Asset Allocation under Distribution Uncertainty

Marcin Kacperczyk*  Paul Damien†

April 2011‡

ABSTRACT

This paper shows how uncertainty about the type of return distribution (distribution uncertainty) can be incorporated in asset allocation decisions by using a novel, Bayesian semiparametric approach. To evaluate the economic importance of distribution uncertainty, the extent of changes in ex-ante optimal asset allocations of investors who factor in distribution uncertainty into their portfolio model is examined. The key findings are: (a) distribution uncertainty is highly time varying; (b) compared to investors facing parameter uncertainty, investors under distribution uncertainty, on average, allocate less money to risky assets; their allocations are less variable; and their certainty-equivalent losses from ignoring distribution uncertainty can be economically significant; (c) portfolio strategies of such investors generate statistically higher returns, even after controlling for common factors.

*Department of Finance Stern School of Business and NBER, New York University, mkacper@stern.nyu.edu.
†Department of Statistics McCombs School of Business, University of Texas, pdamien@mccombs.utexas.edu.
‡We thank Raymond Kan, Richard Kihlstrom, Lutz Kilian, Zbigniew Kominek, Wojciech Kopczuk, Łuboś Pástor, Jacob Sagi, Amit Seru, Tyler Shumway, Clemens Sialm, Rob Stambaugh, Jeremy Stein, Pietro Veronesi, Jessica Wachter, Stephen Walker, Neng Wang, Tan Wang, Yihong Xia, Kathy Yuan, conference participants at the AFA and NFA meetings, and seminar participants at GSB Chicago, Florida State, Fordham, Georgia State, HEC Paris, Miami, Michigan, UBC, and Wharton for helpful comments.
Since the seminal paper by Markowitz (1952), many studies have recognized the importance of the future return distribution for optimal asset allocation. Most of the studies assume that investors know the distributional form of future returns.\(^1\) In practice, however, empirical evidence from financial markets seems to suggest the opposite. For example, distributions of asset returns tend to switch between different regimes. Also, the presence of rare events may perturb beliefs about the form of distributions. Finally, investors tend to differ in their assessment of future returns;\(^2\) hence, any consensus about the precise nature of the underlying stochastic process driving returns seems quite difficult to reach.

The objective of this paper is two fold. From a theoretical perspective, the paper proposes a novel Bayesian semiparametric method to incorporate uncertainty about the type of return distribution (distribution uncertainty) to obtain an optimal mix between a risky and a riskless asset. From an application perspective, the paper characterizes the ex-ante asset allocation decisions of investors who factor in distribution uncertainty into their portfolio model.

Several studies have noted that when constructing economic forecasts, any prediction should necessarily account for parameter uncertainty, generally defined as estimation risk.\(^3\) But such studies assume that the form of the distribution of returns is known. In contrast, this paper addresses the problem of distribution uncertainty.

This paper compares the portfolio decisions of investors accounting for both parameter and distribution uncertainty, henceforth referred to as distribution uncertainty, with the portfolio decisions of investors considering parameter uncertainty only. The latter approach postulates that the distribution of future log returns is approximately Gaussian, primarily because almost all studies on parameter uncertainty assume such a distributional form (e.g., Kandel and Stambaugh (1996), Barberis (2000)). In contrast, the type of distribution uncertainty considered in this paper utilizes a semiparametric model. The conditional distribution of the asset return is

---

\(^1\)The most common description of realized returns involves the Normal or Log-Normal distribution. Significant departures from normality have been illustrated, among others, by Fama (1965), Affleck-Graves and McDonald (1989), and Richardson and Smith (1993).

\(^2\)For evidence on the above three points, see 1) Pástor and Stambaugh (2001), Ang and Bekaert (2004), Guidolin and Timmermann (2007); 2) Liu et al. (2003), Liu et al. (2005); 3) Welch (2000).

\(^3\)Markowitz (1952) was the first to recognize the presence of estimation risk in finance. Later studies on this subject include, among others, Bawa et al. (1979), Barberis (2000), and Tu and Zhou (2004) for portfolio with i.i.d. asset returns, and Kandel and Stambaugh (1996), Barberis (2000), and Xia (2001) for portfolio with predictable asset returns.
assumed to be unimodal, with any degree of skewness and kurtosis modeled nonparametrically using the Dirichlet Process mixture model. Simultaneously, the mean and variance of the asset returns are modeled using parametric regressions.

A Bayesian nonparametric approach to modeling distribution uncertainty is developed in this paper because the Bayesian paradigm lends itself readily to treating both parameters and distributions as random quantities. The economic motivation for using a nonparametric prior comes from the apparent difficulties in characterizing real data, suggesting that investors may not know the underlying distribution of asset returns. Moreover, accounting for distribution uncertainty seems to better describe economic data. For example, Barro (2006) shows that incorporating rare events into asset pricing models helps explain a number of puzzling economic phenomena. Hence, investors may want to allow maximum flexibility in forming their beliefs about the predictive distribution of asset returns. Operationally, modeling such beliefs of investors requires nonparametric priors, such as the Dirichlet Process prior, to allow for the random form of the entire distribution function.

To assess the importance of distribution uncertainty in the data and obtain direct comparisons to earlier studies, we consider investors who make their portfolio decisions for one period ahead. Since investors’ beliefs about the stability of the relationships in the underlying data may differ substantially, the following assumptions are made: (a) investors consider the entire past available data, and (b) investors consider a sample of the most recent 10 years of data. Subsequently, optimal portfolio allocations of investors who believe that returns are conditionally independently and identically distributed (i.i.d.) are studied. The i.i.d. assumption, though somewhat restrictive if one believes in predictability, allows us to quantify the direct impact of distribution uncertainty on asset allocation. Nevertheless, since predictability, especially in variance, may have a significant impact on the predictive distribution this study also contrasts the i.i.d. results with those that additionally factor in volatility timing.

To gauge the ex-ante impact of distribution uncertainty over time, we consider a series of portfolio allocations in S&P 500 index and Treasury bills for each January between 1964 and 2004. Results indicate that investors who incorporate distribution uncertainty into their decision processes tend to invest considerably less in the stock market than investors who
merely consider parameter uncertainty. For example, for investors with risk aversion of 3, the difference in average allocations equals approximately 35 percentage points. Further, the variability of their portfolio weights is smaller, and thus economically more plausible. The difference between the two alternatives is especially large for investors who condition their decisions on a rolling estimation window. Also, it mainly originates in the differences between distributions of future returns rather than in the choice of a particular utility function. Finally, incorporating volatility timing, though statistically important, does not lead to considerably different allocations. These results suggest that distribution uncertainty, above and beyond parameter uncertainty, may play an important role in the decision process.

To understand the source of the results, we focus on describing the important differences between two types of uncertainty with respect to mean, variance, and kurtosis of predictive distributions. Under parameter uncertainty, the expected value of the predictive distribution is equal to the sample mean. In contrast, under distribution uncertainty this value may shift away from the sample mean. The magnitude of the shift varies over time and depends on the degree of uncertainty in the data. For the sample analyzed in this paper, the average effect is a decrease in the first moment. Statistically, the extent of the uncertainty is captured by higher-order moments of the predictive distribution. While under parameter uncertainty these higher moments are time invariant, and fail to match the moments of the empirical distribution, higher moments obtained under distribution uncertainty are on average higher, vary greatly over time, and strongly correlate with their counterparts in the data.

The departure from the parameter uncertainty setting also introduces interesting implications for the variance of the distribution. As mentioned above, parameter uncertainty is captured by the increase in the variance of the predictive distribution. Under distribution uncertainty, investors may fit uncertainty by assigning greater weight to extreme observations without necessarily changing their beliefs about the variance. This effect manifests itself with larger higher-order moments of the distribution. In addition, investors’ perception of the variance may be affected by the presence of the “nonstationarity risk”. Under the presence of such risk, investors accounting for distribution uncertainty have less information about the future and thus face a greater level of uncertainty. This uncertainty leads to a higher perceived
level of variance.\footnote{Barry and Winkler (1976) are among the first to study the importance of “nonstationarity risk” in the portfolio context.} Consequently, the variance perceived under distribution uncertainty is a combination of two effects: the negative effect coming from flexible higher-order moments and the positive effect coming from nonstationarity. The resulting value of variance depends on which of the two is stronger. For the sample examined in this paper, nonstationarity plays a bigger role for investors who base their decisions on a 10-year rolling estimation window, while flexible moments have a dominating role for investors who use a cumulative window.

The differences in optimal portfolio allocations lead to economically significant utility effects. In particular, investors with beliefs centered on distribution uncertainty, when forced to allocate wealth according to a menu optimal for investors ignoring distribution uncertainty, could suffer economically significant losses measured in terms of certainty-equivalent returns. The magnitude of the losses is highly time varying and, for investors with a 10-year rolling estimation window, can be as low as 0.03%, but also as high as 8.5% per year. This relatively large spread in returns suggests that the salient feature of distribution uncertainty is its time variation. As an example, distribution uncertainty has a higher impact in periods that are marked with significant market activity, such as the oil crisis of the 70s or a “dot-com mania”.

Further, the quality of predictions generated is assessed by analyzing the out-of-sample performance of investment strategies which factor in distribution uncertainty. Results show that the strategies perform remarkably better than the corresponding strategies under parameter uncertainty. The superior performance is starkly evident for investors with a rolling estimation window. Moreover, strategies based on distribution uncertainty deliver a significantly better risk-adjusted performance when compared to both strategies based on parameter uncertainty, and the strategy tracking the market portfolio.

The literature on asset allocation under Bayesian uncertainty has largely focused on analyzing parameter uncertainty;\footnote{The empirical evidence on a static portfolio choice includes, among others, Klein and Bawa (1976), Barberis (2000), Polson and Tew (2000), and Kan and Zhou (2007) in the context with no predictability, and Kandel and Stambaugh (1996), Brennan (1998), Barberis (2000), and Xia (2001) in the context with predictability.} the distributions of future returns in these studies are fixed and typically approximately Gaussian. Few papers are exceptions to this paradigm. Bawa (1979) outlines the important theoretical concepts with respect to portfolio selection of multi-
ple risky assets in a purely nonparametric setting, yet he does not deliver extensive empirical results. Tu and Zhou (2004) model uncertainty in the data-generating process by measuring the distance between decisions based on Normal distribution and decisions based on various \(t\) distributions. Their framework, however, is still anchored in a parameter uncertainty world since investors are restricted to choosing among various forms of \(t\) distributions. Here the focus is on a richer class of all possible unimodal distributions, with arbitrary levels of skewness and kurtosis, which includes the \(t\) distribution family as a special case. Harvey et al. (2010) relax the normality assumption by providing a framework in which investors also account for higher-order moments, but they focus explicitly on the significance of skew in the stock-level data. Brandt et al. (2005) consider the dynamic portfolio choice problem under arbitrary distribution functions. However, in each iteration, the distribution function in their model is fixed. Also, all the studies investigate optimal allocations taking into account only one estimation window. As such, they ignore the fact that distribution uncertainty is highly varying. This paper extends the literature in that it simultaneously considers parameter and distribution uncertainty via the semiparametric approach and positions distribution uncertainty in a richer empirical context.

This study is also closely related to the Bayesian literature on model uncertainty. Model uncertainty, as defined in the literature, is the inability to specify the true model; hence, one has to place priors on each of the possible subsets of models. The existing literature, however, assumes that the distribution of the error term under each of these possible models is parametric and its functional form, typically, is fully specified – usually as a Normal distribution (e.g., Pástor and Stambaugh (1999), Avramov (2002), and Cremers (2002)). In contrast to the above studies, we consider model uncertainty from the perspective of the data-generating process, where the distribution of the error term is not fully specified and is treated as a random parameter. This is clearly distinct from the Bayesian model averaging approach utilized under the former approach.\(^6\)

\(^6\)An alternative approach to analyzing model uncertainty which is close in spirit to this study is that based on ambiguity aversion (e.g., Epstein and Wang (1994), Anderson et al. (1998), and Hansen et al. (1999)). Of special interest for the current study are papers on asset allocation. Dow and Werlang (1992) and Maenhout (2004) study a one-period portfolio choice problem. Uppal and Wang (2003), Garlappi et al. (2007), and Leippold et al. (2008) study the asset allocation with multiple risky assets.
The remainder of this paper proceeds as follows. Section 1 describes the framework to model distribution uncertainty. *Ex-ante* implications resulting from incorporating distribution uncertainty, such as the level and variability of portfolio allocations as well as certainty-equivalent losses, are studied in Section 2. Section 3 investigates the *ex-post* performance of various related strategies, while Section 4 reviews the main conclusions of the paper.

1. Methodology

For expositional reasons, this section mostly focuses on intuition and main components of the framework, while the Appendices present the technical details.

1.1. Economy

Consider investors with a one-period investment horizon facing the decision of allocating their wealth at time $T$, $X_T$, between a risky and a riskless asset, with their respective one-period, continuously compounded returns of $R_{T+1}$ and $R_{fT}$. For simplicity, consider investors with standard mean-variance preferences, with the coefficient of relative risk aversion $A$, who derive utility from their terminal wealth, $V(X_{T+1})$. Assume that investors do not have any labor or real estate income. The objective of the investors is to find an optimal allocation of resources to maximize expected utility at time $T$:

$$\hat{\omega}_T = \arg\max_E E_T[V(X_{T+1})] ,$$

where

$$X_{T+1} = X_T(\omega_T \exp(r_{T+1} + R_{fT}) + (1 - \omega_T) \exp(R_{fT})) ,$$

$r_{T+1} \equiv R_{T+1} - R_{fT}$ denotes excess market return, and $\hat{\omega}_T$ represents an optimal allocation of wealth into a risky asset at period $T$. Throughout the paper, we assume that investors can take both short and margin positions in the risky asset, and their starting wealth position, $X_T$, equals one. Further, in the short run, the risk-free rate, $R_{fT}$, is assumed to be deterministic.
and inflation risk is small enough that one can ignore it. Also, there are no taxes and no transaction costs. Finally, investors have access to all past information up to time $T$.

If we combine (1) and (2), and rewrite the conditional expectation in (1) in terms of the predictive distribution of future returns, $p(r_{T+1} | \cdot)$, the optimization problem of the investors can be represented as:

$$\hat{\omega}_T = \arg \max V(\omega_T \exp(r_{T+1} + R_{FT}) + (1 - \omega_T) \exp(R_{FT})) \ p(r_{T+1} | y_T, \lambda) \ dr_{T+1}, \quad (3)$$

where $\lambda$ is a set of parameters describing the predictive distribution of the excess returns, and $y$ subsumes the information contained in past data. In order to describe the predictive distribution, $p$, we assume that the excess market return is generated by the following linear random process:

$$r_{t+1} = \mu(t) + e_{t+1} \quad t = 1, \ldots, T, \quad (4)$$

where $\mu(t)$ means that $\mu$ could either be constant or could depend on a set of predictor variables up to time $t$. The important aspect of this equation is that it does not impose any distributional form on the error term. For the moment, assume merely that $e$ may come from any distribution with mean 0 and variance $\sigma^2$. The subsequent sections show how this particular component of the return process can be used to generate different forms of distribution uncertainty.

1.2. Parameter Uncertainty

In their decisions, the specification in equation (3) requires investors to use the distribution of future excess market returns, $r$. From (4), it is clear that this distribution depends on past data and a set of parameters describing the distribution. Klein and Bawa (1976) and later studies note that true parameters of the distribution are not known to investors as they are estimated with error. Hence, using mean values of the parameters to construct portfolios induces an estimation risk. Therefore, in order to properly assess the risk, investors must account for the underlying parameter uncertainty in their strategies. This is done using a Bayesian approach wherein each uncertain and unknown parameter is assigned a prior distribution. The most widely developed class of models in this context is based on the work of Zellner (1971). Briefly,
such models start with the usual multiple regression model with Normal errors. Parameter uncertainty is captured by assigning the regression coefficients a multivariate Normal prior, while the variance of the error term is assigned a Gamma prior. The resulting predictive distribution is then a student-$t$.

1.3. Distribution Uncertainty

The key aspect to modeling distribution uncertainty can be illustrated using equation (4). Specifically, in the previous section it was noted that the main assumption behind modeling estimation risk is that the conditional distribution of the asset return is Normal (or has some other finite dimensional parametric distribution), leading to a parametric form for the predictive distribution of the returns; this form is fixed. What varies over time are the moments of the predictive distribution. Under distribution uncertainty, this assumption is relaxed by stating that the conditional distribution of the returns itself is uncertain. Operationally this means one has to place a prior distribution on a wide class of distribution functions. Any member of this class could potentially be the conditional distribution of returns. It is this random feature in the modeling framework that is labeled distribution uncertainty in this paper.

To model distribution uncertainty, we adapt the Semiparametric Scale Mixture of Betas (SSMB) family of models discussed in Kacperczyk et al. (2011). Their approach is based primarily on the notion of “scale mixture representation”, which is an idea dating back to Feller (1971) and the references therein. Before delving into the mathematics, a brief intuitive understanding of the representation is as follows. Consider equation (4). Typically, one writes the sampling distribution of $r$, say as a Normal$(0, \sigma^2)$, or some other distribution $f(r)$. Feller generalized this by first introducing an auxiliary variable $U$. The distribution of $U$, $f(U)$, could be any probability density. Suppose we have a random draw, $u$, from $f(U)$. Armed with this $u$, Feller writes the conditional distribution of $r$ as a uniform distribution, Uniform$(a, b)$, where $b$ is now a product of $u$ and the standard deviation of the sampling distribution $f(r)$; hence the phrase, “scale mixture representation”. This is a remarkable idea since now one could induce much more flexibility in modeling the higher-order moments of the sampling distribution $f(r)$; specifically, skewness and high levels of kurtosis are readily handled by appropriately choosing
In the next paragraph, we show it is easy to obtain the Normal distribution as a special case for the form of \( f(r) \).

For the moment, we assume that investors believe that the conditional distribution of the excess returns is unimodal. Since we want a unimodal density for the returns, we use uniform and beta distributions in the scale mixture model. With normal kernels we could get a multimodal density, which does not make sense in our context since outliers will be modeled incorrectly. Also, we want heavier tails rather than another mode. Additionally, we simplify the intuition underlying our model by first addressing symmetric distributions for the sampling distribution of the data. This simplification is practically relevant as our empirical application considers monthly excess market returns. Campbell et al. (1997) point out that the observed deviations from normality observed in the monthly returns are more pronounced as a result of excess kurtosis than skewness. In principle, the framework with symmetric distributions still appeals to the notion of distribution uncertainty in that one does not have to assume a particular form for the underlying sampling distribution of the data. With \( r \) denoting observed data, and \( U \) denoting a latent mixing random variable, Feller’s (1971) formulation of the conditional distribution of \( r \) is given by:

\[
f(r|U = u) \sim \text{Uniform}(\mu - \sigma \sqrt{u}, \mu + \sigma \sqrt{u}),
\]

\[ u \sim F, \tag{5} \]

for some distribution function \( F \) with support on \((0, \infty)\). As \( F \) ranges over all such distribution functions, the density of \( r \) ranges over all unimodal and symmetric density functions. Consequently, with flexible \( F \), such a model can capture wide ranges of kurtosis in the data. To ensure maximum flexibility we model \( F \) nonparametrically. Kacperczyk et al. (2011) show that \( \sqrt{u} \) rather than \( u \) in the formulation above is helpful since one can express higher moments for \( r \) in terms of lower moments for \( U \). They rewrite the model as \( f(r|U) = \sigma \sqrt{U} (1 - 2\text{beta}(1, 1)) \) which will suggest the form of generalizations to asymmetric or skewed densities. An interesting fact that is used later on is noted here: if \( F \) is distributed Gamma\((3/2, 1/2)\), then the distribution of \( r \) is Normal\((0, \sigma^2)\). Similarly, by changing the specifications of the parameters
in $F$, one can obtain other commonly used distributions, such as $t$, generalized exponential, etc.

To obtain an intuitive picture of how distribution uncertainty affects decisions, consider return observations $r_1, \ldots, r_n$ from some unknown distribution $F$. Let $F$ be assigned a Dirichlet Process prior, denoted $\text{Dir}(c, F_0)$. Then Ferguson (1973) showed that the posterior process $F$ has parameters given by $c+n$ and $(cF_0 + nF_n)/(c+n)$. $F_n$ is the empirical distribution function for the data, that is the step function with jumps of $1/n$ at each $r_i$. The classical maximum likelihood estimator is given by $F_n$. (See, also, Appendix A for details on the Dirichlet Process.)

The posterior mean of $F(A)$ for any set $A \in (-\infty, \infty)$ is given by

$$E\{F(A) | \text{data}\} = p_n F_0(A) + (1 - p_n) F_n(A),$$

(6)

where $p_n = c/(c+n)$. Thus, when a new data point arrives, it either shares the component of some previously drawn value or it uses a newly generated component realized from a distribution $F_0$. The frequency with which new components are generated is controlled by a parameter $c > 0$. In a simplified example, if $c = 1$, we place a weight of $1/(n+1)$ on our prior $F_0$ and $n/(n+1)$ on $F_n$. As $c$ increases, the posterior mean is influenced more by the prior, that is, we have more faith in our prior choice, $F_0$. If one took the prior $F_0$ to be Normal, then as $c \to \infty$, one would converge to the standard parameter uncertainty result, that is, the predictive distribution would be approximately Gaussian. On the other hand, if $c \to 0$, the predictive distribution would be tilted more towards the empirical cdf, which could very likely not be Normal. As a result, one can see that the predictive distribution can take any random shape between the prior and the empirical distribution. This is exactly the mechanism that pushes the predictive under distribution uncertainty away from the Gaussian approximation under parameter uncertainty. See, also, Appendix A.

Apart from endogenizing kurtosis (like we did above), one might be interested in modeling the extent of skewness. The extension to asymmetric densities is quite straightforward. Since the uniform density is a beta(1,1) density, we can introduce skewness by having instead a
\[ \beta(1, a) \] (or a \( \beta(a, 1) \)) density, for some parameter \( a > 0 \). The first equation of the model in (5) becomes
\[ f(r|U, a) = a^{-1}(1 + a)\sigma \sqrt{U} \{1/(1 + a) - \beta(1, a)\}. \] (7)

We recover (5) when \( a = 1 \). We prefer to work with (5) for reasons discussed in the data section. This does impact the Markov chain Monte Carlo (MCMC) scheme. We provide a new algorithm in Appendix B for the following semiparametric model. With \( \mathcal{F}_t \) denoting all the data up to and including time \( t \), for each \( t \),
\[ f(r_t|\mathcal{F}_{t-1}, U_t = u_t) \sim \text{Uniform}(\mu_t - \sigma_t \sqrt{u_t}, \mu_t + \sigma \sqrt{u_t}), \]
\[ U_1, \ldots, U_t \sim p(u_1, \ldots, u_t), \] (8)

where \( p(u_1, \ldots, u_t) \) is taken to be a Dirichlet process.

The parameter uncertainty component: Equation (4) defines a random process for the excess market return. Assuming that both the mean and variance of this process can be explained by a linear combination of various predictor variables, one can specify the following parametric structure. Consider the mean regression,
\[ \mu_t = \beta_0 + \sum_{k=1}^{K} \beta_k Z_{kt}, \] (9)

where \( \beta_0, \ldots, \beta_K \) are parameters to be estimated and the \{\( Z_{kt} \)\} are observed predictor variables affecting the mean process up to and including time \( t \). The variance regression is given by,
\[ \sigma_t = \exp(\theta_0 + \sum_{l=1}^{L} \theta_l W_{lt}), \] (10)

where \( \theta_0, \ldots, \theta_L \) are parameters to be estimated and the \{\( W_{lt} \)\} are observed predictor variables affecting the variance process up to and including time \( t \).

While the model above is general, in this paper, we assume expected returns are not predictable, because we prefer not to confound the impact of mean predictors with that of
distribution uncertainty. Likewise, we use the volatility regression in only one empirical segment where it plays a key role. There, we use the squared past log returns calculated as $W_t = (\ln(S_t/S_{t-1}))^2$, where $S$ denotes the value of the S&P 500 index. We emphasize that our results are not significantly dependent on this particular choice of regressor.

1.3.1. Prior Distributions

This section describes the various priors used in the empirical analysis. In specifying priors, the focus is on two components: the distribution uncertainty and the parameter uncertainty components.

**The distribution uncertainty component:** The latent variable $U$ is sampled from the Dirichlet Process with scale parameter $c$ and centered around $F_0$. In a Bayesian context, one has to assign prior distributions for $c$ and $F_0$. Specifically, we assign the scale parameter, $c$, a Gamma($a_0, b_0$) distribution with $a_0 = b_0 = .01$, to be noninformative. Since most of the existing studies tend to assume normality of the predictive distribution, we center the prior $F_0$ around the Normal distribution, that is, the location parameter, $F_0$, of the Dirichlet Process is assigned a Gamma($3/2, 1/2$) distribution.\(^7\)

**The parameter uncertainty component:** We assume prior distributions for each $\beta_k$ to be independent Normal distributions with zero means and variances $\psi_k^2$. Similarly, we assign Normal priors for $\theta_l$. In this paper, we do not use predictors for the mean regression. Hence, we only have a prior for the intercept, $\beta_0$, which is taken to be a Normal(0, 10), an extremely vague prior choice.

Note first that one can readily reduce the above class of models to the case of conditional i.i.d. returns, based on a Bayesian nonparametric model, by merely eliminating all predictors, which is what we do in all but one empirical analysis. Second, if interest is also on one-period ahead mean predictability, then one could eliminate predictors in the variance re-

\(^7\)The distributional form for $F_0$ comes from the mixture of uniforms specification with $[r | U = u] \sim U(\mu - \sigma u^{\alpha}, \mu + \sigma u^{\alpha})$ and $U \sim Gamma(3/2, 1/2)$, which results in the marginal distribution for $r$ being Normal.
gression. Lastly, if the variance process is also of interest then the above general representation encapsulates that facet as well.\(^8\)

Since we wish to compare distribution uncertainty to parameter uncertainty, for the latter we use the standard Normal-Inverse Gamma model, assuming normal errors. Under this set-up the predictive distribution is obtained as in Barberis (2000); these sampled values are then used to make inferences about portfolio selection. Again, vague priors were employed, as in other studies such as Barberis (2000).

The Gibbs sampler for our SSMB model is presented in Appendix B. For both parameter and distribution uncertainty simulations, a total of 750,000 samples are drawn. A “burn-in” of 50,000 values is employed, and then every 10th observation from the remaining values is selected to form the basis for all posterior inferences. Standard MCMC convergence diagnostic checks are implemented; see for example, Smith and Roberts (1993). The unreported results indicate that the MCMC chain attained sound convergence.

1.3.2. Predictive Distributions and Portfolio Weights

In the construction of the SSMB model, one can readily obtain the predictive distribution of the dependent variable \(r\) (the log return on the excess market return); indeed, the conditional structure of the time series in equation (8) provides the expression for obtaining values of the future returns. Clearly, there is no closed-form representation of the predictive distribution; rather, it must be approximated using the sampled values from the Gibbs sampler. This procedure is detailed in Appendix B and Appendix C.

The optimal portfolio weight is derived as the value maximizing the expected utility of terminal wealth. The expected utility is represented as an integral function of the portfolio weight and the predictive distribution of returns as shown in equation (3). Under distribution uncertainty, the predictive distribution does not possess a closed form; hence, it is not possible

\(^8\)It is noteworthy that, for simplicity, the current framework does not introduce any dynamics into the mean and variance regression that could allow to model strategic portfolio decisions. Introducing distribution uncertainty in the dynamic context is technically more involved and would unnecessarily cloud the purpose of this paper.
to solve the integral in equation (3) analytically. Instead, the integral is evaluated numerically at several grid points of possible weights from the specified domain. Such an integral can be approximated, with considerable accuracy, as

\[
E(V) = \frac{1}{N} \sum_{i=1}^{N} V(\omega, r^{(i)}),
\]

where \( r^{(i)} \) is the simulated excess market return from iteration \( i \); \( N \) is the number of simulations, and the grid of weights is set at 0.01%. The time subscripts have been omitted for brevity. The optimal weight, \( \omega \), is selected to be a maximizer of expression (11).

In line with our previous discussion, the scale of the estimated differences in optimal allocations strictly depends on the differences between the respective predictive distributions of returns. Under parameter uncertainty, with diffuse priors, the predictive distribution of investors accounting for parameter uncertainty is approximately Gaussian.\(^9\) As a result, all observations are given equal weight in the estimation process. In contrast, under distribution uncertainty, as presented above, investors factor in uncertainty by changing the shape of their predictive distribution. This change strictly depends on whether new observations approximate the prior Gaussian distribution or not. Mathematically, the predictive distribution of such investors is the closest approximation to the ‘true’ distribution function. In the context of distributions considered here, any deviation from the prior distribution, \( F_0 \), is going to cause extreme observations to be weighted less than the ones in the middle. As a result, the predictive distribution may look quite different than that of investors accounting for parameter uncertainty only. This change in the shape of the distribution has important consequences for the optimal portfolio allocations. Finally, given that the uncertainty depends on the extent of new information in the aggregate data, the impact of distribution uncertainty is likely to be stronger for investors considering a shorter series of data.

\(^9\)Precisely, it will have \( t \) distribution, with large number of degrees of freedom.
1.4. Data

The data used in the remaining part of the paper include monthly observations of the continuously compounded (log) excess market returns on the S&P 500 Index obtained from Standard and Poors’. Since focus is on one-month predictions, the excess market return is calculated as a difference between the return on the monthly index and the return on the one-month T-bills. The latter variable is obtained from the data set of Ibbotson and Sinquefield. As many authors have noted, before the Treasury Accord of 1951 the interest rates were held almost constant by the Federal Reserve Board. Consequently, to avoid possible structural breaks, we restrict the sample to the period of January 1954 to December 2003. This amounts to 600 months of data. Table 1 presents summary statistics of the data. Note that the data have higher kurtosis than a normal distribution, but there is not much skew. This would imply that the model in equation (5) is best suited for our purposes.

2. Impact of Distribution Uncertainty on Asset Allocation

This section studies empirically the importance of distribution uncertainty for asset allocation. One could argue that Bayesian investors should use all available data. However, since investors’ beliefs about the stability of the relationships in the underlying data may differ substantially, in all the subsequent tests we assume that, in forecasting returns, investors consider a sample which includes either all past available data (cumulative estimation window) or the last 10 years of data (rolling window). In all the tests, the coefficient of risk aversion in the utility is set to 3, 5, 10, or 20. For both windows, we compare portfolios of investors who are uncertain about the entire distribution of returns with portfolios of investors who are merely uncertain about parameters of the return-generating process.

10 The rolling-window approach can be also understood as a special case of the cumulative approach, in which a very small weight is assigned to distant observations (here, older than 10 years). This assumption has a nice intuitive interpretation within the context of decision theory with short memory, such as the one investigated in Mullainathan (2002).
2.1. Portfolio Selection

To illustrate the specifics of the investment process, first consider investors with one-month investment horizon who stand at the end of 2003 and distribute their wealth between a stock market, represented by S&P 500 Index, and a riskless asset, represented by one-month T-bills. Investors can differ with respect to the type of uncertainty they account for. They can be either be parameter or distribution uncertain. They condition their decisions either on the entire available past data (1954:1-2003:12) or on the last 10 years of past data (1994:1-2003:12). Based on the above, one obtains four different types of strategies.

The optimal portfolio composition for each of the four types of strategies is calculated using the formula in (3). By construction, the resulting optimal allocations for each of them depend critically on investors’ risk aversion and the form of the perceived predictive distribution. These distributions can differ with respect to the first, second, and higher ‘even’ moments. The presence of such differences naturally leads to contrasts regarding the optimal portfolio choices. Higher moments may have a nontrivial indirect impact on the behavior of lower moments and one should not discount their importance. For that reason, Figure 1 presents the respective predictive distributions and their moments.

The top-panel (bottom-panel) graphs represent investors who use a 10-year (cumulative) series of past data. We first assess the statistical difference between respective distributions using the Kolmogorov-Smirnov test. In both cases, the distributions are statistically significantly different from each other at the 1% level. We further look at the first two moments of the distributions individually. The results using a 10-year window indicate that investors who consider distribution uncertainty expect a much lower equity premium (3.36% per year) when compared to investors who consider only parameter uncertainty (6.00%). The respective difference for investors using a cumulative window, as illustrated in the bottom-panel graphs, is smaller: 2.76% vs. 3.96%. Both differences are statistically significant at the 1% level.

The differences in means in 2004 can be explained using the following argument. Over the previous ten years some return observations were abnormally high, which Bayesian investors could treat as an extreme event. Under parameter uncertainty, extreme observations are given
equal weight in the estimation process. As a result, with diffuse priors, the expected value of the predictive distribution is equal to the sample mean. In contrast, under distribution uncertainty, investors can account for extreme events by increasing the “thickness” of the tails of their predictive distribution. This effect causes the extreme observations to be weighted less than the middle ones. Given the occurrence of the positive shock, the expected value of the predictive distribution drops relative to the sample mean.

The smaller magnitude of changes for a cumulative window can be explained by the fact that for investors who consider a long window a series of lower returns largely dominates the series of abnormally high returns of the 90s. Hence, the marginal impact of the latter series is not nearly as large as it is the case for a relatively more balanced 10-year window.

Flexibility in higher moments also has important consequences for the variance. If investors perceive that distributions are nonstationary, which is more likely to be the case under distribution uncertainty, they will increase their expectation regarding variance. The results show that the impact of nonstationarity is stronger for investors who consider 10-year estimation window. In this case, even though kurtosis is about 1.9 times higher under distribution uncertainty, due to “nonstationarity risk” the respective variances do not differ much. For the cumulative window, the variance of the predictive distribution under parameter uncertainty is higher, suggesting that the impact of higher moments is stronger, possibly because investors do not perceive a significant nonstationarity in returns, and thus they do not increase their expectations about the variance.

Consistent with the above, under distribution uncertainty optimal stock allocations are about half as large as allocations under parameter uncertainty. For example, for investors with risk aversion of $A = 3$ optimal allocations in stock market equal 56.68% and 95.48%, respectively.

For a cumulative window, the resulting portfolio allocations under both types of uncertainty are likely to be more similar. For investors with $A = 3$, the respective optimal allocations in stock market are 66.67% and 77.07%. Table 2 presents detailed information of the portfolio choices assuming four different values of risk aversion: 3, 5, 10, and 20.
To gauge the economic magnitude of these differences, the certainty-equivalent loss (CEL) is calculated, as in Kandel and Stambaugh (1996). Mathematically,

\[ CEL = CER(\text{PU}) - CER(\text{DU}) \] (12)

where \( CER \) – the certainty-equivalent return – is a solution to the equation:

\[ V(X_T(1 + CER_i)) = EV_i, \] (13)

\( i \) is an indicator, which denotes either distribution or parameter uncertainty, and \( EV \) is the expected utility of the mean-variance investor, assuming that predictions are obtained under distribution uncertainty. For relative risk aversion of 3, the difference in \( CER \)s for a rolling window amounts to an annualized return of 0.61\%. Likewise, for a cumulative window, this difference, on the annual basis, equals 0.03\%. Even though, the two numbers do not seem to be of great economic magnitude, as shown later, at times the \( CER \)s exhibit a significant time variation. Depending on the degree of distribution uncertainty in the returns data they can reach economically significant levels.

### 2.2. Time-Series Allocations

The results obtained so far for January 2004 indicate that investors would have invested less of their wealth into risky assets had they additionally incorporated distribution uncertainty as opposed to merely considering parameter uncertainty. Note however that over time investors’ decisions, as the data change, are likely to be different. For example, nothing in the model precludes the situation in which investors would actually invest more in the stock market. This decision entirely depends on investors’ perception of the future. As will be shown, time-series inference offers additional interesting findings which are impossible to obtain once one restricts the analysis to a fixed estimation window.

In order to illustrate the effect of distribution uncertainty over time, we consider the sample for the period of 1954-2004 broken up in December of each year from 1963 to 2003. As a result,
one obtains January allocations for each of those years. Given that under both approaches one needs at least 10 years of data, the predictions for the period of 1954-1963 drop out. Figure 2 illustrates the optimal allocations into risky assets for investors incorporating either distribution or parameter uncertainty.

The results show that the stock market allocations are lower for almost each period of the strategies irrespective of whether investors use a rolling or a cumulative estimation window. Table 3 presents the average allocations in the stock market with the relative risk aversion fixed at 3, 5, 10, and 20. As an example, for investors with $A = 3$, the average allocation for the strategy based on the entire past history (10 years of most recent history) equals 73.44% (46.58%) and 103.22% (83.77%) under distribution and parameter uncertainty, respectively. The respective differences are statistically significant at the 1% level.

Again, we assess the economic importance of the ex-ante differences in stock market allocations by calculating certainty-equivalent losses in each period of the analysis. For the rolling estimation window, the average annualized return loss from ignoring distribution uncertainty equals a respectable statistically significant value of 1.33%. What is also interesting is the substantial variation of that measure; in our sample, the annualized standard deviation of the losses equals 2.46%. The values cover the range from 0.0002% to 8.46%. Certainly, the latter value is of great economic significance. Hence, the economic significance of departures from the framework of parameter uncertainty should necessarily be considered in a time-series setting. Similar analysis for the cumulative estimation window indicates a significantly smaller difference in certainty-equivalent returns. The average annualized certainty-equivalent excess return is again statistically significantly different from 0 and equals 0.44% with the standard deviation of 0.74%. This finding is in line with other results reported so far, which indicate more significant differences in the case of the rolling estimation window. Table 4 presents the detailed information based on the time series of the certainty-equivalent losses.

It is important to delineate the factors which cause the differences in optimal allocations. We study this via a time series of the mean, variance and kurtosis. The results are presented in Table 5.
Most of the driving forces identified for 2004 are also present in other periods. In particular, both for the rolling and cumulative estimation window the mean returns are statistically significantly lower under distribution uncertainty. The respective monthly comparisons are 0.30% versus 0.19% for a rolling, and 0.40% versus 0.24% for a cumulative window. In both cases, the standard errors of the estimates are very similar which suggests that the shift effect is permanent rather than specific to the estimation period.

Similarly, most of the economic intuition initially provided for the variance finds its justification in the multiple period data. In particular, for the rolling estimation window, the time-series average of standard deviations under distribution uncertainty is significantly higher than that under parameter uncertainty. This may suggest that the impact of nonstationarity dominates the impact of accounting for extreme events. This evidence is consistent with the hypothesis signifying the existence of regimes in predictive distributions. In contrast, the average of standard deviations under the cumulative estimation window is smaller for the method based on distribution uncertainty. This is what one would expect in light of the previous discussion that “nonstationarity risk” is less important if one considers a longer series of data.

Given that one can observe a significant shift in the mean and the variance of the distribution, it is not surprising that for both types of estimation windows, the average value of kurtosis under distribution uncertainty is much higher. The difference is especially pronounced for the rolling window scenario. To shed more light on the important differences between two types of uncertainty, Figure 3 presents the variation of kurtosis under both types of uncertainty and further relates it to the variation of this parameter in the data.

The results indicate that investors’ perception of distribution uncertainty is time varying. Moreover, this variation is highly consistent with the variation of this parameter in the data: estimates of kurtosis from the model with distribution uncertainty are highly correlated with the values of kurtosis in the data. These results are consistent with the paper’s main premise that distribution uncertainty plays a significant role in the decision-making process. Altogether, the patterns observed for the moments of the predictive distributions point to significantly lower stock market allocations under distribution uncertainty, especially if one considers a rolling estimation window.
An interesting feature of investment strategies based on distribution uncertainty is that the optimal allocations are less volatile compared to those derived under parameter uncertainty. Specifically, the standard deviation of portfolio weights for investors that consider distribution uncertainty using a cumulative estimation window is about 5 percentage points lower than the respective value for investors who only account for parameter uncertainty. Similar difference for investors using a rolling estimation window equals approximately 25 percentage points. This result stands in strong contrast to the results reported using for example the regime-switching models, such as the study by Guidolin and Timmermann (2007). In their study, the variability of optimal stock allocations is much more extreme. Although this feature may reflect a better fit of time-varying expected returns it is highly undesirable from the perspective of modeling the actual behavior of investors. In particular, the allocations to stocks observed in reality are much less volatile, and do not exceed 100%, which, in turn, is typical for the short-term investments under regime-switching models.

2.3. Time-Varying Distribution Uncertainty

In their seminal paper, Campbell and Cochrane (1999) suggest that time-varying risk aversion may drive time variation in expected returns. This idea is interesting as most of the existing models assume that the risk aversion is constant. The results established so far indicate that, with a constant relative risk aversion, one can observe a significant variation in portfolio weights. Part of the reason for the variation in weights may be the existence of a generally defined distribution uncertainty. In particular, one could construct a measure of such uncertainty using the portfolio weights obtained under two types of uncertainty. To accomplish this, we calculate the differences between optimal allocations of investors taking into account parameter versus distribution uncertainty to generate an index of perceived distribution uncertainty in the stock market, in short Index of Distribution Uncertainty (IDU). In order to capture the direct impact of distribution uncertainty, risk-aversion is set constant. Algebraically, the index is calculated as

\[ IDU = \omega_p - \omega_d, \]  

(14)
where $\omega_p$, and $\omega_d$ denote optimal weights under parameter and distribution uncertainty, respectively.

An increase in the value of such an index could be considered as evidence of a higher perceived level of distribution uncertainty. In contrast, values of the index close to zero would indicate a low level of perceived uncertainty among investors. Figure 4 presents the levels of the uncertainty index for investors with $A = 3$ based on a rolling-window approach to estimation, since it is for this approach that optimal allocations differ considerably more.

The results indicate that distribution uncertainty is highly time varying. Although the levels of the index do not vary much between 1980 and 1990, one can observe a considerable variation in the 70s and late 90s. Interestingly, the periods of high volatility in the index correspond to well-known macro events. Specifically, the index exhibits significant spikes, especially during the first oil crisis in 1973, and also during financial crises in the emerging markets of the late 1990s and a recent stock market boom. These three events undoubtedly have had the highest impact on investors’ behavior. A minor peak also shows up during the second oil crisis in 1978. Surprisingly, not much of the impact can be observed around the crash of 1987. The reason may be that the crash was mainly a one data-point event, and thus investors considering monthly data might not have perceived it as a change in the distribution, compared to the events where more than one month showed abnormal performance. In summary, given that the index is highly related to economic events, one could argue that the time-varying distribution uncertainty may play a similar role as a time-varying risk aversion. Whether it additionally generates time variation in expected returns is a question for future research.

2.4. The Choice of the Utility Function and Higher-Order Moments

This paper evaluates decisions of investors with mean-variance preferences. This form of preferences is chosen partly for simplicity reasons and partly because it fits well in the context in which investors are allowed to take margin or short positions. At the same time, given that in modeling distribution uncertainty the importance of higher-order moments is emphasized, it
may appear that using such preferences may be inappropriate. We argue that the two aspects
of the analysis are entirely consistent with each other.

To demonstrate this point, note that the optimal asset allocation depends on two factors:
(1) predictive distribution of asset returns; and (2) the utility function of the representative
investor. In this paper, accounting for distribution uncertainty significantly alters the predic-
tive distribution of asset returns. As a result, the value of utility and the optimal allocations
of investors with mean-variance preferences change, because both the mean and the variance
change. We term this an \textit{indirect} effect. At the same time, investors in their utility may also
\textit{directly} account for aversion to higher-order moments. This effect, however, is not captured by
mean-variance preferences.

To evaluate the magnitude of this direct effect, we re-estimate all models assuming that
investors have CRRA preferences (captured by power utility), which in turn are sensitive to
higher-order moments.\(^1\) In untabulated results, we find that the allocations of such investors
remain qualitatively similar to those of investors with mean-variance preferences. In particular,
for CRRA investors with relative risk aversion of 3 who base their decisions on a rolling estima-
tion window, the average allocation to risky asset equals 44.91\% as compared to the baseline
46.58\%. This suggests that in our sample the indirect effect is relatively more important than
the direct effect.

\subsection*{2.5. Distribution Uncertainty and Volatility Timing}

The basic model in this paper assumes that the distribution of returns is conditionally i.i.d.
This assumption, though slightly restrictive if one believes in predictability, allows us to quan-
tify the direct impact of distribution uncertainty on asset allocation. Nevertheless, the choice
of this approach should not imply that predictability is not important. We argue that pre-
dictability, especially in volatility, may play an important role. First, Fleming et al. (2001)
show that volatility timing is an important determinant of portfolio allocations. Second, as

\(^1\)The main reason for not using power utility as a default specification is that power utility approaches
negative infinity if wealth becomes negative. In our setting, this could happen if investors are allowed to take
leveraged positions. In turn, mean-variance utility function is not subject to this problem.
already argued, volatility is related to higher-order moments. Finally, the asset allocation decision itself is a function of volatility. Hence, volatility timing may significantly affect inference about the role of distribution uncertainty.

To this end, we estimate our model, taking into consideration the predictability in volatility. To keep the modeling simple, consider volatility as an AR(1) process, which seems quite adequate, especially when applied to lower-frequency monthly data; see, also, Ghysels et al. (2006) who argue in favor of this point. The estimation process in this setting is enhanced by the sampling for the parameter $\theta_1$ that governs the predictability in volatility. The resulting predictive distribution is subsequently used to form optimal asset allocations. Table 6 presents the results for optimal asset allocations. For the ease of comparisons, column (1) repeats the allocations obtained in the base case model with no timing, while column (2) reports the results with timing.

The results suggest that volatility timing does not seem to significantly alter optimal asset allocations. For example, the average allocation to risky asset for investors with relative risk aversion of 3, estimation window of 10 years, and who additionally account for volatility timing equals 52.68% as compared to 46.58% for investors who ignore timing. The respective differences are not statistically significantly different from zero. The two series have also qualitatively similar variability. Nevertheless, these results do not mean that volatility timing is not important; in fact, the coefficient on the variance regressor is positive and statistically significant at the 1% level. Moreover, the seemingly weak contribution of timing to optimal allocations may stem from the fact that our i.i.d. model already allows for variance to be varying over time.

3. Out-of-Sample Performance

So far the analysis of portfolio weights has considered *ex-ante* decisions of investors. In this section, we evaluate the model in terms of the quality of its out-of-sample predictions, which in turn underlie investment strategies.
Assume that investors begin investing in December 1963 and center their beliefs either on parameter uncertainty or on distribution uncertainty. Each of the two scenarios additionally depends on the estimation window considered by investors. In addition, we construct four other alternative strategies. The first two are buy and hold strategies that assume a constant allocation into the stock market, fixed at the average value of allocations obtained under distribution uncertainty, taking into account either a cumulative or a rolling estimation window. The remaining two strategies track the average stock market investment observed for U.S. households and institutional investors.\textsuperscript{12}

All beliefs are assumed fixed over time. In each December of a particular year, investors form a portfolio for January of the subsequent year. Due to data limitations as well as computational constraints it is reasonable to assume that investors do not vary their allocations significantly during the entire investment year. It is very likely that investors would, in fact, turn over their portfolios less often than every month due to, for example, the presence of considerable transaction costs.

The above procedure results in a time series of monthly weights, which are subsequently used to calculate monthly realized returns under each strategy. Although the optimal weights do not vary within a year, the respective realized returns may vary significantly. Since interest is on the performance of each strategy, we assume that in December 1963 each strategy is endowed with $1 of wealth. From that point onwards, the wealth of each strategy is cumulated based on the previously calculated returns. Figure 5 presents a time series of wealth levels for strategies based on uncertainty and strategies based on observed holdings. For clarity of exposition, the graphs based on fixed allocations are not presented, especially since they do not affect qualitative aspects of the analysis.

The results obtained under each strategy are quite striking. Among the strategies considered, the best performing strategy, by far, is the strategy based on distribution uncertainty and a 10-year rolling estimation window. One dollar invested in such a strategy in 1963 would

\textsuperscript{12}The average allocation in the stock market, be it for households or for institutions, is calculated as a percentage of the equity holdings in the total level of equity and bond holdings. Bond holdings are proxied by a sum of open market papers, Treasury securities, agency securities, municipal bonds, and corporate bonds. The respective values are obtained from the Federal Reserve Flow of Funds data.
have amounted to almost $11.54 at the end of 2004. The second best strategy, based on a fixed
investment of 45.43%, returns around $5.29. The strategies based on average stock market
allocation of U.S. institutions and households generate approximately $3.79 and $2.75, respec-
tively. The corresponding strategies perform worse if instead of a rolling estimation window
one considers a cumulative estimation window.

Several conclusions are noted. First, the respective strategies of investors accounting for
distribution uncertainty perform better than strategies of investors solely relying on parame-
ter uncertainty. Second, the strategies based on a rolling estimation window perform better
than the strategies based on a cumulative window. A resulting practical consequence is that
investors in their decisions seem to be better off if they account for possible nonstationarity in
a time series of returns. Third, given that the best strategy based on time-varying allocations
outperforms a strategy based on the respective fixed average allocation, in certain situations,
market timing may be profitable. Finally, the actual behavior of U.S. households and institu-
tions may not be optimal. Obviously, within a group of U.S. households or institutions one
could expect that some investors might perform better than others, even better than those
who account for distribution uncertainty in their decisions. An interesting observation is that,
as a group, households/institutions are worse off than investors who account for certain type
of distribution uncertainty.

One has to be careful when extrapolating the above results into other periods as they may
be specific to the sample period used in this study. In particular, one cannot say for certain
whether the same pattern would still hold if one considered a different draw of data. Also,
it is difficult to determine an optimal length of the data one should consider to maximize
investment’s performance. All that the above analysis says is that the length of the estima-
tion horizon may have an important impact on the resulting performance. In summary, we
can conclude that, at the very least, the out-of-sample properties of strategies incorporating
distribution uncertainty are no worse than the properties of strategies based on parameter
uncertainty.

The analysis of wealth accumulation assumes that all strategies do not differ in terms of
their riskiness. Thus, it is important to assess whether the apparent differences in terminal
wealth are due to different levels of risk taken by investors following each of the above strategies. To evaluate this possibility, we use the time series of returns to calculate the average return, its standard deviation, and *ex-post* Sharpe ratio of each strategy. In addition, for each strategy the average *ex-ante* Sharpe ratios based on the values of the first two moments of the predictive distribution is calculated. Finally, we estimate the risk-adjusted return of strategies going long in a distribution uncertainty portfolio and short in a parameter uncertainty portfolio by taking into account four common factors: market (MKT), size (SMB), value (HML), and momentum (MOM). MKT is the monthly return of the CRSP value-weighted portfolio net of the risk-free rate. SMB is the monthly return of a portfolio that is long small stocks and short large stocks. HML is the monthly return of a portfolio that is long value stocks and short growth stocks. MOM is the return of a portfolio long past one-year return winners and short past one-year return losers. Table 7 presents the results for investors with risk aversion of $A = 3$, whose expectations are based on either a rolling or a cumulative estimation window.

Again, the best strategy in terms of the mean return and the *ex-post* Sharpe ratio, is the strategy of investors who account for distribution uncertainty and use a rolling window. This strategy has statistically higher returns when compared both to the respective strategy based on parameter uncertainty and the strategy based on U.S. households’ stock market allocation. Its *ex-post* annualized Sharpe ratio, equal to 51.06%, is 4 percentage points higher than the Sharpe ratio of a strategy based on institutional holdings, and nearly two times higher than the Sharpe ratio of the next best strategy, namely the strategy based on household investments. Also, strategies based on parameter uncertainty perform quite poorly.

One can also obtain a direct comparison of investing in above strategies by analyzing their alphas from the four-factor model. This framework is suited for the problem considered here only if one assumes that investors exhibit some uncertainty about the model. If there was no uncertainty, one should expect any deviations of four-factor alpha from zero to be attributed to estimation error. For both estimation windows, strategies accounting for distribution uncertainty generate better average returns than strategies assuming that the distributional form is known. The respective differences are statistically significant at the 1% level of significance.
The results are also economically significant. Their magnitudes equal 2.4% and 1.8% per year for rolling and cumulative window, respectively.

The results indicate that the superior performance of strategies based on distribution uncertainty is likely due to effective market timing. In particular, strategies based on distribution uncertainty outperform strategies based on fixed allocations both before and after adjusting for risk. Moreover, these strategies also outperform strategies tracking US households and institutions. This suggests that the simple rule-of-thumb strategies are not superior to an effective market-timing strategy developed under distribution uncertainty.

We have also considered the role of volatility timing. Fleming et al. (2001) argue that ex post this strategy significantly improves the Sharpe ratio, relative to the constant volatility framework. We find that, in the context of distribution uncertainty, this strategy performs only slightly better than the strategy based on i.i.d. returns. This result is thus consistent with the results reported in Section 2.5 and suggests that predictability in volatility plays a secondary role in the model that allows for time-varying variance.

Finally, one could argue that the existing differences in performance among various strategies come from the abnormally high returns at a particular time. If this is the case, one should expect the returns in the period including such an outlier to be significantly different from the returns in some other period. This is evaluated by dividing a sample into halves and testing whether returns in both subperiods are statistically different from each other. Table 8 presents the results of this analysis.

For most strategies, one cannot reject the hypothesis of no significant differences in returns across two periods. The exception is the strategy using a cumulative estimation window and accounting for parameter uncertainty for which the returns in the second subperiod are statistically significantly different (at the 10% level) from returns in the first subperiod. It is unlikely that such a difference is driven by an outlier, at least based on the wealth process in Figure 5. Hence, the observed patterns are robust to different sample specifications.
4. Concluding Remarks

In this paper, a novel methodology to study the asset allocation problem for investors who are uncertain about the *entire* distribution of future returns is proposed. The theoretical foundation of the model is based on the flexible properties of a Bayesian semiparametric approach that allows one to jointly consider uncertainty about the distribution and its parameters. The consequence of this approach is that the predictive distribution of investors accounting for additional uncertainty may be different than any distribution typically considered under parameter uncertainty.

We apply this framework to a static asset allocation decision between a risky and a riskless asset. For most periods, introducing distribution uncertainty leads to significantly different allocations to the risky asset. The average allocation in risky asset is on average lower for investors accounting for distribution uncertainty. This result is mainly driven by the fact that the investors’ perceived distribution of future returns, on average, has a lower mean and higher variance as compared to the Normal, implied by only parameter uncertainty.

To gauge the economic significance of distribution uncertainty, we calculate the certainty-equivalent losses of investors who would be forced to allocate their wealth according to the model ignoring distribution uncertainty. The losses vary substantially over time; depending on the extent of distribution uncertainty in the data, they can be very close to zero but can also attain values of great economic significance. This further implies that distribution uncertainty is time varying both from a statistical and economic perspective.

The economically significant differences in portfolio allocations, under certain simplifying assumptions, offer interesting implications for the discussion related to the equity premium puzzle of Mehra and Prescott (1985). In particular, the average portfolio allocations of investors accounting for distribution uncertainty are lower than the allocations of investors solely accounting for estimation risk. This effect, however, is primarily a result of lower expected returns rather than a result of a higher risk aversion or an increase in risk.
One potential avenue for future work would be to investigate the implications of distribution uncertainty for long-term asset allocation. The theoretical extension would necessarily require a more elaborated modeling technique. In this context, it may also be useful to allow for additional predictors in the conditional mean and variance equations. Numerous studies on predictability indicate a poor performance of most of the regressors in the short horizon, but perhaps a different methodology might have a beneficial impact for the predictive power of well-known models.

Appendix A. The Dirichlet Process

To motivate the Dirichlet Process (DP), consider a simple example. Suppose $X$ is a random variable which takes the value 1 with probability $p$ and the value 2 with probability $1 - p$. Uncertainty about the unknown distribution function $F$ is equivalent to uncertainty about $(p_1, p_2)$, where $p_1 = p$ and $p_2 = 1 - p$. A Bayesian would put a prior distribution over the two unknown probabilities $p_1$ and $p_2$. Of course here we essentially have only one unknown probability since $p_1$ and $p_2$ must sum to one. A convenient prior distribution is the Beta distribution given, up to proportionality, by:

$$f(p_1, p_2) \propto p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1},$$

where $p_1, p_2, \alpha_1, \alpha_2 \geq 0$, and $p_1 + p_2 = 1$. It is denoted Beta($\alpha_1, \alpha_2$). Different prior opinions can be expressed by different choices of $\alpha_1$ and $\alpha_2$. Set $\alpha_i = cq_i$ with $q_i \geq 0$ and $q_1 + q_2 = 1$. We have,

$$E(p_i) = q_i$$

and

$$\text{Var}(p_i) = \frac{q_i(1 - q_i)}{c + 1}, \quad (A1)$$

If $q_i = 0.5$ and $c = 2$ we obtain a non-informative prior. We denote our prior guess $(q_1, q_2)$ by $F_0$. The interpretation is that the $q_i$ center the prior and $c$ reflects our degree of belief in the prior: a large value of $c$ implies a small variance, and hence strong prior beliefs.

The Beta prior is convenient in the example above. Why? Suppose we obtain a random sample of size $n$ from the distribution $F$. This is a binomial experiment with the value $X = 1$ occurring $n_1$ times (say) and the value $X = 2$ occurring $n_2$ times, where $n_2 = n - n_1$. The posterior distribution of $(p_1, p_2)$
is once again a Beta distribution with parameters updated to $\text{Beta}(\alpha_1 + n_1, \alpha_2 + n_2)$. Since the posterior distribution belongs to the same family as the prior distribution, namely, the Beta distribution, such a prior to posterior analysis is called a *conjugate update*, with the prior being referred to as a *conjugate prior*. The above example is the well-known Beta-Binomial model for $p$.

We now generalize the conjugate Beta-Binomial model to the conjugate Multinomial-Dirichlet model. Now the random variable $X$ can take the value $X_i$ with probability $p_i$, $i = 1, \ldots, K$, with $p_i \geq 0$, and $\sum_{i=1}^{K} p_i = 1$. Now, uncertainty about the unknown distribution function $F$ is equivalent to uncertainty about $p = (p_1, \ldots, p_K)$. The conjugate prior distribution in this case is the Dirichlet distribution (not to be confused with the Dirichlet Process) given, up to proportionality, by

$$f(p_1, \ldots, p_K) \propto p_1^{\alpha_1-1} \times \ldots \times p_K^{\alpha_K-1}, \quad (A2)$$

where $\alpha_i, p_i \geq 0$ and $\sum_{i=1}^{K} p_i = 1$. If we set $\alpha_i = cq_i$ then we obtain the same interpretation of the prior, and, in particular, the mean and variance are again given by equation A1. As before, $(q_1, \ldots, q_K)$ represents our prior guess ($F_0$) and $c$ the certainty in this guess. A random sample from $F$ now constitutes a Multinomial experiment and when this likelihood is combined with the Dirichlet prior, the posterior distribution for $p$ is once again a Dirichlet distribution with parameters $\alpha_i + n_i$, where $n_i$ is the number of observations in the $i$th category.

We now make a jump from a discrete $X$ to a continuous $X$ (by imagining $K \to \infty$). In traditional parametric Bayesian analysis, the distribution of $X$, say $F$, would be assumed to belong to a particular family of continuous probability density functions. For example, if $X$ can take on any real value the family of distributions is often assumed to be the Normal distribution, denoted $N(\mu, \sigma^2)$. A Bayesian analysis would then proceed by first placing a prior distribution on $\mu$ and $\sigma^2$, and then obtaining the resultant posterior distributions of these two finite-dimensional parameters.

We enter the realm of Bayesian nonparametrics when $F$ (in the last paragraph) itself is treated as a random variable; that is, one must now assign a prior distribution to $F$. Since $F$ is infinite-dimensional, we need a stochastic process whose sample paths index the entire space of distribution functions. In the main text in this paper, we noted that our focus is on ensuring greater levels of kurtosis in the conditional distribution of the asset. It is now easy to see that by treating this conditional distribution, $F$, itself as a random quantity, we allow the process that is driving the data to take on any degree of kurtosis. Parametric models impose a fixed degree of kurtosis, including fat-tailed distributions like the Student-$t$ that is used in GARCH models. But the Bayesian nonparametric model, loosely speaking, allows the data to determine the level of kurtosis in the distribution of the asset, which could be much larger than...
under a parametric model. Indeed, since the goal of this paper is on the predictive distribution of the asset, allowing for larger degrees of kurtosis is critical because uncertainty in forecasts increases over time. Thus, a model that makes minimal assumptions about the distribution of the asset is likely to better capture this greater uncertainty over time.

We are now ready to define the Dirichlet Process, but first some notation. A partition \( B_1, \ldots, B_k \) of the sample space \( \Omega \) is such that \( \bigcup_{i=1}^{K} B_i = \Omega \) and \( B_i \cap B_j = \emptyset \) for all \( i \neq j \). That is, we have a group of sets that are disjoint and exhaustive. Stated differently, the sets cover the whole sample space and are nonoverlapping.

**Definition:** A Dirichlet Process prior with parameter \( \alpha \) generates random probability distributions \( F \) such that for every \( k = 1, 2, 3, \ldots \) and partition \( B_1, \ldots, B_k \) of \( \Omega \), the distribution of \( (F(B_1), \ldots, F(B_k)) \) is the Dirichlet distribution with parameter \( \alpha(B_1), \ldots, \alpha(B_k) \). Here \( \alpha \) is a finite measure on \( \Omega \) and so we can put \( \alpha(.) = cF_0(.) \), where \( c > 0 \) and \( F_0 \) is a probability distribution on \( \Omega \).

**Example** Consider a random variable \( X \) with distribution function \( F \) defined on the real line. Now consider the probability \( p = \text{pr}(X < x^*) \) and suppose we specify a DP prior with parameter \( \alpha \) for \( F \). If we put \( B_1 = (-\infty, x^*] \) and \( B_2 = (x^*, \infty) \), then we see from the above definition that, \textit{a priori}, \( p \) has a Beta distribution with parameters \( \alpha_1 = \alpha(B_1) = cF_0(x^*) \) and \( \alpha_2 = \alpha(B_2) = c(1 - F_0(x^*)) \), where \( c = \alpha(-\infty, \infty) \). This prior is such that \( E(p) = F_0(x^*) \) and \( \text{var}(p) = F_0(x^*)(1 - F_0(x^*))/(c + 1) \), thus showing the link to equation (A1). The variance of \( p \) and hence the level of fluctuation of \( p \) about \( F_0(x^*) \) depends on \( c \). In particular, if \( c \) is large then we have strong belief in \( F_0(x^*) \) and \( \text{var}(p) \) is small. Note that this is true for all partitions \( B_1, B_2 \) of the real line, or, equivalently, all values of \( a \).

Now consider observations \( X_1, \ldots, X_n \) from \( F \). Let \( F \) be assigned a DP prior, denoted \( \text{Dir}(c, F_0) \). Then Ferguson showed that the posterior process \( F \) has parameters given by

\[
c + n \quad \text{and} \quad \frac{cF_0 + nF_n}{c + n},
\]

where \( F_n \) is the empirical distribution function for the data, namely the step function with jumps of \( 1/n \) at each \( X_i \). The classical maximum likelihood estimator is given by \( F_n \).

The posterior mean of \( F(A) \) for any set \( A \in (-\infty, \infty) \) is given by

\[
E\{F(A)|\text{data}\} = p_nF_0(A) + (1 - p_n)F_n(A),
\]
where $p_n = c/(c+n)$. Thus if $c = 1$ we have a weight of $1/(n+1)$ on our prior $F_0$ and $n/(n+1)$ on the “data” $F_n$. As $c$ increases, the posterior mean is influenced more by the prior, that is, we have more faith in our prior choice, $F_0$.

**Appendix B. Estimation**

The key to implementing a Gibbs sampler is to be able to obtain the conditional distributions, up to proportionality, of the random variables of interest.

The $s$th iteration in the Gibbs sampler, indexed by superscript ($s$), successively samples from the following full conditional distributions: $p(u_t | \text{everything else}), t = 1, \ldots, T$; $p(\beta_k | \text{everything else}), k = 1, \ldots, K$; $p(\theta_l | \text{everything else}), l = 1, \ldots, L$; and $p(c | \text{everything else})$, where $T$ equals the number of observations in the sample, and $K$ and $L$ are the numbers of independent variables in the mean and variance regressions, respectively.

1. $p(u_t | \cdot); t = 1, \ldots, T$. Given that $[r|U, \mu, \sigma] \sim \text{Uniform}(\mu - \sigma\sqrt{U}, \mu + \sigma\sqrt{U})$, and $U$ is assigned a Dirichlet process prior of Ferguson (1973), we can write the full conditional density for $u_t$ as:

$$p(u_t | \cdot) \propto p(r_t | u_t) p(u_t | u_{-t}).$$  \hspace{1cm} (B1)

The first term on the right-hand side is the uniform density. Hence,

$$p(r_t | u_t) \propto u_t^{-1/2}$$  \hspace{1cm} (B2)

The second term is the Dirichlet Process. Its derivation is standard and can be found for example in Escobar and West (1995).

$$p(u_t | u_{-t}) \propto c^{(s-1)} f^*(u_t) + \sum_{j \neq t}^{\delta u_j(u_t)},$$  \hspace{1cm} (B3)

where $\delta u$ is the point mass 1 at $u$, and $u_{-t}$ denotes all values of $u$ except for $u_t$ and $f^*(u_t)$ is the transition density, specified later.

Thus, overall, $p(u_t | \cdot)$ is a mixture distribution comprising a truncated exponential distribution and a discrete distribution made up of those $u_t$ for which $u_t > r_t^2 \exp(-2Z_t\theta)$, that is,
\[
f(u) \propto u_t^{-1/2} p(u_t|u_{-t}) I(u_t > r_t^2/\sigma_{t-1}^2 (s-1)).
\] (B4)

To simplify the sampling from the above conditional, recall that in our case \( f^*(u) \propto u^{1/2} \exp(-u/2) \), that is, Gamma(3/2,1/2) distribution. Let \( a = r_t^2/\sigma_{t-1}^2 (s-1) \). Substituting equation (B3) into equation (B4) we can write:

\[
f(u) \propto e^{(s-1)} u_t^{-1/2} \frac{1}{\Gamma(3/2)} (1/2)^{3/2} \exp(-u_t/2) I(u_t > a) + u_t^{-1/2} \sum_{j \neq i} \delta_{u_j}(u_t) I(u_t > a).
\] (B5)

Next, we rewrite the first term of the above equation as:

\[
\begin{align*}
e^{(s-1)} u_t^{-1/2} & \frac{1}{\Gamma(3/2)} (1/2)^{3/2} \exp(-u_t/2) I(u_t > a) \\
= & \frac{e^{(s-1)} (1/2)^{3/2} \exp(-a/2)}{0.5 \sqrt{\pi}} 0.5 \exp\left\{-\frac{(u_t - a)}{2}\right\} I(u_t > a) \\
= & \frac{e^{(s-1)} (1/2)^{3/2} \exp(-a/2)}{0.25 \sqrt{\pi}} 0.5 \exp\left\{-\frac{(u_t - a)}{2}\right\} I(u_t > a) \\
= & \frac{e^{(s-1)} (1/2)^{-1/2} \exp(-a/2)}{\sqrt{\pi}} 0.5 \exp\left\{-\frac{(u_t - a)}{2}\right\} I(u_t > a)
\end{align*}
\] (B6)

where \( \Gamma(3/2) = 0.5 \sqrt{\pi} \). Therefore,

\[
f(u) \propto e^{(s-1)} \frac{(1/2)^{-1/2} \exp(-a/2)}{\sqrt{\pi}} 0.5 \exp\left\{-\frac{(u_t - a)}{2}\right\} I(u_t > a) + u_t^{-1/2} \sum_{j \neq i} \delta_{u_j}(u_t) I(u_t > a).
\] (B7)

Consequently, we either sample \( u_t \) from a truncated exponential distribution or take \( u_t \) to be \( u_j \), for those \( u_j > r_t^2/\sigma_{t-1}^2 (s-1) \), according to probabilities which follow from equation (B7). Thus,

\[
u_t^{(s)} \begin{cases} 
\sim f^*(u) \text{ with probability } \tau \exp(-a/2) \\
= u_j \text{ with probability } 1/\sqrt{\pi}
\end{cases},
\] (B8)

where \( \tau = e^{(s-1)} (1/2)^{-1/2} / \sqrt{\pi} \) and \( f^*(u) = 0.5 \exp\left\{-\frac{(u^{(s-1)} - a)}{2}\right\} I(u^{(s-1)} > a) \).

2 \( p(\theta|\cdot \cdot \cdot) \).

Define \( \gamma = -\theta \) and write the distribution of \( r_t \) as follows:

\[
r_t \sim \text{Uniform}\left(\mu - \sqrt{u_t/e^{\gamma W_t}}, \mu + \sqrt{u_t/e^{\gamma W_t}}\right).
\] (B9)

Then the posterior conditional distribution of \( \gamma \) can be written using the following general truncated distribution:
\[ P(\gamma | rt, ut, \beta) = \pi(\gamma) \prod_t P(rt, ut, \beta | \gamma) \propto \pi(\gamma) \prod_t I(A_t, B_t), \quad (B10) \]

where \( A \) and \( B \) define bounds of the truncated distribution and \( \pi(.) \) is a prior distribution function for \( \gamma \), taken as Normal in the empirical setting. To find the bounds, we use equation (B9), which implies that

\[ (rt - \sum_{j=1}^{K} \beta_j^{(s-1)} Z_{jt})^2 < \frac{u_t^{(s)}}{\exp(2\gamma W_t)}. \quad (B11) \]

Define

\[ \lambda_t = \ln \frac{u_t^{(s)}}{0.5(rt - \sum_{j=1}^{K} \beta_j^{(s-1)} Z_{jt})^2} - \sum_{j \neq t} \gamma_j^{(s-1)} W_{jt}, \quad (B12) \]

then

\[ \left[ \gamma^{(s)} \right] \propto \pi(\gamma_t) I \left( \gamma_t \in \left[ \max_{W_{it} < 0} \{ \lambda_t/W_{it} \}, \min_{W_{it} > 0} \{ \lambda_t/W_{it} \} \right] \right). \quad (B13) \]

If \( W_{it} > 0 \) for all \( t \) then

\[ \max_{W_{it} < 0} \{ \lambda_t/W_{it} \} = -\infty \quad (B14) \]

and if \( W_{it} < 0 \) for all \( t \) then

\[ \min_{W_{it} > 0} \{ \lambda_t/W_{it} \} = \infty. \quad (B15) \]

The distribution of \( \theta \) is then obtained by using an initial transformation \( \theta = -\gamma \).

3 \( p(\beta_k | \cdots ) \).

Using a similar logic to that we used for \( \theta \) we can sample the distribution of \( \beta \).

Define

\[ \lambda_t = \frac{r_t - \sigma_t^{(s-1)} \sqrt{u_t^{(s)} - \sum_{j \neq k} \beta_j^{(s-1)} Z_{jt}}}{\lambda_t}, \quad (B16) \]

where

\[ \sigma_t^{(s)} = \exp \left( \sum_{j=1}^{L} \theta_j^{(s)} W_{jt-1} \right) \quad (B17) \]

and \( \pi(.) \) to be a prior distribution function for \( \theta \), so

\[ \left[ \beta_k^{(s)} \right] \propto \pi(\beta_k) I \left( \beta_k \in \left[ \max_{Z_{kt} < 0} \{ \lambda_t/Z_{kt} \}, \min_{Z_{kt} > 0} \{ \lambda_t/Z_{kt} \} \right] \right). \quad (B18) \]

If \( Z_{kt} > 0 \) for all \( t \) then

\[ \max_{Z_{kt} < 0} \{ \lambda_t/Z_{kt} \} = -\infty \quad (B19) \]
and if $Z_{kt} < 0$ for all $t$ then

$$\min_{Z_{kt} > 0} \{\lambda_t / Z_{kt}\} = \infty.$$  \hfill (B20)

4. $p(c|\cdots)$.

If the prior for $c$ is a gamma distribution, or a mixture of gamma distributions, we can use a data augmentation technique of Escobar and West (1995) for sampling $c$. We assume that $c$ has a gamma prior with a shape parameter $a_0$ and scale parameter $b_0$. When both $a_0$ and $b_0$ are small, then this prior puts weight on both large and small values of $c$. A data augmentation technique can be used to generate conditional samples.

More specifically, the sampling for $c$ proceeds as follows. In the first step, sample from the beta distribution for the new latent parameter $\tau(j; c)$

$$[c | \eta] \sim \text{beta} (c^{(s-1)} + 1, T).$$  \hfill (B21)

Then $c$ is sampled from the mixture of gamma distributions, with the weights defined as below,

$$[c^{(s)} | \eta, h] \sim \pi_\eta \text{Gamma}(a_0 + h, b_0 - \ln \eta) + (1 - \pi_\eta) \text{Gamma}(a_0 + h - 1, b - \ln \eta).$$  \hfill (B22)

Here $\text{Gamma}(a_0, b_0)$ is the prior distribution for $c$ with $a_0 = b_0 = 0.01$; $h$ is the number of distinct values of the parameter $u$ (usually $h < T$); and $\pi_\eta$ is the solution of the equation

$$\pi_\eta / (1 - \pi_\eta) = (a_0 + h - 1) / T(b_0 - \ln(\eta)).$$  \hfill (B23)

### Appendix C. Predictive Distribution

In order to construct predictive distribution of equity premium one period ahead, consider the following extension of the Gibbs sampler detailed in Appendix B.

If $T$ denotes the time at which we construct forecasts, for the predictions for period $T + 1$, at each iteration $(s)$, sample the following components

$$r_T^{(s)} \sim U(\mu_T^{(s)} - \sigma_T^{(s)} \sqrt{u_{T+1}^{(s)}}, \mu_T^{(s)} + \sigma_T^{(s)} \sqrt{u_{T+1}^{(s)}}) + \sum_{j=1}^{K} \beta_j^{(s)} Z_j T$$  \hfill (C1)

where $\mu_T$ and $\sigma_T$ are defined as

$$\mu_T^{(s)} = \beta^{(s)} Z_T.$$  \hfill (C2)
\[ \sigma_T^{(s)} = e^{W_T \theta^{(s)}} \]  

and \( Z_T, W_T \) are the values of the covariates at the time of constructing portfolio. The sampling of \( u_{T+1} \) component proceeds as follows,

\[
\begin{cases}
  u_T^{(s)} & \sim f(u^{(s)}) \quad \text{with probability } \propto c^{(s)} \\
  u_j^{(s)} & = u_j^{(s)} \quad \text{with probability } \propto 1 \quad j = 1, \ldots, T
\end{cases}
\]

Here \( f(U) \) is Gamma(1.5, 0.5). \( r_{T+1}^{(s)} \) can be obtained from each iteration of the Gibbs sampler, using the current \((c, \beta, \theta)\). \((c, \beta, \theta)\) are set at the values obtained in each iteration from the estimation as shown in Appendix B.

References


38


Predictive distribution with a rolling estimation window for January 2004

Predictive distribution with a cumulative estimation window for January 2004

Figure 1. Predictive distributions of market excess returns. This figure depicts the predictive distributions of market excess returns one-month ahead (January 2004) for investors that account for (a) parameter uncertainty (dotted line) or (b) distribution uncertainty (solid line). The top graph assumes a 10-year rolling estimation window (1994:1-2003:12), while the bottom graph considers a cumulative estimation window (1954:1-2003:12). Each graph includes the respective mean, standard deviation and kurtosis of the distribution. All distribution functions have been obtained from simulated data using a kernel smoothing approach.
Figure 2. Optimal time-series allocations to stocks: 1964-2004 This figure depicts a time-series of optimal allocations to stocks for mean-variance investors that account either for (a) parameter uncertainty (dotted line) or (b) distribution uncertainty (solid line). Investors maximize utility over terminal wealth with $A$ denoting the coefficient of relative risk aversion. The allocations are obtained using either a 10-year rolling or a cumulative estimation window, starting with the period of 1954:1-1963:12. Both scenarios are representative for investors with a relative risk aversion of $A=3$. The horizontal lines illustrate the average allocations over the entire horizon for the respective strategies.
Figure 3. Kurtosis of the distribution of future returns: 1964-2004. This figure depicts a time series of kurtosis of distribution of future returns obtained from the models that incorporate parameter uncertainty and distribution uncertainty or parameter uncertainty only. In addition, the figure shows kurtosis calculated using the data. The top panel presents the results for investors who base their decisions using a 10-year rolling estimation window, while the bottom panel presents the results formed using a cumulative estimation window.
Figure 4. The levels of Index of Distribution Uncertainty: 1964-2004. This figure depicts a time series of levels of Index of Distribution Uncertainty. The Index has been defined as a difference between optimal portfolio weights for mean-variance investors who only account for parameter uncertainty and optimal portfolio weights of investors who additionally account for distribution uncertainty. Investors maximize utility over terminal wealth with $A$ denoting the coefficient of relative risk aversion. The graph presents the results for investors with $A=3$ who consider a 10-year rolling estimation window.
Figure 5. Wealth cumulation of strategies. This figure depicts a time series of wealth derived from $1 invested in December 1963 and cumulated until 2004 for investment strategies based on optimal allocations to S&P 500 and one-month Treasury bills. The first two strategies control for parameter- and distribution uncertainty, while the other two replicate the allocations of U.S. households and institutional investors. The latter two strategies are generated using Flow of Funds data from the Board of Governors of the Federal Reserve. The top panel considers mean-variance investors that use a 10-year rolling estimation window while the bottom panel assumes a cumulative estimation window. Investors maximize utility over terminal wealth with $A$ denoting the coefficient of relative risk aversion. In both panels, a dotted line illustrates a strategy based on parameter uncertainty; a solid line - a strategy based on distribution uncertainty; a dash/dot line - a strategy based on actual allocations of U.S. households; a dashed line - a strategy based on actual allocations of U.S. institutional investors. The y-axis has been presented in the logarithmic scale with a base of 10. All graphs are representative for investors with a relative risk aversion of $A = 3$. 
Table 1
Summary statistics

This table reports the summary statistics of the data for the period of 1954:1-2003:12. The values include the mean, standard deviation, minimum value, maximum value, skewness and kurtosis of the excess S&P 500 Index above the risk-free asset. Annualized percentage values have been provided, where appropriate. The S&P 500 Index has been obtained from Standard & Poors’. The risk free rate is based on the one-month T-bill rate, obtained from the data set of Ibbotson and Sinquefield.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.39</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>14.78</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.37</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.60</td>
</tr>
<tr>
<td>Min</td>
<td>-268.29</td>
</tr>
<tr>
<td>Max</td>
<td>189.54</td>
</tr>
</tbody>
</table>
Table 2  
Optimal allocations to stocks under parameter and distribution uncertainty

This table reports the optimal one-month ahead allocations to stock market (in percentage) in January 2004. We consider mean-variance investors who account either for: (a) parameter uncertainty or (b) distribution uncertainty. They maximize utility over terminal wealth with $A$ denoting the coefficient of relative risk aversion. $A$ is assigned four possible values: 3, 5, 10 and 20. The respective horizons used for estimation are 1994:1-2003:12 and 1954:1-2003:12.

<table>
<thead>
<tr>
<th>A</th>
<th>Parameter uncertainty</th>
<th>Distribution uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>95.48</td>
<td>56.68</td>
</tr>
<tr>
<td>5</td>
<td>77.07</td>
<td>46.24</td>
</tr>
<tr>
<td>10</td>
<td>57.29</td>
<td>34.02</td>
</tr>
<tr>
<td>20</td>
<td>28.64</td>
<td>17.01</td>
</tr>
<tr>
<td></td>
<td>14.31</td>
<td>8.51</td>
</tr>
<tr>
<td></td>
<td>11.55</td>
<td>10.00</td>
</tr>
</tbody>
</table>
This table reports time-series averages and standard deviations of optimal allocations into stock market (in percentage terms) for each subsequent January between 1964 and 2004. We consider mean-variance investors who account either for parameter uncertainty (PU) or for distribution uncertainty (DU). They maximize utility over terminal wealth with $A$ denoting the coefficient of relative risk aversion. $A$ is assigned four possible values: 3, 5, 10, and 20. The estimation windows include either 10 years of the most recent past data (rolling window) or all available past data (cumulative window). Columns 4 and 7 provide the differences between the two types of uncertainties along with their standard errors (in parentheses).

<table>
<thead>
<tr>
<th>$A$</th>
<th>Rolling window</th>
<th>Cumulative window</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PU</td>
<td>DU</td>
</tr>
<tr>
<td>3</td>
<td>83.77</td>
<td>46.58</td>
</tr>
<tr>
<td></td>
<td>(101.35)</td>
<td>(77.13)</td>
</tr>
<tr>
<td>5</td>
<td>50.37</td>
<td>28.22</td>
</tr>
<tr>
<td></td>
<td>(61.20)</td>
<td>(46.78)</td>
</tr>
<tr>
<td>10</td>
<td>25.20</td>
<td>14.18</td>
</tr>
<tr>
<td></td>
<td>(30.65)</td>
<td>(16.63)</td>
</tr>
<tr>
<td>20</td>
<td>12.60</td>
<td>7.10</td>
</tr>
<tr>
<td></td>
<td>(15.33)</td>
<td>(11.79)</td>
</tr>
</tbody>
</table>

*** - 1% significance; ** - 5% significance; * - 10% significance
Certainty-equivalent losses under parameter and distribution uncertainty

This table reports the percentage annualized mean, standard deviation, minimum and maximum value of the certainty-equivalent losses. The certainty-equivalent loss (CEL) is calculated as a difference between the certainty-equivalent return (CER), obtained for mean-variance investors who account for distribution uncertainty but are forced to hold the optimal allocation obtained under parameter uncertainty, and the certainty-equivalent return for investors who hold optimal allocations under distribution uncertainty. Mathematically, \( CEL = CER(\text{PU}) - CER(\text{DU}) \), where CER – the certainty-equivalent return – is a solution of the equation \( U(W_T(1+CER_i)) = EU_i \). \( i \) is an indicator, which denotes either the distribution or parameter uncertainty, and \( EU \) is the expected utility of the mean-variance investor and assuming that predictive returns are obtained under distribution uncertainty. Investors maximize utility over terminal wealth of with \( A \) denoting the coefficient of relative risk aversion. The value of the coefficient is set to 3. Investors derive their portfolios using either a rolling 10-year window or a cumulative window. The data span the period of 1954:1-2003:12.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Rolling window</th>
<th>Cumulative window</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.33***</td>
<td>0.44***</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.46</td>
<td>0.74</td>
</tr>
<tr>
<td>Min</td>
<td>0.0002</td>
<td>0.003</td>
</tr>
<tr>
<td>Max</td>
<td>8.46</td>
<td>2.81</td>
</tr>
</tbody>
</table>

*** - 1% significance; ** - 5% significance; * - 10% significance
Table 5
Parameters of predictive distributions under parameter and distribution uncertainty

This table reports the time-series averages of monthly mean, standard deviation, and kurtosis of predictive distributions obtained for each subsequent January between 1964 and 2004. Mean and standard deviation are expressed in percentage terms. Investors account either for parameter (PU) or for distribution uncertainty (DU). Estimation windows incorporate either 10 years of the most recent past data (rolling window) or all available past data (cumulative window). The sample spans the period of 1954:1-2003:12. Standard deviations of the estimates have been provided in parentheses. Columns 4 and 7 provide differences across types of uncertainty with the respective standard errors (in parentheses).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Rolling window</th>
<th>Cumulative window</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PU</td>
<td>DU</td>
</tr>
<tr>
<td>Mean</td>
<td>0.298</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td>(0.469)</td>
<td>(0.475)</td>
</tr>
<tr>
<td>St. dev.</td>
<td>4.217</td>
<td>4.723</td>
</tr>
<tr>
<td></td>
<td>(0.508)</td>
<td>(0.598)</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.050</td>
<td>5.703</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(2.227)</td>
</tr>
</tbody>
</table>

*** - 1% significance; ** - 5% significance; * - 10% significance
Table 6
Optimal allocations under distribution uncertainty and volatility timing

This table reports time-series averages and standard deviations of optimal allocations into stock market (in percentage terms) for each subsequent January between 1964 and 2004. We consider mean-variance investors who account for distribution uncertainty and allow for predictability in volatility using an AR(1) structure. They maximize utility over terminal wealth with \( A \) denoting the coefficient of relative risk aversion. \( A \) is assigned four possible values: 3, 5, 10, and 20. The estimation window includes 10 years of the most recent past data. Column 3 reports the difference between weights coming from two models along with the respective standard errors (in parentheses).

| \( A \) | i.i.d. volatility timing |  |  |  |
|-------|-------------------------|-------|-------|
|       | (1)                     | (2)   | (1)-(2) |
| 3     | 46.58                   | 52.68 | -6.10  |
|       | (77.13)                 | (86.06)| (6.11) |
| 5     | 28.22                   | 32.04 | -3.82  |
|       | (46.78)                 | (52.29)| (3.66) |
| 10    | 14.18                   | 16.10 | -1.91  |
|       | (16.63)                 | (26.28)| (1.83) |
| 20    | 7.10                    | 8.06  | -0.96  |
|       | (11.79)                 | (13.16)| (0.91) |

*** - 1% significance; ** - 5% significance; * - 10% significance
Table 7
Time-series performance of strategies based on parameter, distribution uncertainty, households’ and institutions’ market participation

This table reports the percentage annualized average return, its standard deviation, Sharpe ratio (ex-ante and ex-post), and the risk-adjusted returns using the CAPM model for the strategies based either on parameter- or distribution uncertainty using a rolling 10-year or the cumulative estimation window, and for the strategies based on the observed U.S. household and institutional investors’ flows of funds, as reported by Board of Governors of the Federal Reserve. The average returns are calculated using monthly returns obtained for each strategy in each year based on optimal allocations obtained for January of the subsequent year. Ex-ante Sharpe ratios are calculated using January values alone. Annualized differences in mean returns between strategies based on distribution uncertainty (DU) and strategies based on parameter uncertainty (PU) and market participation have been provided along with their respective significance based on t-test. The data set spans the period of 1954:1-2003:12. We consider mean-variance investors who maximize utility over terminal wealth with $A$ denoting the coefficient of relative risk aversion, set to $A = 3$. ***, **, * denote 1%, 5%, and 10% level of significance, respectively.

<table>
<thead>
<tr>
<th>A=3</th>
<th>Parameter uncertainty</th>
<th>Distribution uncertainty</th>
<th>Households</th>
<th>Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rolling</td>
<td>Cumulative</td>
<td>Rolling</td>
<td>Cumulative</td>
</tr>
<tr>
<td>Mean</td>
<td>4.13</td>
<td>1.32</td>
<td>6.99</td>
<td>2.30</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>16.45</td>
<td>14.84</td>
<td>13.56</td>
<td>12.12</td>
</tr>
<tr>
<td>Sharpe Ratio ex-ante</td>
<td>38.49</td>
<td>33.53</td>
<td>39.87</td>
<td>23.71</td>
</tr>
<tr>
<td>Sharpe Ratio ex-post</td>
<td>25.08</td>
<td>8.92</td>
<td>51.06</td>
<td>18.94</td>
</tr>
<tr>
<td>DU-PU</td>
<td>-</td>
<td>-</td>
<td>2.80***</td>
<td>0.97</td>
</tr>
<tr>
<td>Uncertainty - Households</td>
<td>1.09</td>
<td>-1.72*</td>
<td>3.88*</td>
<td>-0.74</td>
</tr>
<tr>
<td>Uncertainty - Institutions</td>
<td>1.07</td>
<td>-2.28*</td>
<td>3.40*</td>
<td>-1.22</td>
</tr>
<tr>
<td>Alpha(DU) - Alpha( PU)</td>
<td>-</td>
<td>-</td>
<td>2.40***</td>
<td>1.80***</td>
</tr>
</tbody>
</table>

*** - 1% significance; ** - 5% significance; * - 10% significance
Table 8
The differences between returns in two sub-periods

This table reports the percentage sample means (standard deviations) of the differences between monthly returns in two equal sub-periods (1964:1-1983:12 and 1984:1-2003:12) starting with the estimation period of 1954:1-1963:12. The respective differences have been obtained for strategies based on parameter (PU) or distribution uncertainty (DU); a 10-year rolling or cumulative estimation window. The optimal weights used to calculate the above returns have been obtained for mean-variance investors who maximize utility over terminal wealth, with $A$ denoting the coefficient of relative risk aversion. The value of $A$ is set to 3. t statistics indicate the significance of the differences.

<table>
<thead>
<tr>
<th>Predictability</th>
<th>Rolling window</th>
<th>Cumulative window</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of uncertainty</td>
<td>PU</td>
<td>DU</td>
</tr>
<tr>
<td>Mean (1st half - 2nd half)</td>
<td>-0.20</td>
<td>-0.53</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>(6.83)</td>
<td>(6.26)</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-0.47</td>
<td>-1.33</td>
</tr>
</tbody>
</table>

*** - 1% significance; ** - 5% significance; * - 10% significance