Asset Allocation under Distribution Uncertainty*

Marcin Kacperczyk
Michigan Business School
701 Tappan Street
Ann Arbor, MI-48109
Email: mkacpe@bus.umich.edu

Job Market Paper

First draft: August, 2003
This version: November 13, 2003

*Special thanks go to my dissertation committee: Paul Damien, Gautam Kaul, Lutz Kilian, Tyler Shumway, Clemens Sialm and Stephen Walker for their constant encouragement and helpful discussions. I am also indebted to Sugato Bhattacharya, Bob Dittmar, Art Durnev, Radhakrishnan Gopalan, Zbyszek Kominek, Wojciech Kopczuk, Qin Lei, Adair Morse, Lüboš Pástor, Amit Seru, Sophie Shive, Anjan Thakor, Kathy Yuan, Lu Zheng, and the seminar participants at the Michigan Business School for helpful comments. All errors are my own.
Investors do not know the distribution of future returns, but much of the finance literature assumes they do. This paper shows how uncertainty about the type of return distribution (distribution uncertainty) can be incorporated in deciding one’s asset allocation. I find that distribution uncertainty is highly time-varying. To evaluate the importance of distribution uncertainty, I address the following two questions. Will ex ante optimal portfolios change if investors factor in distribution uncertainty into their portfolio model? Will investors benefit ex post from taking into account such distribution uncertainty? Compared to investors facing a standard parameter uncertainty, investors in my paper, on average, allocate less money to the stock market, their stock market allocations are less variable, and their certainty-equivalent losses from ignoring distribution uncertainty can be economically significant. Also, portfolio strategies of such investors generate statistically higher returns, even after controlling for market risk.
Since the seminal paper by Markowitz (1952), many studies have recognized the importance of the predictive distribution of returns for optimal asset allocation. Most of the existing studies assume that investors know the distributional form of future returns.\(^1\) In practice, however, empirical evidence seems to suggest the opposite. For example, the literature provides evidence that distributions of asset returns tend to switch between different regimes. Also, the existence of rare events may perturb beliefs about the form of predictive distributions. Finally, existing evidence from financial markets is rich in examples which show that investors differ in their assessment of future returns; hence, any consensus about the precise nature of the underlying stochastic process driving returns seems well-nigh impossible.\(^2\)

This paper shows how uncertainty about the type of return distribution (distribution uncertainty) can be incorporated to obtain an optimal mix between a risky and a riskless asset. In particular, the objective is to characterize the asset allocation decisions from the \textit{ex ante} and \textit{ex post} perspectives by providing answers to the following two questions. Will \textit{ex ante} optimal portfolios change if investors factor in distribution uncertainty into their portfolio model? Will investors benefit \textit{ex post} from taking into account such distribution uncertainty?

When constructing economic forecasts, a great majority of studies rely on the assumption that the sample estimates of the parameters of the predictive distribution are sufficient statistics to derive optimal portfolios. A major criticism raised against this assumption is that parameters of the predictive distribution are estimated with error, and thus they should not be treated as known. In fact, any prediction should necessarily account for such parameter uncertainty, generally defined as estimation risk. Markowitz (1952) was the first to recognize the presence of estimation risk. However, it was not until the work of Klein and Bawa (1976) that the estimation risk has been incorporated rigorously in the portfolio context.\(^3\) Nevertheless,

\(^1\)The most common description of realized returns involves the Normal or Log-Normal distribution. Significant departures from normality have been illustrated, among others, by Mandelbrot (1963), Fama (1965), Affleck-Graves and McDonald (1989), and Richardson and Smith (1993).

\(^2\)For evidence on each of the above three points, see 1) P\'astor and Stambaugh (2001), Ang and Bekaert (2003), Guidolin and Timmermann (2003); 2) Liu, Pan, and Wang (2002), Liu, Longstaff, and Pan (2003); 3) Welch (2000), respectively.

\(^3\)Later studies on this subject include, among others, Bawa, Brown, and Klein (1979), Frost and Savarino (1986), Barberis (2000), and Tu and Zhou (2003) for portfolio with i.i.d. asset returns, and Kandel and Stambaugh (1996), Barberis (2000), and Xia (2001) for portfolio with predictable asset returns.
existing studies which factor in parameter uncertainty assume that the form of the distribution of returns is known. This is the problem addressed in this paper.

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, assumed to be unimodal and symmetric, is modeled nonparametrically, while simultaneously the mean and variance of the asset are modeled using parametric regressions.

A Bayesian 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. Indeed, it is for this reason that all the papers which develop parameter uncertainty adopt a Bayesian perspective; as examples, Kandel and Stambaugh (1996), Stambaugh (1999), Barberis (2000), Pástor (2000), and Lewellen and Shanken (2002).

To obtain direct comparisons to earlier studies (e.g., Barberis (2000)), I consider investors with power utility who make their decisions for the next period. Also, since investors’ beliefs about the stability of the relationships in the underlying data may differ substantially, I proceed under two following assumptions: (a) investors consider the entire past available data, and (b) investors consider a sample of the most recent 10 years of data. To gauge the overall impact of distribution uncertainty, the focus is on ex ante and ex post implications. From an ex ante perspective, I examine optimal portfolio allocations of investors who believe that returns are conditionally independently and identically distributed (i.i.d.). The i.i.d. assumption, though slightly restrictive, allows me to quantify the direct impact of distribution uncertainty on asset allocation. Next, I assess the economic significance of such allocations in terms of the certainty-equivalent losses incurred by forcing investors that account for distribution uncertainty to hold optimal allocations obtained under parameter uncertainty. From an ex post perspective, I study a time-series performance of investment strategies based on distribution uncertainty with and
without adjusting for market risk. Looking at performance of respective strategies enables me
to examine their economic feasibility, at the same time indicating practical implications for
investors.

To gauge the ex ante impact of uncertainty over time, I examine a time-series of portfolio
allocations into S&P 500 index and Treasury bills for each January between 1964 and 2001.
Results indicate that investors who incorporate distribution uncertainty into their decision
process tend to invest considerably less in the stock market than investors who merely consider
parameter uncertainty. For example, for investors with a reasonable level of risk aversion of
3, the difference in average allocations equals approximately 30 percentage points. This result
suggests that distribution uncertainty, above and beyond parameter uncertainty, may play an
important role in the decision process.

The scale of the estimated differences in optimal allocations strictly depends on the differ-
ences between respective predictive distributions. A salient feature of the current study is that
each of the two approaches accounts for extreme events in a distinct way. Under parameter
uncertainty, extreme observations are given equal weight in the estimation process. As a re-
sult, 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. Consequently, dependent on
the observed data, the expected value of the predictive distribution may get shifted to the left
or to the right of the sample mean. For the sample analyzed in this paper, the average effect is
a decrease in the first moment. Importantly, with the flexibility in other higher-order moments
of the predictive distribution, the magnitude of the shift will vary over time. The departure
from the typical parameter uncertainty setting also introduces interesting implications for the
variance of the distribution. Under parameter uncertainty, in order to fit the data, extreme
observations are captured by the increase in the variance of the predictive distribution. In
fact, this feature is representative of many models with parameter uncertainty. Allowing for

---

4Similar logic has been presented in the recent studies by Kan and Zhou (2003), and Tu and Zhou (2003);
however, both papers assume a fixed form of the distribution, thus restricting a flexibility coming from time-
varying higher-order moments.
distribution uncertainty brings about significant changes in the resulting estimate of variance. First, extreme observations can be fitted by the appropriate adjustment of higher moments without a significant increase in the variance. Interestingly, however, apart from the impact introduced by the flexibility in higher-order moments, investors’ perception of the variance of the distribution is additionally 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 is reflected in a higher perceived level of variance. Consequently, the variance perceived under distribution uncertainty is a combination of two effects: the negative effect coming from flexible higher moments, and the positive effect coming from nonstationarity. The resulting value of variance depends on which of the two effects is stronger. Indeed, 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 higher moments have a dominating role for investors who, in turn, use a cumulative window of all available data. Since in both cases the means of the distributions move down, the optimal allocations for the rolling window are significantly lower than for the cumulative window.

In another set of results, I find that the variability of portfolio weights for investors considering distribution uncertainty is smaller compared to the variability of weights of investors who only account for parameter uncertainty. The difference between the two alternatives is especially large for investors who condition their decisions on a rolling estimation window. The intuition for the variability result is as follows. Under distribution uncertainty, new information may add uncertainty regarding the values of the parameters of a new distribution and the type of the distribution. Under parameter uncertainty, each additional piece of information adds uncertainty only about the parameters and thus is more informative about the future. As a result, at each instant, investors adjust their portfolio allocations significantly more. Consequently, the overall variability of allocations is higher.

The above-presented differences in optimal portfolio allocations lead to economically significant utility effects. In particular, investors accounting for distribution uncertainty, forced

---

5 Barry and Winkler (1976) are among the first to study the importance of “nonstationarity risk” in the portfolio context.
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 0.04%, 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. This is an interesting observation as it suggests that uncertainty should be studied as a time-series phenomenon. This fact, by and large, has been ignored by the existing literature on asset allocation under Bayesian uncertainty.

To assess the econometric quality of predictions generated by my model, I analyze the performance of investment strategies which factor in distribution uncertainty. Results show that the strategies of investors who account for distribution uncertainty perform remarkably better than the corresponding strategies under parameter uncertainty. The superior performance is starkly evident for investors with a rolling estimation window. Further, strategies based on distribution uncertainty deliver a significantly better risk-adjusted performance both compared to respective strategies based on parameter uncertainty, and the strategy tracking the market portfolio. The above results seem to point to a practical applicability of such strategies.

The remainder of this paper proceeds as follows. Section I reviews the related academic literature. Section II 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 III. Section IV investigates the ex post performance of the strategies that account for distribution uncertainty. Section V reviews the main conclusions of the paper.

I. Related Literature

The literature on asset allocation under Bayesian uncertainty has been largely limited to analyzing the impact of parameter uncertainty. The empirical evidence using this setting includes Klein and Bawa (1976), Bawa, Brown, and Klein (1979), and Barberis (2000) in the context with no predictability, and Kandel and Stambaugh (1996), Brennan (1998), Barberis (2000),
and Xia (2001) in the context with predictability. To my knowledge, the only study related to
distribution uncertainty in the portfolio context has been Bawa (1979). His study outlines the
important theoretical concepts with respect to portfolio selection of multiple risky assets, yet
it does not deliver extensive empirical results. This paper extends the above-mentioned litera-
ture in that it simultaneously considers parameter and distribution uncertainty in an empirical
context from a Bayesian perspective.

This paper 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, you
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: for example, see
Pástor and Stambaugh (1999) and Pástor and Stambaugh (2000), and later Avramov (2002)
and Cremers (2002). In contrast to the above studies, this paper considers 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.\footnote{An alternative approach to analyzing model uncertainty, often defined in this context as ambiguity, has been proposed following the seminal work of Knight (1921). The research based on this approach has yielded several interesting results. Of special interest for the current study are papers on asset allocation. Dow and Werlang (1992), Maenhout (2001) study a single period portfolio choice problem. Uppal and Wang (2003) study the asset allocation with multiple risky assets. Cao, Wang, and Zhang (2002) consider asset allocation from the perspective of a limited stock market participation.}

The problem of incorporating uncertainty in the data-generating process has been a topic of
a recent study by Tu and Zhou (2003). They model uncertainty in the data-generating process
by measuring the distance between decisions based on Normal distribution and decisions based
on various t distributions. However, in contrast to the current study, their framework is still
anchored in a world with a finite number of parameters and hence, from the methodological
viewpoint, limits the notion of distribution uncertainty. Specifically, their study is restricted
to choosing among various forms of t distribution, while here the focus is on a richer class of
all possible unimodal and symmetric distributions, which leads to estimating the model with
an infinite number of parameters. Also, they consider a slightly different portfolio selection
problem with multiple risky assets, of the type studied by Pástor and Stambaugh (2000), while
here the focus is on asset allocation between a risky and a riskless asset. Finally, their study investigates optimal allocations taking into account only one estimation window. As will be shown, this somewhat limited approach renders conclusions which are significantly different from the conclusions of this paper where one considers multiple data ranges.

II. Methodology

This section outlines a theoretical framework for analyzing the problem of portfolio allocation under distribution uncertainty. The first section defines the economic set-up and emphasizes the importance of the predictive distribution of returns for decision-making. The assumptions from this section are utilized in the next two sections, which consider two particular forms of uncertainty that investors may face: parameter uncertainty (estimation risk) and distribution uncertainty. The emphasis is put on developing a framework for incorporating distribution uncertainty, which is shown to include parameter uncertainty as well.

A. Economy

I consider investors with a one-period investment horizon facing the decision of allocating their wealth at time $T$, $W_T$, between a risky and a riskless asset, with their respective one-period, continuously-compounded returns of $R_{T+1}$ and $R_{fT}$. For simplicity, I assume that investors do not have any labor income or a real estate income. Further, their utility, derived from terminal wealth $W_{T+1}$, is represented by the standard Constant Relative Risk Aversion (CRRA) power function with the coefficient of relative risk aversion, $A$:

$$ U(W_{T+1}) = \begin{cases} \frac{W_{T+1}^{1-A} - 1}{1-A} & \text{if } A > 1 \\ \ln(W_{T+1}) & \text{if } A = 1 \end{cases} $$

(1)

An example of a risky asset could be a well-diversified equity fund, while a riskless asset could be proxied by a short-term money market fund.

The selection of this particular utility is motivated by its convenient properties, such as decreasing absolute risk aversion and constant relative risk aversion. Several studies indicate that these particular properties closely approximate the behavior of ‘normal’ investors (e.g., Huang and Litzenberger (1988) and Campbell and Viceira (2002)).
The objective of the investors is to find an optimal allocation of resources to maximize expected utility at time $T$:

$$\hat{\omega}_T = \underset{\omega}{\operatorname{argmax}} E_T[U(W_{T+1})],$$  \hspace{1cm} (2)

where

$$W_{T+1} = W_T(\omega_T \exp(r_{T+1} + R_{fT}) + (1 - \omega_T) \exp(R_{fT})),$$  \hspace{1cm} (3)

$r_{T+1} \equiv R_{T+1} - R_{fT}$ denotes excess market return, and $\hat{\omega}_T$ represents an optimal allocation of wealth into a risky asset at period $T$. Throughout the paper, I assume that investors can take both short and margin positions in the risky asset.\footnote{It is well-known (e.g. Kandel and Stambaugh (1996)) that under power utility one cannot take margin/short positions in either of the assets because, theoretically, there exists a very small chance that the wealth may drop to zero, and thus the expected utility may drop to a negative infinity. This would, however, require a presence of a very extreme realization. In the current study, the results based on a very large number of simulated values do not indicate such a problem. In Section III.B, I provide additional justification that this problem may not be that severe for the results.} Since for the power utility the optimal allocation to risky assets is wealth-independent, without loss of generality, I let $W_T = 1$. 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 (2) and (3), and rewrite the conditional expectation in (2) in terms of the predictive distribution of future returns, $p(r_{T+1} \mid \cdot)$, the optimization problem of the investors can be represented as:

$$\hat{\omega}_T = \underset{\omega}{\operatorname{argmax}} \int U(\omega_T \exp(r_{T+1} + R_{fT}) + (1 - \omega_T) \exp(R_{fT})) \ p(r_{T+1} \mid x_T, \lambda) dr_{T+1},$$  \hspace{1cm} (4)

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

$$r_{T+1} = \mu(t) + \epsilon_{T+1},$$  \hspace{1cm} (5)
where $\mu(T)$ means that $\mu$ could either be constant or could depend on a set of predictor variables up to time $T$.\footnote{Although the empirical part of this paper ignores the evidence of predictability, the theoretical framework is general enough to accommodate the case with predictability. For the focus of this paper is merely on distribution uncertainty, the objective for future research is to discuss the issues related to predictability. In fact, this is a topic of chapter 4 of my dissertation.} The important aspect of this equation is that it does not impose any distributional form on the error term. For the moment, I assume merely that $\epsilon$ 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.

**B. Parameter uncertainty**

The specification in equation (4) requires investors to use the predictive distribution of the excess market return, $r$, in their decisions. From (5), it is clear that this distribution depends on past data and a set of parameters describing 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 what is commonly defined as 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. With the Normality assumption the set of parameters describing distribution is spanned entirely by $\mu$ and $\sigma^2$. The resulting predictive distribution is then a student-$t$;\footnote{In fact, such $t$-distribution, due to its high number of degrees of freedom, is almost identical to the Normal distribution. This will be evident in the empirical part of the paper.} for details on these types of models and other variations, see for example, Kandel and Stambaugh (1996), Barberis (2000), and Pástor (2000).
C. Distribution uncertainty

The key difference in modeling distribution uncertainty, as compared to mere modeling parameter uncertainty, can be illustrated using equation (5). 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 Normally distributed (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, the main 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. The main consequence of this approach is that the predictive distribution of returns will no longer be fixed; it will be random; technically it is this random feature in the modeling framework that is labeled *distribution uncertainty* in this paper.\(^{12}\)

For tractability, I assume that investors believe that the conditional distribution of the excess returns is unimodal and symmetric. This is not a very restrictive assumption, especially in relation to the literature on Bayesian uncertainty, because, for starters, it includes most of the well-known parametric uncertainty models that have been used in the literature as a special case. Thus, for example, the Normal and student-t distributions are members of this family, as well as, for example, Cauchy. Also, as pointed out by Campbell, Lo, and MacKinlay (1997) (pp.19-20), the observed deviations from normality observed in the monthly market returns are more pronounced by means of excess kurtosis than skewness. Importantly, the framework with symmetric distributions still appeals to the notion of distribution uncertainty in that one does not have to assume a particular form of the distribution. Finally, as I will show later, although this class ignores different levels of skewness, it includes a wide range of distributions with extremely different levels of kurtosis and other higher ‘even’ moments. Importantly, allowing for time-varying higher moments has significant consequences for lower

\(^{12}\)Technically speaking, distribution uncertainty can be understood as an extension of the parameter uncertainty case with an infinite number of parameters.
moments of the distribution such as mean and variance. This impact will be illustrated in the subsequent sections of the paper.

In the introduction it was noted that the parameter uncertainty approach is useful because it imposes structure via a regression framework. The distribution uncertainty approach, if it is to be practically useful, must also include such a regression structure. This is obtained by using a class of models that have been called semiparametric scale mixture of uniforms (SSMU) by Kacperczyk, Damien, and Walker (2003). The approach in this paper complements and extends the SSMU in their paper in the following way. In Kacperczyk, Damien, and Walker (2003), the evolution of the diffusion process is modeled by fixing the mean process at zero, while modeling the variance process. Here, the mean process is modeled by a linear regression; the variance process can also be modeled simultaneously. Although the inference mechanism in this paper differs significantly, the general theoretical ideas remain similar in spirit to Kacperczyk, Damien, and Walker (2003). It is important to note that the SSMU framework is superior to any purely nonparametric framework, be it classical or Bayesian one, in that one can model the impact of other regressors both in the mean and variance regression. Such a flexibility is absent in the latter setting.

The SSMU class of models is based on a nonparametric family of prior distributions, namely: the Dirichlet Process, scale mixture representation, linear regression, and Gibbs sampling.\textsuperscript{13} The rest of this section will illustrate the importance of each of these components in defining the SSMU framework.

Consider the following return generating process.

\[
r_t = \mu_t + e_t, \quad t = 1, \ldots, T,
\]

where \(r_t\) is the excess market return, \(\mu_t\) is the market premium, and the \(e_t\) are i.i.d. error terms. In this framework, \(\mu\) can be held constant (i.i.d. returns) or can depend on other economically motivated regressors (predictable returns). The implications of both assumptions for modeling

\textsuperscript{13}For discussion of the Dirichlet process see, for example, Ferguson (1973) and Walker, Damien, Laud, and Smith (1999), scale mixture models have been presented, for example, in Feller (1971), for Gibbs sampling consult, for example, Smith and Roberts (1993).
purposes will be shown in subsequent paragraphs. Also, the following assumptions about the errors apply:

A1. $E(e_t) = 0$, and the $e_t$ are symmetric about 0.

A2. $\text{var}(e_t) = \sigma^2 < \infty$.\(^{14}\)

Note that one can write the model in (6) in a different way. Introduce the latent variable $u = (u_1, \ldots, u_T)$ and consider a stochastic process

$$r_t | u_t = \mu_t + \tau t \sqrt{u_t}, \ t = 1, \ldots, T,$$

(7)

where $\tau_t$ are i.i.d. from the uniform distribution on $(-1, +1)$, and $u_t$ are i.i.d. from some distribution $G$ defined on $(0, \infty)$. Kacperczyk, Damien, and Walker (2003) show that, provided $E(u_t) = 3\sigma^2$, conditions A1 and A2 are satisfied, i.e. the model in (7) is equivalent to the model in (6). Also, they show that the representation in (7) embeds many types of standard regression models. An important fact to observe is that a linear representation in (6) can be reformulated using a ‘scale mixture of uniforms’ approach of (7). In such a framework, the mean and variance of $u$ determine the variance and kurtosis of $r$. For example, to obtain $\text{var}(r) = \sigma^2$ and kurtosis $(r) = \tau$, we require $E(u) = 3$ and $\text{var}(u) = 5\tau + 6$. This implies $\tau$ has to be greater than $-6/5$, which is the kurtosis of the uniform density. Kacperczyk, Damien, and Walker (2003) show that a particular distribution that satisfies these requirements is given by

$$f(u) = \text{Gamma}(9\alpha, 3\alpha),$$

(8)

where $\alpha = (5\tau + 6)^{-1}$. This new family of distributions has parameters $(\mu, \sigma^2, \tau)$, where each element in the vector corresponds to the mean, variance and kurtosis, respectively. Clearly, the Normal distribution is recovered when $\tau = 0$ ($\alpha = 1/6$). Similarly, by changing the specifications of the parameters one can obtain other commonly used distributions, such as t, generalized exponential, etc.

\(^{14}\)One could also apply the proposed methodology to model the mean of distributions with infinite variances, e.g. Cauchy.
So far, the above discussion has focused on modeling parameter uncertainty using a uniform scale mixture representation. To introduce distribution uncertainty, consider the following canonical form of the SSMU class of models.

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

\[u \sim F,\tag{9}\]

If \( F \) is taken to be Normal, one is still in the realm of parameter uncertainty, where one has to impose prior beliefs on the parameter space. On the other hand, if \( F \) is assigned a stochastic process prior, then it is tantamount to saying that the conditional distribution of the returns is no longer fixed; it is random; this is conventionally called Bayesian nonparametrics in the statistical literature. The literature on Bayesian nonparametrics suggests that such randomness in the priors can be obtained using, for example, the Dirichlet process.\(^{15}\) The Dirichlet process is flexible in that priors are assigned to the set of probability distributions on a sample space rather than just to the parameter space as is the case under parameter uncertainty. \( F \sim \text{Dir}(c, F_0) \) means, \( F \) is assigned a Dirichlet process prior with mean \( F_0 \) and scale parameter \( c \). \( c \) is a measure of strength of belief in your prior guess at the shape of \( F \).

Note, as an example, one could center the location parameter, \( F_0 \), around any member of the exponential power family, the most famous one being the Normal distribution. In this paper, I use the Dirichlet process for two reasons: (a) the theoretical properties of the process are very appealing; see Ferguson (1973); (b) implementing the overall model is highly simplified; see MacEachern (1998) and Appendix B. Many studies have recognized the difficulties in sampling directly the Dirichlet process. The scale mixture representation is such that simulating the Dirichlet process is bypassed; that is, the computational burden is substantially reduced; see Appendix B for details.

\(^{15}\)Let \( \tilde{Z}_1, \ldots, \tilde{Z}_k \) be independent random variables with \( \tilde{Z}_j \) having a Gamma distribution with shape parameter \( c_j \geq 0 \) and scale parameter 1, for \( j = 1, 2, \ldots, k \). Let \( c_j > 0 \) for some \( j \). The Dirichlet distribution with parameter \((c_1, c_2, \ldots, c_k)\), denoted by \( D(c_1, c_2, \ldots, c_k) \), is defined as the distribution of \((\theta_1, \theta_2, \ldots, \theta_k)\), where \( \theta_j = \frac{\tilde{Z}_j}{\sum_{i=1}^k \tilde{Z}_i}, j = 1, 2, \ldots, k \); Ferguson (1973).

Let \( c \) be a finite nonnull measure (nonnegative and finitely additive) on \((R^m, B)\). \( P \) is a Dirichlet process with parameter \( c \), denoted by \( P \in D(c) \), if for every finite measurable partition \( \{B_1, \ldots, B_n\} \) of \( R^m \) (i.e., the \( B_i \) are measurable, disjoint and \( \bigcup_{i=1}^n B_i = R^m \)), the random vector \((P(B_1), P(B_2), \ldots, P(B_n))\) has a Dirichlet distribution with parameter \((c(B_1), c(B_2), \ldots, c(B_n))\); Ferguson (1973).
To summarize the development thus far,

\[ [r_t | \mathcal{F}_{t-1}] \sim U \left( \mu_{t-1} - \sigma_{t-1} \sqrt{u_t}, \mu_{t-1} + \sigma_{t-1} \sqrt{u_t} \right), \]

and

\[ u_1 \ldots u_T \sim p(u_1, \ldots, u_T), \quad (10) \]

where \( \mathcal{F}_t = [\mu(r_1, \ldots, r_T); \sigma(r_1, \ldots, r_T)] \) denotes a standard filtration. Equation (10) is essentially a scale mixture of uniforms representation of the observed data conditioned on past values of the mean and the variance.

Taking \( p(u_1, \ldots, u_T) \) to be based on a Dirichlet process prior induces distribution uncertainty. Clearly, the above representation also includes parameter uncertainty. The key point here is that the predictive distribution of the asset returns based on the above representation is no longer fixed; it will be random as well, having no closed form solution; see Appendix C for details.

C.1. Prior Distributions

This section describes the various priors used in the empirical analysis. Where necessary, a conjugate hyper-prior is used.\(^\textsuperscript{16}\) In specifying priors, the focus is on two components: the distribution uncertainty and the parameter uncertainty components.

The distribution uncertainty component: As described before, 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, I assign the scale parameter, \( c \), a Gamma\((a,b)\) hyper-prior distribution. Since most of the existing studies tend to assume normality of the predictive distribution, I also center the prior \( F_0 \) around the Normal distribution, i.e. the location parameter, \( F_0 \), of the Dirichlet process is assigned a Gamma\((3/2, 1/2)\) distribution; see Kacperczyk, Damien, and Walker (2003) for further details.

\(^\textsuperscript{16}\)Given recent advances in Bayesian computation, one could readily employ non-conjugate prior distributions if needed; for details, see MacEachern (1998), and Mira, Moller, and Roberts (2001).
The parameter uncertainty component: Equation (10) defines a stochastic 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,

$$\mu_t = \beta_0 + \sum_{k=1}^{K} \beta_k Z_{kt},$$  

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 $t$, and the variance regression,

$$\sigma_t = \exp(\theta_0 + \sum_{l=1}^{L} \theta_l W_{lt}),$$  

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 $t$.

I assume prior distributions for each $\beta_k$ to be independent Normal distributions with zero means and variances $\psi_k^2$. Similarly, I assign priors for $\theta_l$.

Note first that one can readily reduce the above class of models to the case of conditional i.i.d. returns by merely eliminating all predictors. Second, if interest is also on mean predictability, then one could eliminate predictors in the variance regression. Lastly, if the variance process is also of interest then the above general representation encapsulates that facet as well. The empirical part of the analysis, for the aforementioned reasons, will focus only on the conditionally i.i.d. case.

C.2. Predictive Distributions and Portfolio Weights

In the construction of the SSMU 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 (10) provides the expression for obtaining future values of the 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 C.

According to formula (2), 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 (4). Under distribution uncertainty, the predictive distribution does not possess a closed-form; hence, it is not possible to solve the integral in equation (4) analytically. Instead, I evaluate the integral numerically at several grid points of possible weights from the specified domain. Such an integral can be approximated, with considerable accuracy, as

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

or equivalently as

\[ E(U) = \frac{1}{N} \sum_{i=1}^{N} \left[ \omega \exp(r^{(i)} + R_f) + (1 - \omega) \exp(R_f) \right]^{1-A} - 1, \]  

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 (14).

D. 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, I restrict the
sample to the period of January 1954 to December 2000.\textsuperscript{17} This amounts to 564 months of data. Table I presents summary statistics of the data.

Insert Table I about here

III. The Impact of Distribution Uncertainty on Asset Allocation

This section studies empirically the importance of distribution uncertainty for asset allocation. One can feasibly conduct such an analysis under the assumption of conditionally i.i.d. returns. Since investors’ beliefs about the stability of the relationships in the underlying data may differ substantially, all the subsequent empirical tests assume that, in forecasting returns, investors consider a sample which includes either all past available data (cumulative window) or the last 10 years of data (rolling window). Another rationale for choosing these particular windows is to maintain consistency with the analysis presented in Barberis (2000). The subsequent tests consider the coefficient of risk aversion in the power utility to be greater than 1.\textsuperscript{18} This assumption is consistent with the existing evidence, which suggests that the coefficient is approximately equal to $A = 2.8$. For both windows, I 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.

A. Portfolio Selection

I consider investors who believe that, when forming predictions about future returns, there is nothing to learn from the past data beyond the return generating process implicit in the sequence of past returns, i.e., returns are conditionally i.i.d. However, investors can differ in

\textsuperscript{17} A similar rule has been applied in Barberis (2000), and Cremers (2002), among others.

\textsuperscript{18} The difference between power utility with $A > 1$ and its special case – log utility – is more important if one considers investors with long horizons. In such a case, as is also the case with i.i.d. returns, Merton (1969) and Samuelson (1969) show that the optimal allocation under log utility is consistent with the myopic, short-term investors.
their prior beliefs regarding the form of the conditional distribution of the returns, leading to different predictive distributions. Here, investors either account for, or ignore distribution uncertainty.

Under parameter uncertainty, one can obtain the posterior predictive distribution of returns by combining Normal-Inverse Gamma conjugate priors under the assumption of Normally distributed errors. Using a sampling method given in Barberis (2000), the predictive distribution from such a model is sampled; these sampled values are then used to make inferences about portfolio selection. Under distribution uncertainty, to obtain predicted values of the equity premium, MCMC is employed. Under conditionally i.i.d. returns, \( \mu \) and \( \sigma \) do not depend on any predictor covariates. Hence, the sampling process outlined in Section II becomes simpler; see Appendix B for details.

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 the remaining values form the basis for all posterior inferences. I implement standard MCMC convergence diagnostic checks; see for example, Smith and Roberts (1993).

To reflect vague prior knowledge of investors regarding the distribution of future returns, I assume non-informative priors under both types of uncertainty. This assumption reflects a notion that investors generally do not have much information about the precise nature of future returns and is consistent with other related studies (e.g. Barberis (2000)). Thus, under parameter uncertainty, I impose a non-informative prior for both parameters. Under distribution uncertainty, as described before, parameters are sampled from the Normal distribution with mean zero and significantly large variance, to make the prior noninformative. Besides, the scale parameter of the Dirichlet process, \( c \), is assigned a \( \text{Gamma}(0.01, 0.01) \) prior. This prior gives significant weight to both small and large values of \( c \) and thus imposes considerable uncertainty regarding the choice of the family of transition densities, which, from earlier, is taken to be Normal; i.e., \( F_0 \) is assigned a \( \text{Gamma}(3/2, 1/2) \) prior. Given the nature of Gibbs
sampling one has to assign a starting value to \( c \). In the subsequent analysis this value is set equal to \( c = 20.19 \)

To illustrate the specifics of the investment process, I consider first investors with one-month investment horizon who stand at the end of 2000 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 parameter or distribution uncertain. They condition their decisions either on the entire available past data (1954:1-2000:12) or on the last 10 years of past data (1991:1-2000:12). 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 a formula in (4). 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. The direct impact of each of the moments for portfolio selection is marginally smaller as we move higher in the moments, i.e., the impact of the mean is first-order, of the variance is second-order, etc. Higher moments, however, 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.

Insert Figure 1 about here

The top-panel graphs, representative of investors who use a 10-year series of past data, indicate that investors who consider distribution uncertainty expect a much lower equity premium (0.64% per month) when compared to investors who consider only parameter uncertainty (1.26%). The respective difference for investors using a cumulative window, as illustrated in the bottom-panel graphs, is smaller: 0.26% vs. 0.38%.

\[19\] The qualitative aspects of the analysis are not sensitive to the selection of this value. I have verified the sensitivity of the analysis for a wide range of starting values of \( c \), between 5 and 50, and the results remained qualitatively similar. For brevity, I am not reporting them.
The differences in means in 2001 can be explained using the following argument. Over the last 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 has also important consequences for the variance. In particular, under distribution uncertainty, one should expect that the variance would be lower because, in contrast to parameter uncertainty with Normal distribution, the possibility of changes in higher moments does not require the variance to be increased to capture the existence of tails in the distribution. However, there is an additional effect which may play the opposite role in the estimation of 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 this parameter. As a result, the final impact on the variance will be a combination of the two above-mentioned effects. 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 1.5 times higher under distribution uncertainty, due to “nonstationarity risk” the respective variances do not differ much.\textsuperscript{20} 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.

\textsuperscript{20}Given that for a rolling window the estimation process requires a considerable number of 120 observations, it is unlikely that such a difference would reflect low estimation precision.
Given that, for investors considering distribution uncertainty, the mean of the predictive distribution is lower and its variance unchanged, one should expect the optimal stock allocation to be lower. Indeed, with distribution uncertainty optimal stock allocations are half as large as allocations under parameter uncertainty. For example, for investors with risk aversion of $A = 3$ optimal allocations in stock market equal 152.11% and 295.38%, respectively.

For a cumulative window, since a decrease in mean of the distribution is partially offset by a decrease in variance, and higher moments differ to a lesser degree, the resulting portfolio allocations under both types of uncertainty are likely to be more similar than for a 10-year window. For investors with risk aversion of $A = 3$, the respective optimal allocations in stock market are 75.03% and 88.36%. Table II presents detailed information of the portfolio choices assuming four different values of risk aversion: 3, 5, 10 and 20.

Insert Table II about here

To gauge the economic magnitude of this difference, I calculate the certainty-equivalent loss (CEL) as in Kandel and Stambaugh (1996). Briefly, the certainty-equivalent loss is defined as a difference between the certainty-equivalent return (CER), obtained for 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 $\text{CER} -$ 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 calculated using equation (1) and assuming that predictions are obtained under distribution uncertainty. For relative risk aversion of 3, the difference in CERs for a rolling
window amounts to an annualized return of 6.58%. This result is economically significant. In contrast, for a cumulative window, this difference, on the annual basis, equals 0.04% and is economically negligible.

One could argue, using the argument of observational equivalence, that the ex-ante predictive distribution obtained under distribution uncertainty would be indistinguishable from the distribution with similar parameters obtained under parameter uncertainty. But investors are not clairvoyant. There is no a priori way of judging the magnitude of higher-order moments, an example being kurtosis. Modeling distribution uncertainty obviates the problem of forcing any particular degree of higher moments, and, in a loose sense, allows the data to speak for itself. Next section, as an example, will present direct evidence on time-varying kurtosis under distribution uncertainty. Another possibility for observational equivalence would be that in the limit any distribution has to converge to the Normal. But this is unlikely to be the case here, for as we have claimed above, the underlying predictive distribution may be non-stationary and thus may not converge to one particular distributional form. Instead, the model behaves more like a regime-switching process with infinitely many regimes. Also, the convergence theorems, typically in this context, apply to means while here focus is on the entire distribution. The upshot is that decisions made under distribution uncertainty cannot be directly mapped to decisions made under parameter uncertainty.

B. Time-Series Allocations

The results obtained so far for January 2001 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 where investors would actually invest more in the stock market. This decision entirely depends on investors’ perception of the future. The analysis using multiple periods is also an extension of most of the previous studies on parameter uncertainty. Most of those studies analyze decisions using one data point, i.e. they fix an estimation window and analyze all the effects using that one window. 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, I consider the sample for the period of 1954-2001 broken up in December of each year from 1963 to 2000. As a result, I obtain 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. For investors with a risk aversion of $A = 3$, the average allocation for the strategy based on the entire past history (10 years of most recent history) equals 75.92% (45.43%) and 98.47% (82.46%) under distribution and parameter uncertainty, respectively. Table III presents the average allocations in the stock market with the coefficient of relative risk aversion fixed at 3, 5, 10 and 20.

Again, I 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 1.40%. This value is not very large, which is also a conclusion from the study by Tu and Zhou (2003). Such claim, however, ignores the substantial variation of that measure; in my sample, the annualized standard deviation of the losses equals 2.54%. 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 equals 0.36% with the standard deviation of 0.57%. This finding is in line with other results reported so far, which indicate more significant differences in case of the rolling estimation window. The value of the loss reported for the cumulative window is similar in magnitude to the results reported in Tu and Zhou (2003) for the analysis using a period of 1963-1997. Table IV presents the detailed information based on the time-series of the certainty-equivalent losses.

Insert Table IV about here

It is important to delineate the factors which cause the differences in optimal allocations. When analyzing the allocations for 2001, I noted the importance of mean, variance and kurtosis. In order to investigate whether the same forces drive the differences in allocations over time, I obtained a time-series of the above parameters. The results are presented in Table V.

Insert Table V about here

Most of the findings described for 2001 are also present in other periods. In particular, both for the rolling and cumulative estimation window the mean returns are lower under distribution uncertainty. The respective monthly comparisons are 0.28% versus 0.18% for a rolling, and 0.37% versus 0.25% for a cumulative window. In both cases, the standard errors of the estimates are very similar which indicates 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 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 window of data.

Given that one can observe a significant shift in the mean 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. Importantly, a significant time variation of kurtosis, as compared to no variation under parameter uncertainty, suggests that investors’ perception of distribution uncertainty is time-varying. 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. This is exactly the result I have reported in Table III.

An interesting feature of the data related to investment strategies based on distribution uncertainty is that the derived 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 6 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 recent study by Guidolin and Timmermann (2003). 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 observed allocations to stocks are much less volatile, and do not exceed 100%, which, in turn, is typical for the short-term investments under regime-switching models.

One can explain the variability result using the following economic intuition. Suppose investors at each subsequent period had perfect information about the future. Then they would invest their money solely by comparing the returns on assets in their portfolios, and allocate an infinite amount of money into an asset with higher returns. The volatility of such a portfolio would equal infinity. In contrast, if investors had no additional knowledge about the
future when moving from one period to another, they would always invest the same proportion of their wealth in each specific asset. With consideration of distribution uncertainty, investors, at each point of time, learn less about the future as they do not know whether a new piece of information results in the shift of the parameters of the predictive distribution or in the change of the entire distribution. Investors facing such uncertainty will adjust their allocations less than investors who would not face such additional uncertainty (i.e. in the case of parameter uncertainty only). Hence, the variability of optimal portfolio allocations under distribution uncertainty is likely to be lower than under parameter uncertainty.

A striking feature of the above investment process under uncertainty is that the optimal allocations often require taking a margin or a short position in a risky asset. This problem, as can be seen from the results, is more severe for investors who account for parameter uncertainty, but ignore distribution uncertainty. Several prior studies (e.g. Kandel and Stambaugh (1996), Barberis (2000)) do not expose the scale of this problem, as they restrict investors from engaging in such an activity. The restriction on short selling and margin positions is usually done via assumption on the utility. In fact, under power utility, investors should always avoid taking leveraged positions as this may lead to a negative expected utility. I argue, however, that the cases with negative utility are extremely rare. Indeed, the results from the simulations in this study suggest that extreme events that could potentially lead to infinitely large negative expected utility do not occur too often. Virtually, in none of the several simulations in this study does this take place. Moreover, since uncertainty has an impact on the predictive distribution and not on the utility itself, the results related to uncertainty should not be entirely driven by the utility one specifies. In particular, imposing trading constraints on investors, as suggested in a setting with a power utility, may not properly reflect their risk attitudes. To support this argument, in the unreported results, I calculate allocations for investors with a quadratic utility, which apparently does not suffer from the same problem as a power utility. The allocations are qualitatively similar. This suggests that the qualitative aspects of my results are not driven by the choice of utility. From an economic perspective, one could argue that in the very extreme cases, if investors were running out of wealth, they could either protect themselves with some derivative contracts, such as put options, or their wealth could be protected by the government. This would simply indicate that the wealth
always stays positive. In sum, it is unlikely that the results here suffer because of the power utility specification.

C. Distribution uncertainty and the representative investor

The analysis of asset allocation in this paper is not a general equilibrium argument. Specifically, the Bayesian investor considered in this paper is not the representative investor; otherwise, he would just hold the market. However, one could take one step back and ask the question whether the Bayesian investor considered in this study looks like a representative investor. This question seems to be relevant especially since rationality, which underlies the concept of representative investor, is based on Bayes rule.

To that effect, this section takes a slightly different approach and investigates how much the Bayesian investor who considers distribution uncertainty would differ from the “typical” representative investor. To solve this problem, I compare the stock market allocation of such an investor to the stock market allocation, held by the representative investor.\footnote{A different approach could be to find the coefficient of risk aversion which would match the optimal allocation of representative investors (100%). This approach, however, has its limitations as it does not return sensible values of risk aversion coefficient in case of a negative equity premium. In particular, in the current study, such a situation is present for the Bayesian investor with the rolling estimation window.} To make the comparisons compelling, one has to utilize portfolio allocations of Bayesian investors derived under a reasonable level of risk aversion. Here, consistent with the existing studies, the risk aversion coefficient is assumed equal to 3. To compare the two types of investors, I again consider Figure 2. This time, the focus is on the differences between allocations in a risky asset obtained for parameter and distribution uncertainty and the fixed allocation in a risky asset of 100%.\footnote{The fact that I consider 100% as a benchmark implicitly assumes that the equity is the only asset in a positive net supply. If one relaxed this assumption and, say, additionally allowed corporate bonds this would require a lower value of the benchmark.}

It is apparent from the graphs that a Bayesian investor’s strategy is quite different from a strategy of the representative investor. Interestingly, with the conservative risk aversion level, he occasionally shorts the market. For instance, an investor who accounts for distribution uncertainty and uses a 10-year rolling estimation window would have shorted a risky asset
each year between 1972 to 1984. This is a surprising result as, in equilibrium, the risk-averse investor should always hold a positive amount of wealth in risky assets (Merton (1980)).

Several implications can be discussed in light of the above developments. First, the representative investor, by and large, does not look like a Bayesian investor. Second, the representative investor may not be the only investor in the market. Depending on investors’ beliefs, one can observe substantial heterogeneity in investment decisions. Third, assuming equilibrium setting, the results in the paper may shed some light on the well-documented equity premium puzzle. Several studies have taken the approach where they try to justify the observed realized excess returns using different economic specifications. Here, I look at the premium from the perspective of expected returns. In particular, as demonstrated in Table V, accounting for distribution uncertainty has decreased the equity premium. This result suggests that the premium by itself may not be as large as it is often claimed. The observed decrease in the premium is entirely possible in light of the fact that many studies estimate the premium at the mean without taking into consideration the significant risk related to its estimation. As a result, a proper assessment of estimation risk may be a desirable feature in estimating the magnitude of equity premium.

D. 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 very 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 for the case with parameter and distrib-

23 Similar result of negative allocation has been found by Liu (2001) for the problem of dynamic portfolio choice.
24 Important distinctions between one and the other approach have been presented by Elton (1999) in his presidential address.
25 Similar effect of a decrease in the premium has been obtained by Fama and French (2002) using models based on the dividend and earnings growth.
26 A standard CAPM model is an example of the model with constant risk aversion.
bution uncertainty. To accomplish this, I 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. In order to capture the direct impact of distribution uncertainty, risk-aversion is set constant. Algebraically, the Index can be calculated as

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

(17)

where \( IDU \) is the Index of Distribution Uncertainty, \( \omega_p \), and \( \omega_d \) denote optimal weights under parameter and distribution uncertainty, respectively.

An increase in the value of such an index in the direction of distribution uncertainty 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 3 presents the levels of the uncertainty index for investors with risk aversion of \( A = 3 \), and \( A = 5 \) 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 can also be observed 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 above-mentioned 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.

IV. Performance

So far the analysis of portfolio weights has considered ex ante decisions of investors. The question I ask in this section is whether the strategies based on distribution uncertainty generate feasible predictions measured by their ex post performance. Perhaps, their perform very poorly and investors would never select them. Also, since such strategies are one of the possible alternatives available for investors, it is important to evaluate their practical value.

I 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, I 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.27

All beliefs are assumed fixed over time. In each December of a particular year, investors form a portfolio for January of the subsequent year. This portfolio is held for the remainder of the year until the end of next December when it is again rebalanced. Due to data limitations as well as computational constraints, such a strategy assumes that investors do not vary their allocations significantly during the entire investment year. From the statistical perspective, one would expect that the allocations would not be very volatile within each year, for the set

27The 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.
of additional new data available throughout the year is proportionately smaller than the fixed data available at the outset. Also, any differences should be less pronounced in the process of wealth formation due to a long-term averaging. From the practical perspective, 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, I 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 4 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.

Insert Figure 4 about here

The results obtained under each strategy are quite striking. Among the strategies I consider, 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 have amounted to almost $13 at the end of 2001. The second best strategy, based on a fixed investment of 45.43%, returns around $5. The strategies based on average stock market allocation of U.S. institutions and households return approximately $3.59 and $2.66, respectively. The corresponding strategies perform worse if instead of a rolling estimation window one considers a cumulative estimation window.\(^{28}\)

Several conclusions are noted. First, the respective strategies of investors accounting for distribution uncertainty perform better than strategies of investors solely relying on parameter uncertainty. This indicates that incorporating distribution uncertainty into the decision framework may offer positive wealth benefits. Second, the strategies based on a rolling esti-

\(^{28}\)As will be shown later qualitatively, similar conclusions hold even for risk-adjusted returns.
mation 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 a 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 institutions may not be optimal. The last point does not have to apply to all U.S. investors. 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 err on the side of caution 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 period of time. Also, it is difficult to determine what is an optimal length of the data one should consider to maximize investment’s performance. All that the above analysis says in this respect is that the length of the estimation horizon may have an important impact on the resulting performance.

The analysis of wealth cumulation assumes that all strategies do not differ in terms of their riskiness. In fact, it is very unlikely they do not. Hence, 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, I use the time-series of returns to calculate the average return, its standard deviation, and ex post Sharpe ratio under each strategy. In addition, for each strategy, I calculate the average ex ante Sharpe ratios based on the values of the first two moments of the predictive distribution. Finally, I estimate the risk-adjusted returns based on CAPM model and compare them across different strategies. Table VI presents the results for investors with risk aversion of $A = 3$, whose expectations are formed based on either a rolling or a cumulative estimation window.
Again, the best strategy investors should pursue, 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 56.24%, is 8 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 uncertainty-based strategies and investing in the market by analyzing estimated alphas from the CAPM model. This framework is suited for the problem considered here only if one assumes that investors exhibit some uncertainty about CAPM. If there was no uncertainty, one should expect any deviations of CAPM alpha from zero to be attributed to estimation error.\textsuperscript{29} Of all strategies under consideration, only the strategy based on distribution uncertainty and a rolling estimation window has a statistically significant positive alpha. This strategy is clearly superior to the strategy that follows the market. Other strategies do not differ significantly from the market even though some of them have positive abnormal returns. Interestingly, 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.

The analysis conducted until now does not indicate the driving forces behind the superior performance of strategies based on distribution uncertainty. I posit that the superior performance may be due to investors’ superior ability to time the market. This conjecture can be evaluated by comparing performance of strategies under distribution uncertainty to performance of strategies using fixed weights set at an average value of the time-series allocations under distribution uncertainty. If investors can effectively time the market, they should be able to benefit from varying their stock market allocations while keeping their average exposure to a risky asset at the same level as in the case of a fixed allocation. The results indicate

\textsuperscript{29}P\textsuperscript{ástor (2000) provides the detailed analysis of the sensitivity of portfolio allocations to the degree of model mispricing, alpha, in the context of a domestic and international asset.}
that the superior performance of strategies based on distribution uncertainty is likely to be due to effective market-timing. In particular, strategies based on distribution uncertainty, as was illustrated earlier, outperform strategies based on fixed allocations both before and after adjusting for risk. Interestingly, these strategies also outperform strategies tracking US households and institutions. This seems to suggest that the simple rule-of-thumb strategies are not superior to an effective market timing strategy developed under distribution uncertainty.

One could argue that an alternative explanation of the superior performance of the strategy based on distribution uncertainty could be related to volatility timing, in line with Fleming, Kirby, and Ostdiek (2001). This does not seem to be the case in here. The explanation related to volatility timing assumes that the expected returns are constant and the variation in the portfolio weights is purely driven by the changes in the conditional variance covariance matrix. The current study allows for both parameters to be time-varying. Also, as presented before, most of the differences in weights are not driven by the differences in the variances but rather by the differences in the means. Hence, the volatility timing does not seem to be a main driving force for my empirical results.

One could also argue that the existing differences in performance among various strategies come from the abnormally high returns in 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 the other period. One can test such a conjecture by dividing a sample into two halves and testing whether returns in both subperiods are statistically different from each other. Table VII presents the results of such an analysis.

Insert Table VII about here

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 distribution uncertainty where the returns in the second subperiod are significantly different from returns in the first subperiod. It is unlikely that such a difference is driven by an outlier, at least it seems to be suggested by observing the wealth process in
V. Concluding Remarks and Future Research

The optimal allocation of wealth into various financial assets lies at the heart of every investment decision. This process, among others, depends on the observed data and investors’ beliefs about future returns. In this paper, I study asset allocations into a risky and a riskless asset for investors who are uncertain about the predictive distribution of market returns and contrast them with asset allocations of investors who take into account parameter uncertainty, but consider the type of distribution known. The focus is on investors with short horizons, such as financial institutions and asset managers, or some individual investors whose interest is tactical asset allocation. In particular, it can be observed that, for most periods, introducing distribution uncertainty leads to a lower allocation in the risky asset. This result is mainly driven by the fact that investors’ perceived distribution of future returns, on average has a lower mean and higher variance as compared to the Normal. The difference in the first two moments, in turn, is additionally affected by the high and time-varying levels of higher-order moments.

The differences in allocations offer an opportunity to calculate the certainty-equivalent losses of investors who would be forced to allocate their wealth according to the model ignoring distribution uncertainty. Interestingly, the losses vary substantially over time. They can be very close to zero but can also achieve values of great economic significance. This further implies that distribution uncertainty is time-varying both from the 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 re-
turns rather than a result of a higher risk aversion or an increase in risk. This suggests that, in expectations, the difference between a risky and a riskless asset may not be that problematic.

Introducing the notion of distribution uncertainty also offers practical implications. First, among the strategies considered in this paper, the strategy based on distribution uncertainty is by far the best-performing strategy, even controlling for its riskiness, as measured by the Sharpe ratio or risk-adjusted returns. Second, conditioning on a shorter window of past data is more beneficial than otherwise. This may serve as evidence against the stationarity of returns.\textsuperscript{30}

A potential avenue for future work would be to investigate the implications of distribution uncertainty for long-term (strategic) asset allocation. Short-term (tactical) asset allocation ignores, for example, hedging motives of investors, which, in turn, are crucial in modeling strategic decisions. In this context, it may be useful to allow for additional predictors in the conditional mean equation. 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 (Goyal and Welch (2003)). Finally, one could also allow for possibly time-varying skewness in the predictive distribution of returns.

\textsuperscript{30}Similar conclusion has been drawn by Guidolin and Timmermann (2003) based on the model with regime switching.
Appendix A. The Gibbs Sampler

A particular, Markov Chain Monte Carlo (MCMC) method, the Gibbs sampler, is used to obtain the posterior distributions for all the critical random quantities in the model. Briefly, the Gibbs sampler is a Markov chain Monte Carlo method of sampling from conditional distributions, which, in the limit, induces samples from the appropriate posterior marginal distributions of interest. The key to implementing a Gibbs sampler therefore is to be able to obtain the conditional distributions, up to proportionality, of the random variables of interest.

In this paper, the following full conditional distributions are sampled,

\[ p(u_t \mid \text{everything else}), \quad t = 1, \ldots, T, \]
\[ p(\beta_k \mid \text{everything else}), \quad k = 1, \ldots, K, \]
\[ p(\theta_l \mid \text{everything else}), \quad l = 1, \ldots, L, \]
\[ p(c \mid \text{everything else}), \]

where \( N \) 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. Appendix B provides the details of this Gibbs sampler.

Appendix B. Estimation

This section provides details of the estimation of the SSMU model. The Gibbs sampler, indexed by superscript \((s)\), successively samples from the following full conditional distributions.

1. \( p(u_t \mid \cdots); t = 1, \ldots, T. \) The full conditional here is given by

\[ u_t^{-1/2} I(u_t > r_t^2/\sigma^2) \, (s-1) \, p(u_t \mid u_{-t}). \]

Now

\[ [u_t \mid u_{-t}] \propto c^{(s-1)} \, f(u_t) + \sum_{j \neq i} \delta_{u_j}(u_t), \]

For details, see Smith and Roberts (1993).
where $\delta_u$ is the point mass 1 at $u$, and $u_{-t}$ denotes all values of $u$ except for $u_t$. Recall that $f(u) \propto u^{1/2} \exp(-u/2)$. Consequently, we either sample $u_t$ from a truncated exponential distribution or take $u_t$ to be $u_j$, for those $u_j > r_t^2/\sigma^2(s-1)$, according to probabilities which are straightforward to compute. In fact,

$$
\left\{ \begin{array}{l}
\begin{align*}
& u_t^{(*)} \sim f^*(u) \quad \text{with probability } \tau \exp(-a/2) \\
& u_j \quad \text{with probability } 1/u_j 
\end{align*}
\end{array} \right.
$$

where $\tau = c(s-1)/(4\sqrt{\pi})$, $a = r_t^2/\sigma^2(s-1)$ and $f^*(u) = 0.5\exp\{(u(s-1) - a)/2\}I(u(s-1) > a)$.

2. $p(\theta_t|\cdots)$. Define

$$
\lambda_t = \frac{0.5 \ln(r_t - \sum_{j=1}^K \hat{\beta}_j^{(s-1)} Z_{jt})^2}{u_t^{(*)}} - \sum_{j \neq i} \theta_j^{(s-1)} W_{jt} 
$$

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

$$
[\theta_t^{(*)} | \cdots] \propto \pi(\theta_t) I \left( \theta_t \in \left[ \max_{W_{it}<0} \{\lambda_t/W_{it}\}, \min_{W_{it}>0} \{\lambda_t/W_{it}\} \right] \right). 
$$

If $W_{it} > 0$ for all $t$ then

$$
\max_{W_{it}<0} \{\lambda_t/W_{it}\} = -\infty 
$$

and if $W_{it} < 0$ for all $t$ then

$$
\min_{W_{it}>0} \{\lambda_t/W_{it}\} = \infty. 
$$

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

$$
\lambda_t = r_t - \sigma^{(s-1)} \sqrt{u_t^{(*)}} - \sum_{j \neq k} \beta_j^{(s-1)} Z_{jt}, 
$$

where

$$
\sigma^{(s)} = c \exp\left( L \sum_{j=1}^L \theta_j^{(s)} W_{jt} \right) 
$$

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

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

If $Z_{kt} > 0$ for all $t$ then

$$
\max_{Z_{kt}<0} \{\lambda_t/Z_{kt}\} = -\infty 
$$

and if $Z_{kt} < 0$ for all $t$ then

$$
\min_{Z_{kt}>0} \{\lambda_t/Z_{kt}\} = \infty. 
$$
The sampling for \( c \) proceeds as follows. In the first step, sample from the beta distribution for the new latent parameter \( \eta \in (0, 1) \):

\[
[\eta \mid c, k] \sim \text{beta} \left( c^{(s-1)} + 1, T \right).
\]  

(B10)

Then \( c \) is sampled from the mixture of gamma distributions, where the weights are defined as below,

\[
(c^{(s)} \mid \eta, k) \sim \pi_\eta \ Ga(k + a, b - \ln \eta) + (1 - \pi_\eta) \ Ga(a + k - 1, b - \ln \eta).
\]  

(B11)

Here \( Ga(a, b) \) is the prior distribution for \( c \) with \( a = b = 0.01 \); and \( \pi_\eta \) is the solution of the equation

\[
\pi_\eta/(1 - \pi_\eta) = (a + k - 1)/(b - \ln(\eta)).
\]  

(B12)

**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 \( N \) denotes the time at which we construct forecasts, for the predictions for period \( N + 1 \), at each iteration \( (s) \), sample the following components

\[
r_T^{(s)} \sim U(\mu_T^{(s)} - \sigma_T^{(s)} \sqrt{u_{T+1}^{(s)}}, \mu_T^{(s)} + \sigma_T^{(s)} \sqrt{u_{T+1}^{(s)}}) + \sum_{j=1}^{K} \beta_j^{(s)} Z_{jT}
\]  

(C1)

where \( \mu_T \) and \( \sigma_T \) are defined as

\[
\mu_T^{(s)} = \beta^{(s)} Z_T
\]  

(C2)

\[
\sigma_T^{(s)} = e^{W_T \theta^{(s)}}
\]  

(C3)

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

\[
u_{T+1}^{(s)} \begin{cases} 
\sim f(u^{(s)}) & \text{with probability } \propto c^{(s)} \\
= u_j^{(s)} & \text{with probability } \propto 1 & j = 1, \ldots, T
\end{cases}
\]
Here $f(u)$ is $Ga(1.5, 0.5)$. A $r_{T+1}^{(s)}$ 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


Cao, Henry H., Tan Wang, and Harold H. Zhang, 2002, Model uncertainty, limited market participation and asset prices, unpublished working paper.


Guidolin, Massimo, and Allan Timmermann, 2003, Strategic asset allocation under multivariate regime switching, unpublished working paper.


Figure 1. Predictive distributions of market excess returns. This figure depicts the predictive distributions of market excess returns one-period ahead (January 2001) 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 (1991:1-2000:12), while the bottom graph considers a cumulative estimation window (1954:1-2000:12). 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. This figure depicts a time-series of optimal allocations to stocks for investors that account either for (a) parameter uncertainty (dotted line) or (b) distribution uncertainty (solid line). Investors maximize a power utility over terminal wealth of

\[ U(W) = \frac{W^{1-A}}{1-A}, \]

where \( A \) denotes 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. The levels of Index of Distribution Uncertainty for 1964-2001. This figure depicts a time series of levels of Index of Distribution Uncertainty for 1964-2001. The Index has been defined as a difference between optimal portfolio weights for investors who only account for parameter uncertainty and optimal portfolio weights of investors who additionally account for distribution uncertainty. Investors maximize a power utility over terminal wealth of $U(W) = \frac{W^{1-A} - 1}{1-A}$, where $A$ denotes the coefficient of relative risk aversion. The top panel presents the results for investors with $A=3$, while the bottom panel assumes the value of $A=5$. Both graphs assume a 10-year rolling estimation window.
Figure 4. Wealth cumulation of strategies. This figure depicts a time series of wealth derived from $1 invested in December 1963 and cumulated until 2001 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 investors that use a 10-year rolling estimation window while the bottom panel assumes a cumulative estimation window. Investors maximize a power utility over terminal wealth of \( U(W) = \frac{W^{1-A} - 1}{1-A} \), where \( A \) denotes 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 \( A = 3 \).
This table reports the summary statistics of the data for the period of 1954:1-2000:12. 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>4.60</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>14.51</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.37</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.92</td>
</tr>
<tr>
<td>Min</td>
<td>-268.00</td>
</tr>
<tr>
<td>Max</td>
<td>190.27</td>
</tr>
</tbody>
</table>
Table II
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 2001. Investors account either for: (a) parameter uncertainty or (b) distribution uncertainty. They maximize a power utility over terminal wealth of $U(W) = \frac{W^{1-A} - 1}{1-A}$, where $A$ denotes the coefficient of relative risk aversion. $A$ is assigned four possible values: 3, 5, 10 and 20. The respective horizons used for estimation are 1991:1-2000:12 and 1954:1-2000:12.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>295.38</td>
<td>88.36</td>
<td>152.11</td>
<td>75.03</td>
</tr>
<tr>
<td>5</td>
<td>178.40</td>
<td>53.01</td>
<td>91.69</td>
<td>45.04</td>
</tr>
<tr>
<td>10</td>
<td>89.37</td>
<td>26.50</td>
<td>45.95</td>
<td>22.52</td>
</tr>
<tr>
<td>20</td>
<td>44.69</td>
<td>13.25</td>
<td>22.99</td>
<td>11.26</td>
</tr>
</tbody>
</table>
Table III
Parameters of time-series allocations under parameter and distribution uncertainty

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 2001. Investors account either for parameter or for distribution uncertainty. They maximize a power utility over terminal wealth of $U(W) = \frac{W^{1-A}}{1-A}$, where $A$ denotes 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). The sample spans the period of 1954:1-2000:12.

<table>
<thead>
<tr>
<th>A</th>
<th>Parameter uncertainty</th>
<th>Distribution uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rolling</td>
<td>Cumulative</td>
</tr>
<tr>
<td>3</td>
<td>82.46</td>
<td>98.47</td>
</tr>
<tr>
<td></td>
<td>(105.04)</td>
<td>(56.61)</td>
</tr>
<tr>
<td>5</td>
<td>49.59</td>
<td>59.11</td>
</tr>
<tr>
<td></td>
<td>(63.43)</td>
<td>(34.14)</td>
</tr>
<tr>
<td>10</td>
<td>24.82</td>
<td>29.56</td>
</tr>
<tr>
<td></td>
<td>(31.77)</td>
<td>(17.06)</td>
</tr>
<tr>
<td>20</td>
<td>12.41</td>
<td>14.78</td>
</tr>
<tr>
<td></td>
<td>(15.89)</td>
<td>(8.53)</td>
</tr>
</tbody>
</table>
Table IV
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 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 calculated using equation (1) and assuming that predictive returns are obtained under distribution uncertainty. Investors maximize a power utility over terminal wealth of \( U(W) = \frac{W^{1-A} - 1}{1-A} \), where A denotes 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 1954:1-2000:12.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Rolling window</th>
<th>Cumulative window</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.40</td>
<td>0.36</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2.54</td>
<td>0.57</td>
</tr>
<tr>
<td>Min</td>
<td>0.0002</td>
<td>0.004</td>
</tr>
<tr>
<td>Max</td>
<td>8.46</td>
<td>2.81</td>
</tr>
</tbody>
</table>
Table V  
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 2001. Mean and standard deviation are expressed in percentage terms. Investors account either for parameter or for distribution uncertainty. 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 1954:1-2000:12. Standard deviations of the estimates have been provided in parentheses.

<table>
<thead>
<tr>
<th>Estimation Window</th>
<th>Parameter uncertainty</th>
<th>Distribution uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. dev.</td>
</tr>
<tr>
<td>Rolling</td>
<td>0.284</td>
<td>4.204</td>
</tr>
<tr>
<td></td>
<td>(0.484)</td>
<td>(0.522)</td>
</tr>
<tr>
<td>Cumulative</td>
<td>0.370</td>
<td>4.001</td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td>(0.243)</td>
</tr>
</tbody>
</table>
Table VI
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 1954:1-2000:12. Investors maximize a power utility over terminal wealth of $U(W) = \frac{W^{1-A}}{1-A}$, where $A$ denotes the coefficient of relative risk aversion, set to $A = 3$.

<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>4.50</td>
<td>1.37</td>
<td>7.70</td>
<td>2.44</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>21.18</td>
<td>15.08</td>
<td>13.69</td>
<td>12.27</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>21.26</td>
<td>9.11</td>
<td>56.24</td>
<td>19.85</td>
</tr>
<tr>
<td>CAPM Alpha</td>
<td>0.20</td>
<td>-0.08</td>
<td>0.56***</td>
<td>0.06</td>
</tr>
<tr>
<td>DU-PU</td>
<td>-</td>
<td>-</td>
<td>3.20**</td>
<td>1.07</td>
</tr>
<tr>
<td>Uncertainty - Households</td>
<td>1.35</td>
<td>-1.77</td>
<td>4.55**</td>
<td>-0.71</td>
</tr>
<tr>
<td>Uncertainty - Institutions</td>
<td>0.95</td>
<td>-2.27</td>
<td>4.05**</td>
<td>-1.21</td>
</tr>
<tr>
<td>Alpha(DU) - Alpha(PU)</td>
<td>-</td>
<td>-</td>
<td>0.36***</td>
<td>0.14***</td>
</tr>
</tbody>
</table>

*** - 1% significance; ** - 5% significance; * - 10% significance
Table VII
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 based on the estimation period of 1954:1-2000:12. 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 under assumption that investors maximize a power utility over terminal wealth of \( U(W) = \frac{W^{1-A} - 1}{1-A} \), where \( A \) denotes 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>-0.34</td>
<td>0.08</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>(8.69)</td>
<td>(5.64)</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-0.59</td>
<td>0.20</td>
</tr>
</tbody>
</table>

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