Investing for the Long Run when Returns are Predictable

Nicholas Barberis

First version January 1996
This version January 1999 *

Abstract

We examine how the evidence of predictability in asset returns affects optimal portfolio choice for investors with long horizons. Particular attention is paid to estimation risk, or uncertainty about the true values of model parameters. We find that even after incorporating parameter uncertainty, there is enough predictability in returns to make investors allocate substantially more to stocks, the longer their horizon. Moreover, the weak statistical significance of the evidence for predictability makes it important to take estimation risk into account; a long-horizon investor who ignores it may over-allocate to stocks by a sizeable amount.

*Graduate School of Business, University of Chicago. I am indebted to John Campbell and Gary Chamberlain for guidance and encouragement. I also thank an anonymous referee, the editor René Stulz, and seminar participants at Harvard, the Wharton School, Chicago Business School, the Sloan School at MIT, UCLA, Rochester, NYU, Columbia, Stanford, INSEAD, HEC, the LSE, and the London Business School for their helpful comments.

One of the more striking empirical findings in recent financial research is the evidence of predictability in asset returns.[1] In this paper we examine the implications of this predictability for an investor seeking to make sensible portfolio allocation decisions.

We approach this question from the perspective of horizon effects: given the evidence of predictability in returns, should a long horizon investor allocate his wealth differently from a short horizon investor? The motivation for thinking about the problem in these terms is the classic work of Samuelson (1969) and Merton (1969). They show that if asset returns are i.i.d., an investor with power utility who rebalances his portfolio optimally should choose the *same* asset allocation, regardless of investment horizon.

In light of the growing body of evidence that returns are predictable, the investor's horizon may no longer be irrelevant. The extent to which the horizon *does* play a role serves as an interesting and convenient way of thinking about how predictability affects portfolio choice. Moreover, the results may shed light on the common but controversial advice that investors with long horizons should allocate more heavily to stocks.[2]

On the theoretical side, it has been known since Merton (1973) that variation in expected returns over time can potentially introduce horizon effects. The contribution of this paper is therefore primarily an empirical one: given actual historical data on asset returns and predictor variables, we try to understand the magnitude of these effects by computing optimal asset allocations for both static buy-and-hold and dynamic optimal rebalancing strategies.

An important aspect of our analysis is that in constructing optimal portfolios, we account for the fact that the true extent of predictability in returns is highly uncertain. This is of particular concern in this context because the evidence of time variation in expected returns is sometimes weak. A typical example is the following.[3] Let r_t be the continuously compounded real return on the value-weighted portfolio of the New York Stock Exchange in month t, and $(\frac{d}{p})_t$ be the portfolio's dividend-price ratio, or dividend yield, defined as the sum of dividends paid in months $t-11$ through t divided by the value of the portfolio at the end of month t. An OLS regression of the returns on the lagged dividend yield, using monthly returns from January 1927

[1] See for example Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,b), Fama and French (1988, 1989), and Campbell (1991).

[2] See Siegel (1994), Samuelson (1994) and Bodie (1995) for some recent discussions of this debate.

[3] Kandel and Stambaugh (1996) use a similar example to motivate their related work on portfolio choice.

to December 1995 gives

$$r_t = \underset{(0.0064)}{-0.0056} + \underset{(0.1428)}{0.2580} \left(\tfrac{d}{p}\right)_{t-1} + \epsilon_t, \qquad (1)$$

where standard errors are in parentheses and the R^2 is 0.0039.

The coefficient on the dividend yield is not quite significant, and the R^2 is very low. Some investors might react to the weakness of this evidence by discarding the notion that returns are predictable; others might instead ignore the substantial uncertainty regarding the true predictive power of the dividend yield and analyze the portfolio problem assuming the parameters are known precisely. We argue here that both these views, while understandable, are flawed. The approach in this paper constitutes what we believe is an appropriate middle ground: the uncertainty about the parameters, also known as *estimation risk,* is accounted for explicitly when constructing optimal portfolios.

We analyze portfolio choice in discrete time for an investor with power utility over terminal wealth. There are two assets: Treasury Bills and a stock index. The investor uses a VAR model to forecast returns, where the state vector in the VAR can include asset returns and predictor variables. This is a convenient framework for examining how predictability affects portfolio choice: by changing the number of predictor variables in the state vector, we can compare the optimal allocation of an investor who takes return predictability into account to that of an investor who is blind to it.

How is parameter uncertainty incorporated? It is natural to take a Bayesian approach here. The uncertainty about the VAR parameters is summarized by the posterior distribution of the parameters given the data. Rather than constructing the distribution of future returns conditional on fixed parameter estimates, we integrate over the uncertainty in the parameters captured by the posterior distribution. This allows us to construct what is known in Bayesian analysis as the *predictive* distribution for future returns, conditional only on observed data, and not on any fixed parameter values. By comparing the solution in the cases where we condition on fixed parameters, and where we integrate over the posterior, we see the effect of parameter uncertainty on the portfolio allocation problem.

Our first set of results relate to the case where parameter uncertainty is ignored: that is, the investor allocates his portfolio taking the parameters as fixed at their estimated values in equation (1). We analyze two distinct portfolio problems: a static buy-and-hold problem and a dynamic problem with optimal rebalancing.

In the buy-and-hold case, we find that predictability in asset returns leads to strong horizon effects: an investor with a horizon of 10 years allocates significantly more to stocks than someone with a one year horizon. The reason is that time-variation in expected returns such as that in equation (1) induces mean-reversion in returns, lowering the variance of cumulative returns over long horizons. This makes stocks appear less risky to long horizon investors and leads them to allocate more to equities than would investors with shorter horizons.

We also find strong horizon effects when we solve the dynamic problem faced by an investor who rebalances optimally at regular intervals. However, the results here are of a different nature. Investors again hold substantially more in equities at longer horizons, but only when they are more risk-averse than log utility investors. The extra stock holdings of long horizon investors are "hedging demands" in the sense of Merton (1973). Under specification (1), the available investment opportunities change over time as the dividend yield changes: when the yield falls, expected returns fall. Merton shows that investors may want to hedge these changes in the opportunity set. In our data, we find that shocks to expected stock returns are negatively correlated with shocks to realized stock returns. Therefore, when investors choose to hedge, they do so by increasing their holdings of stocks.

As argued earlier, it may be important that the investor take into account uncertainty about model parameters such as the coefficient on the predictor variable in equation (1), or the regression intercept. The standard errors in equation (1) indicate that the true forecasting ability of the dividend yield may be much weaker than that implied by the raw parameter estimate. The investor's portfolio decisions can be improved by adopting a framework which recognizes this.

We find that in both the static buy-and-hold and the dynamic rebalancing problem, incorporating parameter uncertainty changes the optimal allocation significantly. In general, horizon effects are still present, but less prominent: a long horizon investor still allocates more to equities, but the magnitude of the effect is smaller than would be suggested by an analysis using fixed parameter values. In some situations, we find that uncertainty about parameters can be large enough to reverse the direction of the results. Instead of allocating more to stocks at long horizons, investors may actually allocate *less* once they incorporate parameter uncertainty properly.

While parameter uncertainty has similar implications for both buy-and-hold and rebalancing investors, the mechanism through which it operates differs in the two cases. Incorporating uncertainty about the regression intercept and about the coefficient on the predictor variable increases the variance of the distribution for cumulative returns, particularly at longer horizons. This makes stocks look riskier to a long term buy-and-hold investor, reducing their attractiveness. In the case of dynamic rebalancing, the investor needs to recognize that he will learn more about the uncertain parameters over time; we find that the possibility of learning can also reduce the stock allocation of a long term investor, possibly to a level below that of a short horizon investor. The lower allocation to stocks serves as a hedge against changes in perceived investment opportunities as the investor updates his beliefs about the parameters.

Parameter uncertainty also affects the sensitivity of the optimal allocation to the predictor variable. When it is ignored, the optimal allocation to stocks is very sensitive to the value of the dividend yield: if the yield falls, predicting low stock returns, the investor lowers his allocation to stocks sharply. Behavior of this kind makes for a highly variable allocation to stocks over time. When we acknowledge that the parameters are uncertain, the allocation becomes less sensitive to changes in the dividend yield, leading to more gradual shifts in portfolio composition over time.

There is surprisingly little empirical work on portfolio choice in the presence of time-varying expected returns. To our knowledge, Brennan, Schwartz, and Lagnado (1997) is the first attempt on this problem. Working in continuous time, they analyze the dynamic programming problem faced by an investor who rebalances optimally, for a small number of assets and predictor variables. Their approach is to solve the partial differential equation derived originally by Merton (1973) for this problem. Motivated by their results, Campbell and Viceira (1997) are able to find an analytical approximation to the more general problem of deriving both optimal consumption and portfolio rules for an infinite horizon investor with Epstein-Zin utility. Kim and Omberg (1996) make a related theoretical contribution by deriving exact analytical formulae for optimal portfolio strategies when investors have power utility and expected returns are governed by a single mean-reverting state variable. All these papers ignore estimation risk.

The issue of parameter uncertainty was first investigated by Bawa, Brown, and Klein (1979) in the context of i.i.d returns.[4] Kandel and Stambaugh (1996) were the first to point out the importance of recognizing parameter uncertainty in the context of portfolio allocation with predictable returns. Using a Bayesian framework similar in spirit to ours, they show that for a short horizon investor, the optimal allocation can be sensitive to the current values of predictor variables such as the dividend yield, even though regression evidence for such predictability may be weak. By examining a wider range of horizons, both short and long, rather than the one month horizon of Kandel and Stambaugh (1996), we hope to uncover a broader set of phenomena and a more substantial role played by parameter uncertainty.

[4]Other papers in this vein include Jobson and Korkie (1980), Jorion (1985), and Frost and Savarino (1986).

Section I introduces the framework we use for incorporating predictability and parameter uncertainty into the optimal portfolio problem. Sections II and III construct distributions for long horizon returns, and use them to solve a buy-and-hold investor's portfolio problem. The focus is on how these distributions and the resulting optimal allocation are affected by predictability and parameter uncertainty. Rather than introduce both effects at once, we bring them in one at a time: Section II considers parameter uncertainty in the context of an i.i.d. model; Section III then allows for predictability. Section IV turns to the dynamic problem of optimal rebalancing and contrasts the results with those in the buy-and-hold case. We analyze estimation risk in a dynamic context, including the possibility of learning more about the parameters over time. Section V concludes.

I. A Framework for Asset Allocation

This section presents a framework for investigating how predictability in asset returns and uncertainty about model parameters affect portfolio choice. The framework is based on that of Kandel and Stambaugh (1996), who in turn draw on models originally proposed by Zellner and Chetty (1965). Since much of our analysis focuses on investors with long horizons, it is important to be precise about the choices these investors are allowed to make. We distinguish between three different ways of formulating the portfolio problem.

One possibility is a buy-and-hold strategy. In this case, an investor with a 10 year horizon chooses an allocation at the beginning of the first year, and does not touch his portfolio again until the 10 years are over.

The second strategy, we call myopic rebalancing. In this case, the investor chooses some arbitrary rebalancing interval, say one year for the 10 year investor. He then chooses an allocation at the beginning of the first year, knowing that he will reset his portfolio to that *same* allocation at the start of every year. This is myopic in that the investor does not use any of the new information he has once a year has passed.

The final, most sophisticated strategy is optimal rebalancing. Assume again that the rebalancing interval is one year. In this case, the investor chooses his allocation today, knowing that at the start of every year, he will reoptimize his portfolio using the new information at each time.

This paper presents results for both the buy-and-hold and the optimal rebalancing cases. The results for myopic rebalancing are too similar to those for the buy-and-hold strategy to justify reporting them separately. In the next few paragraphs, we describe the asset allocation framework from the perspective of a buy-and-hold investor. We postpone a detailed discussion of the dynamic rebalancing problem until Section IV; most of the issues described below, however, remain highly relevant for that case as well.

A. Asset Allocation Framework for a Buy-and-Hold Investor

Suppose we are at time T and want to write down the portfolio problem for a buy-and-hold investor with a horizon of $\hat{T}$ months. There are two assets: Treasury Bills and a stock index. For simplicity, we suppose that the continuously compounded real monthly return on Treasury Bills is a constant r_f. We model excess returns on the stock index using a VAR framework similar to that in Kandel and Stambaugh (1987), Campbell (1991), and Hodrick (1992). It takes the form

$$z_t = a + Bx_{t-1} + \epsilon_t, \qquad (2)$$

with $z_t^{'} = (r_t, x_t^{'})$, $x_t = (x_{1,t}, \ldots, x_{n,t})'$ and $\epsilon_t \sim$ i.i.d. $N(0, \Sigma)$. The first component of z_t, namely r_t, is the continuously compounded excess stock return over month t.[5] The remaining components of z_t, which together make up the vector of explanatory variables x_t, consist of variables useful for predicting returns, such as the dividend yield. This VAR framework therefore neatly summarizes the dynamics we are trying to model. The first equation in the system specifies expected stock returns as a function of the predictor variables. The other equations specify the stochastic evolution of the predictor variables.

If initial wealth $W_T = 1$ and ω is the allocation to the stock index, then end of horizon wealth is given by

$$W_{T+\hat{T}} = (1-\omega)\exp(r_f\widehat{T}) + \omega\exp(r_f\widehat{T} + r_{T+1} + \ldots + r_{T+\hat{T}}). \qquad (3)$$

The investor's preferences over terminal wealth are described by constant relative risk-aversion power utility functions of the form

$$v(W) = \frac{W^{1-A}}{1-A}. \qquad (4)$$

Writing the cumulative excess stock return over $\widehat{T}$ periods as

$$R_{T+\hat{T}} = r_{T+1} + r_{T+2} + \ldots + r_{T+\hat{T}}, \qquad (5)$$

[5] A portfolio's "excess" return is defined as the rate of return on the portfolio minus the Treasury Bill rate, where both returns are continuously compounded.

the buy-and-hold investor's problem is to solve

$$\max_{\omega} E_T \left(\frac{\{(1-\omega)\exp(r_f \widehat{T}) + \omega \exp(r_f \widehat{T} + R_{T+\hat{T}})\}^{1-A}}{1-A} \right). \tag{6}$$

E_T denotes the fact that the investor calculates the expectation conditional on his information set at time T. At the heart of this paper is the issue of which distribution the investor should use in calculating this expectation. The distribution may be very different, depending on whether the investor accounts for parameter uncertainty or recognizes the predictability in returns.

To see whether predictability in returns has any effect on portfolio choice, our strategy is to compare the allocation of an investor who recognizes the predictability to that of an investor who is blind to it. The VAR model provides a way of simulating investors with different information sets: we simply alter the number of predictor variables included in the vector x_t.

Once the predictors have been specified, a standard procedure is to estimate the VAR parameters $\theta = (a, B, \Sigma)$, and then iterate the model forward with the parameters fixed at their estimated values. This generates a distribution for future stock returns conditional on a set of parameter values, which we write as $p(R_{T+\hat{T}}|z, \hat{\theta})$, where $z = (z_1, \ldots, z_T)'$ is the data observed by the investor up until the start of his investment horizon. The investor then solves

$$\max_{\omega} \int v(W_{T+\hat{T}})\ p(R_{T+\hat{T}}|z, \hat{\theta})\ dR_{T+\hat{T}}. \tag{7}$$

The problem with this approach is that it ignores the fact that θ is not known precisely. There may be substantial uncertainty about the regression coefficients a and B. For a long horizon investor in particular, it is important to take the uncertainty in the estimation - estimation risk - into account. A natural way to do this is to use the Bayesian concept of a posterior distribution $p(\theta|z)$, which summarizes the uncertainty about the parameters given the data observed so far. Integrating over this distribution, we obtain the so-called *predictive distribution* for long horizon returns. This distribution is conditioned only on the sample observed, and not on any fixed θ:

$$p(R_{T+\hat{T}}|z) = \int p(R_{T+\hat{T}}|\theta, z)\ p(\theta|z)\ d\theta.^{6} \tag{8}$$

A more appropriate problem for the investor to solve is then

$$\max_{\omega} \int v(W_{T+\hat{T}})\ p(R_{T+\hat{T}}|z)\ dR_{T+\widehat{T}}. \tag{9}$$

[6]It is clearly an abuse of notation to use the same notation p for all the different distributions, but hopefully by giving all the arguments explicitly, there will be no confusion.

This framework allows us to understand how parameter uncertainty affects portfolio choice. We simply compare the solution to problem (7), which ignores parameter uncertainty, with the solution to problem (9), which takes this uncertainty into account.

How can problems (7) and (9) be solved? We calculate the integrals in problems (7) and (9) for $\omega = 0, .01, .02, \ldots, .98, .99$, and report the ω which maximizes expected utility. Throughout the paper, we restrict the allocation to the interval $0 \leq \omega \leq 1$, precluding short selling and buying on margin.[7]

The integrals themselves are evaluated numerically by simulation. To illustrate the idea behind simulation methods, imagine that we are trying to evaluate

$$\int g(y)p(y)dy,$$

where $p(y)$ is a probability density function. We can approximate the integral by

$$\frac{1}{I}\sum_{i=1}^{I} g(y^{(i)}),$$

where $y^{(1)}, \ldots, y^{(I)}$ are independent draws from the probability density $p(y)$. To ensure a high degree of accuracy, we take $I = 1,000,000$ throughout.

In the examples considered in this paper, the conditional distribution $p(R_{T+\hat{T}}|z, \theta)$ is Normal. Therefore the integral in (7) is approximated by generating 1,000,000 independent draws from this Normal distribution, and averaging $v(W_{T+\widehat{T}})$ over all the draws.

In the case of (9), it is helpful to rewrite the problem as

$$\begin{aligned} & \max_{\omega} \int v(W_{T+\hat{T}})\ p(R_{T+\hat{T}}, \theta|z)\ dR_{T+\hat{T}}\ d\theta \qquad (10) \\ = & \max_{\omega} \int v(W_{T+\hat{T}})\ p(R_{T+\hat{T}}|z, \theta)\ p(\theta|z)\ dR_{T+\hat{T}}\ d\theta. \end{aligned}$$

The integral can therefore be evaluated by sampling from the joint distribution $p(R_{T+\hat{T}}, \theta|z)$, and then averaging $v(W_{T+\hat{T}})$ over those draws. As the decomposition in equation (10) shows, we sample from the joint distribution by first sampling from the posterior $p(\theta|z)$ and then from the conditional $p(R_{T+\hat{T}}|z, \theta)$. Sections II and III

[7] To be exact, we restrict the allocation ω to the range $0 \leq \omega \leq 0.99$. We do not calculate expected utility for $\omega = 1$ because in this case the integral in (9) equals $-\infty$. The problem is that when $\omega = 1$, wealth can be arbitrarily close to zero, while the left tail of the predictive distribution does not shrink fast enough to ensure that expected utility is bounded from below.

give detailed examples of this.[8]

This section has shown how to vary the degree of predictability observed by the investor - by changing the set of predictor variables included in the regression model - and how to incorporate parameter uncertainty into the analysis, by integrating over the posterior distribution of the parameters. Sections II and III use this framework to examine how the optimal portfolio changes when predictability in returns and estimation risk are accounted for.

B. The Data

The empirical work in this paper uses post-war data on asset returns and predictor variables. The stock index is the value-weighted index of stocks traded on the NYSE, as calculated by the Center for Research in Security Prices (CRSP) at the University of Chicago. In calculating excess returns, we use U.S. Treasury Bill returns as provided by Ibbotson and Associates. We also use the dividend yield as a predictor variable. The dividend yield in month t is defined as the dividends paid by the firms in the stock index during months $t-11$ through t, divided by the value of the index at the end of month t.

Monthly data is used throughout, spanning 523 months from June 1952 to December 1995. We restrict the data to this post-war period so as to avoid the time before the Treasury Accord of 1951 when interest rates were held almost constant by the Federal Reserve Board. Since the regression attempts to model the stochastic behavior of stock returns in excess of Treasury Bills, it is important to avoid structural breaks in the time series for the latter variable.

II. The Effect of Parameter Uncertainty

Rather than incorporate both predictability and parameter uncertainty at once, we introduce them one at a time. That is, we start out by considering the special

[8]The integral in problem (7) is only one dimensional and therefore quadrature methods are a reasonable alternative to simulation. This is not the case for problem (9). To write down a closed-form expression for $p(R_{T+\hat{T}}|z)$, we would need to integrate out the parameters θ, as shown in equation (8), and this is not possible for $\widehat{T} > 1$. We therefore need to integrate over the parameter space as well, and an integral of this size can only be tackled by simulation. For the special case of $\hat{T} = 1$, the parameters θ *can* be integrated out, giving a closed form t-distribution for the one period ahead predictive distribution. The integral is therefore again one dimensional and can be accurately handled by quadrature. By looking only at the $\hat{T} = 1$ case, Kandel and Stambaugh (1996) are able to take advantage of this.

case of the model in Section I where no predictor variables are included in the VAR – and hence where asset returns are i.i.d. – and look at how parameter uncertainty *alone* affects portfolio allocation. In Section III, we move to the more general case which allows for predictability in returns.

A. Constructing the Predictive Distribution

Suppose then that stock index returns are i.i.d., so that

$$r_t = \mu + \epsilon_t, \tag{11}$$

where r_t is the continuously compounded excess return on the stock index over month t, and where $\epsilon_t \sim$ i.i.d. $N(0, \sigma^2)$.

As described in Section I, a buy-and-hold investor with a horizon $\widehat{T}$ months long starting at time T solves the problem stated in (6). In our numerical work, we set r_f, the continuously compounded real monthly T-Bill return equal to 0.0036, the real return on T-Bills over December 1995, the last month of our sample.[9]

The investor has two choices of distribution for calculating the expectation in (6). He may incorporate parameter uncertainty and use the predictive distribution for returns $p(R_{T+\hat{T}}|r)$, where $r = (r_1, \ldots, r_T)$. Alternatively, he may ignore parameter uncertainty and calculate the expectation over the distribution of returns conditional on fixed parameter values, $p(R_{T+\hat{T}}|r, \mu, \sigma^2)$. The effect of parameter uncertainty is revealed by comparing the optimal portfolio allocations in these two cases.

We approximate the integral for expected utility by taking a sample $(R^{(i)}_{T+\hat{T}})^{I=1{,}000{,}000}_{i=1}$ from one of the two possible distributions, and then computing

$$\frac{1}{I}\sum_{i=1}^{I} \frac{\{(1-\omega)\exp(r_f \widehat{T}) + \omega \exp(r_f \widehat{T} + R^{(i)}_{T+\hat{T}})\}^{1-A}}{1-A}. \tag{12}$$

In Section II.B, we present the optimal allocations ω which maximize (12) for a variety of risk aversion levels A and investment horizons $\widehat{T}$, and for each of the two cases where the investor either ignores or accounts for parameter uncertainty. Since sampling from the distributions $p(R_{T+\hat{T}}|r)$ and $p(R_{T+\hat{T}}|r, \mu, \sigma^2)$ is an important step in computing these optimal allocations, we devote the next few paragraphs to explaining the sampling procedure in more detail.

As indicated in (10), there are two steps to sampling from the *predictive* distribution for long horizon returns $p(R_{T+\hat{T}}|r)$. First, we generate a large sample from the

[9] The nominal return on T-Bills in December 1995 is deflated using the rate of change in the Consumer Price Index, provided by Ibbotson and Associates.

posterior distribution for the parameters $p(\mu, \sigma^2|r)$. Second, for each of the (μ, σ^2) pairs drawn, we sample once from the distribution of long horizon returns conditional on both past data and the parameters, $p(R_{T+\hat{T}}|\mu, \sigma^2, r)$, a Normal distribution. This produces a large sample from the predictive distribution. We now provide more detail about each of these steps.

To construct the posterior distribution $p(\mu, \sigma^2|r)$, a prior is required. Throughout this section, we use a conventional uninformative prior,

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}.^{10}$$

Zellner (1971) shows that the posterior is then given by

$$\begin{aligned} \sigma^2|r &\sim \mathrm{IG}\left(\frac{T-1}{2}, \frac{1}{2}\sum_{t=1}^{T}(r_t - \overline{r})^2\right) \\ \mu|\sigma^2, r &\sim N\left(\overline{r}, \frac{\sigma^2}{T}\right), \end{aligned}$$

where $\overline{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$.

To sample from the posterior $p(\mu, \sigma^2|r)$, we therefore first sample from the marginal $p(\sigma^2|r)$, an Inverse Gamma distribution and then, given the σ^2 drawn, from the conditional $p(\mu|\sigma^2, z)$, a Normal distribution. Repeating this many times gives an accurate representation of the posterior distribution.

Table I presents the results of this procedure. The left panel uses monthly data on stock index returns from June 1952 to December 1995. The right panel uses the subsample of data from January 1986 to December 1995. An investor who believes that the mean μ and variance σ^2 of stock returns are changing over time may feel more comfortable estimating those parameters over this second more recent data sample.

In each case, the data is used to generate a sample of size 1,000,000 from the posterior distribution for μ and σ^2. Table I gives the mean and standard deviation of the posterior distribution for each parameter. For example, for an investor using the full sample from 1952 to 1995, the posterior distribution for the mean monthly excess stock return μ has mean 0.005 and standard deviation 0.0018. This appears to be an important source of parameter uncertainty for the investor. The posterior distribution for the variance σ^2 is much tighter and is centered around 0.0017. An investor confining his attention to the shorter data set will be more uncertain about

[10] Another reasonable approach would be to use a more informative prior which puts zero weight on negative values of μ, reflecting the observation in Merton (1980) that the expected market risk premium should be positive.

the parameters; the standard deviation of the posterior for μ is now a substantial 0.0039.

The second step in sampling from the predictive distribution is to sample from the distribution of returns conditional on fixed parameter values $p(R_{T+\hat{T}}|\mu, \sigma^2, r)$. Since

$$\begin{aligned} r_{T+1} &= \mu + \epsilon_{T+1}, \\ &\vdots \\ r_{T+\hat{T}} &= \mu + \epsilon_{T+\hat{T}}, \end{aligned} \tag{13}$$

the sum $R_{T+\hat{T}} = r_{T+1} + r_{T+2} + \ldots + r_{T+\hat{T}}$ is Normally distributed conditional on μ and σ^2 with mean $\hat{T}\mu$ and variance $\hat{T}\sigma^2$. Therefore, for each of the 1,000,000 pairs of μ and σ^2 drawn from the posterior $p(\mu, \sigma^2|r)$, we sample one point from the Normal distribution with mean $\hat{T}\mu$ and variance $\hat{T}\sigma^2$. This gives a sample of size 1,000,000 from the predictive distribution $p(R_{T+\hat{T}}|r)$, which we can use to compute the optimal allocation when taking parameter uncertainty into account.

Our strategy for understanding the effect of parameter uncertainty is to compare the allocation of an investor who uses the predictive distribution when forecasting returns with the allocation of an investor who ignores estimation error, sampling instead from the distribution of returns conditional on fixed parameters $p(R_{T+\hat{T}}|r, \mu, \sigma^2)$. For the latter case, we assume that the investor takes the posterior means of μ and σ^2 given in Table I as the fixed values of the parameters, and then draws 1,000,000 times from a Normal distribution with mean $\hat{T}\mu$ and variance $\hat{T}\sigma^2$.

We are now ready to present optimal portfolio allocations. We compute the quantity in (12) for ω ranging from zero to 0.99 in increments of 0.01, and report the ω maximizing this quantity. The procedure is repeated for several possible investment horizons, ranging from one year to 10 years in one year increments, for several values of risk aversion A and for the two possible distributions for cumulative returns, one of them ignoring estimation risk, the other incorporating it.

B. Results

Figure 1 shows the optimal percentage 100ω percent allocated to the stock index, plotted against the investment horizon in years. The two upper graphs show the optimal allocations chosen by investors who use the full data set from 1952 to 1995; the two lower graphs are for investors who use only the subsample from 1986 to 1995. The two graphs on the left are based on a risk-aversion of $A = 5$, while those on the right are for $A = 10$. The dash/dot line shows the allocation conditional on fixed parameter values, and the solid line shows the allocation when we account for parameter uncertainty.

The dash/dot line is completely horizontal in all the graphs. In other words, an investor ignoring the uncertainty about the mean and variance of asset returns would allocate the same amount to stocks, regardless of his investment horizon. This sounds similar to Samuelson's famous horizon irrelevance result, although it is important to note that the two results are different. Samuelson (1969) shows that with power utility and i.i.d. returns, the optimal allocation is independent of the horizon. However, he proves this for an investor who optimally rebalances his portfolio at regular intervals, rather than for the buy-and-hold investor we consider here.

The main point of this exercise though, is to show how the allocation differs when parameter uncertainty is explicitly incorporated into the investor's decision-making framework. Interestingly, Figure 1 shows that in this case, the stock allocation *falls* as the horizon increases. In other words, parameter uncertainty can introduce horizon effects even within the context of an i.i.d. model for returns. This point does not appear to have been noted before in the earlier literature on this topic, such as Bawa, Brown, and Klein (1979). That research focuses more on how estimation risk varies with the size of the data sample, while keeping the investor's horizon fixed. The results here are concerned with the effects of estimation risk as we vary the investor's horizon, while keeping the sample size fixed.

The magnitude of the effects induced by parameter uncertainty are substantial. For an investor using the full data set, and with $A = 5$, the difference in allocation at a 10 year horizon is approximately 10 percent. For another investor with the same risk-aversion of $A = 5$, but who uses only the more recent subsample of data, the effect is dramatically larger, a full 35 percent at the 10 year horizon! This reflects the greater impact of the higher parameter uncertainty faced by an investor who uses such a short data sample.

The fact that parameter uncertainty makes a difference is often confusing at first sight. When parameter uncertainty is ignored, the investor uses a Normal distribution with mean $\widehat{T}\mu$ and variance $\widehat{T}\sigma^2$ for his forecast of log cumulative returns. Both the mean, and more importantly the variance, grow linearly with the investor's horizon $\widehat{T}$. Figure 1 shows that this leads to the same stock allocation, regardless of the investor's horizon.

Accounting for estimation risk changes this. The investor's distribution for long horizon returns now incorporates an extra degree of uncertainty, increasing its variance. Moreover, this extra uncertainty makes the variance of the distribution for cumulative returns increase *faster* than linearly with the horizon $\widehat{T}$. This makes stocks look riskier to long horizon investors, who therefore reduce the amount they allocate to equities.

The reason variances increase faster than linearly with the horizon is because in the presence of parameter uncertainty, returns are no longer i.i.d. from the point of view of the investor, but rather positively serially correlated. To understand this more precisely, recall that an important source of uncertainty in the parameters surrounds the mean of the stock return. If the stock return is high over the first month, then it will probably be high over the second month because it is likely that the state of the world is one with a high realization of the uncertain stock mean parameter μ. This is the sense in which stock returns are positively serially correlated from the investor's perspective.

An important issue we have not yet discussed is the accuracy of the numerical methods used to obtain the optimal portfolios. In an effort to maintain high accuracy, we use samples containing 1,000,000 draws from the sampling distribution when calculating expected utility. In the Appendix, we attempt to convey the size of the simulation error that is present; the results there suggest that using 1,000,000 draws does indeed provide a high degree of accuracy.

III. The Effect of Predictability

A. Constructing the Predictive Distribution

Now that the impact of parameter uncertainty alone has been illustrated, predictability can be introduced as well. We return to the regression model (2) discussed in Section I. In the calculations presented in this section, the vector z_t contains only two components: the excess stock index return r_t, and a single predictor variable, the dividend yield $x_{1,t}$, which captures an important component of the variation in expected returns.[11]

As explained in Section I, the problem faced at time T by a buy-and-hold investor with a horizon of $\widehat{T}$ months is given by (6). There are a number of possible distributions the investor can use when computing the expectation in (6). An investor who ignores the uncertainty in the model parameters uses the distribution of future returns conditional on both past data and fixed parameter values θ, $p(R_{T+\hat{T}}|\theta, z)$, where $z = (z_1, \dots, z_T)'$. In contrast, the investor who takes parameter uncertainty

[11] Many papers demonstrate the dividend yield's ability to forecast returns. See for example Keim and Stambaugh (1986), Fama and French (1988), and Campbell (1991).

into account samples from the predictive distribution, conditional only on past data and not on the parameters, $p(R_{T+\hat{T}}|z)$.

We approximate the integral for expected utility by taking a sample $(R^{(i)}_{T+\hat{T}})_{i=1}^{I=1,000,000}$ from one of the two possible distributions, and then computing

$$\frac{1}{I}\sum_{i=1}^{I}\frac{\{(1-\omega)\exp(r_f\widehat{T})+\omega\exp(r_f\widehat{T}+R^{(i)}_{T+\hat{T}})\}^{1-A}}{1-A}. \tag{14}$$

In Section III.B, we present the optimal allocations ω which maximize the quantity in (14) for a variety of risk aversion levels A and investment horizons $\widehat{T}$, and for different cases where the investor either ignores or accounts for parameter uncertainty. The next few paragraphs explain how we sample from $p(R_{T+\hat{T}}|z)$ and $p(R_{T+\hat{T}}|\theta,z)$, an important step in computing these optimal allocations.

The procedure for sampling from the predictive distribution is similar to that in Section II. First, we generate a sample of size 1,000,000 from the posterior distribution for the parameters $p(a,B,\Sigma|z)$. Second, for each of the 1,000,000 sets of parameter values drawn, we sample once from the distribution of returns conditional on both past data and the parameters, a Normal distribution. This gives us a sample of size 1,000,000 from the predictive distribution for returns, conditional only on past returns, with the parameter uncertainty integrated out. We now provide more detail about each of these steps.

To compute the posterior distribution $p(a,B,\Sigma|z)$, rewrite the model as

$$\begin{pmatrix} z_2^{'} \\ \vdots \\ z_T^{'} \end{pmatrix} = \begin{pmatrix} 1 & x_1^{'} \\ 1 & \vdots \\ 1 & x_{T-1}^{'} \end{pmatrix} \begin{pmatrix} a^{'} \\ B^{'} \end{pmatrix} + \begin{pmatrix} \epsilon_2^{'} \\ \vdots \\ \epsilon_T^{'} \end{pmatrix}, \tag{15}$$

or

$$Z = XC + E, \tag{16}$$

where Z is a $(T-1,n+1)$ matrix with the vectors $z_2^{'},\ldots,z_T^{'}$ as rows; X is a $(T-1,n+1)$ matrix with the vectors $(1\ x_1^{'}),\ldots,(1\ x_{T-1}^{'})$ as rows, and E is a $(T-1,n+1)$ matrix with the vectors $\epsilon_2^{'},\ldots,\epsilon_T^{'}$ as rows. Finally C is an $(n+1,n+1)$ matrix with top row $a^{'}$ and the matrix $B^{'}$ below that. Here $n=1$ because we use only one predictor variable, the dividend yield.

Zellner (1971) discusses the Bayesian analysis of a multivariate regression model in the traditional case with exogenous regressors. His analysis carries over directly to our dynamic regression framework with endogenous regressors; the form of the likelihood function is the same in both cases, so long as we condition on the first observation in the sample, z_1. A standard uninformative prior here is

$$p(C, \Sigma) \propto |\Sigma|^{-\frac{n+2}{2}}.^{12}$$

The posterior $p(C, \Sigma^{-1}|z)$ is then given by

$$\begin{aligned} \Sigma^{-1}|z &\sim \text{Wishart}(T - n - 2, S^{-1}) \\ vec(C)|\Sigma, z &\sim N(vec(\hat{C}), \Sigma \otimes (X'X)^{-1}) \end{aligned}$$

where $S = (Z - X\hat{C})'(Z - X\hat{C})$ with $\hat{C} = (X'X)^{-1}X'Z$. We sample from the posterior distribution by first drawing from the marginal $p(\Sigma^{-1}|z)$, and then from the conditional $p(C|\Sigma, z)$.

Table II presents the mean and standard deviation of the posterior distribution for a, B, and Σ, generated by sampling 1,000,000 times from that posterior. The left panel uses the full data sample covering the period 1952 to 1995, while the results on the right are obtained using the subsample from 1986 to 1995. An investor who believes that the relationship between the dividend yield and stock returns is changing over time may prefer to estimate the regression over this more recent sample.

Look first at the left panel of Table II. The B matrix shows the well-documented predictive power of the dividend yield for stock returns: the posterior distribution for that coefficient has mean 0.5118 and standard deviation 0.2129. The dividend yield is highly persistent. The variance matrix shows the strong negative correlation between innovations in stock returns and the dividend yield, estimated here at -0.9351; this has an important influence on the distribution of long horizon returns. Note also the greater parameter uncertainty faced by investors using only the recent subsample; in particular, the predictive power of the dividend yield for returns is now estimated much less accurately.

The second step in sampling from the predictive distribution is to sample from $p(R_{T+\hat{T}}|\theta, z)$. Note that since $z_t = a + Bx_{t-1} + \varepsilon_t$, we can write $z_t = a + B_0 z_{t-1} + \varepsilon_t$ where $B_0 = \left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \; B \right]$. Therefore

[12] Stambaugh (1998) discusses the use of alternative priors and of using the unconditional likelihood instead of conditioning on z_1.

$$
\begin{aligned}
z_{T+1} &= a + B_0 z_T + \epsilon_{T+1} \\
z_{T+2} &= a + B_0 a + B_0^2 z_T + \epsilon_{T+2} + B_0 \epsilon_{T+1} \\
&\vdots \\
z_{T+\hat{T}} &= a + B_0 a + B_0^2 a + \ldots B_0^{\hat{T}-1} a \\
&\quad + B_0^{\hat{T}} z_T \\
&\quad + \epsilon_{T+\hat{T}} + B_0 \epsilon_{T+\hat{T}-1} + B_0^2 \epsilon_{T+\hat{T}-2} + \ldots + B_0^{\hat{T}-2} \epsilon_{T+2} + B_0^{\hat{T}-1} \epsilon_{T+1}.
\end{aligned}
\tag{17}
$$

Conditional on a, B, and Σ, the sum $Z_{T+\hat{T}} = z_{T+1} + z_{T+2} + \ldots + z_{T+\hat{T}}$ is Normally distributed with mean and variance given by

$$
\begin{aligned}
\mu_{sum} &= \hat{T} a + (\hat{T}-1) B_0 a + (\hat{T}-2) B_0^2 a + \ldots + B_0^{\hat{T}-1} a \\
&\quad + (B_0 + B_0^2 + \ldots + B_0^{\hat{T}}) z_T
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
\Sigma_{sum} &= \Sigma \\
&\quad + (I + B_0) \Sigma (I + B_0)' \\
&\quad + (I + B_0 + B_0^2) \Sigma (I + B_0 + B_0^2)' \\
&\quad \vdots \\
&\quad + (I + B_0 + \ldots + B_0^{\hat{T}-1}) \Sigma (I + B_0 + \ldots + B_0^{\hat{T}-1})'.
\end{aligned}
\tag{19}
$$

For each of the 1,000,000 realizations of the parameters in the sample from the posterior $p(a, B, \Sigma | z)$, we draw one point from the Normal distribution with mean and variance given by the above expressions, thereby giving a sample of size 1,000,000 from the predictive distribution.

The aim of this section is to understand how predictability in asset returns and parameter uncertainty affect portfolio choice. To do this, we compute optimal allocations using four different choices for the distribution of future returns. These distributions differ in whether they take into account predictability and parameter uncertainty. For instance, the investor may choose to take predictability into account when forecasting returns. He does so by including the dividend yield in the VAR he uses to forecast returns. Alternatively he may ignore the predictability in returns, simply by excluding the dividend yield from the VAR. In this case, the model for returns reduces to the i.i.d. model of Section II.

So far, this gives two different ways of forecasting future returns. However, for each of these two ways, there is a further choice to be made. The investor may account for the parameter uncertainty in the model, and use a predictive distribution constructed in the manner described earlier. Alternatively, he may ignore the parameter uncertainty in the model; in this case, we assume that the distributions for future returns are constructed using the posterior means of a, B, and Σ given in Table II (or of μ and σ^2 in Table I if predictability is also ignored) as the fixed values of the parameters, and then drawing 1,000,000 times from the Normal distribution with mean and variance given by equations (18) and (19) above. This extra choice about whether to incorporate parameter uncertainty gives a total of four possibilities for the distribution of future returns.

In Section III.B below, we present optimal portfolio allocations. We evaluate the quantity in (14) for ω ranging from zero to 0.99 in increments of 0.01, and report the ω maximizing this quantity. We do this calculation for several values of investor risk-aversion A; for several investment horizons, ranging from one year to 10 years at one year intervals; and for the four possible distributions for future returns. By comparing how the optimal portfolios differ depending on which distribution we use for forecasting returns, we can understand how the predictive power of the dividend yield and parameter uncertainty affect portfolio choice.[13]

The investor's distribution for future returns will of course depend on the value of the dividend yield at the beginning of the investment horizon, $x_{1,T}$.[14] If the yield is low, this forecasts low returns, lowering the mean of the distribution for future returns and reducing the allocation to equities. In our initial set of results, we abstract from this effect by setting the initial value of the dividend yield to its mean in the sample, namely $x_{1,T} = 3.75\%$, and investigate how the optimal allocation changes with the investor's horizon for this fixed initial value of the predictor. Later, we look at how the results are affected when the initial dividend yield takes values above or below its sample mean.

B. Results

Figure 2 presents the solutions to the allocation problem. Each graph corresponds to a different level of risk-aversion. Within each graph, each line shows the percentage 100ω percent allocated to stocks plotted against the investment horizon ranging from one to 10 years. The four lines on each graph correspond to the four possible distributions the investor could use to forecast future returns.[15] All of the computations

[13] The optimal portfolios for the cases where predictability is ignored are computed in Section II, so we simply carry those results over to this section.

[14] The initial value of the predictor $x_{1,T}$ enters equation (18) through $z_T = (r_T \ x_{1,T})'$.

[15] Sometimes, the optimal allocation 100ω percent lies outside the (0,100) range for all horizons between one and ten years. This is why there may be fewer than four lines on some graphs.

are based on the full data sample from 1952 to 1995; results based on the more recent subsample are presented in Section III.C.

We focus on the graph for a risk aversion level of 10, which presents the results most clearly. The two lower lines in this graph represent the cases where the investor ignores predictability, excluding the dividend yield from the VAR. Of course, in this case, the model for returns simply reduces to the i.i.d. model discussed in Section II. These two lines are therefore exactly the same as those in the top right graph in Figure 1.

The main results of this section center on the two upper lines. These two lines correspond to the cases where the dividend yield is included in the analysis. The graph shows that when we ignore uncertainty about the model parameters (the dashed line), the optimal allocation to stocks for a long horizon investor is much higher than for a short horizon investor. When we take the uncertainty about the parameters into account, (the solid line), the long horizon allocation is again higher than the short horizon allocation – but not nearly as much higher as when we ignore estimation risk. The rest of this section explores these results in more detail.

We start with the case where parameter uncertainty is ignored. Why does the allocation to stocks in this case rise so dramatically at long horizons when the dividend yield is included in the regression?

Recall that when asset returns are modelled as i.i.d., the mean and variance of cumulative log returns grow linearly with the investor's horizon $\widehat{T}$. In Section II, we find that this leads to identical allocations to stocks, regardless of the investor's horizon.

When we acknowledge that returns may be predictable rather than i.i.d., this is no longer the case. The variance of cumulative log stock returns may grow *slower* than linearly with the investor's horizon, making stocks look relatively *less* risky at longer horizons and hence leading to higher allocations to stocks in the optimal portfolio.

This point can be seen mathematically. Write the regression model in full as

$$
\begin{aligned}
r_{t+1} &= \alpha + \beta x_{1,t} + \varepsilon_{1,t+1} && (20)\\
x_{1,t+1} &= \gamma + \phi x_{1,t} + \varepsilon_{2,t+1}, && (21)
\end{aligned}
$$

where

$$
\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N(0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}).
$$

The conditional variances of one- and two- period cumulative stock returns are

$$
var_T(r_{T+1}) = \sigma_1^2 \tag{22}
$$

$$var_T(r_{T+1} + r_{T+2}) \;=\; 2\sigma_1^2 + \beta^2\sigma_2^2 + 2\beta\sigma_{12}. \tag{23}$$

For the parameter values estimated from the data, in other words the posterior means in Table II, $\beta^2\sigma_2^2 + 2\beta\sigma_{12} < 0$, which implies that the conditional variance of two period returns is *less* than twice the conditional variance of one period returns. When we consider the predictive power of the dividend yield, conditional variances grow slower than linearly with the investor's horizon, lowering the perceived long run risk of stocks and increasing their optimal weight in the investor's portfolio.

Some numbers may help to make this point clearer. Table I shows that the variance of monthly excess stock returns is estimated at 0.0017, implying a standard deviation of $\sqrt{0.0017} = 4.12$ percent. In a model specifying i.i.d. returns, this implies a standard deviation for cumulative log excess returns over 10 years of $(0.0412)\sqrt{120} = 45.2$ percent. However, the standard deviation of the distribution for 10 year cumulative log excess returns generated by an investor who models returns using the VAR in Table II, ignoring parameter uncertainty, is from equation (19) equal to 23.7 percent, much lower than 45.2 percent! This demonstrates the extent to which conditioning on the dividend yield can slow the evolution of the variance of cumulative returns.

The intuition behind this effect is the following: suppose that the dividend yield falls unexpectedly. Since $\sigma_{12} < 0$, this is likely to be accompanied by a contemporaneous *positive* shock to stock returns. However, since the dividend yield is lower, stock returns are forecast to be *lower* in the future, since $\beta > 0$. This rise, followed by a fall in returns generates a component of negative serial correlation in returns which slows the evolution of the variance of cumulative returns as the horizon grows.

The results obtained here should not be viewed as being specific to the particular way we have modelled returns, nor to the particular parameter values estimated from the data. There is a strong economic intuition behind the general idea that time variation in expected returns induces mean-reversion in realized returns. If there is a positive shock to expected returns, it is very reasonable that realized returns should suffer a contemporaneous negative shock, since the discount rate for discounting future cashflows has suddenly increased. This negative shock to current realized returns, followed by the higher returns predicted in the future, are the source of mean-reversion, which in turn makes stocks more attractive to long run investors.

While mean-reversion provides a simple way of interpreting our results, it is important to note that horizon effects can be present even without negative serial correlation in returns. In other words, the predictability in returns may be sufficient to make stocks more attractive at long horizons, without being strong enough to induce mean-reversion in returns. One way to see this is to note that in our simplified model,

$$cov(r_t, r_{t+1}) = \frac{\beta^2 \phi \sigma_2^2}{1 - \phi^2} + \beta \sigma_{12} \tag{24}$$

$$cov(r_t, r_{t+i}) = \phi^{i-1} cov(r_t, r_{t+1}) \tag{25}$$

It is straightforward to note that we can choose parameters so that returns are serially uncorrelated at all lags and yet, by equation (23), so that the two-period conditional variance is less than twice the one-period conditional variance. In this situation, a two-period investor would allocate more to stocks than a one-period investor and yet there is no mean-reversion in returns.

The second important result in Figure 2 is that incorporating parameter uncertainty can substantially reduce the size of the horizon effect. For $A = 10$, ignoring this uncertainty can lead to an over-allocation to stocks of over 30 percent at a 10 year horizon!

Introducing parameter uncertainty has a number of effects. First, the investor acknowledges that he is uncertain about the mean stock return. In exactly the same way as in Section II, incorporating the uncertainty about the mean makes conditional variances grow more *quickly* as the horizon grows, tending to make stocks look more risky. Therefore the allocation to stocks is lower than in the case where estimation risk is ignored.

The investor also recognizes that the true predictive power of the dividend yield is uncertain; therefore it is also uncertain whether the dividend yield really does slow the evolution of conditional variances, and hence whether stocks really are less risky at long horizons. The investor is therefore again more cautious about stocks and allocates less to them.[16]

In the absence of estimation risk, we saw that predictability makes stocks look *less* risky at long horizons; incorporating the estimation risk makes them look *more* risky. These two effects therefore battle it out, leading to stock allocations that are not necessarily monotonic as a function of the investment horizon. The solid line in the bottom left graph in Figure 2 shows that for horizons up to eight years, predictability wins out, and the allocation to stocks rises; from that point on, the line falls slightly,

[16] When the investor takes estimation risk into account, he acknowledges both that the predictive power of the dividend yield may be weaker than the point estimates suggest – in which case he would certainly be reluctant to allocate more to stocks at long horizons – *and* that it may in fact be stronger, in which case he would be even keener to allocate more to equities at longer horizons. These effects go in opposite directions; on net, the investor invests less at long horizons because he is risk-averse and hence dislikes the mean-preserving spread that accounting for estimation risk adds to the distribution of future returns.

suggesting that estimation risk has caught up with the investor, making stocks look less attractive.

C. The Role of the Predictor Variable

Up to this point we have focused on just one effect of including the dividend yield as a predictor in the VAR. Conditioning on the dividend yield reduces the variance of predicted long horizon cumulative returns, leading to a higher allocation to stocks for long horizon investors.

Conditioning on the dividend yield has another, more direct implication for portfolio allocation. By its very nature as a predictor, the dividend yield also affects the mean of the distribution for future returns. When the dividend yield is low relative to its historical mean, an investor forecasts lower than average stock returns and hence reduces his allocation to stocks. This effect has not been prominent so far in the paper because the initial value of the dividend yield has been kept fixed at its unconditional mean in the sample period.

We now repeat the earlier analysis of this section for different initial values of the dividend yield.[17] Figure 3 presents the results. Each graph corresponds to a different level of risk-aversion. The graphs on the left illustrate the optimal allocations when parameter uncertainty is ignored; the graphs on the right incorporate it. Within each graph, we plot the optimal stock allocation as a function of the investor's horizon for five different initial values of the dividend yield. The five values we use are the historical mean of the dividend yield in our sample, namely $x_{1,T} = 3.75$ percent, and the values one and two standard deviations on either side of that.[18]

Look first at the graphs on the left side of Figure 3. They show that for all the initial values of the predictor that we consider, the earlier result of this section continues to hold: the allocation to stocks rises with the investor's horizon. Of course, for any *fixed* horizon, the optimal allocation is higher for higher values of the dividend yield, since the investor expects higher future returns. For any fixed initial value of the dividend yield, however, the 10 year allocation is higher than the one year allocation. Moreover, the optimal allocation of an investor with a 10 year horizon is just as sensitive to the initial value of the dividend yield as the optimal allocation of a one

[17]In effect, we are producing optimal portfolio recommendations for an investor who has observed a hypothetical sample with different $x_{1,T}$ but the same posterior distribution for the parameters as in the actual sample.

[18]For some of the initial values of the dividend yield, the optimal allocation 100ω percent lies outside the (0,100) range for all horizons between one and ten years. This is why there may be fewer than five lines on any one graph.

year horizon investor. In other words, the allocation lines show no sign of converging.

The picture is remarkably different when the investor properly accounts for the fact that he is uncertain about the parameters governing asset returns. These results are shown on in the graph on the right hand side of Figure 3. For lower initial values of the dividend yield, the optimal allocation once again rises with the investor's horizon. For higher values of the predictor, though, the allocation to stocks *falls* with the investment horizon. Another way of looking at this is to note that the allocation lines converge, resulting in a 10 year allocation which is less sensitive to the initial dividend yield than the allocation of a one year investor, and much less sensitive than the allocation of an investor with a 10 year horizon who ignores estimation risk. This result is intuitive: if the true forecasting power of the dividend yield is uncertain, the allocation of a long horizon investor should be less sensitive to the initial value of the predictor.

While Figure 3 shows that the impact of parameter uncertainty is substantial, its effect can be even more dramatic. Suppose that an investor believes that the true predictive power of the dividend yield changes over time and is therefore wary of running a regression over the full 1952 to 1995 period, preferring instead to estimate the relationship over the shorter 10 year period from 1986 to 1995. The posterior distribution of the VAR parameters over this data sample was summarized in Table II.

Figure 4 repeats the calculations of Figure 3 for the case where the investor uses the more recent subsample in making his decisions. Once again, the optimal allocation rises sharply with the horizon for the investor who takes the parameters as fixed. When parameter uncertainty is incorporated, however, the recommended portfolios are completely different! The allocations for investors with 10 year horizons are now largely insensitive to the initial value of the predictor: the convergence in the allocation lines, already pronounced in Figure 3, is now much more dramatic. Just as in Figure 3, estimation risk is sometimes so strong as to cause the stock allocation for a 10 year investor to be lower than that for a one year investor. In this case, this is even true when the dividend yield is at its sample mean, $x_{1,T} = 3.36$ percent!

Another intriguing result in Figures 3 and 4 is the fact that for a given investment horizon and risk-aversion level, the optimal stock allocation is not necessarily increasing in the initial value of the predictor variable. This is a surprising fact at first sight: if the initial value of the dividend yield is five percent rather than four percent, the distribution for future returns forecast by the investor has a higher posterior mean, which should lead to a higher allocation to the stock index. Moreover, the variance of the distribution for future returns is insensitive to the initial value of the predictor, so this cannot explain the non-monotonicity result. Stambaugh (1998) demonstrates that it is in fact the third moment of the return distribution, skewness, that is important here. Incorporating parameter uncertainty generates positive skewness in the predictive distribution for low initial values of the predictor, and negative skewness for higher initial values. This negative skewness makes stocks less attractive, the higher the dividend yield, and makes the optimal allocation non-monotonic in the initial value of the predictor.

A reader trying to interpret the results in Figures 2, 3, and 4 will obviously be concerned about the accuracy of the numerical methods. As mentioned at the end of Section II, the Appendix contains a discussion of simulation error. We do not dwell on it any more here, other than to say that the results there suggest that by using 1,000,000 draws in our simulations, we can be comfortable that the level of accuracy is high.

IV. Dynamic Allocation

Up until this point, we have focused on the buy-and-hold investment problem. We now examine portfolio choice when the investor optimally rebalances over his investment horizon. Specifically, consider an investor who is allowed to rebalance annually using the new information at the end of each year. We analyze how the optimal allocation depends on the investor's horizon. To begin, we work with the simpler case where parameter uncertainty is ignored. Then we look at how the results change when the investor incorporates parameter uncertainty.

A. An Asset Allocation Framework with Dynamic Rebalancing

We use the same regression model as in Section III, originally introduced as equation (2) in Section I, with $z_t = (r_t \; x_t)'$, where $x_t = x_{1,t}$ is the dividend yield. We also maintain the earlier simplification that the continuously compounded real return on T-Bills is a constant r_f per period. As before, we set $r_f = 0.0036$ in all our numerical work.

The investor who optimally rebalances his portfolio at regular intervals faces a dynamic programming problem. To solve this problem, we employ the standard technique of discretizing the state space and using backward induction. The next few paragraphs formalize this.

Suppose we are at time T, and the investor has a horizon $\widehat{T}$ months long. Divide the horizon into K intervals of equal length, $[t_0, t_1]$, $[t_1, t_2], \ldots [t_{K-1}, t_K]$, where the start and end of the investor's horizon are $t_0 = T$, and $t_K = T + \widehat{T}$, respectively. The investor adjusts his portfolio K times over the horizon, at points $(t_0, t_1, \ldots, t_{K-1})$. The control variables at the investor's disposal are $(\omega_0, ..., \omega_{K-1})$, his allocations to the stock index at times $(t_0, ..., t_{K-1})$ respectively. To make the notation less cumbersome, we write W_k in place of W_{t_k} for the investor's wealth at time t_k, z_k in place of z_{t_k}, and x_k in place of x_{t_k}. The investor's problem is then

$$\max_{t_0} E_{t_0} \left(\frac{W_K^{1-A}}{1-A} \right), \tag{26}$$

where $\max_{t_0}$ means that the investor maximizes over all remaining decisions from time t_0 on, and where

$$W_{k+1} = W_k \{ (1-\omega_k) \exp(r_f \frac{\widehat{T}}{K}) + \omega_k \exp(r_f \frac{\widehat{T}}{K} + R_{k+1}) \}, \tag{27}$$

$$R_{k+1} = r_{t_k+1} + r_{t_k+2} + \ldots + r_{t_{k+1}}, \tag{28}$$

for $k = 0, \ldots, K-1$. Note that the return on T-Bills between rebalancing points is now $\exp(r_f \frac{\widehat{T}}{K})$ because the full $\widehat{T}$ month horizon is broken into K intervals. The cumulative excess stock return between rebalancing points t_k and t_{k+1} is R_{k+1}.

Define the derived utility of wealth

$$J(W_k, x_k, t_k) = \max_{t_k} E_{t_k} \left(\frac{W_K^{1-A}}{1-A} \right), \tag{29}$$

where $\max_{t_k}$ means a maximization over all remaining decisions from time t_k on. Note that the value function J does *not* depend on r_k, the stock return over month t_k, because in our model, the current value of the predictor variable alone characterizes the investment opportunity set. The Bellman equation of optimality is

$$J(W_k, x_k, t_k) = \max_{\omega_k} E_{t_k} \{ J(W_{k+1}, x_{k+1}, t_{k+1}) \}. \tag{30}$$

A simple induction argument shows that derived utility may be written

$$J(W_k, x_k, t_k) = \frac{W_k^{1-A}}{1-A} Q(x_k, t_k), \tag{31}$$

for $A \neq 1$, or in the case of $A = 1$,

$$J(W_k, x_k, t_k) = \log(W_k) + Q(x_k, t_k), \tag{32}$$

so that the Bellman equation can be rewritten (for $A \neq 1$) as

$$Q(x_k, t_k) = \max_{\omega_k} E_{t_k}\{[(1-\omega_k)\exp(r_f \frac{\widehat{T}}{K}) + \omega_k \exp(r_f \frac{\widehat{T}}{K} + R_{k+1})]^{1-A} Q(x_{k+1}, t_{k+1})\}. \tag{33}$$

Since we are ignoring estimation risk in this section, the expectation in (33) is taken over the Normal distribution $p(R_{k+1}, x_{k+1}|\theta, x_k)$, conditioned on parameter values fixed at the posterior means in Table II.

The usual technique for solving a Bellman equation, which we adopt here, is to discretize the state space and then use backward induction. In particular, we take the interval ranging from three standard deviations below the historical mean of the dividend yield to three standard deviations above, and discretize this range with 25 equally spaced grid points, which we write as $(x_k^j)_{j=1,\ldots,25}$.

Suppose that $Q(x_{k+1}, t_{k+1})$ is known for all $x_{k+1} = x_{k+1}^j$, $j = 1, \ldots, 25$. Clearly this is true in the last period as $Q(x_K, t_K) = 1$, $\forall x_K$. Then we can use equation (33) to obtain $Q(x_k^j, t_k)$.

Specifically, for each x_k^j, $j = 1, \ldots, 25$, we draw a large sample $(R_{k+1}^{(i)}, x_{k+1}^{(i)})_{i=1}^{I=1,000,000}$ from the Normal distribution $p(R_{k+1}, x_{k+1}|\theta, x_k^j)$, and set $Q(x_k^j, t_k)$ equal to

$$\max_{\omega_k} \frac{1}{I} \sum_{i=1}^{I} [(1-\omega_k)\exp(r_f \frac{\widehat{T}}{K}) + \omega_k \exp(r_f \frac{\widehat{T}}{K} + R_{k+1}^{(i)})]^{1-A} Q(x_{k+1}^{(i)}, t_{k+1}), \tag{34}$$

Of course in general, we only know $Q(x_{k+1}, t_{k+1})$ for $x_{k+1} = x_{k+1}^j$, so we approximate $Q(x_{k+1}^{(i)}, t_{k+1})$ by $Q(x_{k+1}^j, t_{k+1})$, where x_{k+1}^j is the closest element of the discretized state space to $x_{k+1}^{(i)}$. This calculation gives $Q(x_k^j, t_k)$ for all $j = 1, \ldots, 25$. Backward induction through all K rebalancing points eventually gives $Q(x_0^j, t_0)$ and hence the optimal allocations ω_0.

In the Appendix, we find that using a sample size of 1,000,000 from the distribution for cumulative returns provides a high degree of accuracy, and we therefore continue to use this sample size here. Another variable which affects the accuracy of our results is the number of grid points we use to discretize the state variable. We repeated the calculation of the optimal portfolio policies with increasingly fine discretizations until the results remained unchanged. The results presented here are for the finest discretization tried, namely with 25 grid points.

B. Results

The graphs on the left hand side of Figure 5 present optimal allocations for investors with horizons ranging from one to 10 years, who rebalance optimally every year. Each graph corresponds to a particular risk aversion level, and within each graph, each line corresponds to a different initial value of the dividend yield. The data sample used is the more recent subsample from 1986 to 1995; using this shorter sample illustrates the effects of estimation risk more clearly when we reintroduce it in Section IV.C. Results are displayed for five different initial values of the predictor: its historical mean in the sample, and the values one and two standard deviations above and below that.

The graphs show that for the risk-aversion levels presented here, the results appear similar to those in the buy-and-hold case in Figure 4, which is based on the same data sample: the optimal allocation to stocks rises with the investor's horizon. Although the results appear similar, the effect driving them is different. The increase in allocation across horizons for a rebalancing investor is due to the so-called hedging demands first described by Merton (1973). In our model, the dividend yield is the state variable governing expected returns. As it changes over time, it also changes the investment opportunity set faced by the investor. Merton shows that investors may want to hedge these changes by investing in such a way that gives them higher wealth precisely when investment opportunities are unattractive, in other words, when expected returns are low. Table II shows that shocks to expected returns are reliably negatively correlated with shocks to realized returns; this makes holding more in stocks an ideal way of hedging against movements in expected returns.

The results in these graphs should not be seen as obtaining only for the particular model of stock returns used here, but as holding more generally whenever expected returns vary. We would expect any state variable positively related to expected returns to be negatively correlated with realized stock returns, making stocks a good hedge for that state variable, should investors choose to hedge.

Kim and Omberg (1996) examine this intertemporal problem analytically and show that all investors with risk-aversion $A > 1$ will want to hedge in this way. Hedging against movements in state variables governing expected returns will be particularly relevant for long horizon investors; the fact that these investors, according to Figure 5, hold substantially more in stocks indicates that the effect in the theoretical work of Kim and Omberg is empirically quite large. Kim and Omberg also show that the opposite effect holds when $A < 1$. In that case, investors choose to have more wealth when investment opportunities are good. This makes the hedging demand negative, with the result that long horizon investors hold less in equities than those with a shorter term outlook.

The result that longer horizon investors hold more in stocks to hedge against movements in expected returns may no longer hold if the investor has a broader range of asset classes to choose from. In that case, the best hedge against the state variable risk may be a portfolio quite different from the stock index alone. Long term investors with $A > 1$ would then hold more of this hedging portfolio, but not necessarily more in the stock index.

C. Parameter Uncertainty in a Multiperiod Setting

In a dynamic context, parameter uncertainty has two effects. The first effect is analogous to the one we faced in the buy-and-hold problem. When calculating the value function in (33), the expectation should be taken over a distribution which incorporates the uncertainty about the parameters. There is a second effect, however: the uncertainty about the parameters may *change* over time. As new data is observed, the investor updates his posterior distribution for the parameters. Therefore the investment opportunity set perceived by the investor may change over time not simply because the dividend yield changes, but because the investor's beliefs about the relationship between the dividend yield and stock returns - about the model parameters - have changed. This new source of variation in the investment opportunity set may itself generate a hedging demand.

The effect of "learning" about parameters on portfolio holdings is studied theoretically by Williams (1977) who notes the possibility of a learning-based hedging demand. Gennotte (1986) also examines portfolio allocation with incomplete information, and finds a similar hedging demand, albeit in a slightly different framework. In his model, the state variable governing expected returns (which we call $x_{1,t}$) is unobservable while the predictive power of the state variable for returns (which we call β) is known. In our model of course, the state variable is taken to be the fully observable dividend yield, and it is the forecasting power of this variable for returns that is unknown.[19]

While the papers of Williams and Gennotte have analyzed the incomplete information problem theoretically, we attempt to quantify *empirically* the importance of estimation risk in a dynamic context. This is a formidable problem. The reason is that the investment opportunity set is no longer characterized by one variable alone,

[19]While the setup in Gennotte's model appears different at first sight, the methodology of this paper can be applied to his case. His model can be estimated on past data, giving a posterior distribution for the state variable $x_{1,t}$ conditional on past data, and hence a predictive distribution for future returns, incorporating the uncertainty about the state variable.

One difference between Gennotte's framework and our own is that while investors in our economy eventually learn the true value of β, Gennotte's investors never completely learn the ever-changing state variable. However, even this difference disappears if we extend our model to allow for an unknown and changing regression coefficient β_t.

namely the dividend yield, but also by variables summarizing the investor's beliefs about the parameters $\theta = (a, B, \Sigma)$. Adding in these variables, which might include estimated means and variances for the posterior distributions of a, B, and Σ, dramatically increases the size of the state space, making the dynamic programming problem difficult to solve.

In spite of these difficulties, we are able to make progress. For the simple i.i.d. model of Section II, we are able to present a full analysis of the effects of estimation risk, including the issue of learning. We leave this to Section IV.C.1. For the remainder of this section, we keep the more general model which allows for predictability, and analyze estimation risk with some simplifying assumptions. The simplification we make is to suppose that while the investor acknowledges that he is uncertain about model parameters, he ignores the impact on today's optimal allocation of the fact that his beliefs about those parameters may change. In other words, he solves the dynamic problem assuming that his beliefs about the parameters remain the same as they are at the start of his investment horizon. These beliefs are summarized by the posterior distribution calculated conditional only on data up until the start of his horizon.

As a result of the simplification we make, the investor's opportunity set is still described by the dividend yield alone. Hence we can still use equation (33) to calculate the value function. Since the investor now accounts for parameter uncertainty, the expectation E_{t_k} is taken over the predictive distribution $p(R_{k+1}, x_{k+1}|x_k)$ rather than over $p(R_{k+1}, x_{k+1}|\theta, x_k)$. The investor constructs a sample from the predictive distribution by taking a large sample of 1,000,000 draws from the posterior distribution $p(\theta|z_1, \ldots z_T)$ - conditional *only* on data up until the horizon start date - and then for each set of parameter values drawn, makes a draw from $p(R_{k+1}, x_{k+1}|\theta, x_k)$, a Normal distribution.[20]

The graphs on the right side of Figure 5 show the results of this procedure. Recall that the graphs on the left are for the case where parameter uncertainty is ignored; presenting the graphs side by side makes the contrast more striking. Figure 5 demonstrates clearly that when the uncertainty in the predictive power of the dividend yield is taken into account, the allocation lines are much flatter and the hedging demands much smaller. Once the investor acknowledges this uncertainty, he becomes more

[20]One possible objection to this experiment is that it is *too* simple, in that the investor must at least recognize that the precision of his parameter estimates will improve over time, while we are keeping the dispersion of the posterior fixed in our calculations. One way to justify this is to think of it in the context of a more sophisticated model which allows the true parameter to change over time. In that case, it is no longer true that the posterior becomes tighter as more data is received; it may, in fact become more dispersed.

skeptical about whether the investment opportunity set really is changing over time, and hence about whether he really should invest more heavily in stocks as a hedge.

Another result that stands out from the side-by-side comparison of the graphs in Figure 5 is the reduced sensitivity of the optimal allocation to the initial value of the dividend yield, once we account for parameter uncertainty. An investor who takes the model parameters as known reacts strongly to changes in the dividend yield, leading to a wildly gyrating allocation to stocks over time; as the dividend yield falls for instance, the investor needs to reduce his allocation to stocks dramatically, according to the graphs on the left. Once we acknowledge that the predictive power of the dividend yield is weak, the optimal allocation becomes less sensitive to the predictor, particularly at long horizons. Changes in portfolio composition occur more gradually over time.

These results suggest that analyses of dynamic strategies which ignore estimation risk, such as those in Brennan, Schwartz, and Lagnado (1997) and Campbell and Viceira (1997) may need to be interpreted with some caution. Since they give the investor no way of incorporating parameter uncertainty into the decision framework, they may recommend allocations to stocks that are too high, and too sensitive to the variables parameterizing expected returns.

C.1. Learning

One feature of parameter uncertainty in a dynamic context is that the degree of uncertainty about the parameters changes over time as more data is received. If the investor anticipates this learning, it may affect his portfolio holdings. After all, the investor's beliefs about the model parameters determine the perceived investment opportunity set; as these beliefs change, the opportunity set also changes, and there may be a demand for stocks based on hedging against such changes.

In this section, we demonstrate that learning can indeed affect the investor's decisions. To keep the problem tractable, we confine our analysis to the simple i.i.d. model of Section II. However, the results are also relevant for investors using the more general VAR model of Sections III and IV.[21]

Suppose then that continuously compounded excess stock index returns r_t are described by equation (11). The continuously compounded real monthly T-Bill return r_f, is set equal to 0.0036 as usual. The numbers in Table I suggest that the main source

[21]Brennan (1997) also studies learning and portfolio choice in a similar context to ours, although he uses a methodology quite different from the discrete time backward induction we employ here. He works in continuous time and obtains optimal portfolios by applying optimal filtering theory and then solving a partial differential equation.

of parameter uncertainty facing the investor revolves around the mean excess stock return μ. To simplify even further, we focus exclusively on this source of uncertainty by assuming that σ^2 is known and equal to 0.0019.[22]

Suppose we are at time T. The investor has a horizon of $\widehat{T}$ months, and rebalances his portfolio every year. How should he allocate his portfolio today, given the uncertainty about the parameter μ, and recognizing that he will update his beliefs about μ every year?

Suppose that before observing any data, the investor's prior beliefs about the parameter μ are diffuse, $p(\mu) \propto 1$. After observing data $r = (r_1, \ldots r_t)$, the investor's posterior beliefs about the mean stock return are

$$\begin{aligned} \mu | r &\sim N(m_t, V_t) \\ m_t &= \frac{1}{t} \sum_{\tau=1}^{t} r_\tau && (35) \\ V_t &= \frac{\sigma^2}{t}. && (36) \end{aligned}$$

The two variables m_t and V_t summarize the investment opportunity set perceived by the investor. However, V_t is a deterministic function of time, so the state space at time t is fully described by wealth W_t, and m_t, the investor's beliefs about the mean excess stock return. Using the same framework and notation as in Section IV.A, define

$$J(W_k, m_k, t_k) = \max_{t_k} E_{t_k} \left(\frac{W_K^{1-A}}{1-A} \right). \quad (37)$$

The Bellman equation is

$$J(W_k, m_k, t_k) = \max_{\omega_k} E_{t_k} \{ J(W_{k+1}, m_{k+1}, t_{k+1}) \}, \quad (38)$$

and an induction argument shows that we can write

$$J(W_k, m_k, t_k) = \frac{W_k^{1-A}}{1-A} Q(m_k, t_k), \quad (39)$$

for $A \neq 1$, so that the Bellman equation becomes

$$Q(m_k, t_k) = \max_{\omega_k} E_{t_k} \{ [(1-\omega_k) \exp(r_f \frac{\widehat{T}}{K}) + \omega_k \exp(r_f \frac{\widehat{T}}{K} + R_{k+1})]^{1-A} Q(m_{k+1}, t_{k+1}) \}. \quad (40)$$

[22]This is the mean of the posterior distribution for σ^2 calculated using the subsample of data from 1986 to 1995. It is taken from Table I.

The expectation here is taken over $p(R_{k+1}|m_k)$. Since

$$p(R_{k+1}|m_k) = \int p(R_{k+1}|\mu, m_k) p(\mu|m_k)\ d\mu, \tag{41}$$

where

$$\begin{aligned} R_{k+1}|\mu, m_k &\sim N(\mu \frac{\widehat{T}}{K}, \sigma^2 \frac{\widehat{T}}{K}) \\ \mu|m_k &\sim N(m_k, \frac{\sigma^2}{T + (\frac{\widehat{T}}{K})k}), \end{aligned}$$

we obtain after some algebra,

$$R_{k+1}|m_k \sim N(m_k \frac{\widehat{T}}{K}, \sigma^2 (\frac{\widehat{T}}{K} + \frac{(\frac{\widehat{T}}{K})^2}{T + (\frac{\widehat{T}}{K})k})).$$

To solve the Bellman equation, we discretize the state space and use backward induction, in exactly the same way as in Section IV.A. We let the variable m_k take 25 discrete values, ranging from zero to 0.013 at equally spaced intervals, which we call m_k^j, $j = 1, \ldots, 25$.[23] For each m_k^j, $j = 1, \ldots, 25$, we draw a large sample $(R_{k+1}^{(i)})_{i=1}^{I=1{,}000{,}000}$ from $p(R_{k+1}|m_k)$, and set $Q(m_k^j, t_k)$ equal to

$$\max_{\omega_k} \frac{1}{I} \sum_{i=1}^{I} [(1 - \omega_k) \exp(r_f \frac{\widehat{T}}{K}) + \omega_k \exp(r_f \frac{\widehat{T}}{K} + R_{k+1}^{(i)})]^{1-A} Q(m_{k+1}^{(i)}, t_{k+1}), \tag{42}$$

where $m_{k+1}^{(i)}$ is the investor's updated belief about the mean excess stock return having observed the new data $R_{k+1}^{(i)}$, and is given by

$$\begin{aligned} m_{k+1}^{(i)} &= \frac{1}{T + (k+1)\frac{\widehat{T}}{K}} \sum_{\tau=1}^{T+(k+1)\frac{\widehat{T}}{K}} r_\tau \\ &= \frac{1}{T + (k+1)\frac{\widehat{T}}{K}} (r_1 + \ldots + r_{T+k(\frac{\widehat{T}}{K})} + R_{k+1}^{(i)}) \\ &= \frac{m_k (T + k(\frac{\widehat{T}}{K})) + R_{k+1}^{(i)}}{T + (k+1)\frac{\widehat{T}}{K}}. \end{aligned} \tag{43}$$

This calculation gives $Q(m_k^j, t_k)$ for all $j = 1, \ldots, 25$. Backward induction through all K rebalancing points eventually gives $Q(m_0^j, t_0)$ and hence the optimal allocations ω_0.

[23] In other words, we let the investor's belief about the mean monthly excess stock return m_k lie between zero percent and 1.3 percent. The mean excess return in the sample, which from Table I is 0.0065, lies at the midpoint of this range.

The solid lines in the graphs in Figure 6 present the optimal allocations for investors with horizons ranging from one to ten years, and who rebalance annually. These investors take the uncertainty in the mean stock return μ into account, update their beliefs about μ every year, and anticipate learning more about the parameter over time. The dynamic programming framework above produces optimal allocations as a function of the state of the world, m_k; for simplicity, in Figure 6 we present the stock allocation when $m_k = 0.0065$, the historical mean excess monthly stock return in the data, which here is the recent subsample from 1986 to 1995.

The two graphs in Figure 6 correspond to different risk-aversion levels. To make the effect of parameter uncertainty clearer, we have included in each graph a dashed line, which shows the optimal allocation that would be chosen by an investor who ignores the uncertainty in μ, taking its value to be fixed at 0.0065, the posterior mean from Table I. In this case, we know from the theoretical work of Samuelson (1969) that the optimal allocation does not depend on the investor's horizon.

The striking result in Figure 6, also obtained by Brennan (1997), is that the investor who acknowledges the uncertainty in μ will allocate *less* to stocks at longer horizons. This appears very similar to the buy-and-hold results in the two lower graphs from Figure 1, which are based on the same sample period. However, the effect driving them is very different. The decrease in allocation across horizons for a rebalancing investor is due to Merton-type hedging demands. At the risk-aversion levels shown in Figure 6, investors are trying to hedge changes in the state variable m_t that governs their investment opportunities. Since realized stock returns are positively correlated with m_t – if realized stock returns are high, the investor raises his beliefs about the mean stock return m_t – holding less in stocks is a simple way to hedge.

Although we have obtained these results for the i.i.d. model, the lessons also apply to investors using the full VAR model. Those investors also face uncertainty about the mean stock return, and they will also want to allocate less to stocks in anticipation of learning more about that mean.

These results suggest that just as in the buy-and-hold case, there are two effects driving the relative magnitude of short and long horizon equity allocations. On the one hand, an investor with $A > 1$ who believes that there are state variables such as the dividend yield driving expected returns will want to hold more in stocks at long horizons as a hedging demand. On the other hand, the presence of estimation risk and the possibility of learning pushes down the long horizon allocation relative to the short run allocation.

V. Conclusion

The evidence of time-variation in expected returns is among the more intriguing empirical findings in finance. To date, very little has been said about the implications of this feature of asset returns for investors making portfolio decisions.

This paper addresses this question, using the sensitivity of the optimal portfolio allocation to the investor's horizon as a way of thinking about the effects of predictability. Our analysis shows that sensible portfolio allocations for short and long horizon investors can be very different in the context of predictable returns.

Take first the case where the parameters in the model describing asset returns are treated as if known with complete precision. In this case, a buy-and-hold investor invests substantially more in risky equities in the presence of predictability, the longer his horizon. Time variation in expected returns induces mean-reversion in returns, slowing the growth of conditional variances of long horizon returns. This makes equities appear less risky at long horizons, and hence more attractive to the investor. In a dynamic setting with optimal portfolio rebalancing, an investor more risk-averse than a log utility investor would also allocate substantially more to equities, the longer his horizon. In this case, the higher allocation to equities provides the investor with a hedge against changes in available investment opportunities.

Investment advisors often maintain that long run investors should allocate more aggressively to equities. This view finds little support in a world with i.i.d. asset returns, a point made forcefully by Samuelson (1969). The results presented here suggest that time-variation in asset returns may provide a rationale for practitioners' recommendations after all.

This conclusion may be too hasty, however. The investor faces substantial uncertainty about model parameters: in our regression model, both the intercept – an important component of the mean return on stocks – and the coefficient on the state variable are estimated imprecisely.

Our results suggest that portfolio calculations can be seriously misleading if the allocation framework ignores the uncertainty surrounding parameters such as these. When this source of uncertainty is accounted for, long horizon investors in general still allocate more to equities than short horizon investors, but the difference is not as large. In some cases, the estimation risk can be so severe as to make the optimal stock allocation *decrease* with the investor's horizon. Moreover, parameter uncertainty makes the optimal allocation much less sensitive to the current value of the predictor. This suggests that analyses which ignore estimation risk may lead the investor to take positions in stocks which are both too large and too sensitive to the predictor.

We have tried to used the simplest possible structure to illustrate our findings. The framework can be extended though, to examine other issues of interest to investors. We could include more assets, such as long-term government bonds, more predictor variables, and could introduce variation in conditional volatilities as well as conditional means. The model could also allow for time-variation in the parameters which might magnify the effects of estimation risk even further. Finally, the methodology in this paper could easily be amended for use in other contexts. There is much evidence of *cross-sectional* differences in asset returns, based on sorting stocks by size or book-to-market value. An investor wondering how to allocate his portfolio in light of this evidence – and the statistical uncertainty that comes with it – may also benefit from applying the framework in this paper.

Appendix

In this Appendix, we evaluate the accuracy of the numerical methods used to obtain optimal portfolio allocations.

In an effort to ensure a high degree of accuracy, we use a very large sample of 1,000,000 draws from the appropriate distribution when evaluating the integrals for expected utility. However, there is always the possibility that even samples of this size do not guarantee sufficient accuracy, and that using still larger samples would produce different results for the optimal stock allocation.

To be confident that this is not the case, we do the following check on simulation error. We calculate expected utility, and hence the optimal allocation, using 10 independent samples of 10,000 draws each from the sampling distribution. We then repeat the calculation, this time using 10 independent samples of 100,000 draws. Finally, we repeat the procedure one more time, this time using 10 independent samples of 1,000,000 draws, the sample size actually used for the results presented in the paper.

Figure 7 presents the outcome of two such robustness tests. The graph on the left focuses on optimal allocations for an investor with a horizon of 10 years, risk-aversion $A = 5$, and who uses an i.i.d. model estimated over the full sample from 1952 to 1995 in making his decisions. Figure 1 shows that an investor who ignores parameter uncertainty allocates 69 percent to the stock index, while an investor who accounts for this uncertainty only invests 58 percent in stocks. How sure can we be about the accuracy of these numbers?

Figure 7 shows the allocations obtained using the 30 independent samples described above. The first 10 samples contain 10,000 draws each, samples 11 through 20 contain 100,000 each, and samples 21 through 30 contain 1,000,000 each. The vertical dotted lines are intended to help the reader interpret the results by showing where one set of samples ends and the other begins.

Note first that for samples 21 through 30, the optimal allocation obtained is the same in all ten independent samples. In other words, for the sample size actually used in our results, there does not appear to be any significant variation in the recommended portfolio. Since the variance is so low, it is likely that we have converged to the portfolio that would be obtained if we could perform the integrations exactly.

One caveat to this is that the variance of the estimate of the optimal portfolio may fall faster than its bias. While we cannot rule this possibility out completely, the portfolios obtained using samples one through 20 suggest that it is unlikely. Those samples contain fewer draws from the sampling distribution, so naturally there is more variation in the computed optimal portfolios. However, of more interest is the fact that the "distribution" of recommended portfolios seems to be centered at the same point, whether we use 10,000 draws, 100,000 draws or 1,000,000 draws. There does not seem to be much evidence of bias in the estimate of the optimal portfolio.

The graph on the right presents another example, this time focusing on the optimal allocation of an investor with a horizon of 10 years and risk-aversion $A = 10$, who uses the VAR model of Section III. Suppose that this investor uses only the recent subsample of data from 1986 to 1995 to judge the predictability in returns, and that the initial dividend yield is $x_{1,T} = 2.37$ percent, two standard deviations below the average in the sample. Figure 4 shows that if this investor ignores estimation risk, he allocates 53 percent to stocks, but only 20 percent if he incorporates it. From the graph in Figure 7, we see that once again, the recommended portfolio allocation is the same across all samples of 1,000,000 draws, suggesting that we are very close to the exact optimal portfolio.

REFERENCES

Bawa, Vijay, Stephen Brown, and Roger Klein, 1979, *Estimation Risk and Optimal Portfolio Choice* (North Holland, Amsterdam).

Bodie, Zvi, 1995, On the risk of stocks in the long run, *Financial Analysts Journal* 51, 18-22.

Brennan, Michael, 1997, The role of learning in dynamic portfolio decisions, Working Paper 3-97, Anderson School of Management.

Brennan, Michael, Eduardo Schwartz, and Ronald Lagnado, 1997, Strategic asset allocation, *Journal of Economic Dynamics and Control* 21, 1377-1403.

Campbell, John, 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.

Campbell, John, and Robert Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195-227.

Campbell, John, and Robert Shiller, 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661-676.

Campbell, John, 1991, A variance decomposition for stock returns, *The Economic Journal* 101, 157-179.

Campbell, John, and Luis Viceira, 1997, Consumption and portfolio decisions when expected returns are time varying, Working Paper, Harvard University.

Fama, Eugene, and Kenneth French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3-27.

Fama, Eugene, and Kenneth French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23-49.

Frost, Peter, and James Savarino, 1986, An empirical bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293-305.

Gennotte, Gerard, 1986, Optimal portfolio choice under incomplete information, *Journal of Finance* 41, 733-746.

Hodrick, Robert, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357-386.

Jobson, J.D., and Robert Korkie, 1980, Estimation for Markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544-554.

Jorion, Phillipe, 1985, International portfolio diversification with estimation risk, *Journal of Business* 58, 259-278.

Kandel, Shmuel, and Robert Stambaugh, 1987, Long horizon returns and short horizon models, CRSP Working Paper no. 222, University of Chicago.

Kandel, Shmuel, and Robert Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385-424.

Keim, Donald, and Robert Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357-390.

Kim, Tong Suk, and Edward Omberg, 1996, Dynamic nonmyopic portfolio behavior, *Review of Financial Studies* 9, 141-161.

Merton, Robert, 1969, Lifetime portfolio selection: The continuous-time case, *Review of Economics and Statistics* 51, 247-257.

Merton, Robert, 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.

Merton, Robert, 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323-361.

Samuelson, Paul, 1969, Lifetime portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics* 51, 239-246.

Samuelson, Paul, 1994, The long-term case for equities, *Journal of Portfolio Management,* Fall 1994, 15-24.

Siegel, Jeremy, 1994, *Stocks for the Long Run* (Richard D. Irwin, Burr Ridge, IL).

Stambaugh, Robert, 1998, Predictive regressions, Working Paper, University of Pennsylvania.

Williams, John, 1977, Capital asset prices with heterogeneous beliefs, *Journal of Financial Economics* 5, 219-239.

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley and Sons, New York).

Zellner, Arnold, and V. Karrupan Chetty, 1965, Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association* 60, 608-616.

Table I

Parameter estimates for an i.i.d. model of stock returns

The results in this table are based on the model $r_t = \mu + \epsilon_t$, where r_t is the continuously compounded excess stock index return in month t and $\epsilon_t \sim$ i.i.d. $N(0, \sigma^2)$. The table gives the mean and standard deviation (in parentheses) of each parameter's posterior distribution. The left panel uses data from June 1952 to December 1995; the right panel uses data from January 1986 to December 1995.

1952-1995		1986-1995	
μ	σ^2	μ	σ^2
0.0050	0.0017	0.0065	0.0019
(0.0018)	(0.0001)	(0.0039)	(0.0003)

Table II

Parameter estimates for a VAR model of stock returns

The results in this table are based on the model $z_t = a + Bx_{t-1} + \epsilon_t$, where $z_t = (r_t \; x_t)'$ includes continuously compounded monthly excess stock returns r_t and the dividend yield x_t and where $\epsilon_t \sim$ i.i.d $N(0, \Sigma)$. The table gives the mean and standard deviation (in parentheses) of each parameter's posterior distribution. The figures in bold above the diagonal in the variance matrices are correlations. The left panel uses data from June 1952 to December 1995; the right panel uses data from January 1986 to December 1995.

1952-1995		1986-1995	
a	B	a	B
-0.0143	0.5118	-0.0303	1.0919
(0.0081)	(0.2129)	(0.0281)	(0.8265)
0.0008	0.9774	0.0013	0.9577
(0.0003)	(0.0091)	(0.0010)	(0.0305)
Σ		Σ	
0.0017	**-0.9351**	0.0019	**-0.9323**
(0.0001)	(0.0055)	(0.0003)	(0.0122)
	3.0E-6		2.6E-6
	(1.9E-7)		(3.4E-7)

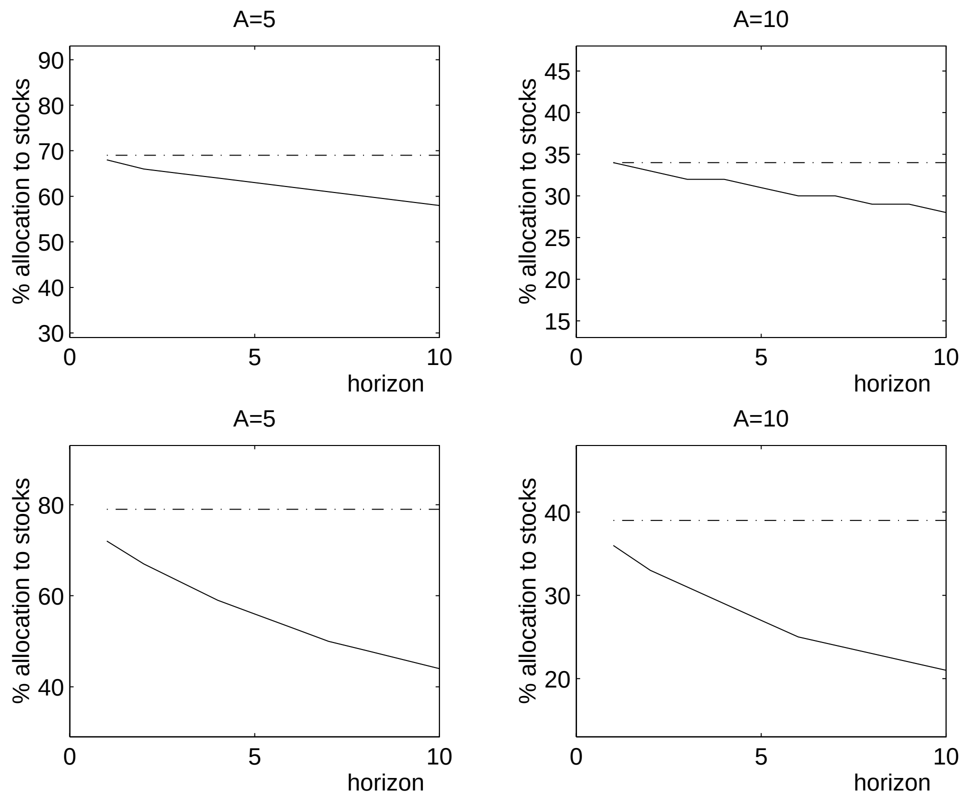

Figure 1. Optimal allocation to stocks plotted against the investment horizon in years. The investor follows a buy-and-hold strategy, uses an i.i.d model for asset returns and has power utility $\frac{W^{1-A}}{1-A}$ over terminal wealth. The dash/dot line corresponds to the case where the investor ignores parameter uncertainty, the solid line to the case where he accounts for it. The top two graphs use data from 1952-1995, while the lower two use data from 1986-1995.

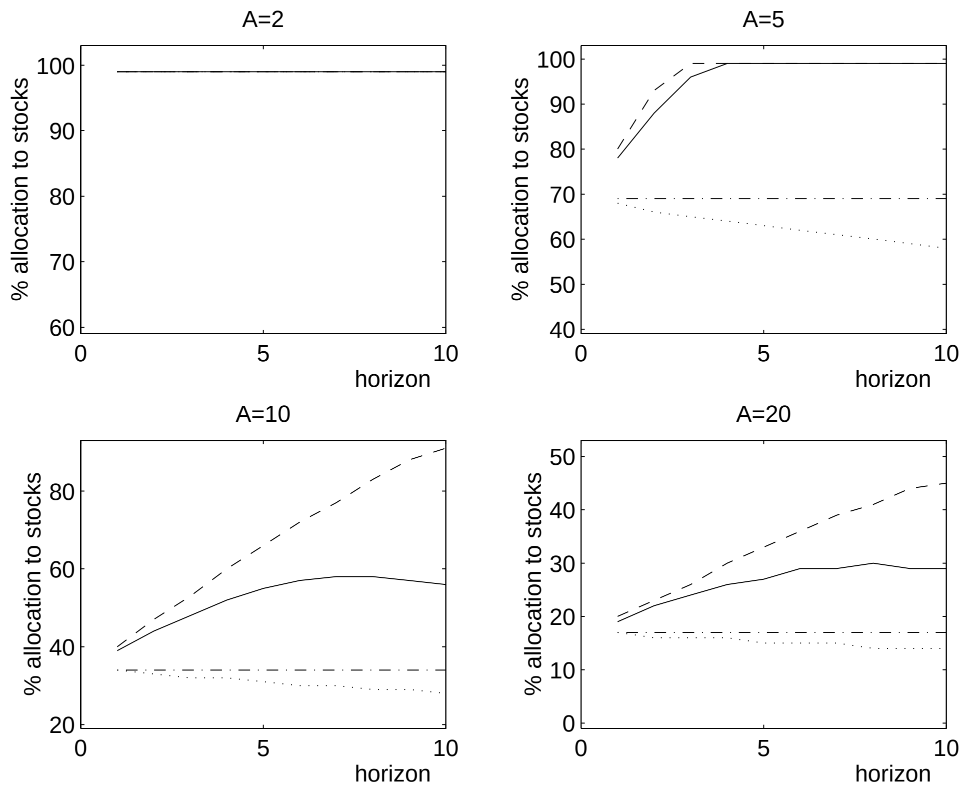

Figure 2. Optimal allocation to stocks plotted against the investment horizon in years. The investor follows a buy-and-hold strategy, uses a VAR model which allows for predictability in returns, and has power utility $\frac{W^{1-A}}{1-A}$ over terminal wealth. The solid and dotted lines correspond to cases where the investor accounts for uncertainty in the parameters, the dashed and dash/dot lines to cases where he ignores it. The solid and dashed lines correspond to cases where the investor takes into account the predictability in returns, the dotted and dash/dot lines to cases where he ignores it. The model is estimated over the 1952-1995 sample period.

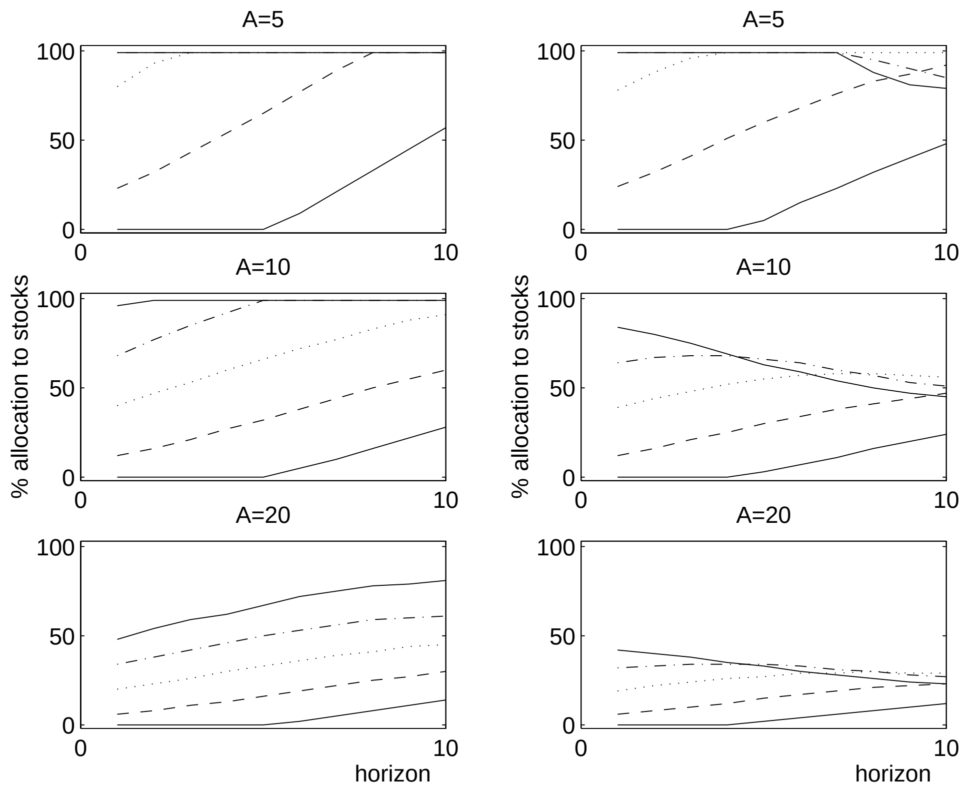

Figure 3. Optimal allocation to stocks plotted against the investment horizon in years. The investor follows a buy-and-hold strategy, uses a VAR model which allows for predictability in returns, and has power utility $\frac{W^{1-A}}{1-A}$ over terminal wealth. The three graphs on the left ignore parameter uncertainty, whereas those on the right account for it. The five lines within each graph correspond to different initial values of the predictor variable, the dividend yield: d/p = 2.06 percent (solid), d/p = 2.91 percent (dashed), d/p = 3.75 percent (dotted), d/p = 4.59 percent (dash/dot), and d/p = 5.43 percent (solid). The model is estimated over the 1952-1995 sample period.

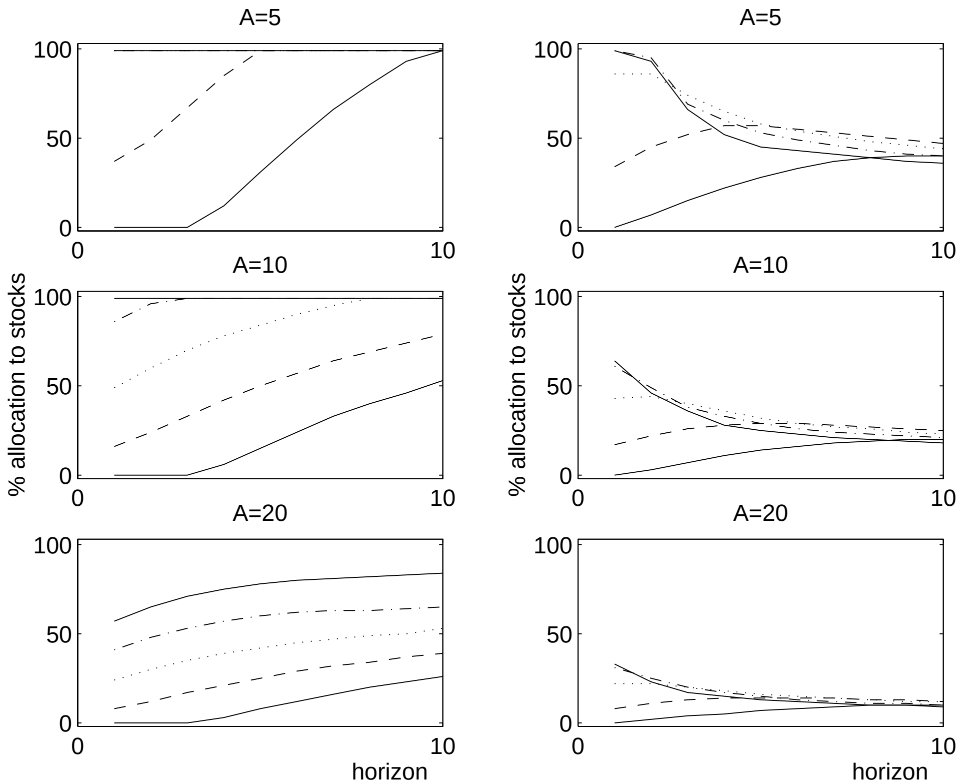

Figure 4. Optimal allocation to stocks plotted against the investment horizon in years. The investor follows a buy-and-hold strategy, uses a VAR model which allows for predictability in returns, and has power utility $\frac{W^{1-A}}{1-A}$ over terminal wealth. The three graphs on the left ignore parameter uncertainty, whereas those on the right account for it. The five lines within each graph correspond to different initial values of the predictor variable, the dividend yield: d/p = 2.37 percent (solid), d/p = 2.86 percent (dashed), d/p = 3.36 percent (dotted), d/p = 3.85 percent (dash/dot), and d/p = 4.35 percent (solid). The model is estimated over the 1986-1995 sample period.

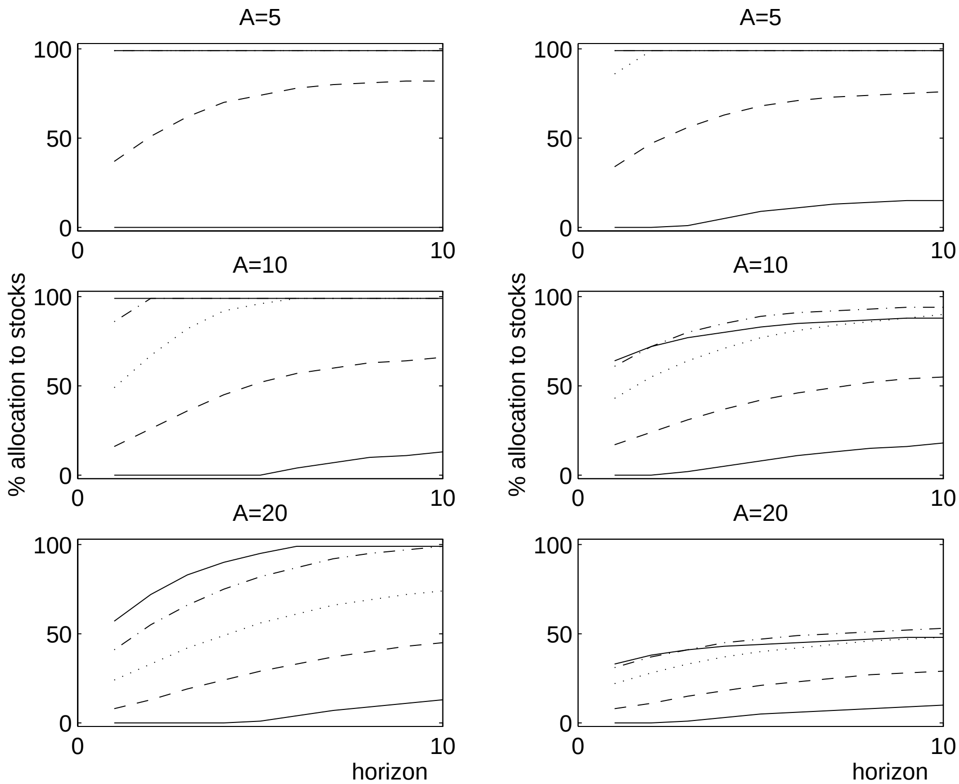

Figure 5. Optimal allocation to stocks plotted against the investment horizon in years. The investor rebalances optimally once a year, uses a VAR model which allows for predictability in returns, and has power utility $\frac{W^{1-A}}{1-A}$ over terminal wealth. The three graphs on the left ignore parameter uncertainty, whereas those on the right account for it. The five lines within each graph correspond to different initial values of the predictor variable, the dividend yield: d/p = 2.37 percent (solid), d/p = 2.86 percent (dashed), d/p = 3.36 percent (dotted), d/p = 3.85 percent (dash/dot), and d/p = 4.35 percent (solid). The model is estimated over the 1986-1995 sample period.

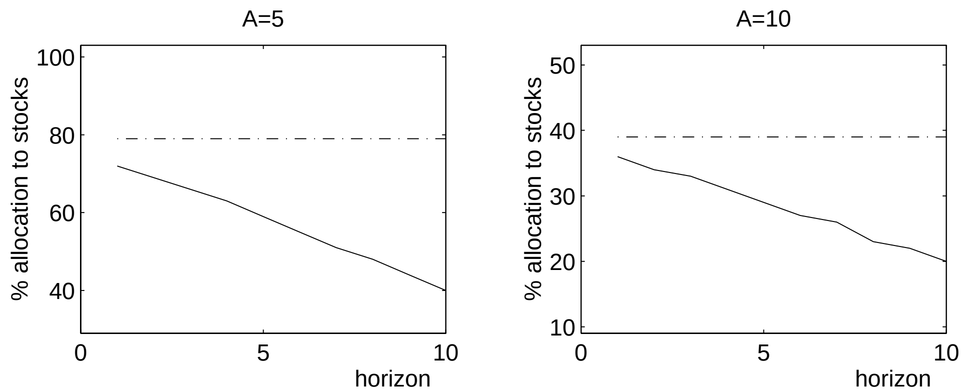

Figure 6. Optimal allocation to stocks plotted against the investment horizon in years. The investor rebalances optimally once a year, uses an i.i.d model for asset returns and has power utility $\frac{W^{1-A}}{1-A}$ over terminal wealth. The dash/dot line corresponds to the case where the investor ignores parameter uncertainty, the solid line to the case where he accounts for it. The model is estimated over the 1986-1995 sample period.

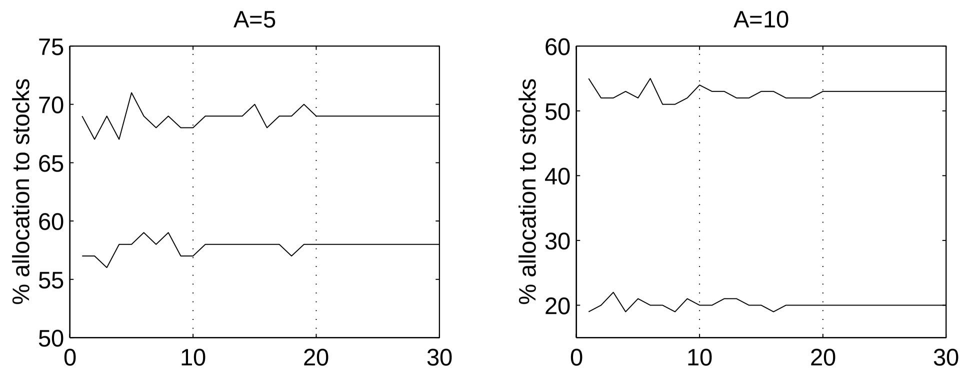

Figure 7. Computed optimal allocation for 30 independent samples from the distribution for future returns. Samples one through 10 contain 10,000 draws, samples 11 through 20 contain 100,000 draws, and samples 21 through 30 contain 1,000,000 draws. The graph on the left is for a buy-and-hold investor with risk-aversion $A = 5$, a horizon of 10 years, and who uses an i.i.d. model for returns, estimated over the 1952-1995 period. The graph on the right is for a buy-and-hold investor with risk-aversion $A = 10$, a horizon of 10 years, and who uses the VAR model for returns, estimated over the 1986-1995 period. The initial value of the dividend yield is 2.37 percent. The upper line ignores estimation risk, while the lower line accounts for it.