0.36	-0.45	-0.53	-0.64
4.5	0.00	-0.24	-0.41	-0.53	-0.62	-0.74

Table V:
Optimal Stock Demand to Hedge Stochastic Predictive Variable
($T = 20$ years and $\gamma = 5.0$)

This table summarizes the effect of parameter uncertainty on the optimal stock allocation for different values of the investor's prior b and predictive variable $s \equiv d/p$. The results are obtained by solving the nonlinear second order differential equations in terms of ϕ with three state variables (b, s, ν) , and the investment strategy is the "dynamic portfolio strategy considering parameter uncertainty". When the investment horizon is one month, the optimal stock allocation is the same as the myopic one, so the comparison is trivial and omitted. Only the comparisons with a twenty year horizon is reported. Panel A reports results with $d/p = 2\%$, panel B has results with $d/p = 4\%$, and Panel C with $d/p = 6\%$. An investor using the whole sample estimation has a $\mu = 4.0$; an investor using Barberis' sample or the sample of first twenty years' data has roughly a $\mu = 6.0$; if the investor ignores parameter uncertainty, then $\nu_0 = 0.0$. The remaining values of ν_0 are randomly chosen.

Levels of Parameter Uncertainty ν_0						
Prior b_0	$\nu_0 = 0.0$	$\nu_0 = 1.0$	$\nu_0 = 2.0$	$\nu_0 = 3.0$	$\nu_0 = 4.0$	$\nu_0 = 6.0$
Panel A: $d/p = 2\%$						
0.0	0.00	0.00	-0.01	-0.02	-0.02	-0.02
1.5	0.15	0.11	0.09	0.07	0.06	0.05
2.5	0.25	0.18	0.15	0.12	0.11	0.09
3.5	0.36	0.22	0.18	0.16	0.14	0.12
4.5	0.40	0.22	0.19	0.17	0.16	0.14
Panel B: $d/p = 4\%$						
0.0	0.00	-0.01	-0.02	-0.03	-0.03	-0.04
1.5	0.19	0.12	0.09	0.07	0.06	0.03
2.5	0.41	0.26	0.20	0.16	0.13	0.09
3.5	0.74	0.43	0.33	0.27	0.23	0.17
4.5	1.16	0.62	0.49	0.40	0.33	0.25
Panel C: $d/p = 6\%$						
0.0	0.00	-0.01	-0.01	-0.01	-0.01	-0.02
1.5	0.24	0.12	0.08	0.06	0.04	0.02
2.5	0.57	0.25	0.17	0.13	0.10	0.06
3.5	1.12	0.44	0.29	0.21	0.17	0.10
4.5	1.92	0.66	0.45	0.33	0.25	0.16

Table VI:
The Effect of Parameter Uncertainty on the Optimal Stock Allocation
($T = 20$ years and $\gamma = 5.0$)

This table summarizes the effect of parameter uncertainty on the optimal stock allocation for different values of the investor's prior b and predictive variable $s \equiv d/p$. The results are obtained by solving the nonlinear second order differential equations in terms of ϕ with three state variables (b, s, ν) , and the investment strategy is the "dynamic portfolio strategy considering parameter uncertainty". When the investment horizon is one month, the optimal stock allocation is the same as the myopic one, so the comparison is trivial and omitted. Only the comparisons with a twenty year horizon is reported. Panel A reports results with $d/p = 2\%$, panel B has results with $d/p = 4\%$, and Panel C with $d/p = 6\%$. An investor using the whole sample estimation has a $\mu = 4.0$; an investor using Barberis' sample or the sample of first twenty years' data has roughly a $\mu = 6.0$; if the investor ignores parameter uncertainty, then $\nu_0 = 0.0$. The remaining values of ν_0 are randomly chosen.

Levels of Parameter Uncertainty ν_0						
Prior b_0	$\nu_0 = 0.0$	$\nu_0 = 1.0$	$\nu_0 = 2.0$	$\nu_0 = 3.0$	$\nu_0 = 4.0$	$\nu_0 = 6.0$
Panel A: $d/p = 2\%$						
0.0	0.55	0.55	0.56	0.56	0.56	0.58
1.5	0.40	0.40	0.40	0.41	0.42	0.45
2.5	0.32	0.29	0.30	0.31	0.32	0.35
3.5	0.23	0.16	0.17	0.19	0.21	0.26
4.5	0.08	-0.02	0.02	0.06	0.09	0.16
Panel B: $d/p = 4\%$						
0.0	0.55	0.54	0.53	0.52	0.52	0.51
1.5	0.74	0.67	0.64	0.62	0.61	0.58
2.5	0.96	0.81	0.75	0.71	0.68	0.64
3.5	1.29	0.97	0.88	0.82	0.76	0.72
4.5	1.71	1.17	1.04	0.95	0.88	0.80
Panel C: $d/p = 6\%$						
0.0	0.55	0.46	0.40	0.36	0.32	0.26
1.5	1.08	0.81	0.69	0.60	0.54	0.43
2.5	1.60	1.10	0.91	0.78	0.69	0.56
3.5	2.34	1.44	1.16	0.99	0.86	0.69
4.5	3.34	1.84	1.45	1.22	1.05	0.83

Figure 1:
The Term Structure of Optimal Portfolio Allocation: Effect of learning
 ($\gamma = 5.0$ and $b_0 = 1.5$)

This figure plots the optimal proportion of wealth allocated to stock for a horizon of 1 month to 240 months. The optimal portfolio allocation when the investor faces a certain predictability is compared with the allocation when the investor faces uncertain predictability.

Legend: $\nu_0 = 0.0$: solid line; $\nu_0 = 4.0$: dashed line.

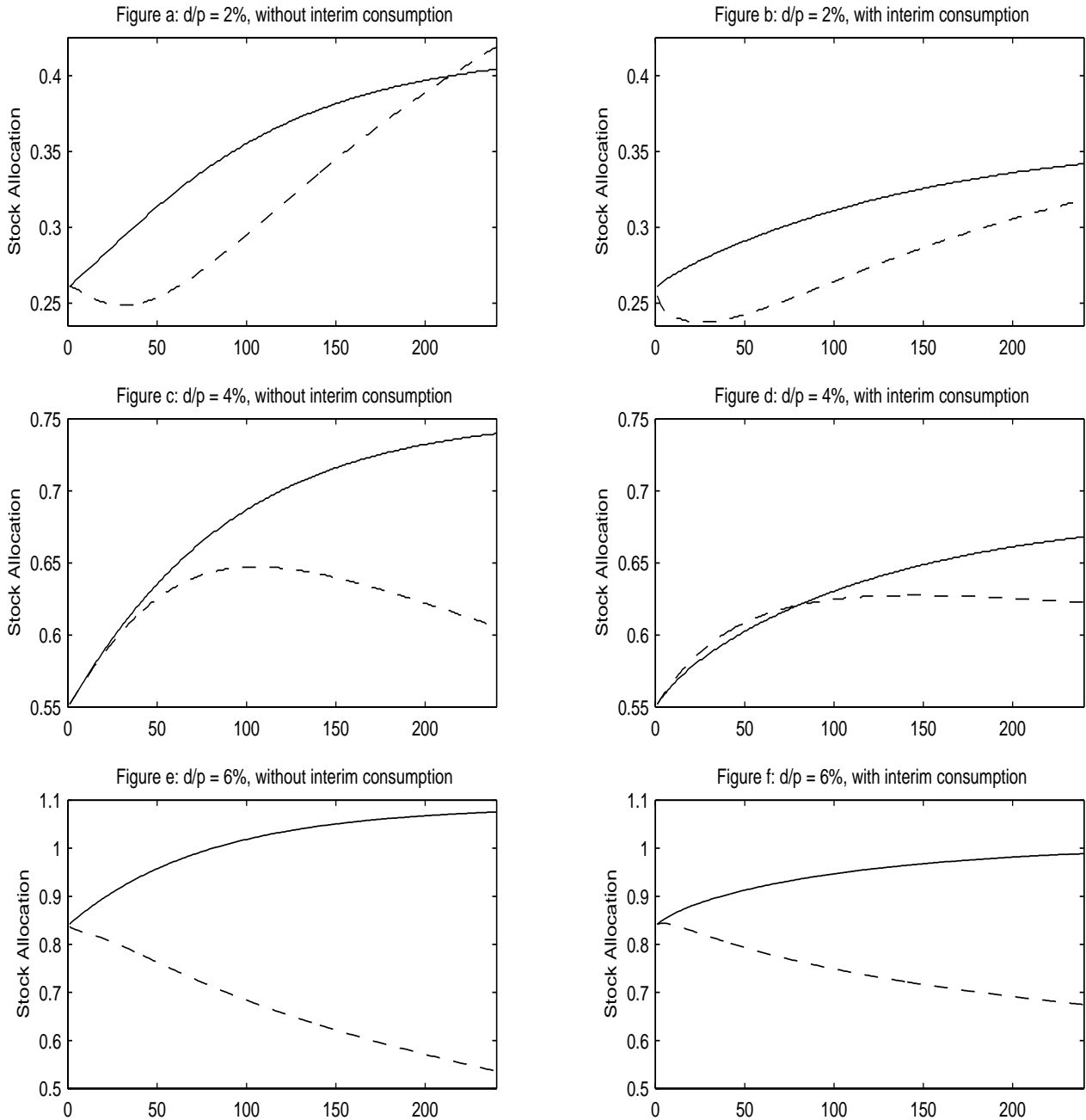


Figure 2:
 Optimal Allocation Conditional On Dividend Yield: Effect of Parameter Uncertainty on Market
 Timing
 ($\gamma = 5.0$ and $T = 20$ Years)

This figure plots the optimal proportion of wealth allocated to stock as a function of the current observation of dividend yield. The first two panels assume that $b_0 = 0$ and the last two assume that $b_0 = 1.5$. The allocations under myopic strategy, dynamic strategy without learning and optimal strategy are compared in each panel.

Legend: optimal strategy: solid line; dynamic strategy without parameter uncertainty: dashed line; myopic strategy: dotted line.

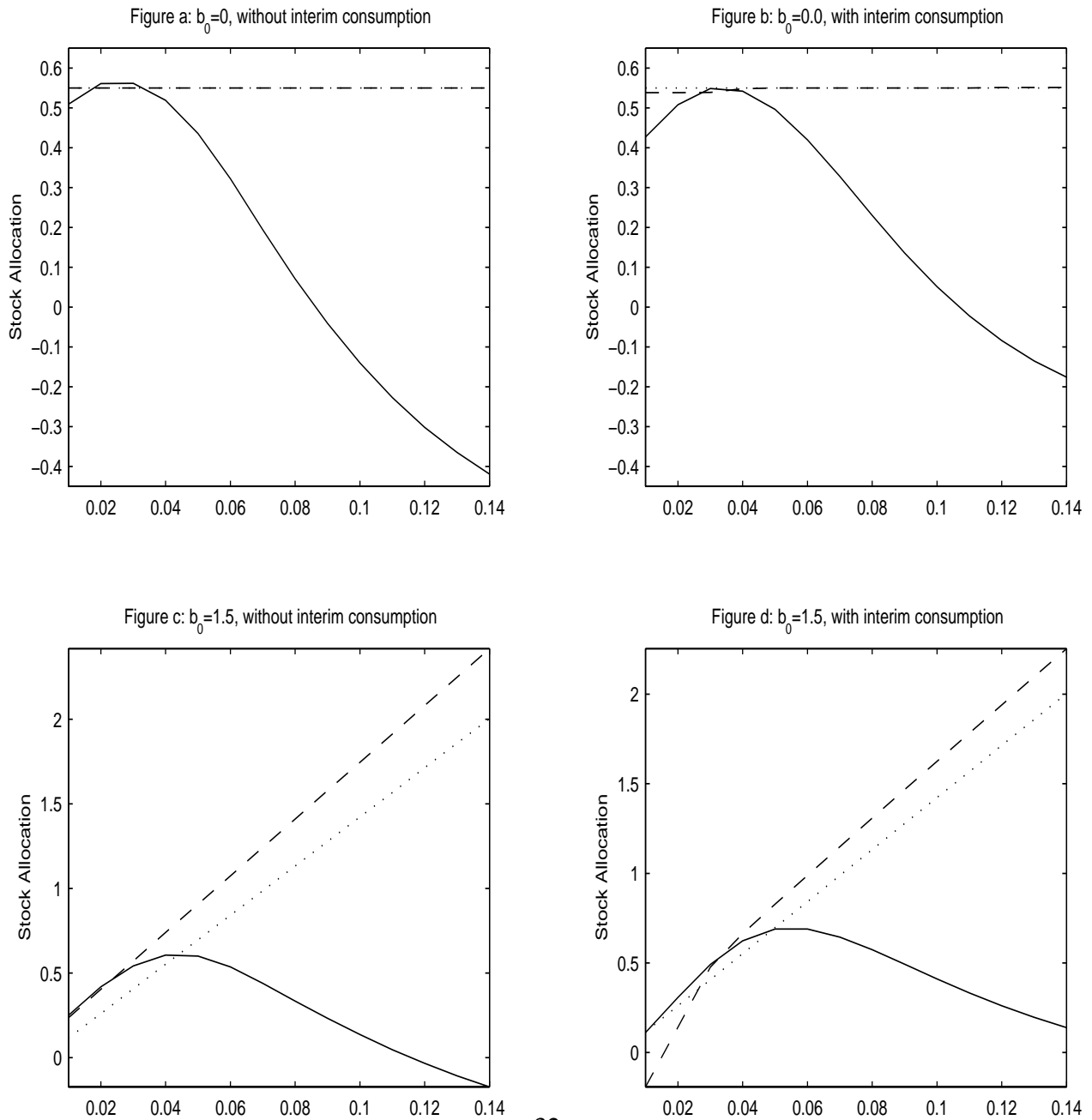


Figure 3:
 Simulated Optimal Portfolio Allocation Using Historical Stock Return and Dividend Yield
 ($\gamma = 5.0$, $b_0 = 6.76$, $\nu_0 = 6.38$ and $T = 20$)

This figure plots the optimal proportion of wealth allocated to stock for each month from January 1978 to December 1997 under the optimal, KS, BSL and iid investment strategies. The portfolios are not constrained to be non-negative. The portfolios are calculated using the historical value-weighted market stock return and the dividend yield to update b and ν through time. The prior b and ν_0 are set to equal the VAR estimation for the sample period of February 1950 to December 1977.

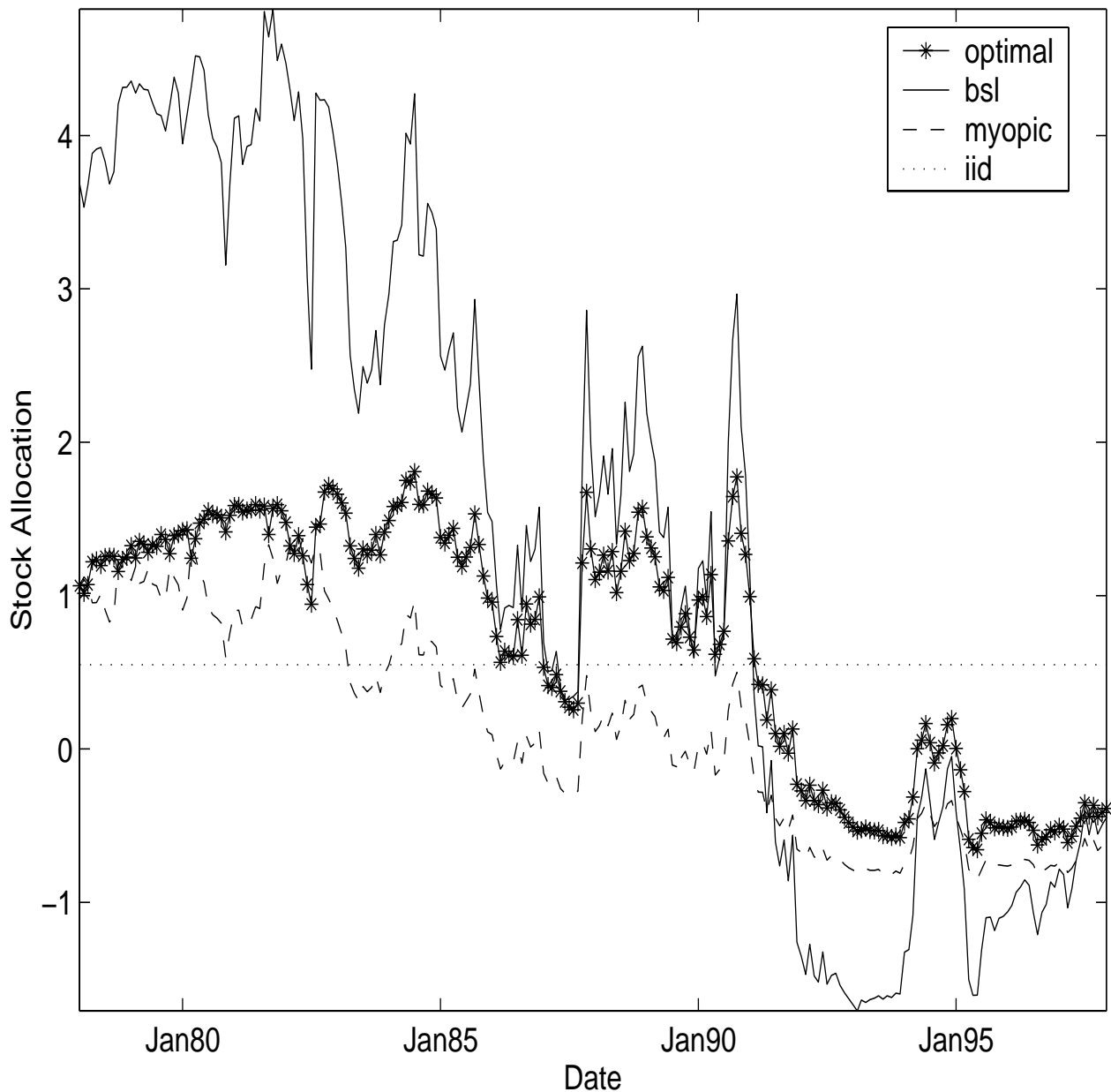
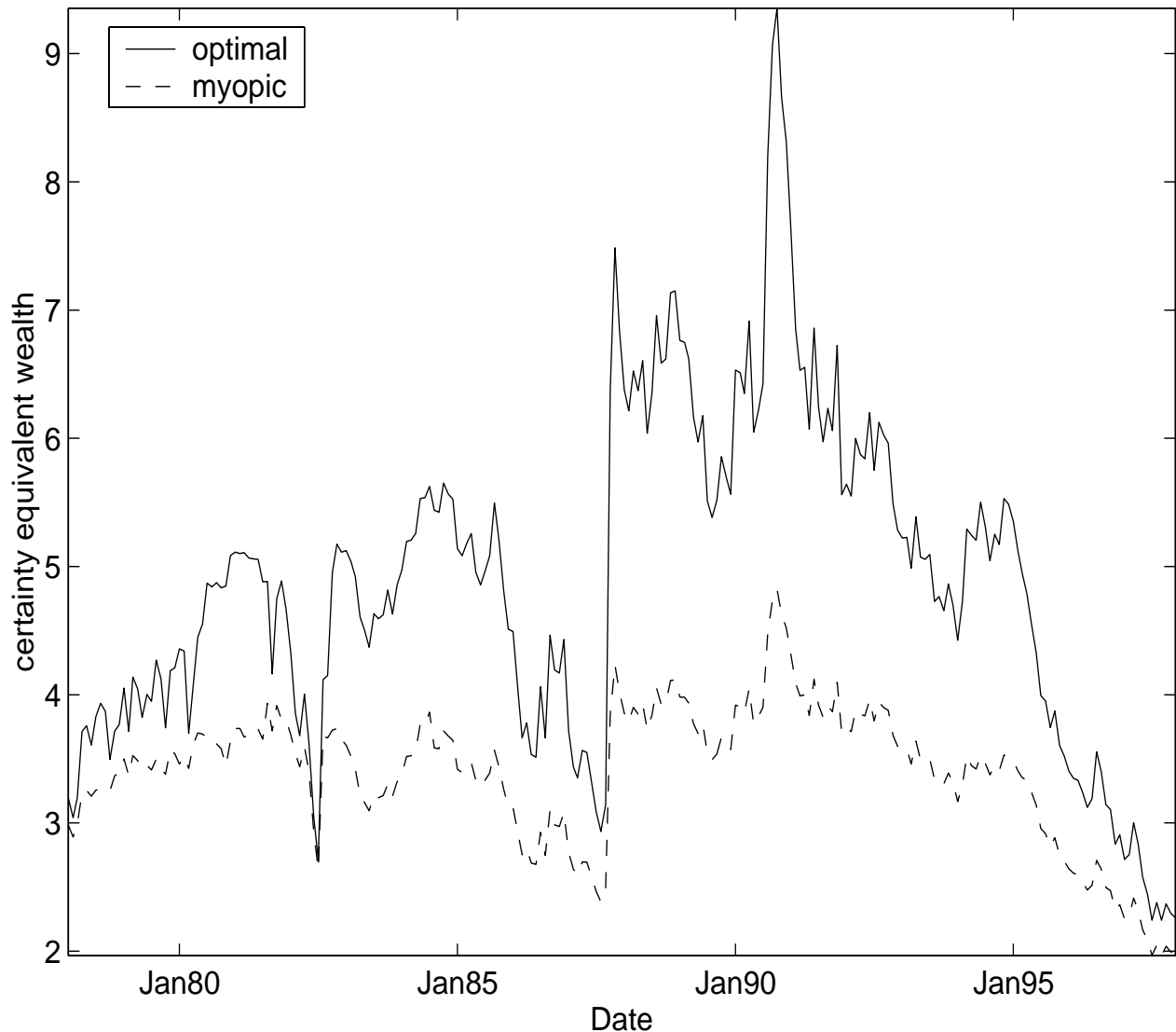


Figure 4:
 Economic Value of Uncertain Predictability: CEW with Historical Stock Return and Dividend Yield

$$(\gamma = 5.0, T = 20 \text{ years}, b_0 = 6.76 \text{ and } \nu_0 = 6.38)$$

This figure compares the economic value of the optimal and KS investment strategies for each month from January 1978 to December 1997. The portfolios are calculated using the historical value-weighted market stock return and the dividend yield to update b and ν through time. The prior $b_0 = 6.76$ and $\nu_0 = 6.38$ are set to equal the VAR estimation for the sample period of February 1950 to December 1977. The economic value is calculated in terms of PVCEW. Define the certainty equivalent wealth (CEW) of a specific strategy δ for an investor as the amount of wealth which makes the investor indifferent between receiving it for sure at the horizon T and having \$1 today to invest up to the horizon using strategy δ . Thus, the present value of the CEW of strategy δ for investment horizon T is defined by

$$PVCEW_{\delta}(b, s, \nu, T) = e^{-(r-\rho)T} \phi(b, s, \nu, T|\delta)^{\frac{1}{1-\gamma}}.$$



Appendix

A The Investor's Inference Problem

The investor's prior distribution over the initial value of β is assumed to be Gaussian, and β is assumed to follow (4). Because, (i) $\frac{dP}{P}$ and dS are following a joint Brownian motion; and (ii) all the parameters in (4), (2) and (5) are linear functions of the unobservable state variable β , the distribution function of β at time t , $F_t(x) = P(\beta \leq x | \mathcal{F}_t^I)$, is (conditionally) Gaussian given the investor's information structure at time t , \mathcal{F}_t^I , which is generated by the joint processes $\mathcal{I}(t) = (P(t), S(t))$. Let $b_t = E(\beta_t | \mathcal{F}_t^I)$, then b will be the optimal (in the mean square sense) estimate of β from $\mathcal{I}(t)$. The knowledge of the variance $\mathcal{U} = E((\beta - b)(\beta - b)' | \mathcal{F}_t^I)$, which is an $n \times n$ matrix, gives us a measure of filtering error and the investor's level of uncertainty. We call the process for $\mathcal{I}(t) = (P(t), S(t))$ the investor's signal used to form expectations of β . Let's write the processes for signals (the measurement equation) as:

$$d\mathcal{I} = (I_0(\mathcal{I}, t) + I_1(\mathcal{I}, t)\beta)dt + \omega(\mathcal{I}, t)dz, \quad (\text{A1})$$

where $I_0(\mathcal{I}, t)$ is an $(n + 1) \times 1$ vector with $\bar{\mu}$ in the first row and A_0 in the rest rows, $I_1(\mathcal{I}, t)$ and $\omega(\mathcal{I}, t)$ are matrices of the dimension of $((n + 1) \times n)$ and $((n + 1) \times n)$ with $S - \bar{S}$ and A_1 form the rows of the former and σ_P and σ_S form the rows of the latter. Rewrite the process for β (the transition equation) as

$$d\beta = [a_0(\mathcal{I}, t) + a_1(\mathcal{I}, t)\beta]dt + \eta(\mathcal{I}, t)dz. \quad (\text{A2})$$

Define $\Gamma(\mathcal{I}, t) = \eta\omega'$ as the covariance between the signals and the state variables, $\Sigma(\mathcal{I}, t) = \eta\eta'$ as the variance-covariance matrix of the state variables, and $\Phi(\mathcal{I}, t) = \omega\omega'$ as the variance-covariance matrix of the signal processes. A direct extension of Theorem 12.1 in Liptser & Shiriyayev (1978) to vectors of measurement and transition equations yields:

$$db(t) = [a_0(\mathcal{I}, t) + a_1(\mathcal{I}, t)b]dt + [\nu(t)I_1(\mathcal{I}, t)' + \Gamma(\mathcal{I}, t)]\Phi(\mathcal{I}, t)^{-1}\omega(\mathcal{I}, t)d\hat{z}(t), \quad (\text{A3})$$

$$d\nu(t) = a_1(\mathcal{I}, t)\nu(t) + \nu(t)a_1(\mathcal{I}, t)' + \Sigma(\mathcal{I}, t) \quad (\text{A4})$$

$$- [\nu(t)I_1(\mathcal{I}, t)' + \Gamma(\mathcal{I}, t)]\Phi(\mathcal{I}, t)^{-1}[\nu(t)I_1(\mathcal{I}, t)' + \Gamma(\mathcal{I}, t)]',$$

where

$$\begin{aligned} d\hat{z}(t) &= \omega^{-1} \left\{ d\mathcal{I} - E \left(d\mathcal{I} | \mathcal{F}_t^I \right) \right\} \\ &= \omega^{-1} \left\{ d\mathcal{I} - [(I_0(\mathcal{I}, t) + I_1(\mathcal{I}, t)b)dt] \right\}. \end{aligned} \quad (\text{A5})$$

The vector of innovation processes, $\hat{z}(t)$, is Wiener processes with respect to the investor's information filtration \mathcal{F}_t^I . Note that $b(t)$ and $\nu(t)$ satisfy equations (A4) and (A5) subject to the conditions $b_0 = E(\beta_0 | \mathcal{F}_0^I)$ and $\nu_0 = E[(\beta_0 - b_0)(\beta_0 - b_0)' | \mathcal{F}_0^I]$.

Next we consider the case of $n = 1$ and $k = 3$, i.e., there is only one predictive variable and three sources of Brownian motion. Let's specify the diffusion process for the predictive variable as

$$ds = \kappa(\bar{s} - s)dt + \sigma_s dz_s, \quad (\text{A6})$$

rewrite the process for stock returns as

$$\frac{dP}{P} = (\bar{\mu} + \beta(s - \bar{s}))dt + \sigma_P dz_P, \quad (\text{A7})$$

and assume that the coefficient β itself is also mean-reverting,

$$d\beta = \lambda(\bar{\beta} - \beta)dt + \sigma_\beta dz_\beta. \quad (\text{A8})$$

Assume that $E(dz_P dz_s) = \rho_{sP} dt$, $E(dz_P dz_\beta) = \rho_{\beta P} dt$, and $E(dz_\beta dz_s) = \rho_{\beta s} dt$. Under these simplifying assumptions, the process satisfied by b and ν reduces to

$$\begin{aligned} db &= \lambda(\bar{b} - b)dt + [\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})] \begin{pmatrix} \sigma_P^2 & \sigma_{sP} \\ \sigma_{sP} & \sigma_s^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_P d\hat{z}_P \\ \sigma_s d\hat{z}_s \end{pmatrix}, \\ &= \lambda(\bar{b} - b)dt + \nu_1 d\hat{z}_P + \nu_2 d\hat{z}_s, \end{aligned} \quad (\text{A9})$$

$$\begin{aligned} \frac{d\nu}{dt} &= -2\lambda\nu + \sigma_\beta^2 - [\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})] \begin{pmatrix} \sigma_P^2 & \sigma_{sP} \\ \sigma_{sP} & \sigma_s^2 \end{pmatrix}^{-1} \\ &\quad [\nu(s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})]', \\ &= -2\lambda\nu + \sigma_\beta^2 - [\nu_1^2 + \nu_2^2 + 2\nu_1\nu_2\rho_{sP}], \end{aligned} \quad (\text{A10})$$

where

$$\bar{b} = \bar{\beta}, \quad (\text{A11})$$

$$\nu_1 = \frac{\nu(s - \bar{s}) + \sigma_P \sigma_\beta (\rho_{\beta P} - \rho_{\beta s} \rho_{sP})}{\sigma_P (1 - \rho_{sP}^2)}, \quad (\text{A12})$$

$$\nu_2 = \frac{-\nu(s - \bar{s}) \rho_{sP} + \sigma_P \sigma_\beta (\rho_{\beta s} - \rho_{\beta P} \rho_{sP})}{\sigma_P (1 - \rho_{sP}^2)}, \quad (\text{A13})$$

$$\begin{aligned} dz_P &= \frac{1}{\sigma_P} \left(\frac{dP}{P} - (\bar{\mu} + b(i - i^*) dt) \right), \\ &= dz_P + \frac{(s - \bar{s})(\beta - b)}{\sigma_P} dt, \end{aligned} \quad (\text{A14})$$

$$\begin{aligned} dz_s &= \frac{1}{\sigma_s} (ds - \kappa(\bar{s} - s) dt), \\ &= dz_s. \end{aligned} \quad (\text{A15})$$

Both the Brownian motions dz_P and dz_s are adapted to the investor's information set, while dz_P is not due to the unknown β .

To understand the intuition behind the updating rule, we can re-write the observation (measurement) equation as:

$$\begin{aligned} y_t &= \begin{pmatrix} \left(\frac{dP}{P}\right)_t \\ s_t \end{pmatrix} = \begin{pmatrix} \bar{\mu} dt \\ (1 - e^{-\kappa dt}) \bar{s} + e^{-\kappa dt} s_{t-dt} \end{pmatrix} + \begin{pmatrix} (s_{t-dt} - \bar{s}) dt \\ 0 \end{pmatrix} \hat{b}_t \\ &+ \begin{pmatrix} (s_{t-dt} - \bar{s})(\beta - \hat{b}_t) dt \\ 0 \end{pmatrix} + \begin{pmatrix} \sigma_P & 0 \\ 0 & \sigma_s \end{pmatrix} \begin{pmatrix} dz_P \\ dz_s \end{pmatrix}; \end{aligned} \quad (\text{A16})$$

and the transition equation for the stochastic state variable β as:

$$\begin{aligned} \beta_t &= (1 - e^{-\lambda dt}) \bar{\beta} + e^{-\lambda dt} \beta_{t-dt} + \sigma_\beta dz_\beta, \\ &\equiv \hat{b}_t + (\beta_t - \hat{b}_t) + \sigma_\beta dz_\beta. \end{aligned} \quad (\text{A17})$$

The observation y_t is normally distributed with mean \bar{y}_t given by the first two terms in (A16) and variance $\Phi = \begin{pmatrix} \sigma_P^2 & \sigma_{sP} \\ \sigma_{sP} & \sigma_s^2 \end{pmatrix}$. Note that the variance, Φ , does not reflect the estimation error given by the third term of (A16), because the estimation error has an order of dt , which is dominated by the term of innovations (dz). In a discrete time model, where dt is not infinitely

small, then the estimation error should make Φ larger than the variance-covariance matrix of the innovations.

This is the usual Kalman filtering state space form with time-varying measurement matrices. From equations (A16) and (A17), we get the covariance between β_t and y_t as $\Gamma = \nu \times (s - \bar{s}, 0) + (\sigma_{\beta P}, \sigma_{\beta s})$. y_t and β_t have conditional joint normal distribution with

$$\begin{pmatrix} y_t \\ \beta_t \end{pmatrix} \sim \left(\begin{bmatrix} \bar{y}_t \\ \hat{b}_t \end{bmatrix}, \begin{bmatrix} \Phi & \Gamma \\ \Gamma & \nu + \sigma_\beta^2 \end{bmatrix} \right). \quad (\text{A18})$$

The best estimate in terms of minimum mean squared error, b_t , of β_t given additional information provided by y_t , is then given by

$$E(\beta_t|y_t) = b_t = (1 - e^{-\lambda dt}) \bar{\beta} + e^{-\lambda dt} b_{t-dt} + \Gamma \Phi^{-1} (y_t - \bar{y}_t); \quad (\text{A19})$$

and

$$\begin{aligned} b_t - b_{t-dt} &= (1 - e^{-\lambda dt}) (\bar{\beta} - b_{t-dt}) + \Gamma \Phi^{-1} (y_t - \bar{y}_t) \\ &\approx \lambda (\bar{b} - b) dt + \Gamma \Phi^{-1} (y_t - \bar{y}_t). \end{aligned} \quad (\text{A20})$$

When the predictive variable is constant, i.e., when $\lambda = 0$ and $\sigma_\beta = 0$, the above equations can be furtherly simplified into:

$$db = \frac{\nu(s - \bar{s})}{\sigma_P(1 - \rho_{sP}^2)} (dz_P - \rho_{sP} dz_s) + \frac{\nu(s - \bar{s})^2 (\beta - b)}{\sigma_P^2 (1 - \rho_{sP}^2)} dt, \quad (\text{A21})$$

$$= \frac{\nu(s - \bar{s})}{\sigma_P(1 - \rho_{sP}^2)} (d\hat{z}_P - \rho_{sP} dz_s). \quad (\text{A22})$$

B Closed Form Solution without Learning

When $\nu = 0$ and β is constant, the investment opportunity set is completely summarized by the dynamics of s , and the investor's optimization problem is described by equation (B1):

$$0 = \frac{1}{2}\sigma_s^2\phi_{ss} + \kappa(\bar{s} - s)\phi_s + \phi_t + \left[(1 - \gamma)r + \frac{1}{2}\gamma(1 - \gamma)\sigma_P^2(x^*)^2 - \rho \right] \phi, \quad (\text{B1})$$

where x^* is given by

$$x^* = \frac{\bar{\mu} + \beta(s - \bar{s}) - r}{\gamma\sigma_P^2} + \frac{\rho_{sP}\sigma_s\sigma_P\phi_s}{\gamma\sigma_P^2\phi}. \quad (\text{B2})$$

Kim and Omberg (1996) provides a closed form solution to this equation when the investor's utility function is from the HARA family.

Define $y = (\bar{\mu} - \beta\bar{s} - r) + \beta s$ as the new state variable to replace s , and let $\tau = T - t$ be the horizon, we conjecture that

$$\phi = \exp \left\{ A(\tau) + B(\tau)y + \frac{1}{2}C(\tau)y^2 \right\}. \quad (\text{B3})$$

With the above conjecture, the optimal portfolio choice x^* is given by

$$x^* = \frac{y}{\gamma\sigma_P^2} + \frac{(B(\tau) + C(\tau)y)\beta\rho_{sP}\sigma_s}{\gamma\sigma_P}. \quad (\text{B4})$$

The expressions of $A(\tau)$, $B(\tau)$ and $C(\tau)$ can be solved recursively from the following ordinary differential equations:

$$C_\tau = a_1C^2(\tau) + a_2C(\tau) + a_3, \quad (\text{B5})$$

$$B_\tau = a_1B(\tau)C(\tau) + \frac{1}{2}a_2B(\tau) + \kappa(\bar{\mu} - r)C(\tau), \quad (\text{B6})$$

$$A_\tau = \frac{1}{2}a_1B^2(\tau) + \frac{1}{2}(\beta\sigma_s)^2C(\tau) + \kappa(\bar{\mu} - r)B(\tau) + (r(1 - \gamma) - \rho), \quad (\text{B7})$$

with boundary conditions of

$$A(\tau) = 0, \quad B(\tau) = 0, \quad \text{and} \quad C(\tau) = 0 \quad \text{at} \quad \tau = 0,$$

where the parameters are:

$$a_1 = (\beta\sigma_s)^2 \left(1 + \frac{1-\gamma}{\gamma} \rho_{sP}^2 \right), \quad (\text{B8})$$

$$a_2 = 2 \left(\frac{(1-\gamma)\beta\sigma_s\rho_{sP}}{\gamma\sigma_P} - \kappa \right), \quad (\text{B9})$$

$$a_3 = \frac{1-\gamma}{\gamma\sigma_P^2}. \quad (\text{B10})$$

Note that $a_1 = 0$ when $\beta = 0$, so that the optimization problem reduces to the case with no predictability. The investment opportunity set is then constant, and the problem has a simple closed form solution as given in Ingersoll (1986) or Merton (1990).

There are five different solutions depending on the parameter values of the model when $\beta \neq 0$. The details are given in the appendix of Kim and Omberg (1996). Because $q = a_2^2 - 4a_1a_3 > 0$ and $a_3 > 0$ for all the values of β and γ used in this study, we have a well-behaved normal solution:

$$C(\tau) = \frac{2a_3(1 - e^{-\eta\tau})}{(\eta - a_2) + (\eta + a_2)e^{-\eta\tau}}, \quad (\text{B11})$$

$$B(\tau) = \frac{4a_3\kappa(\bar{\mu} - r) \left(1 - e^{-\frac{\eta\tau}{2}} \right)^2}{\eta [(\eta - a_2) + (\eta + a_2)e^{-\eta\tau}]}, \quad (\text{B12})$$

$$\begin{aligned} A(\tau) = & \left[a_3 \left(\frac{2\kappa^2(\bar{\mu} - r)^2}{\eta^2} + \frac{(\beta\sigma_s)^2}{\eta - a_2} \right) + r(1 - \gamma) - \rho \right] \tau \\ & + \frac{4a_3\kappa^2(\bar{\mu} - r)^2 \left[(2a_2 + \eta)e^{-\eta\tau} - 4a_2e^{-\frac{\eta\tau}{2}} + 2a_2 - \eta \right]}{\eta^3 [(\eta - a_2) + (\eta + a_2)e^{-\eta\tau}]} \\ & + \frac{2a_3(\beta\sigma_s)^2}{\eta^2 - a_2^2} \ln \left| \frac{(\eta - a_2) + (\eta + a_2)e^{-\eta\tau}}{2\eta} \right|, \end{aligned} \quad (\text{B13})$$

where $\eta = \sqrt{q}$.

The horizon effect on the optimal portfolio allocation can be derived in closed form when there is no parameter uncertainty. We first derive the dependence of $B(\tau)$ and $C(\tau)$ on the horizon,

τ :

$$\frac{\partial C}{\partial \tau} = \frac{4a_3\eta^2 e^{-\eta\tau}}{[\eta - a_2 + (\eta + a_2)e^{-\eta\tau}]^2}, \quad (\text{B14})$$

$$\frac{\partial B}{\partial \tau} = \frac{4a_3\kappa(\bar{\mu} - r) \left(1 - e^{-\frac{\eta\tau}{2}}\right) e^{-\frac{\eta\tau}{2}} \left(\eta - a_2 + (\eta + a_2)e^{-\frac{\eta\tau}{2}}\right)}{[\eta - a_2 + (\eta + a_2)e^{-\eta\tau}]^2}. \quad (\text{B15})$$

Note that we only consider the situation when the investor is more risk averse than the log utility, i.e., $\gamma > 1$. This implies that

$$a_1 > (\beta\sigma_s)^2 (1 - \rho_{sP}^2) \geq 0, \quad (\text{B16})$$

$$a_3 < 0, \quad (\text{B17})$$

$$\eta \geq |a_2|, \quad (\text{B18})$$

and

$$\left(\eta - a_2 + (\eta + a_2)e^{-\frac{\eta\tau}{2}}\right) > 0. \quad (\text{B19})$$

Therefore,

$$\frac{\partial C}{\partial \tau} < 0, \quad (\text{B20})$$

$$\frac{\partial B}{\partial \tau} < 0. \quad (\text{B21})$$

The horizon effect on the optimal allocation can be obtained from

$$\frac{\partial x^*}{\partial \tau} = \frac{\beta\rho_{sP}\sigma_s}{\gamma\sigma_P^2} \left(\frac{\partial B}{\partial \tau} + \frac{\partial C}{\partial \tau} y \right) \quad (\text{B22})$$

In our example, $\rho_{sP} = -0.93 < 0$, so a *sufficient* condition for $\frac{\partial x^*}{\partial \tau} > 0$ is that $y > 0$ or $\beta(s - \bar{s}) + (\bar{\mu} - r) > 0$. This is satisfied for most reasonable parameter values of β and $(s - \bar{s})$. Only when $s < \bar{s}$ and β is sufficiently large then $y < 0$. When $\rho_{sP} < 0$ and for a given value of y , the horizon effect is summarized by

$$\frac{\partial x^*}{\partial \tau} \begin{cases} > 0 & \text{if } \zeta(\tau) < y \\ \leq 0 & \text{if } \zeta(\tau) \geq y \end{cases}, \quad (\text{B23})$$

where

$$\zeta(\tau) = -\frac{\frac{\partial B}{\partial \tau}}{\frac{\partial C}{\partial \tau}} = \frac{\kappa(\bar{\mu} - r) \left(1 - e^{-\frac{\eta\tau}{2}}\right) \left(\eta - a_2 + (\eta + a_2)e^{-\frac{\eta\tau}{2}}\right)}{\eta^2}. \quad (\text{B24})$$

Let the cutoff horizon be τ^* such as $\zeta(\tau^*) = y$, then

$$\tau^* = \frac{2}{\eta} \ln \left(\frac{- \left[y + \frac{2a_2\kappa}{\eta^2} (\bar{\mu} - r) \right] + \sqrt{\Delta}}{\frac{2\kappa}{\eta^2} (\bar{\mu} - r) (\eta - a_2)} \right), \quad (\text{B25})$$

where

$$\Delta = y^2 + 4y \frac{a_2\kappa(\bar{\mu} - r)}{\eta^2} + 4 \frac{\kappa^2(\bar{\mu} - r)^2}{\eta^2}. \quad (\text{B26})$$

Equation (B23) can be simplified as:

$$\frac{\partial x^*}{\partial \tau} \begin{cases} > 0 & \text{if } \tau > \max(0, \tau^*) \\ \leq 0 & \text{if } 0 \leq \tau \leq \max(0, \tau^*) \end{cases}, \quad (\text{B27})$$

Therefore, if $\tau^* > 0$, then the optimal portfolio allocation is decreasing with the horizon when $\tau < \tau^*$ and the optimal portfolio allocation is increasing with the horizon when $\tau > \tau^*$. If $\tau^* \leq 0$, then the optimal portfolio allocation is always increasing with the horizon.

C Moments of the Stock Return With Certain Predictability

The intuition behind the horizon effect comparisons when there is no learning vs. when there is learning can be seen mathematically. First, let's consider the case with no learning with a known constant β . We obtain a closed form solution for the mean and variance of the stock returns over a horizon T .

$$\begin{aligned}
 \ln P_T - \ln P_0 &= \int_0^T \left(\bar{\mu} - \beta \bar{s} - \frac{1}{2} \sigma_P^2 \right) d\tau + \beta \int_0^T s_\tau d\tau + \sigma_P \int_0^T dz_P(\tau) \\
 &= \left(\bar{\mu} - \frac{1}{2} \sigma_P^2 \right) T + \frac{\beta(s_0 - \bar{s})}{\kappa} (1 - e^{-\kappa T}) \\
 &\quad + \sigma_P \int_0^T dz_P(\tau) + \frac{\beta \sigma_s}{\kappa} \int_0^T (1 - e^{\kappa(\tau-T)}) dz_s(\tau). \tag{C1}
 \end{aligned}$$

From equation (C1), we have

$$E = E[\ln(P_T/P_t) | \mathcal{F}_t^I] = \left(\bar{\mu} - \frac{1}{2} \sigma_P^2 \right) (T-t) + \frac{\beta(s_t - \bar{s})}{\kappa} (1 - e^{-\kappa(T-t)}), \tag{C2}$$

$$\begin{aligned}
 V &= Var[\ln(P_T/P_t) | \mathcal{F}_t^I] = \left(\sigma_P^2 + \frac{\beta^2 \sigma_s^2}{\kappa^2} + \frac{2\beta \rho_{sP} \sigma_s \sigma_P}{\kappa} \right) (T-t) \\
 &\quad - 2 \left(\frac{\beta^2 \sigma_s^2}{\kappa^3} + \frac{\beta \rho_{sP} \sigma_s \sigma_P}{\kappa^2} \right) (1 - e^{-\kappa(T-t)}) + \frac{\beta^2 \sigma_s^2}{2\kappa^3} (1 - e^{-2\kappa(T-t)}). \tag{C3}
 \end{aligned}$$

When the stock return is predictable, the expected cumulative return for an interval of $T-t$ depends both on its long run mean and the current observation of the predictive variable. The variance of the return is affected by its own volatility, the volatility of the predictive variable, the correlation between the two series and the predictive coefficient. When $\beta = 0$ or $\sigma_s = 0$, the mean and the variance grow proportionally with the horizon. The mean-variance ratio remains the same regardless of the investment horizon, so the investor would choose to invest the same proportion of his wealth in the stock.

When neither $\beta = 0$ nor $\sigma_s = 0$, the expected return is still increasing with the investment horizon if $s_t \geq \bar{s}$ or $T-t$ is large enough, since

$$\frac{\partial E}{\partial(T-t)} = \left(\bar{\mu} - \frac{1}{2} \sigma_P^2 \right) + \beta(s_t - \bar{s})e^{-\kappa(T-t)} > 0, \tag{C4}$$

when the second term is close to zero or positive. The variance of the cumulative return also

increases with the investment horizon:

$$\frac{\partial V}{\partial(T-t)} = \left[\left(\sigma_P \rho_{sP} + \frac{\beta \sigma_s}{\kappa} \right) - \frac{\beta \sigma_s}{\kappa} e^{-\kappa(T-t)} \right]^2 + \sigma_P^2 (1 - \rho_{sP}^2) > 0. \quad (\text{C5})$$

When $T-t$ is long so that the terms associated with $e^{-\kappa(T-t)}$ become less important or when we consider the relative magnitudes between the mean and the variance of stock return, we generally have $\frac{\partial(E-V)}{\partial(T-t)} > 0$, which implies that the mean increases with the horizon faster than the variance, so a buy-and-hold investor will generally hold more wealth in stock if the horizon is longer.

For an investor who rebalances dynamically, we consider the covariance between returns over periods $(t_1 - t_2)$ and $(t_2 - t)$ where $t_1 > t_2 > t$,

$$\begin{aligned} C &= Cov \left(\ln(P_{t_1}/P_{t_2}), \ln(P_{t_2}/P_t) \mid \mathcal{F}_t^I \right) \\ &= \frac{\beta^2 \sigma_s^2}{2\kappa^3} \left(1 - e^{-\kappa(t_1-t_2)} \right) \left(1 - e^{-\kappa(t_2-t)} \right)^2 \\ &+ \frac{\beta \rho_{sP} \sigma_s \sigma_P}{\kappa^2} \left(1 - e^{-\kappa(t_1-t_2)} \right) \left(1 - e^{-\kappa(t_2-t)} \right). \end{aligned} \quad (\text{C6})$$

In the special case where $t_1 - t_2 = 1$ and $t_2 - t = 1$, then the first order auto covariance of stock returns is given by

$$\begin{aligned} C_1 &= \frac{(1 - e^{-\kappa})^2}{\kappa^2} \left[\frac{\beta^2 \sigma_s^2}{2\kappa} (1 - e^{-\kappa}) + \beta \rho_{sP} \sigma_s \sigma_P \right] \\ &\begin{cases} \geq 0 & : \quad \rho_{sP} \geq -\frac{\beta \sigma_s (1 - e^{-\kappa})}{2\kappa \sigma_P} \quad \text{and} \quad \beta > 0 \\ \geq 0 & : \quad \rho_{sP} \leq -\frac{\beta \sigma_s (1 - e^{-\kappa})}{2\kappa \sigma_P} \quad \text{and} \quad \beta < 0 \\ = 0 & : \quad \beta = 0 \quad \text{or} \quad \sigma_s = 0 \\ < 0 & : \quad \text{Otherwise} \end{cases} \end{aligned} \quad (\text{C7})$$

This shows that the stock return is iid when there is no predictability or the predictive variable is non-stochastic. Return predictability introduces serial correlation into the return series except for the special cases where $\rho_{sP} = -\frac{\beta \sigma_s (1 - e^{-\kappa})}{2\kappa \sigma_P}$. In our study, the dividend yield serves as the predictive variable. The large negative correlation between the stock return and the dividend yield, $\rho_{sP} = -0.93$, introduces a negative first order autocorrelation in the stock return. This mean-reverting pattern in the return calls for a positive hedge demand as shown in Appendix B, so a long horizon investor holds more wealth in stock.

References

- Barberis, N., 1998, Investing for the Long Run when Returns are Predictable, *Journal of Finance*, forthcoming.
- Bawa, V., S. J. Brown and R. W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North-Holland Publishing Company.
- Brandt, M. W., 1998, Estimating Portfolio and Consumption Choice: A Conditional Euler Equation Approach, *Journal of Finance*, forthcoming.
- Brennan, M. J., 1998, The Role of Learning in Dynamic Portfolio Decisions, *European Economic Review* 1, 295-306.
- Brennan, M. J., E. S. Schwartz and R. Lagnado, 1997, Strategic Asset Allocation, *Journal of Economic Dynamics and Control* 21, 1377-1403.
- Brennan, M.J. and Y. Xia, 1998a, Stock Price Volatility, Learning, and the Equity Premium, *Working Paper, Anderson School of Management, UCLA*.
- Brennan, M.J. and Y. Xia, 1998b, Resolution of an Asset Allocation Puzzle, *Working Paper, Anderson School of Management, UCLA*.
- Brennan, M.J. and Y. Xia, 1999, Assessing Asset Pricing Anomalies, *Working Paper, Anderson School of Management, UCLA*.
- Bossaerts, P. and P. Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: what do we learn?, *Review of Financial Studies* 12, 405-428.
- Campbell J. Y. and L. Viceira, 1998, Consumption and Portfolio Decisions When Expected Returns are Time Varying, *Quarterly Journal of Economics*, forthcoming.
- Cochrane, J.H., 1999a, Portfolio Advice for a Multifactor World, *Working Paper, University of Chicago*.
- Cochrane, J.H., 1999b, New Facts in Finance, *Working Paper, University of Chicago*.
- Detemple, Jerome, 1986, Asset Pricing in a Production Economy with Incomplete Information, *Journal of Finance* 41, 383-391.
- Dothan, M. U. and D. Feldman, 1986, Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy, *Journal of Finance* 41, 369-382.
- Fama, E.F., 1981, Stock Returns, Real Activity, Inflation and Money, *American Economic Review* 71, 545-565.
- Fama, E.F., 1991, Efficient Capital Markets: II, *Journal of Finance* 46, 1575-1615.
- Fama, E.F. and K.R. French, 1988a, Dividend Yields and Expected Stock Returns, *Journal of Financial Economics* 23, 3-25.
- Fama, E.F. and K.R. French, 1988b, Permanent and Temporary Components of Stock Prices,

Journal of Political Economy 96, 246-273.

Fama, E.F. and G.W. Schwert, 1977, Asset Returns and Inflation, *Journal of Financial Economics* 5, 115-146.

Feldman, D., 1992, Logarithmic Preferences, Myopic Decisions, and Incomplete Information, *Journal of Financial and Quantitative Analysis* 27, 619-629.

Gennotte, G., 1986, Optimal Portfolio Choice under Incomplete Information, *Journal of Finance* 41, 733-746.

Geske, R. and R. Roll, 1983, The Fiscal and Monetary Linkage between Stock Returns and Inflation, *Journal of Finance* 38, 1-33.

Goetzmann, W.N. and P. Jorion, 1993, Testing the Predictive Power of Dividend Yields, *Journal of Finance* 48, 663-679.

Hodrick, R., 1992, Dividend Yields and Expected Stock Returns, *Review of Financial Studies* 5, 357-386.

Ingersoll, J.E., Jr., 1987, *Theory of Financial Decision Making*, Rowman and Littlefield Publishers, Inc.

Kandel, S. and R. F. Stambaugh, 1991, Asset Returns and Intertemporal Preferences, *Journal of Monetary Economics* 27, 39-71.

Kandel, S. and R. F. Stambaugh, 1996, On the Predictability of Stock Returns: An Asset-Allocation Perspective, *Journal of Finance* 51, 385-424.

Kaul, G., 1987, Stock Returns and Inflation: The role of the Monetary Sector, *Journal of Financial Economics* 18, 253-276.

Kaul, G., 1991, Monetary Regimes and the Relation between Stock Returns and Inflationary Expectations, *Journal of Financial and Quantitative Analysis* 25, 307-321.

Keim, D.B. and R. F. Stambaugh, 1986, Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics* 17, 357-390.

Kim, T. S. and E. Omberg, 1996, Dynamic Nonmyopic Portfolio Behavior, *Review of Financial Studies* 9, 141-161.

Kothari, S.P. and J. Shanken, 1997, Book-to-market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis, *Journal of Financial Economics* 44, 69-203.

Lintner, J., 1975, Inflation and Security Returns, *Journal of Finance* 30, 259-280.

Liptser, R.S. and A.N. Shiriyayev, 1978, *Statistics of Random Processes II: Applications*, Springer-Verlag.

Liu, Jun, 1998, Portfolio Selection in Stochastic Environments, *dissertation, Stanford University*.

Lo, A.W. and A.C. MacKinlay, 1988, Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies* 1, 41-66.

- Lynch, A. and P. Balduzzi, 1998, Predictability and Transaction Costs: The Impact on Rebalancing Rules and Behavior, *Working Paper, Stern School of Business, New York University*.
- Merton, R.C., 1990, *Continuous Time Finance*, Blackwell Publishers Ltd.
- Stulz, R., 1986, Interest Rates and Monetary Policy Uncertainty, *Journal of Monetary Economics* 17, 331-347.
- Stulz, R., 1987, An Equilibrium Model of Exchange Rate Determination and Asset Pricing with Nontraded Goods and Imperfect Information, *Journal of Political Economy* 95, 1024-1040.
- Stambaugh, R.F., 1999, Predictive Regression, *Working Paper, The Wharton School, University of Pennsylvania*.
- Summers, L.H., 1986, Does the Stock Market Rationally Reflect Fundamental Values?, *Journal of Finance* 41, 591-601.
- Veronesi, P., 1999, How does Information Quality Affect Stock Returns?, *Journal of Finance*, forthcoming.
- Williams, J.T., 1977, Capital Asset Prices with Heterogeneous Beliefs, *Journal of Financial Economics* 5, 219-241.

Footnotes

¹Summers (1986) and Fama & French (1988b), for example, argue that the log stock index price may be described by the sum of a random walk and a stationary mean-reverting component. Shiller and Perron (1985) and Poterba and Summers (1987) also propose similar models but motivate their alternative hypotheses as models of investors' fads. However, Lo and MacKinlay (1988) reject both the random walk hypothesis and the mean-reverting alternative using U.S. data.

²Refer to Fama (1991) for a summary of these empirical studies on return predictability. For example, Fama (1981) finds that stock returns are negatively related to expected inflation and the level of short-term interest rate. Keim and Stambaugh (1986) find that several predetermined variables that reflect levels of bond and stock prices appear to predict returns on common stocks of firms of various sizes, long-term bonds of various default risk, and default-free bonds of various maturities. Fama and French (1988a) report that past returns over a long horizon can predict as much as 40% of future returns. Kothari and Shanken (1997) also find that the book-to-market ratio (B/M) has predictive power. While most studies that use daily or weekly data find very low predictability as measured by statistics such as R^2 or p -values, the evidence of predictability using long-horizon returns is much more striking.

³Hodrick (1992) and Goetzmann and Jorion (1993), for example, argue that many findings based on long horizon return regressions may be spurious due to the poor small sample properties of commonly used inference methods.

⁴Many authors have discussed the importance of estimation risk and/or learning. Bawa, Brown and Klein (1979) is a collection of papers about estimation risk and optimal portfolio choice in a one period setting. Stulz (1986, 1987) studies the effect of learning about the monetary policy on interest and exchange rates. Detemple (1986) discusses about estimation risk in a production economy, while Dothan and Feldman (1986) and Feldman (1992) discuss the term structure and interest rate dynamics with Learning. Gennotte (1986) is a good example of the study of signal filtration and learning in a dynamic portfolio choice setting.

⁵An investor with logarithmic utility optimally ignores stochastic variation in the future investment opportunity set.

⁶Brandt (1999) finds that the conventional advice is correct in that the long horizon investor holds more stock. Barberis (1999) found that the static estimation risk reduces the horizon effect, but the stock allocation still increases with horizon, although to a less extent.

⁷This is consistent with the finding of Campbell and Viceira (1999) where no parameter uncertainty is considered.

⁸We only consider investors with a risk aversion parameter greater than the logarithmic case in this paper.

⁹The assumption of a stochastic process for β can be motivated by Bossaerts and Hillion

(1999), who find that predictive relation is not stationary and predictive parameters can vary through time.

¹⁰Stulz (1986, 1987) shows in the context of monetary policy and interest rates that regime shifts have important implications for learning. Veronesi (1999) analyzes a model in which the growth rate of the dividend switches among several discrete states.

¹¹The separation theorem is not affected by the predictability assumed here, because the predictive relation equation (3) only implies that the coefficients in the transition and possibly in the measurement (observation) equation are stochastic and time-varying, so that a non-linear filtering is called for.

¹²We do not model why the investor has such a normal prior and how the prior mean and variance are derived. In the numerical study, we assume that the investor's prior mean and variance are derived from regression studies using historical data.

¹³The assumption that the signal follows an O-U process as in equation (9) is consistent with Barberis (1999) and Stambaugh (1999) among others, who generally assume that the predictor such as dividend yield follows a AR(1) process in discrete time.

¹⁴The partial differential equation is given by

$$\begin{aligned}
0 &= \frac{1}{2}\sigma_s^2\phi_{ss} + \frac{1}{2}(\nu_1^2 + \nu_2^2 + 2\nu_1\nu_2\rho_{sP})\phi_{bb} + \sigma_s(\nu_1\rho_{sP} + \nu_2)\phi_{sP} + \phi_t \\
&+ [\kappa(\bar{s} - s) + (1 - \gamma)\rho_{sP}\sigma_s\sigma_P x^*]\phi_s + [\lambda(\bar{b} - b) + (1 - \gamma)\sigma_P(\nu_1 + \nu_2\rho_{sP})x^*]\phi_b \\
&+ [-2\lambda\nu + \sigma_\beta^2 - (\nu_1^2 + \nu_2^2 + 2\nu_1\nu_2\rho_{sP})]\phi_\nu + \gamma\phi^{\frac{\gamma-1}{\gamma}} \\
&+ [(1 - \gamma)((\bar{\mu} + b(s - \bar{s}) - r)x^* + r) - \frac{1}{2}\gamma(1 - \gamma)\sigma_P^2(x^*)^2 - \rho]\phi,
\end{aligned}$$

where x^* is the optimal portfolio strategy and is given by (22). The PDE will be solved numerically with a boundary condition:

$$\phi(b, \nu, s, T) = 1. \quad (C9)$$

¹⁵See Kandel and Stambaugh (1996) for detailed formulas and discussions.

¹⁶Barberis assumes β to be constant, so that the third term in the bracket is zero.

¹⁷Most powerful predictive variables include past market returns, the dividend yield, and the earnings-price ratio, nominal interest rates and expected inflation.

¹⁸When β follows an O-U process, the model is estimated using Kalman filtering algorithm. The estimates for λ and σ_β are reported in Section IIIC.

¹⁹Inflation rate is calculated from the CPI data from Datastream.

²⁰This is comparable with the yield on US indexed bonds of about 4.0% as of July 1999.

²¹We use annualized stock return and annual dividend yield in the VAR regression. Our estimated coefficients and standard errors are thus exactly 12 times as large as the results from regressions using monthly returns.

²²A linear boundary condition is also tried and the results are the same. The process s has a conditional Gaussian distribution, with

$$s_t|s_0 \sim N\left(e^{-\kappa t}s_0 + (1 - e^{-\kappa t})\bar{s}, \frac{\sigma_s^2}{2\kappa}(1 - e^{-2\kappa t})\right).$$

In the numerical study, we set $\bar{s} = 4\%$, $\kappa = 0.19$ and $\sigma_s = 0.6\%$, which leads to a mean of 4% and a standard deviation of 0.009% for $s_t|s_0$ when t is set to 20 and s_0 is assumed to be at \bar{s} . Therefore, the probability that s reaches the boundary in twenty years is almost zero. The conditional distribution of b_t given ν and s is also Gaussian as indicated by equation (14). The probability that b reaches the boundary of $b = 10.0$ or $b = -10.0$ in twenty years depends on the conditional distribution of b_t . Since the boundary is 20 standard deviations away from zero, the probability of b reaching the boundary in 20 years is also small under the null hypothesis of no predictability.

²³Barberis (1998) finds that the allocation could decrease with horizon when the investor uses a buy-and-hold strategy and takes estimation risk into consideration. However, he does not discuss the state dependence of horizon effect. Brennan (1997) finds that the horizon effect is negative when the true stock return is i.i.d. and there is learning.