Asset Allocation

Jessica Wachter*

University of Pennsylvania and NBER

April 30, 2010

Abstract

This review article describes recent literature on asset allocation, covering both static and dynamic models. The article focuses on the bond–stock decision and on the implications of return predictability. In the static setting, investors are assumed to be Bayesian, and the role of various prior beliefs and specifications of the likelihood are explored. In the dynamic setting, recursive utility is assumed, and attention is paid to obtaining analytical results when possible. Results under both full and limited-information assumptions are discussed.

Key words: Portfolio choice, Predictive regression, Recursive utility

*Department of Finance, The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104; Tel: (215) 898-7634; Email: `jwachter@wharton.upenn.edu`. I thank Itamar Drechsler, Lubos Pastor and Moto Yogo for helpful comments and Jerry Tsai for excellent research assistance.

1 Introduction

The study of portfolio allocation has played a central role in financial economics, from its very beginnings as a discipline.[1] It is not difficult to see why the area has attracted (and continues to attract) the attention that it does: As a field of study it is both highly practical and apparently amenable to the application of sophisticated mathematics.

This study will review the recent academic literature on asset allocation. Two important simplifications will be employed: First, the field has drawn a distinction between the study of allocation to broad asset classes and allocation to individual assets within a class. This article focuses on the former. In fact, the empirical applications in this article assume an even more specific case, namely investor the chooses between a broad stock portfolio and a riskless asset. Second, the surveyed models will assume, for the most part, no financial frictions. That is, I assume that the investor does not face unhedgeable labor income risk or barriers to trading in the assets, such as leverage or short-sale constraints. This is not to deny the importance of other asset classes, or of financial frictions. Recent surveys on portfolio choice (encompassing portfolios of many assets) include Cochrane (1999), Brandt (2009) and Avramov and Zhou (this issue). Campbell (2006) and Curcuru, Heaton, Lucas, and Moore (2009) survey work on asset allocation under realistic frictions faced by households.

I will focus on two broad classes of models: static models (in which the investor looks one period ahead) and dynamic models (in which the investor looks multiple periods ahead and takes his future behavior into account when making decisions). For static models, the solution where investors have full information about asset returns has been known for some time (Markowitz (1952)), so the focus will be on incorporating uncertainty about the return process. In contrast, much has been learned in recent years about dynamic models, even in the full-information case. A barrier to considering dynamic models is often their complexity: for this reason I will devote space to analytical results. These results, besides being interesting in their own right, can serve as a starting point for understanding the

[1] See, Bernstein (1992) for discussion of the origins of financial economics as an academic discipline and its early focus on the asset allocation question.

behavior of models that can only be solved numerically.

Finally, in both the static and dynamic sections, I will consider in detail the model where excess returns on stocks over short-term Treasury bills are in part predictable. A substantial empirical literature devotes itself to the question of whether returns are predictable; the asset allocation consequences of such predictability being striking and well-known in at least a qualitative sense since Graham and Dodd (1934).

Ultimately, the goal of academic work on asset allocation is the conversion of the time series of observable returns and other variables of interest into a single number: Given the preferences and horizon of the investor, what fraction of her wealth should she put in stock? The aim is to answer this question in a "scientific" way, namely by clearly specifying the assumptions underlying the method and developing based on these assumptions a consistent theory. The very specificity of the assumptions and the resulting advice can seem dangerous, imputing more certainty to the models than the researcher can possibly possess. Yet only by being so highly specific does the theory turn into something that can be clearly debated and ultimately refuted in favor of an equally specific and hopefully better theory. This development implies the use of mathematics to model the investment decision; the reader is encouraged to remember throughout that the subject of the modeling is an individual or household making a decision with significant consequences for lifelong financial security.

2 Static models

2.1 The basic decision problem

In this first section, I consider the problem of an investor maximizing wealth as of time T by allocating wealth between a risky and a riskless asset. The portfolio decision takes place at a time $\hat{T} < T$. Let R_{t+1} denote the simple return on the risky asset between times t and $t+1$, and $R_{f,t+1}$ the simple return on the riskless asset between t and $t+1$. Let W_t denote the investor's wealth at time t. The investor solves

$$\max_z E_{\hat{T}} \left[\frac{W_T^{1-\gamma}}{1-\gamma} \right], \tag{1}$$

where

$$W_T = W_{\hat{T}} \left(z \prod_{s=\hat{T}+1}^{T} R_s + (1-z) \prod_{s=\hat{T}+1}^{T} R_{f,s} \right). \tag{2}$$

The parameter γ is assumed to be positive, and $\gamma = 1$ should be interpreted as logarithmic utility. Note that the investor described above decides on the allocation z at time $\hat{T}$ and then does not trade. This is a *buy-and-hold* investor. The implicit weight on the risky security can, and almost certainly will change over time, however, the problem written as above assumes that the investor does not rebalance back to the original weights. For now, it is assumed that z can take on any value: short-sales and borrowing at the riskfree rate are allowed. For the purposes of solving the model, I will assume R_{fs}, for s between $\hat{T}+1$ and T, is known to the investor at time $\hat{T}$. Following much of the literature, I assume that the investor's utility takes a power form, implying that relative risk aversion is constant and that asset allocation does not depend on wealth. The scale-invariance of power utility has broad empirical support in that interest rates have remained stationary despite the fact that wealth has grown.

Define the continuously-compounded excess return to be

$$y_t = \log R_t - \log R_{ft}$$

and assume y_t follows the process

$$y_{t+1} = \alpha + \beta x_t + u_{t+1} \tag{3}$$

$$x_{t+1} = \theta + \rho x_t + v_{t+1}, \tag{4}$$

where

$$\begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix} | \, y_t, \ldots, y_1, x_t, \ldots, x_0 \sim N(0, \Sigma), \tag{5}$$

and

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}. \tag{6}$$

That is, y_{t+1} has a predictable component x_t that follows a first-order autoregressive process. The errors are assumed to be serially uncorrelated, homoskedastic and jointly normally distributed. A substantial and long-standing empirical literature documents predictability in excess returns, in the sense that running regression (3) for observable x_t generates statistically significant coefficients.[2] In what follows, I focus on the case where x_t is the dividend yield because the dividend yield and future expected excess returns are linked through a present value identity (Campbell and Shiller (1988)). Theory suggests, therefore, that if returns are predictable, the dividend yield should capture at least some of that predictability. The general setting that I have laid out here largely follows that of Barberis (2000): it will differ in some important respects however.

I assume the investor does not know the parameters of the system above. Rather, he is a Bayesian, meaning that he has prior beliefs on the parameters and, after viewing the data, makes inferences based on the laws of probability (Berger (1985)). Let $\hat{\beta}$ be the ordinary least squares (OLS) estimate of β in the regression (3). Bayesian analysis turns the standard frequentist analysis on its head: Instead of asking for the distribution of the test statistic $\hat{\beta}$ (which depends on the data) as a function of the true parameter β, Bayesian analysis asks for the distribution of the true parameter β as a function of the data (which often comes down to a function of sufficient statistics, like $\hat{\beta}$).

For notational convenience, stack the coefficients from (3) and (4) into a vector:

$$b = [\alpha, \beta, \theta, \rho]^\top.$$

The investor starts out with prior beliefs $p(b, \Sigma)$. Let $L(D|b, \Sigma)$ denote the likelihood function, where D is the data available up until and including time $\hat{T}$. It follows from Bayes'

[2] See, for example, Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell and Shiller (1988), Fama and French (1989), Cochrane (1992), Goetzmann and Jorion (1993), Hodrick (1992), Kothari and Shanken (1997), Lettau and Ludvigson (2001), Lewellen (2004), Ang and Bekaert (2007), Boudoukh, Michaely, Richardson, and Roberts (2007).

rule that the posterior distribution is given by

$$p(b, \Sigma|D) = \frac{L(D|b, \Sigma)p(b, \Sigma)}{p(D)},$$

where $p(D)$ is an unconditional likelihood of the data in the sense that

$$p(D) = \int_{b,\Sigma} L(D|b, \Sigma)p(b, \Sigma)\, dbd\Sigma,$$

namely that $p(D)$ integrates out b and Σ. It follows that

$$p(b, \Sigma|D) \propto L(D|b, \Sigma)p(b, \Sigma), \tag{7}$$

where $\propto$ denotes "proportional to", because $p(D)$ does not depend on b or Σ. The likelihood function, given $\hat{T}$ years of data is equal to

$$L(D|\Sigma, b) = \prod_{t=0}^{\hat{T}-1} p_{t+1|t}(y_{t+1}, x_{t+1}|x_t, \Sigma, b)p_0(x_0|b, \Sigma). \tag{8}$$

where $p_{t+1|t}(y_{t+1}, x_{t+1}|x_t, \Sigma, b)$ is given by a bivariate normal density function as described in (3–6), and $p_0(x_0|b, \Sigma)$ gives the initial condition of the time series. Given the posterior, the predictive density for returns from time $\hat{T}$ to T is defined as:

$$p(y_{\hat{T}+1}, \ldots y_T|D) = \int p(y_{\hat{T}+1}, \ldots, y_T|b, \Sigma, x_{\hat{T}})p(b, \Sigma|D)\, db\, d\Sigma. \tag{9}$$

The predictive distribution (9) summarizes the agent's beliefs about the return distribution after viewing the data. The expectation in (1) is taken with respect to this distribution.

How might predictability influence an investor's optimal allocation? Kandel and Stambaugh (1996) find that optimal allocation for the single-period case can be approximated by

$$z \approx \frac{1}{\gamma} \frac{E[y_T] + \frac{1}{2}\mathrm{Var}(y_T)}{\mathrm{Var}(y_T)}, \tag{10}$$

where the mean and the variance are taken under the investor's subjective distribution of returns (which is (9) in the Bayesian case). Holding the variance constant, an upward shift in the mean increases the allocation. This is not surprising given that the investor prefers more wealth to less. This approximation is valid only for short horizons and small shocks. However, it is useful as a first step to understanding the portfolio allocation.

2.2 The conditional Bayesian model

The initial condition $p_0(x_0|b,\Sigma)$ in (8) is a bit of a nuisance. Kandel and Stambaugh (1996) show that if this term were to disappear, the system (3–6) would take the form of a classical multivariate regression model (Zellner (1971, Chapter 8)).[3] Assuming the standard non-informative prior for regression:

$$p(b,\Sigma) \propto |\Sigma|^{-3/2}, \tag{11}$$

the posterior distribution for all of the parameters could then be obtained in closed form. Indeed, conditional on Σ, β would be normally distributed around its OLS estimate $\hat{\beta}$.

However, $p_0(x_0|b,\Sigma)$ is there, and something must be done about it. One approach is to let it stay and specify what it should be. I will refer to the resulting set of assumptions and results as the "exact Bayesian model," to be described Section 2.3. Another approach is to make an assumption that makes it irrelevant to the decision problem. The assumption is that x_0 conveys no prior information:

$$p(b,\Sigma|x_0) = p(b,\Sigma). \tag{12}$$

Because the prior is now conditional on x_0, the likelihood can condition on x_0 as well. The posterior is, of course, conditional on x_0 because it is conditional on all the data. That is, the assumption in (12) allows (7) to be replaced by

$$p(b,\Sigma|D) \propto L_c(D|b,\Sigma,x_0)p(b,\Sigma|x_0),$$

where L_c is the likelihood conditional on x_0:

$$L_c(D|\Sigma,b,x_0) = \prod_{t=0}^{T-1} p_{t+1|t}(y_{t+1},x_{t+1}|x_t,\Sigma,b). \tag{13}$$

[3]However, it still would not be a classical regression model. An assumption of classical regression is that the dependent variable is either nonstochastic or independent of the disturbance term u_t at all leads and lags (Zellner (1971, Chapter 3)). As emphasized in Stambaugh (1999), the independence assumption fails in predictive regressions. Under the assumptions of classical regression, Gelman, Carlin, Stern, and Rubin (1996) show that the likelihood function for the regressor is irrelevant to the agent's decision problem (and so, therefore, is the initial condition).

I will refer to this as the "conditional Bayesian model." I comment further on assumption (12) below.

This conditional Bayesian model is used by Barberis (2000) to study asset allocation for buy-and-hold investors. For comparison, Barberis also solves the model in the full-information case, namely when the investor knows the parameters in (3–6). Figure 1 shows the resulting optimal allocation for a risk aversion of 5 at various horizons and values of the dividend yield. The panel on the left assumes the full-information case (parameters are set equal to their posterior means). The panel on the right shows the results from the conditional Bayesian model, thus taking uncertainty about the parameters, otherwise known as estimation risk, into account.

A striking feature of the left panel of Figure 1 is the degree to which portfolio weights respond to changes in the dividend yield. That is, the investor aggressively engages in market timing, with the dividend yield as the signal of how much to allocate to equities. The solid line in each plot corresponds to the optimal allocation to stocks when the dividend yield is at its historical mean. The dash-dotted lines show the optimal allocation when the dividend yield is one standard deviation above and below the historical mean. The dashed lines show the allocation when the dividend yield is two standard deviations above and below the historical mean. For an investor with a one-year horizon, the optimal allocation to is about 80% when the dividend yield is at its mean. When the dividend yield is at one or two standard deviations above the long-run mean, the investor has all of her wealth in stocks. When the dividend yield is at one standard deviations below the long-run mean, the optimal allocation falls to 20%. The optimal allocation is bounded above and below because the power utility investor would never risk wealth below zero. Because the distribution for returns is unbounded from above, this investor would never hold a negative position in stock. The investor would also never hold a levered position in stock, for this too implies a positive probability of negative wealth, since returns could be as low as -100%. Note that these endogenous bounds on the optimal portfolio are an example of errors in the approximation (10), which contains no such bounds.

What about the case where the investor incorporates estimation risk into her decision making? One might think that estimation risk would make a substantial difference, because (as Barberis (2000) reports), the evidence for predictability at a monthly horizon is only borderline significant in the relevant sample. However, the results incorporating estimation risk are quite similar to those that do not at short horizons. Indeed, differences start to become noticeable only at buy-and-hold horizons of five years or more.[4] For the statistical model for stock returns above, parameter uncertainty resulting from the regression is small compared to the measured uncertainty of holding stocks.[5]

Figure 1 also reveals that the optimal allocation is increasing in the horizon in the case of full information, and for all but the longest horizons in the parameter uncertainty case (see Stambaugh (1999) for an explanation of the reversed relation between holdings and the dividend yield at the longest horizons). Because innovations to the dividend yield are negatively correlated with innovations to returns, stocks when measured at long horizons are less risky than stocks measured at short horizons (mean reversion in stock returns is pointed out in earlier work of Poterba and Summers (1988)). Regardless of whether parameter uncertainty is taken into account, the longer the horizon, the greater the allocation to stocks should be. The implications of mean-reversion for long-horizon investors is also the subject of Siegel (1994).

[4]While such long buy-and-hold horizons may characterize the behavior of some investors, from the normative perspective of this article, such infrequent trading seems extreme.

[5]Note that the effect of the dividend yield attenuates at longer horizons both when parameter uncertainty is taken into account and when it is not. This occurs because the dividend yield is mean-reverting and because the investor cannot rebalance as the dividend yield reverts to its mean. In the limit, an investor with an infinite horizon (who cares about wealth at the end of the horizon) would care only about the unconditional distribution of returns and not about the current value of the dividend yield. Because the dividend yield is so persistent, the effect attenuates very slowly as a function of the horizon.

2.3 The exact Bayesian model

The results above implicitly assume (12), namely that prior beliefs are independent of the initial observation x_0. This assumption enables the use of the conditional likelihood, which, combined with prior (11), leads to closed-form expressions for the posterior distributions of the parameters. While (12) is convenient, how realistic is it? Under (12), the agent believes that the value of x_0 conveys no information about the parameters of the process for x or y. For instance, the initial value of the dividend yield would tell you nothing about, say, the average dividend yield. There is nothing mathematically incorrect about specifying such prior beliefs; the agent can in principal believe anything as long as it does not entail a logical inconsistency or require a peek ahead at the data. However, such beliefs do not seem very reasonable.[6]

The question of whether to allow the initial condition appears to be of a technical nature, but it turns out to have unexpectedly deep implications for Bayesian estimation and for the portfolio allocation decision. These implications are the subject of Stambaugh (1999), which also addresses frequentist properties of predictive regressions; namely that OLS estimation of the coefficient β implies results that are upward-biased (i.e. biased toward finding more evidence of predictability than is actually there). As mentioned above, the posterior mean of β implied by the conditional Bayesian model is also given by the OLS estimate $\hat{\beta}$, and thus is also biased, despite the fact that Bayesian estimation explicitly takes the finite-sample properties of the regression into account. The bias in the posterior mean of β is perhaps an indication that all is not right with this model.

Without (12), the conditional likelihood (13) is no longer correct and the so-called exact likelihood (8) must be used.[7] Immediately, a decision must be made about the distribution of the initial observation x_0. Stambaugh (1999) assumes that x_0 is drawn from the stationary distribution of (4). If (and only if) ρ is between -1 and 1, this stationary distribution exists

[6]For instance, the logic of these beliefs would allow the agent to exclude an arbitrary amount of the data from consideration, just by making the prior parameters independent of these data.

[7]The source for this terminology appears to be Box, Jenkins, and Reinsel (1994, Chapter 7)).

and is given by

$$x_0 \sim N\left(\frac{\theta}{1-\rho}, \frac{\sigma_v^2}{1-\rho^2}\right) \tag{14}$$

(Hamilton (1994, p. 53)). The relevant likelihood function is therefore (8), where p_0 is the normal density given by (14).

Under choices (8) and (14) for the likelihood function, the assumption (11) on the prior is no longer possible. This is because (11) allows ρ to take on any value between minus and plus infinity, whereas (14) is only defined for ρ between -1 and 1. Stambaugh (1999) therefore considers the prior

$$p(b, \Sigma) \propto |\Sigma|^{-3/2}, \rho \in (-1, 1) \tag{15}$$

as well as an alternative prior specification

$$p(b, \Sigma) \propto (1-\rho^2)^{-1}|\Sigma|^{-5/2}, \rho \in (-1, 1). \tag{16}$$

What is the rationale for (16), or for that matter, for (15) or (11)? The prior (11) is standard in regression models. Its appeal is best understood by the fact that it embodies three conditions: first, that b and Σ should be independent in the prior; second, for the elements of b, ignorance is best represented by a uniform distribution (which, in the limit, becomes a constant as in (11)); and third, that

$$p(\Sigma) \propto |\Sigma|^{-3/2}, \tag{17}$$

which generalizes the assumption that, for a single system, the log of the standard deviation should have a flat distribution on $-\infty$ and ∞. Jeffreys (1961, p. 48) proposes these rules for cases where there is no theoretical guidance on the values of the parameters. An additional appeal of (11) discussed above, is that when combined with the likelihood (13), explicit expressions for the posterior distributions of the parameters can be obtained.

This discussion would seem to favor prior (15) (because theory now requires a stationary process) in combination with the exact likelihood. However, applying these rules does not constitute the only approach. Jeffreys (1961) proposed an alternative means of defining ignorance: that inference should be invariant to one-to-one changes in the parameter space. This criterion is appealing in the case of the predictive model (3–6), in which the particular parametrization appear arbitrary. The exact form of Jeffreys prior depends on the sample size T and is derived by Uhlig (1994). Stambaugh (1999) derives an approximate Jeffreys prior that becomes exact as the sample size approaches infinity. This approximate Jeffreys prior is equal to (16). Relative to the flat prior for ρ, (15), more weight is placed on values of ρ close to -1 and 1.

Tables 1 shows the implications of these specification choices for the posterior mean of the regressive coefficient β and the autocorrelation ρ.[8] For the conditional likelihood and prior (11), the posterior mean of beta equals the OLS regression coefficient (which is known to be biased upward). When values of ρ are restricted to be between -1 and 1, the posterior mean of β is slightly *higher*. On the other hand, when the exact likelihood is used, the posterior mean of β is lower and the difference is substantial, regardless of whether the uniform prior or the Jeffreys prior is used.

To understand these differences in posterior means, consider the following approximate relation (Stambaugh (1999)):

$$E[\beta|D] \approx \hat{\beta} + E\left[\frac{\sigma_{uv}}{\sigma_v^2}|D\right](E[\rho|D] - \hat{\rho}). \tag{18}$$

Because σ_{uv} is negative, positive differences between the posterior mean of ρ and $\hat{\rho}$ translate into negative differences between β and $\hat{\beta}$. Equation (18) is the Bayesian version of the observation that the upward bias in $\hat{\beta}$ originates from the downward bias in $\hat{\rho}$. OLS estimates the persistence to be lower than what it is in population (this bias arises from the need to estimate both the sample mean and the regression coefficient at the same time; the observations revert more quickly to the sample estimate of the mean than the true mean (Andrews (1993))). Because of the negative correlation, OLS also estimates the predictive

[8]The specifications involving the exact likelihood or the Jeffreys prior do not admit closed-form solutions for the posterior distribution. Nonetheless, the posterior can be constructed using the Metropolis-Hastings algorithm (see Chib and Greenberg (1995, Section 5)). See Johannes and Polson (2006) for further discussion of sampling methods for solving Bayesian portfolio choice problems.

coefficient to be too high.[9]

Compared with the uniform prior over $-\infty$ to ∞, the prior that restricts ρ to be between -1 and 1 lowers (slightly) the posterior mean of ρ because it rules out draws of ρ that are greater than one. For this reason it raises (slightly) the posterior mean of β even above the OLS value. This result is analogous to the fact that imposing stationarity in a frequentist framework implies additional evidence in favor of predictability (Lewellen (2004), Campbell and Yogo (2006), Campbell (2008), Cochrane (2008)). Introducing the exact likelihood leads to an estimate of ρ that is higher that $\hat{\rho}$. This result (which is sample-dependent) comes about because of two sources of evidence on ρ contained in the specification (8). There is the evidence from the covariance between x_t and x_{t+1}, but there is also evidence from the difference between x_0 and the sample mean. If x_0 is relatively far from the sample mean, the posterior of ρ shifts toward higher values. This implies that $\hat{\rho}$ is lower than ρ, and therefore, that $\hat{\beta}$ is higher than β. Introducing the Jeffreys prior, in combination with the exact likelihood further shifts ρ back towards 1; this raises ρ relative to $\hat{\rho}$ and therefore lowers β relative to $\hat{\beta}$.[10]

Table 2 shows the implications for expected returns and asset allocation by reporting these values at various levels of the dividend yield. Comparing the first and last rows of each panel shows that Bayesian estimation with the conditional likelihood and prior (11) have implications that are virtually identical to ignoring parameter uncertainty and using the OLS estimates. For the exact likelihood and prior (15), both the expected returns and allocations are less variable, as one would expect given the lower posterior mean of β. Surprisingly, not

[9]Intuition for this result can be stated as follows: If ρ is above the OLS estimate $\hat{\rho}$, it must be the case that $\hat{\rho}$ is "too low", namely, in the sample, shocks to the predictor variable tend to be followed more often by shocks of a different sign than would be expected by chance. Because shocks to the predictor variable and to the return variable are negatively correlated, this implies that shocks to the predictor variable tend to be followed by shocks to returns of the same sign. This implies that $\hat{\beta}$ will be "too high".

[10]Generally, it appears that the net bias reduction resulting from these modifications is smaller than standard frequentist-based estimates of the bias. Whether this is good, bad or merely neutral depends on one's perspective (Sims and Uhlig (1991)).

only are the expected returns less variable, they are also substantially lower for both values of the dividend yield, leading to lower allocations as well. In fact, the average excess stock return is different in the various cases (Wachter and Warusawitharana (2009b)). As explained in that paper, differences in estimates of average excess stock returns arise from differences in estimates of the mean of the predictor variable. Over this sample, the conditional maximum likelihood estimate of the dividend yield is below the exact maximum likelihood estimate. Therefore, shocks to the predictor variable over the sample period must have been negative on average; it follows that shocks to excess returns must have been positive on average. Therefore, the posterior mean of returns is below the sample mean.

2.4 Informative priors

Introducing a Jeffreys prior and the exact likelihood has the effect of making portfolio choice less sensitive to the dividend yield, as compared with the conditional Bayesian model. However, the agent still engages in market timing to a large degree. As Wachter and Warusawitharana (2009a) show, the priors described above assign a perhaps unrealistically high probability to high R^2 statistics in the regression equation.[11]

Let

$$\sigma_x^2 = \frac{\sigma_v^2}{1-\rho^2}, \tag{19}$$

and note that (19) is the variance of the stationary distribution of x_t. The population R^2 for the regression (3) is defined to be the ratio of the variance of the predictable component of the return to the total variance. It follows from (19) that the R^2 is equal to

$$R^2 = \frac{\beta^2\sigma_x^2}{\beta^2\sigma_x^2 + \sigma_u^2}. \tag{20}$$

Wachter and Warusawitharana (2009a) consider a class of priors that translate into distributions on the population R^2. Specifically, they define a "normalized" β:

$$\eta = \sigma_u^{-1}\sigma_x\beta$$

[11] Wachter and Warusawitharana (2009a) argue that economic theory points toward low levels of the R^2, should predictability exist at all.

and assume that the prior distribution for η equals

$$\eta \sim N(0, \sigma_\eta^2). \tag{21}$$

The population R^2 can be rewritten in terms of η:

$$R^2 = \frac{\eta^2}{\eta^2 + 1}. \tag{22}$$

Equation 22 provides a mapping between a prior distribution on η and a prior distribution on the population R^2. The prior distribution for η implies a conditional prior for β. Namely,

$$\beta | \alpha, \theta, \rho, \Sigma \sim N(0, \sigma_\eta^2 \sigma_x^{-2} \sigma_u^2). \tag{23}$$

Because σ_x is implicitly a function of ρ and σ_v, the prior on β is also a function of these parameters. The approximate Jeffreys prior for the remaining parameters is given by

$$p(\alpha, \theta, \rho, \Sigma) \propto \sigma_x \sigma_u |\Sigma|^{-\frac{5}{2}}. \tag{24}$$

Equations 23–24 form a class of prior distributions indexed by η. For $\sigma_\eta = 0$, the prior dogmatically specifies that there can be no predictability: β is identically equal to zero. For $\sigma_\eta = \infty$, the prior is uninformative, and is in fact equal to the approximate Jeffreys prior in Stambaugh (1999). Because of the relation between η and the R^2, a prior on η translates directly into a prior on the R^2. An appeal of this approach is its scale-invariance: It is hard to imagine putting an economically meaningful prior on β without knowing something about the variance of the predictor variable x.

Figure 2 illustrates the implications of different values of σ_η for the prior distribution on the R^2. The prior with $\sigma_\eta = 0$ implies a dogmatic view that there can be no predictability, which is why the R^2 is a point mass at zero. On the other hand, with $\sigma_\eta = 100$ (which well-approximates the Jeffreys prior), the R^2 is very nearly flat over the single-digit range, dipping down in a region close to 1. Figure 2 shows that uninformative beliefs imply not only that high values of the population R^2 are possible, they are extremely likely. The prior assigns a probability of nearly 100% to the R^2 exceeding any given value, except for values that are an infinitesimal distance from one.

The literature has considered other specifications for informative priors. Kandel and Stambaugh (1996), for example, construct priors assuming that the investor has seen, in addition to the actual data, a hypothetical prior sample of the data such that the sample means, variances and covariances of returns and predictor variables are the same in the hypothetical prior sample as in the actual sample. However, in the hypothetical sample, the R^2 is exactly equal to zero (see also Avramov (2002, 2004)). Cremers (2002) constructs informative priors assuming the investor knows sample moments of the predictive variable. These constructions raise the question of how the investor knows the sample moments of returns and predictive variables (note that it is not sufficient for the investor to make a guess that is close to the sample values). If it is by seeing the data, the prior and the posterior are equal and the problem reduces to the full-information case. An alternative is to assume that the investor has somehow intuited the correct values. According to this latter (somewhat awkward) interpretation, to be consistent these moments would have to be treated as constants (namely conditioned on) throughout the analysis, which they are not.

Figure 2 suggests that priors of the form (23–24) with small σ_η have more reasonable economic properties than uninformative priors. Wachter and Warusawitharana (2009a) investigate the quantitative implications of these priors for portfolio allocation. Not surprisingly, because the posterior mean of β shrinks toward zero, the portfolio allocation under these priors exhibits less dependence on the dividend yield. The priors also can lead to improved out-of-sample performance. This issue has come to the fore based on several papers that critique the evidence in favor of predictability based on out-of-sample performance. Bossaerts and Hillion (1999) find no evidence of out-of-sample return predictability using a number of predictors; Goyal and Welch (2008) find that predictive regressions often perform worse than using the sample mean when it comes to predicting returns. For the empirical researcher, these studies raises the question of how the Bayesian asset allocation strategies perform out-of-sample. Note, however, that from the point of view of the Bayesian investor, such additional information is irrelevant. The predictive distribution for returns, as generated from the likelihood and the prior, is the sole determinant of the portfolio strategy.

Wachter and Warusawitharana (2009a) examine the out-of-sample performance implied by various priors. They show that asset allocation, using the results of OLS regression without taking parameter uncertainty into account, indeed delivers worse out-of-sample performance than a strategy implied by a dogmatic belief in no predictability. Relative to the OLS benchmark, the strategy implied by the uninformative Jeffreys prior (16) performs better, but still worse than the no-predictability prior. Across various specifications, the best-performing prior is an intermediate one, representing some weight on the data and some weight on an economically reasonable view that, if predictability should exist, the R^2 should be relatively small.[12]

2.5 Additional sources of uncertainty

One of the objectives of adopting Bayesian decision theory into the asset allocation problem is to better capture the uncertainty faced by investors. However, despite the uncertain nature of the predictive relation, estimation risk appears to play a minor role in the empirical findings. The disconnect between these results and our intuition may be due to the fact that assuming the model given by (3–6) still, to a large degree, understates the uncertainty actually faced by investors. While investors do not know the parameters of the system, they know in fact that returns and predictor variables obey such a system. With the available data, this turns out to be enough to estimate the parameters quite precisely. In reality, of course, investors do not know that returns obey such a system. That is, while (3) is unrestrictive in the sense that one could always regress returns on the lagged dividend yield, the system itself is restrictive. For instance, it requires that u_{t+1} is not only an error in the traditional regression sense of being uncorrelated with the right-hand-side variable, but that

[12] A second approach to improving out-of-sample performance is adopted by Campbell and Thompson (2008). They show that the out-of-sample performance improves when weak economic restrictions are imposed on the return forecasts: namely requiring that the expected excess return be positive and that the predictor variable has the theoretically expected sign. The Campbell and Thompson paper is non-Bayesian, but it would not be difficult to incorporate these prior views in a Bayesian setting.

it is in fact a *shock*, namely independent of any variable known at time t. The possibility that other likelihood functions are possible is something that undoubtedly would occur to real-world investors.

Pastor and Stambaugh (2009a, 2009b) confront this problem by assuming that returns obey a predictive system:

$$\begin{aligned} y_{t+1} &= \mu_t + u_{t+1} \\ x_{t+1} &= (I - A)E_x + Ax_t + v_{t+1} \\ \mu_{t+1} &= (1 - \rho)E_r + \rho\mu_t + w_{t+1}, \end{aligned} \tag{25}$$

where u, v and w are iid (across time) and jointly normally distributed. Here, μ (unobserved) is the true expected excess return and the agent learns about μ by observing x and y. Under this predictive system, one could still regress y_{t+1} on the observable x_t. However, the error in the regression would be correlated with time-t variables. Pastor and Stambaugh find that this distinction between μ_t and x_t, and particularly the fact that the autocorrelation of x need not equal the autocorrelation of μ, has important consequences for investors.

One could expand the uncertainty faced by investors in other ways. Recent studies (Avramov (2002), Cremers (2002), Wachter and Warusawitharana (2009b)) explore the possibility that an investor assigns some prior probability to (3–6), but also a non-zero probability to a model in which returns are iid. While this represents a form of "model uncertainty", the agent is still Bayesian in the sense that he assigns a probability to each model. One could go further and assume that there are some forms of uncertainty that investors simply cannot quantify. Gilboa and Schmeidler (1989) define a set of axioms on preferences that distinguish between risk (in which the agent assigns probabilities to states of nature) and uncertainty (in which probabilities are not assigned). They show that aversion to uncertainty leads investors to maximizes the minimum over the set of priors that may be true. Uncertainty aversion, also called ambiguity aversion, has been the subject of a fast-growing literature in recent years, much of which has focused on asset allocation (see Chamberlain (1999), Chen and Epstein (2002), Chen, Ju, and Miao (2009), Garlappi, Uppal, and Wang

(2007), Hansen (2007) and Maenhout (2006)).

This notion of additional uncertain facing investors is likely to be a subject of continued active debate. As discussed above, there are a number of complementary approaches, such as the predictive system, model uncertainty with probabilities over the models and finally model uncertainty such that the agent need not even formulate probabilities over the models. The contention of the previously discussed models is that periods of low valuation (e.g. when the dividend yield is high) represent, to some uncertain extent, an opportunity for the investor that is just there for the taking. However, another possibility is that the excess returns earned by this market timing strategy are in fact a compensation for a type of risk that does not appear in the sample; the risk of a rare event.[13]

In Wachter (2008), I show that the level of predictability in excess returns can be captured by a model with a representative investor with recursive preferences (see below), in which there is a time-varying probability of a rare event. Times when this rare event probability are high correspond to times when the dividend yield is high as well. Most of the time, the rare event does not happen, implying higher than average realized returns. Occasionally, the rare event does happen, in which case high dividend yields are followed by quite low returns. The representative agent holds a constant weight in equities (as is required by equilibrium) despite the fact that excess returns do vary in a predictable fashion. Strategies that attempt to time the market, according to this view, are in fact quite risky, though this risk would be difficult to detect in the available time series.

[13]Yet another possibility is that the excess returns represent compensation for greater volatility. Shanken and Tamayo (2005) evaluate this claim directly in a Bayesian setting and find little support for it. A large literature debates the extent to which changes in volatility are linked to changes in expected returns; based on available evidence, however, it does not appear that the fluctuations in expected returns captured by the dividend yield arise correspond to changes volatility. See Campbell (2003) for a discussion of this literature.

3 Dynamic models

I now consider the investor who has a horizon beyond one period, and, at each time point, faces a consumption and portfolio choice decision. I start with a general specification that allows for multiple risk assets and state variables. Let C_t denote the investor's consumption at time t, z_t the $N \times 1$ vector of allocations to risky assets and W_t the investor's wealth. Let R_{t+1} denote the vector of (simple) returns on the risky assets and $R_{f,t+1}$ the (simple) riskfree return. Samuelson (1969) models this problem as

$$\max_{c,z} E \sum_{t=0}^{T} e^{-\beta t} \frac{C_t^{1-\gamma}}{1-\gamma} \tag{26}$$

subject to the budget constraint

$$W_{t+1} = (W_t - C_t) R_{f,t+1} + W_t z_t^\top (R_{t+1} - R_{f,t+1}) \tag{27}$$

and terminal condition $W_T \geq 0$. Here $e^{-\beta t} \frac{C_t^{1-\gamma}}{1-\gamma}$ represents period utility (for simplicity, I have assumed that the investor does not have a bequest motive). An alternative is to consider the problem without the utility flow from consumption, namely the investor just maximizes $\frac{W_T^{1-\gamma}}{1-\gamma}$. This is not as realistic, but it is sometimes a helpful simplification. Another helpful simplification is to take the limit of (26) as T goes to infinity.

The problem above can be solved by backward induction using the Bellman equation (see Duffie (1996, Chapter 3)). Let X_t denote an $n \times 1$ vector of state variables that determine the distribution of returns. Let $\tau = T - t$ denote the horizon. Define the value function as the remaining utility:

$$J(W_t, X_t, \tau) = \max_{c,z} E \sum_{s=0}^{\tau} e^{-\beta s} u(c_{t+s}).$$

Then it follows that J can be defined through backward induction as:

$$J(W_t, X_t, \tau) = u(c_t) + e^{-\beta} E_t \left[J(W_{t+1}, X_{t+1}, \tau - 1) \right] \tag{28}$$

with the boundary condition $J(W, X, 0) = u(W)$. See Brandt (2009) for further discussion of the value function and its properties.

While (28) reduces the multiperiod problem (26) to a series of one-period problems, these one-period problems might look quite different from the problem considered in Section 2 because of the interaction between the state variables X and wealth W. Indeed, when there is no X (so returns are iid), Samuelson (1969) shows that (26) indeed does reduce to a series of one-period problems. In this article, however, I am primarily interested in the case where returns are not iid. Merton (1973) characterizes the solution to this general problem. For technical reasons that will be discussed, it is easier to do this if one assumes that time is continuous.

3.1 Return distribution and the value function

Let B_t denote a $d \times 1$ vector of independent Brownian motions. Let

$$\lambda(X) = [\lambda_1(X), \ldots, \lambda_N(X)]^\top$$

denote the $N \times 1$ vector of instantaneous excess returns and

$$\sigma(X) = [\sigma_1(X)^\top, \ldots, \sigma_N(X)^\top]^\top$$

denote the $N \times d$ matrix of loadings on the Brownian motions. Then the price process for asset i, $i = 1, \ldots, N$, is given by

$$\frac{dP_t^{(i)}}{P_t^{(i)}} = (\lambda_i(X_t) + r_f(X_t))\, dt + \sigma_i(X_t)\, dB_t, \tag{29}$$

where $r_f = \log R_f$. I assume X_t follows a Markov process:[14]

$$dX_t = b(X_t)\, dt + a(X_t)\, dB_t. \tag{30}$$

[14] It is also possible to allow for deterministic time-variation in the investment opportunity set by defining λ, σ and r_f as functions of t as well as X. Alternatively, X could be augmented with the time t. However, many of the results below simplify considerably when X is time-varying but deterministic; from a theory perspective it makes sense to consider this case as distinct from the stochastic case. Because the economic significance of deterministic time-variation is less than of stochastic variation, I will not explicitly consider that case here. In what follows, "time-varying" should be understood to mean stochastic.

Assumptions (29) and (30) imply that the current value of the state variables at time t, fully determine the investment opportunities that are available to the investor; that is, they determine the *investment opportunity set*.[15]

Merton (1971) shows that under the assumptions above, wealth follows the process

$$dW_t = \left(W_t z_t^\top \lambda(X_t) + W_t r_f(X_t) - C_t\right) dt + W_t z_t^\top \sigma(X_t)\, dB_t. \tag{31}$$

Merton (1973) derives a partial differential equation characterizing the value function J. Moreover, he shows that the first-order condition with respect to z leads to the following characterization of z in terms of derivatives of J:

$$z = -\frac{J_W}{J_{WW} W} \left(\sigma\sigma^\top\right)^{-1} \lambda - \frac{1}{J_{WW} W} \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top J_{XW}, \tag{32}$$

where J_W, J_{WW} and J_{XW} refer to first and second partial derivatives of J. Here and in what follows, I eliminate time subscripts and function arguments when not required for clarity. I show in the Appendix that the value function takes the form

$$J(W, X, \tau) = \frac{I(X,\tau)^{1-\gamma} W^{1-\gamma}}{1-\gamma}. \tag{33}$$

Applying (33), it follows that the allocation can be rewritten as

$$z = \frac{1}{\gamma} \left(\sigma\sigma^\top\right)^{-1} \lambda + \frac{1-\gamma}{\gamma} \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top \frac{I_X^\top}{I}. \tag{34}$$

Equation 34 (and, more generally, (32)) provides a gateway to understanding portfolio choice in this rich dynamic context. There are two terms in (34), only one of which depends on the process for X. Note that in the discrete-time setting when one period remains, the value function depends only on wealth, not on X. The same is true in continuous time; in the limit, as the horizon approaches 0 the value function's dependence on X also approaches zero. Therefore, as the horizon approaches 0, only the first term remains. For this reason, Merton (1973) refers to this as what the investor would choose if he behaved *myopically*,

[15] When markets are complete, one can be more specific as Cox and Huang (1989) show. In this case, it suffices to specify a vector of risk prices η, such that $\sigma\eta = \lambda$ and the riskfree rate r_f as functions of X, as well as the evolution of X itself. The actual assets that are available, as in (29), need not be specified.

namely, if, like the discrete-time investor with one period left, he only took into account the very immediate future, and did not look beyond.

Given that myopic demand captures, in a limiting sense, the desired allocation of a one-period investor, how does it compare with the results derived in Section 2? Consider, for simplicity the case of a single risky asset.[16] In this case, λ corresponds to the (instantaneous) expected excess return on the asset and $\sigma\sigma^\top$ to the (instantaneous) variance. Indeed, Ito's Lemma implies that for an asset with price P_t

$$d \log P_t = (\lambda + r_f - \frac{1}{2}\sigma\sigma^\top)\, dt + \sigma\, dB_t,$$

so that, assuming units are the same, $E_t[y_{t+1}] \approx \lambda - \frac{1}{2}\sigma\sigma^\top$ and $\mathrm{Var}_t[y_{t+1}] \approx \sigma\sigma^\top$. Myopic demand therefore closely resembles (10). The main difference is that (10) is approximate, while (34) is exact. Recall that in the setting of Section 2 (indeed in any discrete-time setting) power preferences rule out levered positions or short positions in the stock at any horizon. However, when trading is continuous, the agent can exit these positions in time to avoid negative wealth. This property, which is not without controversy, plays a key role in making the continuous-time model tractable.

Myopic demand, then, is the continuous-time analogue of the static portfolio choice described in Section 2. In contrast, the second term in (34) is completely new. As Merton (1973) shows, this term represents the agent's efforts to hedge future changes in the investment opportunity set. There are two offsetting motives: On the one hand, the investor would like more wealth in states with superior investment opportunities, all the better to take advantage of them; on the other hand, the investor would like more wealth in states with poorer investment opportunities, so as to lessen the overall risk to long-term wealth. The first of these is a substitution effect, the second, an income effect.

To see how these motives are represented by (34), consider the case with a single state variable. Note that the sign of J_X equals the sign of I_X. I will say that an increase in X indicates an improvement in investment opportunities if and only if it increases the agent's

[16] In fact, the parallels hold in the multiple-risky-asset case as well (see Merton (1971)).

utility, namely if and only if $J_X > 0$. (Merton (1973) discusses hedging motives in terms of the consumption-wealth ratio rather than the value function. I will explore the link to the consumption-wealth ratio in what follows.) If an asset positively covaries with the stock return (continue to assume, for simplicity, that there is a single risky asset), hedging demand is negative as long as γ is greater than 1 and positive as long as γ is less than 1. That is, the agent with $\gamma > 1$ reduces his investment to an asset that pays off in states with superior investment opportunities (the income effect dominates), while the agent with $\gamma < 1$ increases his investment to such an asset (the substitution effect dominates). Logarithmic utility ($\gamma = 1$) corresponds to the knife-edge case when these effects cancel each other out.

To go further, it is necessary to learn more about the function $I(X, \tau)$. This function depends on the parameters in (29–30), and so will embody an empirical statement about the distribution of returns. Applying the theory above to estimated processes for returns is one way the literature has built on the insights in Merton (1973). A second source of innovation is in the type of utility function considered. This is the topic of the next section.

3.2 Recursive utility

One limitation of assumption (26) is that it implies that an identical parameter, γ, controls both the agent's attitudes toward the smoothness of consumption over time, and the agent's attitudes toward to the smoothness of consumption over states, namely her attitudes towards risk. Building on work of Kreps and Porteus (1978), Epstein and Zin (1989, 1991) and Weil (1990) develop a class of utility functions that retains the attractive scale invariance of power utility, but allows for a separation between the concepts of risk aversion and the willingness to substitute over time. Such a separation implies that the agent has preferences over the timing of the resolution of uncertainty, which may be attractive in and of itself. The resulting utility function lies outside out of the expected utility framework in the sense that the utility cannot be written explicitly as an expectation of future consumption. Rather, utility is defined recursively.[17]

I use the continuous-time formulation of the Epstein and Zin (1989) utility function developed by Duffie and Epstein (1992a,b). Let V_t denote the remaining utility. Following Duffie and Epstein, I will use the notation V to denote the utility process and the notation J to denote optimized utility as a function of wealth, the state variables and the horizon. At the optimum, $V_t = J(W_t, X_t, T-t)$. Duffie and Epstein specify V_t as follows:

$$V_t = E_t \int_t^T f(C_s, V_s)\, ds, \tag{35}$$

where

$$f(C,V) = \begin{cases} \frac{\beta}{1-\frac{1}{\psi}}\left((1-\gamma)V\right)\left(\left(C\left((1-\gamma)V\right)^{-\frac{1}{1-\gamma}}\right)^{1-\frac{1}{\psi}} - 1\right) & \psi \neq 1 \\ \beta((1-\gamma)V)(\log C - \frac{1}{1-\gamma}\log((1-\gamma)V)) & \psi = 1. \end{cases} \tag{36}$$

Duffie and Epstein (1992a) show that parameter $\psi > 0$ can be interpreted as the elasticity of intertemporal substitution (EIS) and $\gamma > 0$ can be interpreted as relative risk aversion. When $\gamma = 1/\psi$, power preferences, given in (26) are recovered (note that the resulting formulation of V_t may not take the same form as (26), but will imply the same underlying preferences, and therefore the same choices).

Results in Duffie and Epstein (1992a) show that the first-order condition for portfolio allocation (32), and the first-order condition for consumption $f_C = J_W$ derived by Merton (1973) are valid in this more general setting. Below, I use these results to characterize optimal consumption and investment behavior, considering the case of $\psi \neq 1$ and $\psi = 1$ separately.[18]

[17]Kihlstrom (2009) develops an alternative approach to separating the inverse of the elasticity of substitution and risk aversion within an expected utility framework.

[18]Interesting questions of existence and uniqueness of solutions are beyond the scope of this study. Schroder and Skiadas (1999) provide such results assuming bounded investment opportunities and a utility function that generalizes the recursive utility case considered here; Wachter (2002) proves existence in the return predictability case (Section 3.3) under power utility with risk aversion greater than 1. Further existence results under power utility are provided by Dybvig and Huang (1988) and Dybvig, Rogers, and Back (1999).

3.2.1 Characterizing the solution when the EIS does not equal one

As shown in the Appendix, as long as $\psi \neq 1$, the form of the value function (33), and therefore the form of optimal allocation, (34), still holds. Myopic demand takes the same form as under power utility; that is, it is determined by γ alone. The parameter γ also determines whether the income or substitution effect dominates in the portfolio decision. These results support the interpretation of γ as risk aversion in this more general model.

It is also instructive to consider the consumption policy. Define a function H as follows:

$$H(X,\tau) = \beta^{-\psi} I(X,\tau)^{-(1-\psi)}. \tag{37}$$

It follows form the first-order condition for consumption ($f_C = J_W$) that the wealth-consumption ratio is equal to H:

$$\frac{W_t}{C_t} = H(X_t, T-t). \tag{38}$$

It follows from (37) that

$$\frac{I_X}{I} = -\frac{1}{1-\psi}\frac{H_X}{H}. \tag{39}$$

Recall that the sign of I_X equals the sign of J_X, the derivative of the value function with respect to the state variables. As in the asset allocation decision, there are two effects that changes in investment opportunities could have on consumption behavior. On the one hand, an improvement could lead investors to consume less out of wealth, to better take advantage of the opportunities (the substitution effect). On the other, an improvement raises wealth in the long-run, allowing the investor to consume more today (the income effect). Equation (39) shows that for investors who are relatively willing to substitute intertemporally ($\psi > 1$), consumption falls relative to wealth when investment opportunities rise (the substitution effect dominates). For investors who are relatively unwilling to substitute intertemporally ($\psi < 1$), consumption rises (the income effect dominates). These results support the interpretation of ψ as the elasticity of intertemporal substitution.

Substituting into the Bellman equation (53) leads to the following differential equation for H:

$$\frac{1}{1-\psi}\frac{H_\tau}{H} - \frac{1}{1-\psi}\frac{H_X}{H}b + \frac{1}{2}\frac{1}{\gamma}\lambda^\top\left(\sigma\sigma^\top\right)^{-1}\lambda - \frac{1-\gamma}{1-\psi}\frac{1}{\gamma}\frac{H_X}{H}a\sigma^\top\left(\sigma\sigma^\top\right)^{-1}\lambda$$
$$+ r_f + \frac{1}{2}\frac{1}{1-\psi}\left(\frac{1-\gamma}{1-\psi}+1\right)\mathrm{tr}\left(a^\top\frac{H_X^\top}{H}\frac{H_X}{H}a\right) - \frac{1}{2}\frac{1}{1-\psi}\mathrm{tr}(a^\top H_{XX}a)\frac{1}{H}$$
$$+ \frac{1}{2}\left(\frac{1-\gamma}{1-\psi}\right)^2\frac{1}{\gamma}\frac{H_X}{H}a\sigma^\top\left(\sigma\sigma^\top\right)^{-1}\sigma a^\top\frac{H_X^\top}{H}$$
$$- \frac{1}{1-\psi}H^{-1} - \beta\left(1-\frac{1}{\psi}\right)^{-1} = 0, \quad (40)$$

with boundary condition $H(X, 0) = 0$. Equation (40) is useful in considering the special cases below.

Constant investment opportunities

In the special case of constant investment opportunities, portfolio choice is myopic, as explained above. The wealth-consumption ratio can also be derived in closed-form: the differential equation for H (which is now a function of τ alone) is given by:

$$\frac{1}{1-\psi}H' + \left(\frac{1}{2}\frac{1}{\gamma}\lambda^\top\left(\sigma\sigma^\top\right)^{-1}\lambda + r_f - \beta\left(1-\frac{1}{\psi}\right)^{-1}\right)H - \frac{1}{1-\psi} = 0. \quad (41)$$

The solution is

$$H(\tau) = \frac{1}{k(1-\psi)}\left(1 - e^{-k(1-\psi)\tau}\right), \quad (42)$$

where

$$k = \frac{1}{2}\frac{1}{\gamma}\lambda^\top\left(\sigma\sigma^\top\right)^{-1}\lambda + r_f - \beta\left(1-\frac{1}{\psi}\right)^{-1}.$$

The first two terms in k measure of the quality of investment opportunities. For $\psi > 1$, $H(\tau)$ is increasing in k.[19] That is, the greater are investment opportunities, the less the investor consumes out of wealth. Note that the discount rate enters k with a negative sign: while an increase in investment opportunities causes the investor to consume less out of wealth, an increase in the discount rate causes the investor to consume more. For $\psi < 1$, $H(\tau)$ is decreasing in k. The greater are investment opportunities, the more the investor consumes

[19]This follows from the fact that $H(0) = 0$ and that $H'(\tau)$ is increasing in k for any (fixed) $\tau > 0$.

out of wealth. Also, an increase in investment opportunities and an increase in the discount both lead the investor to consume more and save less as a percentage of wealth.

Power utility and complete markets

In this setting without trading restrictions, markets are complete if and only if the diffusion terms for asset prices span the diffusion terms for X. The term $a\sigma^\top \left(\sigma\sigma^\top\right)^{-1}$ represents the projection of the diffusion terms for X on the diffusion terms for $P^{(i)}$; therefore markets are complete if and only if

$$a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \sigma = a,$$

namely if the projection recovers the diffusion terms on X. Further note that, because $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for conforming matrices,

$$\mathrm{tr}\left(a^\top H_X^\top H_X a\right) = \mathrm{tr}\left(H_X a a^\top H_X^\top\right) = H_X a a^\top H_X^\top.$$

Therefore (40) reduces to the much simpler

$$H_\tau - H_X b + \frac{1}{2}\frac{\gamma-1}{\gamma^2}\lambda^\top \left(\sigma\sigma^\top\right)^{-1} \lambda H + \frac{\gamma-1}{\gamma} H_X a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \lambda \\ + \frac{\gamma-1}{\gamma} r_f H - \frac{1}{2}\mathrm{tr}(a^\top H_{XX} a) - 1 + \frac{1}{\gamma}\beta H = 0. \quad (43)$$

Equation (43) has a solution of the form

$$H(X,\tau) = \int_0^\tau F(X,s)\,ds \quad (44)$$

with $F(X,0) = 1$. To see this, note that it follows from integration by parts that

$$\int_0^\tau \frac{\partial F}{\partial s}\,ds = F(X,\tau) - F(X,0) = H_\tau - 1.$$

Substituting in, I find that F satisfies

$$\frac{\partial F}{\partial \tau} - F_X b + \frac{1}{2}\frac{\gamma-1}{\gamma^2}\lambda^\top \left(\sigma\sigma^\top\right)^{-1} \lambda F + \frac{\gamma-1}{\gamma} F_X a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \lambda \\ + \frac{\gamma-1}{\gamma} r_f F - \frac{1}{2}\mathrm{tr}(a^\top F_{XX} a) + \frac{1}{\gamma}\beta F = 0. \quad (45)$$

It is no accident that the differential equation simplifies under the case of power utility and complete markets. This is the case where the conceptually simpler martingale method of Cox and Huang (1989), Karatzas, Lehoczky, and Shreve (1987) and Pliska (1986) is straightforward to apply (for an example, see Wachter (2002)). While this method can be extended to incomplete markets (He and Pearson (1991), Cuoco (1997)) and to recursive utility (Duffie and Skiadas (1994), Schroder and Skiadas (1999), Skiadas (2007)), it is less straightforward in these cases.

3.2.2 Characterizing the solution when the EIS equals 1

In the case of $\psi = 1$, the value function takes the form

$$J(W, X, \tau) = \frac{W^{(1-\gamma)(1-e^{-\beta\tau})} G(X, \tau)^{1-\gamma}}{1-\gamma}.$$

The differential equation for G is given in the Appendix. Equation (32) still holds (see Duffie and Epstein (1992a)), implying that the portfolio allocation is given by

$$z = \frac{1}{1-(1-\gamma)(1-e^{-\beta\tau})} \left(\sigma\sigma^\top\right)^{-1} \lambda + \frac{1-\gamma}{1-(1-\gamma)(1-e^{-\beta\tau})} \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top \frac{G_X^\top}{G}. \qquad (46)$$

As in the case of $\psi \neq 1$ the portfolio allocation separates into two terms, the first of which can be interpreted as myopic demand (because it does not depend on future investment opportunities), and the second as hedging demand.

Myopic demand is horizon-dependent when $\psi = 1$. When the horizon is large (as $\tau \to \infty$), myopic demand approaches the myopic demand in (34): namely it is determined by the coefficient of relative risk aversion only. However, for finite horizons, myopic demand is also determined by the EIS. In fact, myopic demand is determined by a weighted average of $\psi^{-1}(=1)$ and γ, with the horizon determining the weights:

$$1-(1-\gamma)\left(1-e^{-\beta\tau}\right) = e^{-\beta\tau} + \gamma(1-e^{-\beta\tau}).$$

The first-order condition $f_C = J_W$, applied to (46) implies that the wealth-consumption ratio is given by $\frac{W}{C} = \frac{1-e^{-\beta\tau}}{\beta}$. Unlike in the $\psi \neq 1$ case, the wealth-consumption ratio does not depend on investment opportunities. Unit EIS corresponds to the knife-edge case where the substitution and income effects cancel each other out, as far as consumption behavior is concerned. In the limiting case of an infinite horizon, the wealth-consumption ratio is constant and equal to β^{-1}.

3.3 Time-varying risk premia

I now consider a special case of price dynamics in which there is a single risky asset and a single state variable X_t. This example is meant to be illustrative; the solution technique can be extended to other forms of affine dynamics (see Schroder and Skiadas (1999), Liu (2007)). Specifically, assume r_f, σ and $a = \sigma_X$ are constants, and let

$$\begin{aligned} \lambda(X) &= \left(\sigma\sigma^\top\right)^{\frac{1}{2}} X \\ b(X) &= -\kappa(X - \bar{X}). \end{aligned}$$

This specification implies that the expected excess return on the stock is time-varying, and depends linearly on a variable X_t that follows a mean-reverting process. It is therefore the continuous-time equivalent of the process assumed in Section 2.[20] Note that X_t is the Sharpe ratio on the risky asset.

When are exact solutions available?

More explicit solutions for the value function, and therefore for portfolio and consumption choice are available in two special cases of the above analysis: (1) when the EIS is equal to 1, and (2) when power utility obtains ($\gamma = 1/\psi$), and markets are complete. The first case is considered by Schroder and Skiadas (1999) (who also assume complete markets), and Campbell, Chacko, Rodriguez, and Viceira (2004) (who also assume an infinite horizon). Here I consider the first case, allowing markets to be incomplete and the horizon to be finite.[21] The second case is the subject of Wachter (2002). In a related contribution, Kim

[20] Wachter (2002) makes this statement precise.

[21] Duffie and Epstein (1992a) consider a related finite-horizon problem in which they analyze the term structure of interest rates under Cox, Ingersoll, and Ross (1985)-type dynamics.

and Omberg (1996) show that one can also obtain closed-form solutions for portfolio choice when the investor maximizes power utility over terminal wealth.

Indeed, when $\psi = 1$, the value function is given by (60), with G taking the form

$$G(X, \tau) = \exp\left\{A_1^{(1)}(\tau)\frac{X^2}{2} + A_2^{(1)}(\tau)X + A_3^{(1)}(\tau)\right\}, \tag{47}$$

and where $A_i^{(1)}$ satisfy a system of ordinary differential equations with boundary conditions $A_i^{(1)}(0) = 0$. For power utility and complete markets, the wealth-consumption ratio $H(X, \tau)$ is given by (44), where

$$F(X, \tau) = \exp\left\{A_1^{(2)}(\tau)\frac{X^2}{2} + A_2^{(2)}(\tau)X + A_3^{(2)}(\tau)\right\} \tag{48}$$

Substituting into (45) results in a set of ordinary differential equation for $A_i^{(2)}$, with boundary conditions $A_i^{(2)} = 0$.

An approximate solution technique

The affine dynamics above lend themselves to an approximate solution technique developed by Campbell and Viceira (1999) based on earlier work by Campbell (1993). Campbell and Viceira propose log-linearizing the budget constraint around the mean consumption-wealth ratio. They then derive an approximate analytical solution to the above problem, assuming an infinite horizon. As long as the consumption-wealth ratio is not too variable (i.e. the EIS is not far from 1), the approximation error will be small.

Chacko and Viceira (2005) show how to implement this approximation in a continuous-time setting. Consider the differential question for the wealth-consumption ratio (40). Following Chacko and Viceira, I assume that the horizon is infinite, and look for a stationary solution, namely a solution with $H_\tau = 0$. Let $h = \log H$, and consider a first-order approximation of e^{-h} around the mean of $-h$:

$$e^{-h} \approx e^{E[-h]} + (-h - E[-h])e^{E[-h]}. \tag{49}$$

Let $h_1 = e^{E[-h]}$ and $h_0 = h_1(1 - \log h_1)$. Then (49) implies

$$H^{-1} \approx h_0 - h_1 \log H. \tag{50}$$

Substitute (50) and $H_\tau = 0$ into (40) implies

$$-\frac{1}{1-\psi}\frac{H_X}{H}b+\frac{1}{2}\frac{1}{\gamma}\lambda^\top\left(\sigma\sigma^\top\right)^{-1}\lambda-\frac{1-\gamma}{1-\psi}\frac{1}{\gamma}\frac{H_X}{H}a\sigma^\top\left(\sigma\sigma^\top\right)^{-1}\lambda$$
$$+r_f+\frac{1}{2}\frac{1}{1-\psi}\left(\frac{1-\gamma}{1-\psi}+1\right)\operatorname{tr}\left(a^\top\frac{H_X^\top}{H}\frac{H_X}{H}a\right)-\frac{1}{2}\frac{1}{1-\psi}\operatorname{tr}(a^\top H_{XX}a)\frac{1}{H}$$
$$+\frac{1}{2}\left(\frac{1-\gamma}{1-\psi}\right)^2\frac{1}{\gamma}\frac{H_X}{H}a\sigma^\top\left(\sigma\sigma^\top\right)^{-1}\sigma a^\top\frac{H_X^\top}{H}$$
$$-\frac{1}{1-\psi}\left(h_0-h_1\log H\right)-\beta\left(1-\frac{1}{\psi}\right)^{-1}\approx 0. \quad (51)$$

Observe that this differential equation is similar in form to the one for the value function in the $\psi = 1$ case (given by (65)). In fact, it is simpler in that there is no time-dependence. It follows that the approximation method can be implemented any setting where the EIS = 1 yields an exact solution. Under the above assumptions on the asset return process and state variable processes,

$$H(X,\tau)\approx\exp\left\{A_1^{(3)}\frac{X^2}{2}+A_2^{(3)}X+A_3^{(3)}\right\},$$

where $A_i^{(3)}$ can be determined by matching coefficients. Campbell and Viceira (1999) use this approximation to show, among other results, that, in the infinite-horizon problem, portfolio decisions are driven, almost entirely, by risk aversion γ.

Numerical results

When calibrated to reasonable values, what do these dynamic considerations add to the asset allocation problem? In what follows, I present results from Wachter (2002); the near perfect negative correlation between the dividend yield and the stock return makes it reasonable to assume that markets are complete. I calibrate this model using the same parameters as Barberis (2000) and assume a risk aversion of 5, so the results are quantitatively comparable to those discussed in Section 2.[22] Similar results are found using alternative specifications

[22] Wachter (2002) provides details on how the discrete-time results are used to calibrate the continuous-time model. However, that paper calibrates the model mistakenly assuming that the process in Barberis (2000) applies to the net return on equities rather than the excess return. These results correct that mistake.

and methods (e.g. Brennan, Schwartz, and Lagnado (1997), Brandt (1999), Balduzzi and Lynch (1999)).

Figure 3 shows the optimal allocation as a function of horizon for various levels of the dividend yield. As in the static case, there are substantial horizon and market timing effects. However, in this case, rather than decreasing (slowly) in the horizon, the degree to which the allocation varies with the dividend yield is even more marked for long-term investors than for short-term investors. This greater dependence results from hedging demand. Because the dividend yield (proportional to X) is negatively correlated with stock returns, hedging demand leads the investor to allocate more money to stocks for $\gamma > 1$. The greater is the dividend yield the more the investor cares about this hedge, which is why hedging demand makes market timing more extreme. Unlike the simpler horizon effect in Section 2, this effect reverses for $\gamma < 1$. The long-horizon investor with $\gamma < 1$ holds less in stock than the short-horizon investor.

3.4 Parameter uncertainty and learning in dynamic models

So far in this section I have assumed that the investor has full knowledge of the parameters. I now consider the case of parameter uncertainty in dynamic models. One obvious difference between the dynamic and static settings is that the degree to which parameters are uncertain varies over time; that is, the agent learns more about the distribution as time goes on. Less obviously, learning introduces hedging demands for investors with risk aversion not equal to one. A well-studied special case is when returns are iid and the investor learns about the average excess return (see the discussion in Pastor and Veronesi (2009)). Given the assumption of iid returns, it is natural to assume that only the mean is uncertain as, in a continuous-time setting, the volatility can be estimated with effectively infinite precision (Merton (1980)). Moreover, as the time horizon shortens, the role of uncertainty about the mean goes to zero (Detemple (1986), Gennotte (1986)). The estimated mean, will differ from the true mean; however, the uncertainty about this estimated mean has no effect on the allocation.[23]

What does have an effect, and a large, one, is learning. Hedging demand induced by learning is negative and can be substantial (Brennan (1998)). The reason is that the investor's estimate of the average return (effectively a state variable) is positively correlated with realized returns. When a positive shock to prices occurs, the investor updates his beliefs about the average return, estimating it to be higher than before. Thus stocks are less attractive to an investor with $\gamma > 1$ (see (34)).

Uncertainty about parameters other than the mean is harder to address because it does not lend itself to closed-form solutions.[24] Studies have explored this question using numerical methods. Xia (2001) allows the investor to be uncertain about the degree of predictability (the coefficient β in (3)) and assumes the other parameters are known. She decomposes hedging demand into the component to hedge learning about β, and the component to hedge changes in X_t. Learning-induced hedging demand decreases in the difference between the dividend yield and its mean. Moreover, it switches in sign: it is positive when the dividend yield is below its long-run mean, zero when it is at the long-run mean, and negative when it is above the long-run mean. As Xia shows, these properties make the overall allocation less variable then when only predictability is taken into account. However, the allocation is still more variable than implied by the myopic strategy.

Solving the asset allocation problem when there is uncertainty about the full set of parameters is undertaken by Brandt, Goyal, Santa-Clara, and Stroud (2005) and by Skoulakis (2007). The lack of closed-form solutions and the high dimensionality of the problem make this a formidable technical challenge. These studies show that, in addition to the effect noted by Xia (2001), uncertainty about the mean (as in Brennan (1998)) exerts an important influence, driving down the average allocation relative to that discussed above. While the net effect of hedging demand is under dispute, the market timing effect remains alive and well.

[23]Given this limiting result in continuous time, it is perhaps not surprising that estimation risk should have little effect at short horizons as shown in Section 2.

[24]As in the static case, Bayesian learning requires that a prior and likelihood function be specified; the studies discussed below make use of the conditional Bayesian framework (Section 2.2).

4 Concluding remarks

In this study, I have reviewed the literature on static and dynamic asset allocation, with a focus on the implications of return predictability for long-run investors. For both buy-and-hold and dynamically trading investors, the optimal allocation to stocks is greater, the longer the horizon, given reasonable assumptions on preferences. This similarity should not obscure some key differences. In the static case, the effect of any stationary variable on the allocation will diminish as the horizon grows. In the dynamic case, there is no reason for this to happen, and indeed the opposite may be true. That is, for investors who dynamically trade, even short-term variables can have long-term implications.

This survey has also highlighted efforts to introduce parameter uncertainty into the agent's decision process. This, in theory, serves to pass on some of the uncertainty faced by the econometrician to the agent; the agent now incorporates this estimation risk into his decisions. Empirically, however, estimation risk appears to have very little effect, except at long buy-and-hold horizons (at least for the specifications explored herein). This is not to say that the perfect and imperfect-information cases are identical. Indeed, learning can induce important hedging demands in the dynamic setting. Furthermore, I show in the static setting that the choice of prior and likelihood can have a large impact on the results. The notion of uninformative is less than clear in a predictive regression setting, moreover economic theory indicates a possible role for unapologetically informative priors that take this theory into account. While these results do not arise from estimation risk per se, they do incorporate the small-sample nature of the evidence into the decision problem. Our data on financial markets is unavoidably finite; this should influence agents in economic models just as it influences the economists doing the modeling. Despite the progress reported here, it is fair to say that much work along these lines remains to be done.

Appendix: Solving for the value function in the dynamic recursive utility model

In this Appendix I will use the more general form of the aggregator suggested by Duffie and Epstein (1992a) for the $\psi = 1$ case. The formulas in the text result from taking the limit as $\xi \to 0$.

$$f(C,V) = \begin{cases} \frac{\beta}{1-\frac{1}{\psi}}\left((1-\gamma)V\right)\left(\left(C\left((1-\gamma)V\right)^{-\frac{1}{1-\gamma}}\right)^{1-\frac{1}{\psi}} - 1\right) & \psi \neq 1 \\ \beta(\xi^{1-\gamma} + (1-\gamma)V)(\log C - \frac{1}{1-\gamma}\log(\xi^{1-\gamma} + (1-\gamma)V)) & \psi = 1. \end{cases} \tag{52}$$

The continuous-time Bellman equation is derived by Duffie and Epstein (1992b):

$$-J_\tau + J_X b + J_W(Wz^\top \lambda + Wr_f - C) + \frac{1}{2}\mathrm{tr}(\Sigma) + f(C, J(X,W,\tau)) = 0, \tag{53}$$

where

$$\Sigma = \begin{bmatrix} a \\ Wz^\top\sigma \end{bmatrix}^\top \begin{bmatrix} J_{XX} & J_{XW} \\ J_{WX} & J_{WW} \end{bmatrix} \begin{bmatrix} a \\ Wz^\top\sigma \end{bmatrix}.$$

The first order condition for consumption is

$$f_C = J_W.$$

Substituting in from (52), I find (with some abuse of notation), C as a function of W, X and τ:

$$C(W,X,\tau) = \begin{cases} \beta^\psi J_W^{-\psi}((1-\gamma)J)^{\frac{1-\gamma\psi}{1-\gamma}} & \psi \neq 1 \\ \beta J_W^{-1}\left(\xi^{1-\gamma} + (1-\gamma)J\right) & \psi = 1. \end{cases} \tag{54}$$

It follows from (54), it follows that

$$f(C,V) = \begin{cases} \beta^\psi\left(1-\frac{1}{\psi}\right)^{-1} J_W^{1-\psi}((1-\gamma)J)^{\frac{1-\gamma\psi}{1-\gamma}} - \beta\left(1-\frac{1}{\psi}\right)^{-1}(1-\gamma)J & \psi \neq 1 \\ \beta(\xi^{1-\gamma} + (1-\gamma)J)\log(\beta J_W^{-1}\left(\xi^{1-\gamma} + (1-\gamma)J)\right) - \\ \qquad \frac{\beta}{1-\gamma}(\xi^{1-\gamma} + (1-\gamma)J)\log(\xi^{1-\gamma} + (1-\gamma)J) & \psi = 1. \end{cases} \tag{55}$$

Further note that

$$CJ_W = \begin{cases} \beta^\psi J_W^{1-\psi}((1-\gamma)J)^{\frac{1-\gamma\psi}{1-\gamma}} & \psi \neq 1 \\ \beta(\xi^{1-\gamma} + (1-\gamma)J) & \psi = 1. \end{cases} \tag{56}$$

For $\psi \neq 1$, substituting into (53) from (55), (56) and (32) implies

$$
\begin{aligned}
-J_\tau + J_X b - \frac{1}{2}\frac{J_W^2}{J_{WW}}\lambda^\top \left(\sigma\sigma^\top\right)^{-1}\lambda - \frac{J_W}{J_{WW}} J_{XW}^\top a\sigma^\top \left(\sigma\sigma^\top\right)^{-1}\lambda \\
+ J_W W r_f + \frac{1}{2}\mathrm{tr}(a^\top J_{XX} a) - \frac{1}{2}\frac{1}{J_{WW}} J_{XW}^\top a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top J_{XW} \\
- \frac{1}{1-\psi}\beta^\psi J_W^{1-\psi}((1-\gamma)J)^{\frac{1-\gamma\psi}{1-\gamma}} - \beta\left(1-\frac{1}{\psi}\right)^{-1}(1-\gamma)J = 0. \quad (57)
\end{aligned}
$$

The form of J (33), combined with (37) implies

$$
J(W, X, \tau) = \left(\beta^\psi H(X,\tau)\right)^{-\frac{1-\gamma}{1-\psi}}. \quad (58)
$$

Substituting (58) into (57) leads to (40).

The remainder of this section assumes the $\psi = 1$ case. Substituting into (53) from (55), (56) and (32) implies

$$
\begin{aligned}
-J_\tau + J_X^\top b - \frac{1}{2}\frac{J_W^2}{J_{WW}}\lambda^\top \left(\sigma\sigma^\top\right)^{-1}\lambda - \frac{J_W}{J_{WW}} J_{XW}^\top a\sigma^\top \left(\sigma\sigma^\top\right)^{-1}\lambda \\
+ J_W W r_f + \frac{1}{2}\mathrm{tr}(a^\top J_{XX} a) - \frac{1}{2}\frac{1}{J_{WW}} J_{XW}^\top a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top J_{XW} \\
- \beta\left(\xi^{1-\gamma} + (1-\gamma)J\right) + \beta\left(\xi^{1-\gamma} + (1-\gamma)J\right)\left(1 - \frac{1}{1-\gamma}\right)\log\left(\xi^{1-\gamma} + (1-\gamma)J\right) \\
+ \beta\left(\xi^{1-\gamma} + (1-\gamma)J\right)\log(\beta J_W^{-1}) = 0 \quad (59)
\end{aligned}
$$

Guess

$$
\frac{1}{1-\gamma}\log\left(\xi^{1-\gamma} + (1-\gamma)J(W,X,\tau)\right) = q(\tau)\log W + \log G(X,\tau). \quad (60)
$$

Derivatives of J can be found by implicitly differentiating on both sides of (60):

$$
\begin{aligned}
J_\tau &= \left(\xi^{1-\gamma} + (1-\gamma)J\right)\left(q'\log W + \frac{G_\tau}{G}\right) \\
J_W &= \left(\xi^{1-\gamma} + (1-\gamma)J\right) q\frac{1}{W} \quad (61) \\
J_X &= \left(\xi^{1-\gamma} + (1-\gamma)J\right)\frac{G_X}{G}.
\end{aligned}
$$

Second derivatives follow from (61):

$$
\begin{aligned}
J_{WX} &= \left(\xi^{1-\gamma} + (1-\gamma)J\right) q(1-\gamma)W^{-1}\frac{G_X}{G} \\
J_{WW} &= \left(\xi^{1-\gamma} + (1-\gamma)J\right) W^{-2}\left(-q + q^2(1-\gamma)\right) \quad (62) \\
J_{XX} &= \left(\xi^{1-\gamma} + (1-\gamma)J\right)\left(-\gamma\frac{G_X G_X^\top}{G^2} + \frac{G_{XX}}{G}\right).
\end{aligned}
$$

Substituting (60), (61) and (62) into (59) and dividing by $\xi^{1-\gamma} + (1-\gamma)J$ leads to the following:

$$
\begin{aligned}
&- q' \log W - \frac{G_\tau}{G} + \frac{G_X}{G} b + \frac{1}{2}\frac{q}{1-q(1-\gamma)} \lambda^\top \left(\sigma\sigma^\top\right)^{-1} \lambda \\
&\qquad\qquad + \frac{q(1-\gamma)}{1-q(1-\gamma)} \frac{G_X}{G} a \sigma^\top \left(\sigma\sigma^\top\right)^{-1} \lambda \\
&+ q r_f + \frac{1}{2}\mathrm{tr}\left(-\gamma \frac{a^\top G_X^\top G_X a}{G^2} + \frac{a^\top G_{XX} a}{G}\right) + \frac{1}{2}\frac{q(1-\gamma)^2}{1-q(1-\gamma)} \frac{G_X}{G} a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top \frac{G_X^\top}{G} \\
&- \beta - \beta\gamma(q \log W + \log G) + \beta \log(\beta q^{-1}) + \beta \log W - \beta(1-\gamma)(q\log W + \log G) = 0. \quad (63)
\end{aligned}
$$

Matching coefficients on $\log W$ implies that

$$-q' + \beta - \beta q = 0,$$

with boundary condition $q(0) = 0$. Therefore

$$q(\tau) = 1 - e^{-\beta\tau}. \quad (64)$$

The resulting differential equation for G is as follows:

$$
\begin{aligned}
&- \frac{G_\tau}{G} + \frac{G_X}{G} b + \frac{1}{2}\frac{q}{1-q(1-\gamma)} \lambda^\top \left(\sigma\sigma^\top\right)^{-1} \lambda \\
&\qquad\qquad + \frac{q(1-\gamma)}{1-q(1-\gamma)} \frac{G_X}{G} a \sigma^\top \left(\sigma\sigma^\top\right)^{-1} \lambda \\
&+ q r_f + \frac{1}{2}\mathrm{tr}\left(-\gamma \frac{a^\top G_X^\top G_X a^\top}{G^2} + \frac{a^\top G_{XX} a}{G}\right) + \frac{1}{2}\frac{q(1-\gamma)^2}{1-q(1-\gamma)} \frac{G_X}{G} a\sigma^\top \left(\sigma\sigma^\top\right)^{-1} \sigma a^\top \frac{G_X^\top}{G} \\
&\qquad\qquad - \beta - \beta\gamma \log G + \beta \log(\beta q^{-1}) - \beta(1-\gamma)\log G = 0. \quad (65)
\end{aligned}
$$

Table 1: Posterior means of β and ρ under various combinations of the likelihood and the prior

	Posterior means	
Specification	β	ρ
Conditional likelihood; $p(b,\Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-\infty, \infty)$	0.437	0.9800
Conditional likelihood; $p(b,\Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-1, 1)$	0.441	0.9798
Exact likelihood; $p(b,\Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-1, 1)$	0.375	0.9828
Exact likelihood; $p(b,\Sigma) \propto (1-\rho^2)^{-1}\sigma_v^2\|\Sigma\|^{-5/2}, \rho \in (-1, 1)$	0.276	0.9872

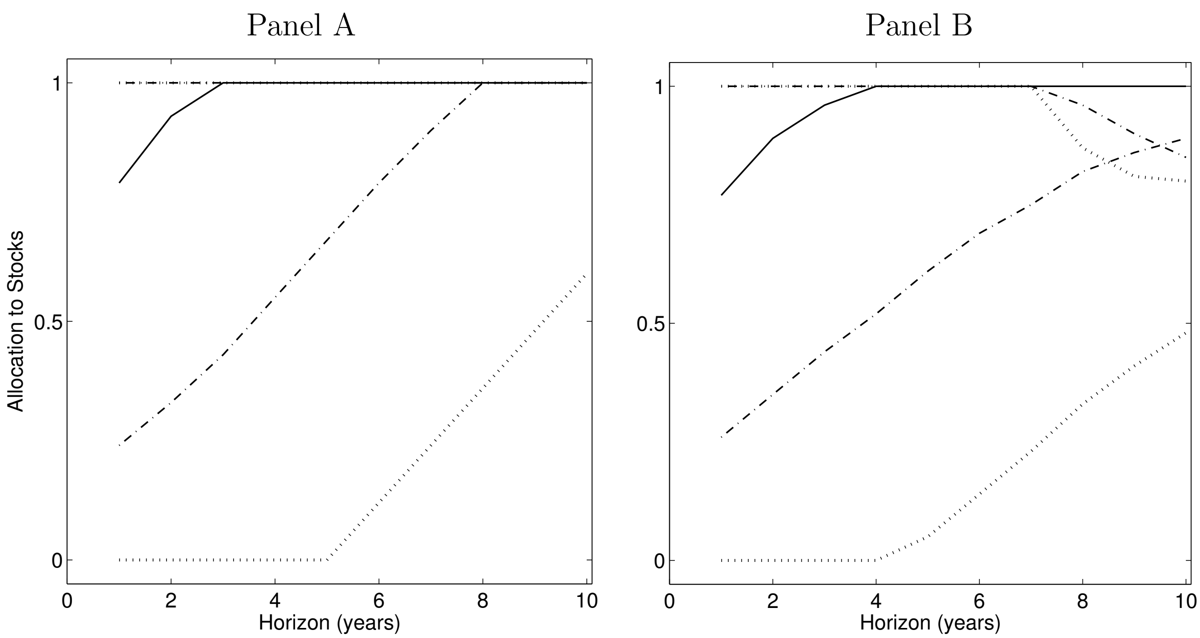
Notes: Results are from Stambaugh (1999, Figure 1). The predictor variable is the dividend-price ratio. Data are monthly from 1952–1996. The conditional likelihood refers to (13); the exact likelihood refers to (8) with initial condition given by (14).

Table 2: Expected returns and optimal allocations under various combinations of the likelihood and prior (monthly horizon)

Specification	Current dividend yield		
	3%	4%	5%
Panel A: Expected excess returns (in percent)			
Conditional likelihood; $p(b, \Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-\infty, \infty)$	2.0	7.3	12.5
Exact likelihood; $p(b, \Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-1, 1)$	1.0	5.5	10.0
Exact likelihood; $p(b, \Sigma) \propto (1 - \rho^2)^{-1}\sigma_v^2\|\Sigma\|^{-5/2}, \rho \in (-1, 1)$	1.9	5.2	8.5
Conditional MLEs as true parameters	2.0	7.3	12.5
Panel B: Stock allocation (in percent)			
Conditional likelihood; $p(b, \Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-\infty, \infty)$	22	61	97
Exact likelihood; $p(b, \Sigma) \propto \|\Sigma\|^{-3/2}, \rho \in (-1, 1)$	15	46	79
Exact likelihood; $p(b, \Sigma) \propto (1 - \rho^2)^{-1}\sigma_v^2\|\Sigma\|^{-5/2}, \rho \in (-1, 1)$	21	45	68
Conditional MLEs as true parameters	22	60	98

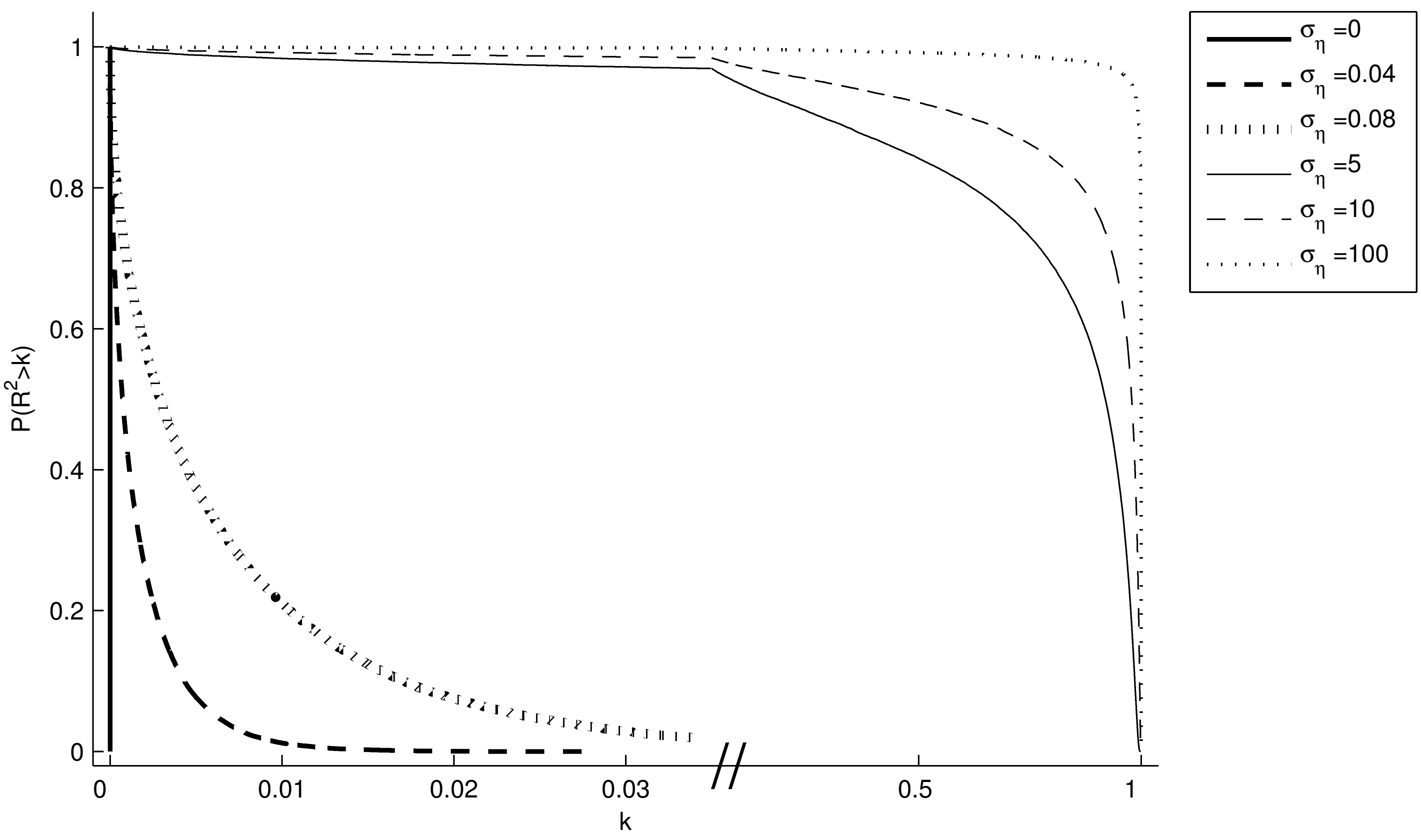
Notes: Results are from Stambaugh (1999, Table 3,4). The predictor variable is the dividend-price ratio. Data are monthly from 1952–1996. The conditional likelihood refers to (13); the exact likelihood refers to (8) with initial condition given by (14). The table assumes that the investor has a horizon of one month and has constant relative risk aversion equal to seven.

Figure 1: Static allocation as a function of horizon assuming return predictability



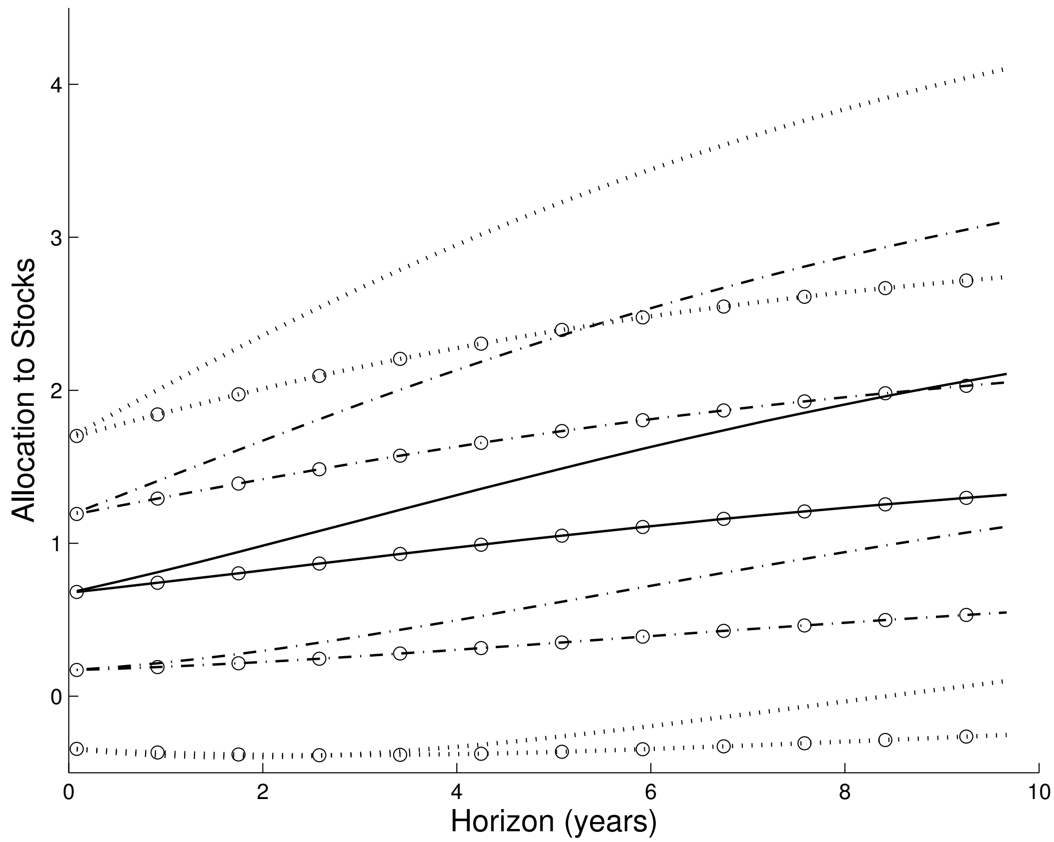
Notes: Panel A shows the allocation when there is no parameter uncertainty; Panel B incorporates parameter uncertainty. The solid line corresponds to the optimal (buy-and-hold) allocation when the dividend yield is at its sample mean (3.75%). The dash-dotted lines correspond to the allocations when the dividend yield is one standard deviation above or below its mean (2.91% and 4.59% respectively). The dotted lines correspond to the allocations when the dividend yield is two standard deviations above or below its mean (2.06% and 5.43% respectively). The agent has power utility over terminal wealth with relative risk aversion equal to 5. Note that some lines may lie on top of each other and that the allocations weakly increase as a function of the dividend yield except at very long horizons in Panel B. The model is estimated over monthly data from 1952 to 1995. This figure is adapted from Barberis (2000, Figure 3).

Figure 2: Prior distributions on the R^2 of a predictive regression



Notes: The prior probability that the R^2 exceeds a value k implied by various prior beliefs. Prior beliefs are indexed by σ_η, the prior standard deviation of the normalized coefficient on the predictor variable. The dogmatic prior is given by $\sigma_\eta = 0$; the diffuse prior by $\sigma_\eta = \infty$. Intermediate priors express some skepticism over return predictability. Note that left portion of the x-axis of the graph is scaled differently than the right portion.

Figure 3: Dynamic allocation as a function of horizon assuming return predictability



Notes: The figure shows the optimal allocation as a function of horizon, assuming the investor can trade continuously. The solid line corresponds to the optimal allocation when the dividend yield is at its sample mean (3.75%). The dash-dotted lines correspond to the allocations when the dividend yield is one standard deviation above or below its mean (2.91% and 4.59% respectively). The dotted lines correspond to the allocations when the dividend yield is two standard deviations above or below its mean (2.06% and 5.43% respectively). The agent has power utility over consumption (lines with circles) or over terminal wealth (lines without circles) with risk aversion equal to five. Note that the allocations increase as a function of the dividend yield. The model is estimated over monthly data from 1952 to 1995.

References

Andrews, Donald W. K., 1993, Exactly median-unbiased estimation of first order autoregressive/unit root models, *Econometrica* 61, 139–165.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Avramov, Doron, 2004, Stock return predictability and asset pricing models, *Review of Financial Studies* 17, 699–738.

Avramov, Doron, and Guofu Zhou, 2010, Bayesian Portfolio Analysis, *Annual Review of Financial Economics* 2.

Balduzzi, Pierluigi, and Anthony W. Lynch, 1999, Transaction costs and predictability: some utility cost calculations, *Journal of Financial Economics* 52, 47–78.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Berger, James O., 1985, *Statistical decision theory and Bayesian analysis.* (Springer New York).

Bernstein, Peter L., 1992, *Capital Ideas.* (Free Press New York, NY).

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing statistical criteria to select return forecasting models: What do we learn?, *Review of Financial Studies* 12, 405–428.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.

Box, George, Gwilym M. Jenkins, and Gregory Reinsel, 1994, *Time Series Analysis: Forecasting and Control, 3rd Edition.* (Holden-Day Inc. S. Francisco, CA).

Brandt, Michael W., 1999, Estimating portfolio and consumption choice: A conditional Euler equations approach, *Journal of Finance* 54, 1609–1645.

Brandt, Michael W., 2009, Portfolio choice problems, in Yacine Ait-Sahalia, and Lars Peter Hansen, eds.: *Handbook of Financial Econometrics* (North Holland, Amsterdam, The Netherlands).

Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud, 2005, A simulation approach to dynamics portfolio choice with an application to learning about return predictability, *Review of Financial Studies* 18, 831–873.

Brennan, Michael, 1998, The role of learning in dynamic portfolio decisions, *European finance review* 1, 295–306.

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

Campbell, John Y., 1993, Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487–512.

Campbell, John Y., 2003, Consumption-based asset pricing, in *Handbook of the Economics of Finance Volume IB* (North-Holland, Amsterdam, The Netherlands).

Campbell, John Y., 2006, Household Finance, *Journal of Finance* 61, 1553 – 1604.

Campbell, John Y., 2008, Viewpoint: Estimating the equity premium, *Canadian Journal of Economics* 41, 1–21.

Campbell, John Y., George Chacko, Jorge Rodriguez, and Luis M. Viceira, 2004, Strategic asset allocation in a continuous-time VAR model, *Journal of Economic Dynamics and Control* 28, 2195–2214.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time-varying, *Quarterly Journal of Economics* 114, 433–495.

Campbell, John Y., and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.

Chacko, George, and Luis Viceira, 2005, Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets, *Review of Financial Studies* 18, 1369–1402.

Chamberlain, Gary, 1999, Econometric applications of maxmin expected utility, Working paper, Harvard University.

Chen, Hui, Nengjiu Ju, and Jianjun Miao, 2009, Dynamic asset allocation with ambiguous return predictability, working paper.

Chen, Zengjing, and Larry Epstein, 2002, Ambiguity, risk and asset returns in continuous time, *Econometrica* 70, 1403–1443.

Chib, Siddhartha, and Edward Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *American Statistician* 49, 327–335.

Cochrane, John H., 1992, Explaining the variance of price-dividend ratios, *Review of Financial Studies* 5, 243–280.

Cochrane, John H., 1999, Portfolio advice for a multifactor world, *Economic Perspectives* Federal Reserve Bank of Chicago 23, 59–78.

Cochrane, John H., 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Cox, John C., and Chi-Fu Huang, 1989, Optimal consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory* 49, 33–83.

Cox, John C., Jonathan C. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–408.

Cremers, K.J. Martjin, 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.

Cuoco, Domenico, 1997, Optimal consumption and equilibrium prices with portfolio constraints and stochastic income, *Journal of Economic Theory* 72, 33 – 73.

Curcuru, Stephanie E., John Heaton, Deborah Lucas, and Damien Moore, 2009, Heterogeneity and portfolio choice: Theory and evidence, in Yacine Ait-Sahalia, and Lars Peter Hansen, eds.: *Handbook of Financial Econometrics* (North Holland, Amsterdam, The Netherlands).

Detemple, Jerome B., 1986, Asset Pricing in a Production Economy with Incomplete Information, *The Journal of Finance* 41, 383–391.

Duffie, Darrell, 1996, *Dynamic Asset Pricing Theory.* (Princeton University Press Princeton, NJ).

Duffie, Darrell, and Larry G Epstein, 1992a, Asset pricing with stochastic differential utility, *Review of Financial Studies* 5, 411–436.

Duffie, Darrell, and Larry G. Epstein, 1992b, Stochastic differential utility, *Econometrica* 60, 353–394.

Duffie, Darrell, and Costis Skiadas, 1994, Continuous-time asset pricing: A utility gradient approach, *Journal of Mathematical Economics* 23, 107–132.

Dybvig, Philip H., and Chi-Fu Huang, 1988, Nonnegative wealth, absence of arbitrage, and feasible consumption plans, *Review of Financial Studies* 1, 377–401.

Dybvig, Philip H., L. C. G. Rogers, and Kerry Back, 1999, Portfolio Turnpikes, *The Review of Financial Studies* 12, 165–195.

Epstein, Larry, and Stan Zin, 1989, Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

Epstein, Larry G., and Stanley E. Zin, 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis, *Journal of Political Economy* 99, 263–286.

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

Fama, Eugene F., and William G. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang, 2007, Portfolio selection with parameter and model uncertainty: A multi-prior approach, *Review of Financial Studies* 20, 41–81.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 1996, *Bayesian Data Analysis.* (Chapman & Hall/CRC London, UK).

Gennotte, Gerard, 1986, Optimal Portfolio Choice Under Incomplete Information, *The Journal of Finance* 41, 733–746.

Gilboa, Itzhak, and David Schmeidler, 1989, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141–153.

Goetzmann, William N., and Philippe Jorion, 1993, Testing the predictive power of dividend yields, *Journal of Finance* 48, 663–679.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Graham, Benjamin, and David L. Dodd, 1934, *Security Analysis*. (McGraw Hill New York) 1 edn.

Hamilton, J. D., 1994, *Time Series Analysis*. (Oxford University Press Princeton, NJ).

Hansen, Lars Peter, 2007, Beliefs, doubts and learning: Valuing macroeconomic risk, *The American Economic Review* 97, 1–30.

He, Hua, and Neil D. Pearson, 1991, Consumption and portfolio policies with incomplete markets and short-sale constraints: The infinite dimensional case, *Journal of Economic Theory* 54, 259–304.

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

Jeffreys, Harold, 1961, *Theory of Probability*. (Oxford University Press Clarenden).

Johannes, Michael, and Nicholas Polson, 2006, MCMC methods for financial econometrics, in Yacine Ait-Sahalia, and Lars Hansen, eds.: *Handbook of Financial Econometrics* (Elsevier, North-Holland).

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

Karatzas, Ioannis, John P. Lehoczky, and Steven E. Shreve, 1987, Optimal portfolio and consumption decisions for a small investor on a finite horizon, *SIAM Journal of Control and Optimization* 25, 1557–1586.

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

Kihlstrom, Richard, 2009, Risk aversion and the elasticity of substitution in general dynamic portfolio theory: Consistent planning by forward looking, expected utility maximizing investors, *Journal of Mathematical Economics* 45, 634 – 663.

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

Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.

Kreps, D., and E. Porteus, 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46, 185–200.

Lettau, Martin, and Sydney C. Ludvigson, 2001, Consumption, aggregate wealth and expected stock returns, *Journal of Finance* 56, 815–849.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

Liu, Jun, 2007, Portfolio selection in stochastic environments, *Review of Financial Studies* 20, 1–39.

Maenhout, Pascal, 2006, Robust portfolio rules and detection-error probabilities for a mean-reverting risk premium, *Journal of Economic Theory* 128, 136–163.

Markowitz, Harry, 1952, Portfolio Selection, *The Journal of Finance* 7, 77–91.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

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

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

Pastor, Lubos, and Robert F. Stambaugh, 2009a, Are stocks really less volatile in the long run?, working paper.

Pastor, Lubos, and Robert F. Stambaugh, 2009b, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583 – 1628.

Pastor, Lubos, and Pietro Veronesi, 2009, Learning in Financial Markets, *Annual Review of Financial Economics* 1, 361–381.

Pliska, Stanley R., 1986, A Stochastic calculus model of continuous trading: Optimal portfolios, *Mathematics of Operations Research* 11, 371–382.

Poterba, James M., and Lawrence H. Summers, 1988, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics* 22, 27–59.

Samuelson, Paul A, 1969, Lifetime Portfolio Selection by Dynamic Stochastic Programming, *The Review of Economics and Statistics* 51, 239–46.

Schroder, Mark, and Costis Skiadas, 1999, Optimal consumption and portfolio selection with stochastic differential utility, *Journal of Economic Theory* 89, 68–126.

Shanken, Jay A., and Ane Tamayo, 2005, Dividend yield, risk, and mispricing: A Bayesian analysis, Working paper, Emory University and London Business School.

Siegel, Jeremy J., 1994, *Stocks for the long run: a guide to selecting markets for long-term growth.* (Irwin Burr Ridge, IL).

Sims, Christopher A., and Harald Uhlig, 1991, Understanding unit rooters: A helicopter tour, *Econometrica* 59, 1591–1599.

Skiadas, Costis, 2007, Dynamic portfolio choice and risk aversion, in John Birge, and Vadim Linetsky, eds.: *Handbooks in operations research and management science: Financial Engineering* (North-Holland, Amsterdam, The Netherlands).

Skoulakis, Georgios, 2007, Dynamic portfolio choice with Bayesian learning, Working paper, University of Maryland.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Uhlig, Harald, 1994, On Jeffreys prior when using the exact likelihood function, *Econometric Theory* 10, 633–644.

Wachter, Jessica A., 2002, Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets, *Journal of Financial and Quantitative Analysis* 37, 63–91.

Wachter, Jessica A., 2008, Can time-varying risk of rare disasters explain aggregate stock market volatility?, NBER working paper #14386.

Wachter, Jessica A., and Missaka Warusawitharana, 2009a, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.

Wachter, Jessica A., and Missaka Warusawitharana, 2009b, What is the chance that the equity premium varies over time? Evidence from predictive regressions, Working paper, Board of Governors of the Federal Reserve and University of Pennsylvania.

Weil, Philippe, 1990, Nonexpected utility in macroeconomics, *Quarterly Journal of Economics* 105, 29–42.

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

Zellner, Arnold, 1971, *An introduction to Bayesian inference in econometrics.* (John Wiley and Sons, Inc. New York, NY).