The Economic Value of Distributional Timing

Eric Jondeau\textsuperscript{a} and Michael Rockinger\textsuperscript{b}

February 2008

Abstract

It is well known that there is a significant value for investors to allocate their wealth using return and volatility forecasts, phenomena usually called market and volatility timing. In this paper, we show that distribution timing, i.e. the ability to use distribution forecasts for asset allocation, also yields a significant economic value. Considering the weekly asset allocation among the three largest international stock markets, distributional timing yields around 140 basis points per year, to be compared with an economic value of volatility timing of around 55 basis points. To control for parameter uncertainty of the model, we cast the model into a Bayesian setting, also considering alternative samples, datasets, frequencies, as well as preference structures. In all cases, the value of distributional timing remains highly significant. Investors with a preference for skewness rationally put more weight on positively skewed assets and under-diversify their portfolio relative to mean-variance investors. Hence they will benefit from distributional timing, if they are able to predict the subsequent returns' distribution.

Keywords: Portfolio allocation, distributional timing, volatility timing, non-normality, GARCH model, parameter uncertainty, Bayesian estimation.

JEL classification: G11, F37, C22, C51.

\textsuperscript{a}Swiss Finance Institute and University of Lausanne, Faculty of Business and Economics, CH 1015 Lausanne, Switzerland. E-mail: eric.jondeau@unil.ch.

\textsuperscript{b}Corresponding author. Swiss Finance Institute, University of Lausanne and CEPR. University of Lausanne, Faculty of Business and Economics, CH 1015 Lausanne, Switzerland. E-mail: michael.rockinger@unil.ch.

Both authors acknowledge help from the Swiss National Science Foundation through NCCR FINRISK (Financial Valuation and Risk Management). We benefited from comments of and discussions with Tim Bollerslev, Peter Bossaerts, Ales Cerny, Darrell Duffie, Bernard Dumas, René Garcia, Steward Hodges, Eric Jacquier, Nour Meddahi, and Raman Uppal, and the participants at the European Finance Association meeting in Ljubljana (2007). The usual disclaimer applies.
1 Introduction

The interest in the higher moments of the asset returns distribution (beyond mean and variance) for asset pricing and portfolio allocation has been essentially driven by two well-established features. First, many financial assets display non-normal returns, in particular in the short run. Second, a rational investor, if asked to chose between two assets with the same mean and variance, will probably invest in the asset with the highest skewness and the lowest kurtosis (Scott, and Horvath, 1980, Dittmar, 2002).

From a portfolio allocation perspective, the standard mean-variance criterion is questionable under non-normal returns or when the investor’s utility depends on the higher moments of the return’s distribution. Evaluating the effect of higher moments on the portfolio allocation involves two main tasks. The first one consists in designing an appropriate allocation criterion that accounts for the shape of the return’s distribution. Some authors have proposed portfolio criteria, based on an extension of the mean-variance criterion. Examples are the higher-dimensional efficient frontier (Athayde and Flöres, 2004) or the allocation based on various downside risk measures (Ang, Chen, and Xing, 2006). Others have proposed alternative utility functions based on prospect theory (Barberis and Huang, 2005), on ambiguity aversion (Ait-Sahalia and Brandt, 2001), or on an approximation of a general utility function based on higher moments (Harvey et al., 2004, Guidolin and Timmermann, 2007). The second task consists in designing the data generating process driving asset returns. Papers in this field have explored models with systematic jumps (Das and Uppal, 2004), with regime shifts (Ang and Bekaert, 2002), or with non-normal conditional distributions (Patton, 2004). Das and Uppal (2004) show that, in presence of (unexpected) jumps occurring at the same time across countries, the loss from the reduction in diversification is not substantial and that the cost of ignoring common jumps is large only for highly levered positions. In a regime-switching framework, hence where changes in moments may persist, Ang and Bekaert (2002) demonstrate the economic relevance of a dynamic strategy as long as a risk-free asset is incorporated in the allocation.1 Patton (2004) demonstrates that a correct evaluation of the skewness and asymmetric dependence across assets can lead to significant portfolio gains.

Our first contribution in this paper is to establish the economic value of distribution timing, i.e. the investor’s ability to forecast the subsequent distribution of asset returns and to invest accordingly. This concept echoes the market timing and volatility timing already explored in the literature. While market timing,

1Although this approach is appropriate for investigating the consequences of time-varying first and second moments, it is less convenient for addressing the consequences of time-varying higher moments. The reason is that the return’s distribution is conditionally normal, so that the dynamics of the return’s higher moments are implied by the regime shifts in the first two moments and therefore they cannot be modeled separately.
involving the expected return predictability, has been extensively studied (Kandel
and many others), the focus on volatility timing is relatively recent. Graham and
Harvey (1996) and Busse (1999) have shown that investors actually design strate-
gies exploiting predictability in volatility. More recently, several papers have shown
that volatility timing has a significant economic value for daily to monthly horizons,
e.g., Fleming, Kirby, and Ostdiek (2001, 2003), Marquering and Verbeek (2001),
Johannes, Polson, and Stroud (2002). These authors construct strategies based on
volatility forecasts and show that such strategies are valuable.\footnote{Fleming, Kirby,
and Ostdiek (2001) show that volatility timing strategies outperform strategies
based on constant expected returns and volatilities. The relevance of volatility timing
with respect to ‘expected return’ timing is discussed by Johannes, Polson, and Stroud
(2002) who argue that strategies based on time-varying expected returns perform
rather poorly, due to estimation risk, while volatility timing strategies do not suffer
from this problem.}

In this paper, we systematically study the incremental value of taking skewness or
kurtosis into account and compare the magnitude of distributional timing relative to
that of volatility timing. For this purpose, we consider a dynamic mean-variance strat-
yegy where investors try to benefit from their ability to predict subsequent volatility.
Then, in a dynamic higher-moment strategy, investors try to benefit from their ability
to predict not only volatility, but also the distribution of returns.

As a second contribution, we extend the dynamic conditional correlations (DCC)
model of Engle (2002) and Engle and Sheppard (2001) to the case of a joint dis-
tribution that allows both asymmetry and fat tails. This framework incorporates
most of the statistical features that are known to characterize the dynamics of asset
returns. Hence it provides a good fit of the various asset returns we consider in this
paper. We then focus on the characteristics of a portfolio composed of these assets.
We derive closed-form solutions for the moments of the distribution of the portfolio.
These moments can be directly used as inputs to a fourth-order approximation of
the expected utility. Within this framework, we show that time variability in higher
moments matter for portfolio allocation, therefore confirming the findings of Patton

Our third contribution is that we perform the estimation of our model in a
Bayesian setting. There are several motivations to this choice. First, it is a natu-
ral way to handle nonlinear models with a large number of parameter constraints.
Second, it directly takes parameter uncertainty into account.\footnote{Given the complex-
ity of a model to forecast not only first and second moments but also higher
moments, including co-moments, one may expect more uncertainty in the estimated
parameters. Fleming, Kirby, and Ostdiek (2001) use a bootstrap approach to evaluate
the effect of uncertainty about expected returns on the economic value of volatility
timing. In our context, the Bayesian approach is not only an estimation technique,
but also as a simulation tool to evaluate the effect of parameter uncertainty.} Third, we can take
advantage of the large number of parameter draws to test economic hypotheses. In particular, we measure the economic value of distributional timing and test its statistical significance. Last, the Bayesian setting provides a very efficient way to investigate the consequences of investor’s aversion to model uncertainty on the optimal asset allocation.\(^4\)

We apply our approach to the weekly allocation of wealth between the three largest stock markets, i.e. the US, Japanese, and UK markets. We show that our model captures the main statistical characteristics of these market returns. In addition, we find that the mean-variance criterion results in an excessive risk taking and a significant opportunity cost, as compared to a strategy based on higher moments. The performance fee an investor would be willing to pay to benefit from the higher-moment dynamic strategy (distributional timing) is much higher than the fee she would be willing to pay to benefit from the mean-variance dynamic strategy (volatility timing). For intermediate levels of risk aversion, the economic value of distributional timing is around 140 basis points per year, to be compared with an economic value of volatility timing of around 55 basis points per year (the expected return is about 5% over the allocation period). We investigate alternative sample periods, preferences, datasets, and frequencies and show that the economic value of distribution timing remains sizeable and comparable to volatility timing.

The outline of the paper is as follows. In Section 2, we formulate our approach for modeling returns with a non-normal multivariate distribution and for measuring distributional timing. In Section 3, we present the data and discuss the results of the estimation of the model. In Section 4, we consider the case of time-varying investment opportunities and comment the main characteristics of the optimal portfolios. In Section 5, we measure the economic value of the dynamic strategies, under alternative preference structures. In Section 6, we provide several robustness checks of our main results. The last section concludes the paper. Several appendices contain the details of the statistical model describing the evolution of returns. A Technical Appendix is available upon request from the authors. It provides further details on the estimation techniques used in the paper as well as additional empirical results.

\(^4\)Several authors have investigated how model uncertainty may be incorporated in the portfolio optimization problem. See Barberis (2000), Polson and Tew (2000), Johannes, Polson, and Stroud (2002), Harvey et al. (2004). In most papers, estimation risk is addressed in a Bayesian framework, which provides a direct way to maximize the expected utility given the distribution of parameters. Wang (2005) and Garlappi, Uppal, and Wang (2007) adopt alternative approaches to incorporate estimation risk in the optimization procedure. In Section 5, we discuss the case of investors with aversion to model uncertainty.
2 Methodology

This section describes the conditional asset allocation problem with non-normal returns. In this context, the standard mean-variance criterion is in general inappropriate to select the optimal portfolio. Incorporating higher moments forecasts in the expected utility of investors is likely to improve the allocation of wealth (Harvey and Siddique, 2000, and Dittmar, 2002). As in Guidolin and Timmermann (2007), we approximate the expected utility up to the fourth moment to obtain the optimal asset allocation. This gives rise to what we call distributional timing. Restrictions on the most general econometric model, shutting down temporal variation in skewness or kurtosis, allows us to gauge the relative importance of the various components.

2.1 The Multivariate Return Process

Given our interest in the effect of higher moments on allocation performances, we build a model that provides a complete description of the multivariate return process. For convenience, we split the returns’ dynamics into various components:

\[ r_t = \mu + \varepsilon_t, \]
\[ \varepsilon_t = \Sigma_t^{1/2} z_t, \]
\[ z_t \sim g(z_t|\eta_t). \]

Equation (1) decomposes the return at time \( t \), \( r_t \), into two \( n \times 1 \) vectors, i.e. the mean, \( \mu \), and the unexpected return, \( \varepsilon_t \). Equation (2) indicates that the unexpected return \( \varepsilon_t \) combines the independent innovation \( z_t \) and the conditional covariance matrix of returns, \( \Sigma_t = E_{t-1}[(r_t - \mu)(r_t - \mu)^\prime] \). The vector of independent innovations, \( z_t \), has zero mean and identity covariance matrix. We denote by \( \Sigma_t^{1/2} \) the Choleski decomposition of \( \Sigma_t \). Last, equation (3) specifies that innovations, \( z_t \), follow a conditional distribution \( g \) with (possibly time-varying) shape parameters, \( \eta_t \).

When the conditional distribution is normal, there is no shape parameter since the normal distribution is entirely characterized by its mean and variance. In more general cases, shape parameters typically involve parameters capturing asymmetry and fat-tailedness of the distribution.

In Appendix 1, we detail the general model we use as a data generating process. The model accounts for the well-known properties of volatility clustering and dynamic conditional correlations (Engle, 2002, and Engle and Sheppard, 2001). We extend this model to the multivariate Sk-\( t \) distribution introduced by Sahu, Dey, and Branco (2003). Thereby, we allow for both asymmetry and fat-tailedness, often found in actual data. An interesting feature of this distribution is that it nests the normal and \( t \) distributions and that the associated parameters have a rather natural interpretation. This distribution appears to fit the data very well. Most

\(^5\text{For a univariate setting, see Hansen (1994) and Jondeau and Rockinger (2003).}\)
importantly, it is possible to derive analytically the moments of a portfolio where
the various component returns follow this distribution. This insight, combined with
a Taylor approximation of the expected utility, allows us to perform allocation in
very efficient manner.\footnote{Other work, e.g. Patton (2004), obtains the expected utility via Monte Carlo integration of a
bivariate distribution. This setting would be very time consuming in a multivariate context. Our
model allows estimation and portfolio allocations for many assets.}

\section{2.2 The Criterion for Asset Allocation}

We consider an investor who allocates her portfolio by maximizing the expected
utility \( E_t[U(W_{t+1})] \) over the end-of-period wealth \( W_{t+1} \).\footnote{We do not consider a multi-period investment problem. The reason is that the available
approaches (Monte Carlo simulation or dynamic programming) are too time consuming. As his
work demonstrates, taking parameter uncertainty, learning, and dynamic allocations into account
with dynamic programming techniques is already difficult in a setting with only one risky asset.
We wish to leave this extension to a multi-period investment for further research.} The initial wealth \( W_t \) is
arbitrarily set equal to one and \( E_t \) denotes the expectations operator where all
information up to time \( t \) is used. There are \( n \) risky assets with return vector \( r_{t+1} = (r_{1,t+1}, \ldots, r_{n,t+1})' \) and a risk-free asset with return \( r_{f,t} \) for the period between \( t \) and \( t+1 \). Return \( r_{i,t+1} \) is defined as the simple rate of return of asset \( i \) from time \( t \) to time \( t+1 \). End-of-period wealth is \( W_{t+1} = 1 + r_{p,t+1} \), where \( r_{p,t+1} = r_{f,t} + \alpha_t' (r_{t+1} - r_{f,t}) \) denotes the portfolio return and \( \alpha_t = (\alpha_{1,t}, \ldots, \alpha_{n,t})' \) the vector of weights allocated
to the various risky assets at time \( t \). Short sales are allowed and the weight of the
risk-free asset, \( \alpha_{0,t} = 1 - \sum_{i=1}^{n} \alpha_{i,t} \), can be negative (borrowing) or positive (lending).
We assume also that the investor has forecasts for the expected mean vector \( \mu_{t+1} \),
the covariance matrix \( \Sigma_{t+1} \), and possibly the higher-order co-moment matrices.

Formally, optimal portfolio weights are obtained by maximizing the expected
utility

\[
\max_{\{\alpha_t\}} E_t[U(W_{t+1}(\alpha_t))] = E_t[U(1 + r_{f,t} + \alpha_t'(r_{t+1} - r_{f,t}e))] \tag{4}
\]

\[
\text{s.t. } \sum_{i=0}^{n} \alpha_{i,t} = 1,
\]

where \( e = (1, \ldots, 1)' \) denotes the \( n \times 1 \) vector of ones.\footnote{This framework could be extended to allow for portfolio adjustments only if changes in the
optimal weights exceed a certain threshold.} In general, equation (4) does
not have a closed-form solution and numerical techniques must be used. Quadrature
rules have been used for normal iid returns (Campbell and Viceira, 1999) or regime-
switching conditionally normal returns (Ang and Bekaert, 2002). Non-normal dis-
tributions would require a number of quadrature points that increases exponentially.
with the number of assets, so that solving the optimization problem with numerical integration is often intractable for more than two or three assets.

Since we are primarily interested in measuring the effect of higher moments on the asset allocation, we follow an alternative approach that approximates the expected utility as a function of the moments of the portfolio return. The utility function can be written as an infinite-order Taylor series expansion of the form

\[ U(W_{t+1}) = \sum_{k=0}^{\infty} \frac{U^{(k)}(W_t)}{k!} (W_{t+1} - W_t)^k, \]

where \( W_{t+1} - W_t = r_{f,t} + \alpha_t' (r_{t+1} - r_{f,t}) = r_{p,t+1} \) denotes the portfolio return at date \( t + 1 \), and \( U^{(k)} \) the \( k \)-th derivative of the utility function. Under rather mild conditions, the expected utility is given by

\[ E_t[U(W_{t+1})] = \sum_{k=0}^{\infty} \frac{U^{(k)}(W_t)}{k!} E_t[(r_{p,t+1})^k], \] (5)

so that it depends on all the moments of the distribution of the portfolio return.\(^9\) The investor’s preference (or aversion) towards the \( k \)-th moment is directly given by the \( k \)-th derivative of the utility function. Scott and Horvath (1980) show that, under certain conditions, the investor has a preference for odd moments and an aversion for even moments.

In the following, we focus on a Taylor series expansion up to the fourth order. The main reason is that our econometric model is designed to capture the dynamics of the third and fourth moments through the modeling of the asymmetry and degree-of-freedom parameters. The dynamics of even higher order moments would not be independent from the dynamics of the third and fourth moments. We verified that, for our data, the allocation obtained from the maximization of the exact expected utility by numerical integration is essentially the same as the one obtained from the maximization of the Taylor series expansion. The computational burden is, however, much heavier.

Focusing on terms up to the fourth order, expected utility (5) rewrites

\[ E_t[U(W_{t+1})] \approx U(W_t) + U^{(1)}(W_t) m_{p,t+1} + \frac{1}{2} U^{(2)}(W_t) m_{p,t+1}^{(2)} \]

\[ + \frac{1}{3!} U^{(3)}(W_t) m_{p,t+1}^{(3)} + \frac{1}{4!} U^{(4)}(W_t) m_{p,t+1}^{(4)}, \] (6)

\(^9\)Such an approach has been adopted in a large number of contributions, see Rubinstein (1973), Kraus and Litzenberger (1976), and Dittmar (2002) among others. Necessary conditions for the infinite Taylor series expansion to converge to the expected utility have been explored by Loistl (1976). The region of convergence of the series depends on the utility function considered. For the power utility function, convergence is guaranteed for wealth levels in the range \([0, 2\bar{W}]\) where \( \bar{W} = E(W_{t+1}) \). Such a range is likely to be large enough for bonds and stocks. In contrast, it may be too small for options, due to their leverage effect. These results hold for arbitrary return distributions.
where \( m_{p,t+1}^{(i)} = E_t[r_{p,t+1}^i] \) is the non-central moments of order \( i \).\(^{10}\) If we consider the power utility function

\[
U(W_{t+1}) = W_{t+1}^{1-\gamma} / (1 - \gamma),
\]

where \( \gamma > 1 \) measures the investor’s constant relative risk aversion, expression (6) becomes

\[
E_t[U(W_{t+1})] \approx \frac{1}{1-\gamma} + m_{p,t+1} - \frac{\gamma}{2} m_{p,t+1}^{(2)} + \frac{\gamma(\gamma+1)}{3!} m_{p,t+1}^{(3)} - \frac{\gamma(\gamma+1)(\gamma+2)}{4!} m_{p,t+1}^{(4)} .
\]

Consistently with the theoretical arguments developed by Scott and Horvath (1980), the effect of the third and fourth moments on the approximated expected utility is unambiguous. Expected utility decreases with large negative skewness (i.e. left-skewed distributions) and large kurtosis (i.e. fat-tailed distributions).

Maximizing expression (7) for each date \( t \) defines a dynamically rebalanced portfolio that maximizes the expected utility of the investor. It clearly shows how forecasts of the higher moments of the portfolio return distribution will affect the optimal weights at date \( t + 1 \). Presently, we are endowed with a model for asset returns and a criterion for asset allocation. We now discuss how to evaluate the economic relevance of distributional timing.

2.3 Allocation Strategies

The models we consider have constant expected returns, as in Fleming, Kirby, and Ostdiek (2001, 2003). The reason for this choice is that there is little information content in past variables regarding future returns at short horizons. For such horizons, estimation uncertainty of predictive regressions may also do more harm than good. Some recent papers have shown that macroeconomic variables have a predictive power for monthly returns (Kandel and Stambaugh, 1996, Brandt, 1999, Marquering and Verbeek, 2001), although this is still a debated topic (Cremers, 2002). To our knowledge, the literature on return’s predictability at short frequency is scarce. Since our focus is on the evaluation of volatility and distributional timing, we leave the investigation of the return’s predictability at short horizons for further research.

Our benchmark strategy is the static strategy. We consider a mean-variance investor who estimates the expected returns and the covariance matrix using sample

\(^{10}\)Dittmar (2002), Guidolin and Timmermann (2007), and Jondeau and Rockinger (2006) also maximize this four-moment expected utility. In Appendix 2, we explain how to obtain the portfolio return’s moments from the asset returns’ ones and give the relation that exists between non-central moments and the usual moments of the distribution.
moments over the estimation period and then keeps these parameter estimates constant over the allocation period. Then optimal portfolio weights vary only when the risk-free rate varies.\textsuperscript{11}

In the second strategy, the investor still adopts a mean-variance criterion, but forecasts the time-varying conditional covariance matrix. Thus a DCC model is estimated under the assumption of a joint normal distribution. This dynamic mean-variance strategy is denoted by $MV^d$. Its performance relative to the static strategy provides a measure of the economic value of volatility timing.

In the last strategy, the investor forecasts the covariance matrix as well as the conditional distribution of asset returns. This is achieved by our full model, where innovations are distributed as a Sk-$t$ distribution with time-varying shape parameters. With this strategy, the investor takes full advantage of volatility and distributional timing. This dynamic higher-moment strategy is denoted $HM^d$.

\subsection*{2.4 Measuring Distributional Timing}

We compare the performance of the various strategies by using several measures. A first measure of performance is the standard Sharpe ratio, which is computed using the ex-post average return $m_p$ and volatility $\sigma_p$, as $SR_p = (m_p - r_f) / \sigma_p$. Since the Sharpe ratio does not provide a measure of out-performance over alternative strategies, we also consider the modified Sharpe ratio $mSR$ introduced by Graham and Harvey (1997)

\begin{equation}
\frac{m_0}{\sigma_0} \left( \frac{m_p - r_f}{\sigma_p} \right) - \left( \frac{m_0 - r_f}{\sigma_0} \right),
\end{equation}

where $m_0$ and $\sigma_0$ are the average return and volatility of the static strategy. This measure corresponds to a scaled difference of the prices of risk of the two allocations under comparison.

These measures have however an obvious drawback in our context, since they do not capture the effect of non-normality. We therefore consider another tool to evaluate the economic value of volatility and distributional timing, namely the performance fee measure proposed by West, Edison, and Cho (1993) and Fleming, Kirby, and Ostdiek (2001). It measures the management fee an investor is willing to pay to switch from the static strategy to a given dynamic strategy. The performance fee (or opportunity cost), denoted $\vartheta$, is defined as the average return that has to be subtracted from the return of the dynamic strategy so that the investor becomes indifferent between both strategies

\begin{equation}
E_t \left[ U (1 + \hat{r}_{p,t+1}) \right] = E_t \left[ U \left( 1 + r^*_{p,t+1} - \vartheta \right) \right],
\end{equation}

\textsuperscript{11}Given the relative performance of the various assets over time, there are also some transaction costs, when the investor has to rebalance her portfolio to maintain constant weights. It turns out that this rebalancing effect is negligible.
where $r_{p,t+1}^*$ is the optimal portfolio return obtained under the dynamic strategy, and $\hat{r}_{p,t+1}$ the optimal portfolio return obtained under the static strategy. This measure is obtained numerically by solving equation (9).

We also compute the success rate, i.e. the percentage of times the return of the dynamic strategy exceeded the static strategy. This measure, denoted $Z$, is useful to check whether the out-performance of the dynamic strategy is due to some very specific events or to a better ability to capture changes in investment opportunities. For instance, if $Z$ is very small, it suggests that the out-performance is essentially due to a few happy instances.

Finally, since we compare static and dynamic strategies, we have to control for the possible effect of transaction costs. Indeed, since the static strategy generates very low turnover, the gain of a dynamic strategy may be partly offset by transaction costs. In practice, it is difficult to estimate the actual transaction costs, since a wide range applies depending on the asset and the type of customer relation. For this reason, we follow the approach of Han (2006) and compute the break-even transaction cost, denoted $\tau_{be}$. It measures the level of transaction costs that makes the investor indifferent between the dynamic and the static strategies. If transaction costs are equal to a fixed fraction $\tau$ of the value traded in all stocks in the portfolio, $tv$, the average weekly transaction cost of this strategy is $\tau \times tv$, where

$$tv = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \left| \alpha_{i,t} - \frac{\alpha_{i,t-1} (1 + r_{i,t})}{1 + r_{p,t}} \right|.$$

Finally, the break-even transaction cost between the dynamic strategy $d$ and the static strategy $s$ is defined as

$$\tau_{be} = \frac{\bar{r}_d - \bar{r}_s}{tv_d - tv_s}. \quad (10)$$

If the actual transaction cost is lower than $\tau_{be}$, the investor will prefer the dynamic strategy to the static strategy.\textsuperscript{13}

\textsuperscript{12}We also considered the certainty equivalent, previously adopted among others by Kandel and Stambaugh (1996), Campbell and Viceira (1999), Ang and Bekaert (2002), and Das and Uppal (2004). It is defined as the compensation (in percentage of initial wealth) an investor must receive so that she is willing to put one dollar in the sub-optimal strategy rather than in the optimal one. Since the performance fee and the certainty equivalent provided the same measure of the economic gain (up to a few basis points), we only report the former in our empirical evidence.

\textsuperscript{13}Actual transaction costs are difficult to estimate. Marquering and Verbeek (2001) consider 1% as high. Balduzzi and Lynch (1999) consider that 0.5% is a reasonable transaction cost for direct trades in stocks, whereas these costs may be as low as 0.01% for an institutional investor trading in futures. See also the long discussion in Han (2006).
3 Data and Preliminary Analysis

3.1 Data Description

We investigate several datasets to evaluate the economic value of distributional timing. The objective is to show that the gain is not due to the selection of a particular dataset but reflects a general result instead. In this section, we concentrate on the first dataset, based on returns of the three largest international markets, the United States, Japan, and the United Kingdom. The asset allocation problem is viewed from the perspective of an unhedged US investor, so that returns are expressed in US dollars. The risk-free rate is the 7-day US Treasury bill rate. The data are weekly and cover the period from January 1977 through December 2005, for a total of 1510 observations. To avoid in-sample overfitting as well as spurious findings, this sample period is broken in two subsamples: the first sample (from 1977 to 2001, 1304 observations) is used for the estimation of the model, while the second sample (from 2002 to 2005, 206 observations) is used for the out-of-sample investigation. This dataset has been selected as the benchmark to establish our results, because it covers large indices over large developed markets. As a consequence, it is less likely to be characterized by extreme behaviors that may drive the results. In Section 6, we report evidence based on alternative subsamples, datasets, or frequencies. It turns out that our main results are not significantly altered.

Table 1 reports several summary statistics for the market returns under investigation (Panel A). Average returns are all positive and significant, ranging between 0.18% and 0.25% per week. Volatilities range between 2.1% and 3% per week. The US market yields a moderate return, with a low volatility, while the Japanese market is characterized by a low return and a high volatility. Skewness measures are dispersed across markets. US returns are negatively skewed, suggesting that crashes occur more often than booms, while the Japanese market has a large positive skewness. Kurtosis are between 4.1 and 5.2, a range that is not consistent

14The data consists in Friday-to-Friday weekly returns, based on closing prices, from Datastream International. At the end of 2005, the US, Japanese, and UK markets represent respectively 41.3%, 18.4%, and 7.5% of the world market capitalization. Market returns are measured by the return on the indices S&P 500, Nikkei 225, and FTSE 100 respectively. Japanese and UK returns are converted into US dollars using the exchange rate on the same day. Non synchronicity of the markets is expected to be softened by the use of the weekly frequency. These market indices are very easy to trade, since they all have tradable futures. The transaction costs on these futures are very low, suggesting that transaction costs will not be a key issue.

15The positive skewness of the Japanese market return suggests that the multivariate model with jumps occurring at the same time across countries, as proposed by Das and Uppal (2004), is probably at odds with our data. In their dataset on developed countries, the skewness of the Japanese market is close to 0.
with the normality assumption. We reject normality with great confidence for all markets. Regarding temporal dependence, we find no systematic evidence for serial correlation in market returns, but squared returns are strongly correlated.

Turning to the multivariate characteristics of market returns, we notice that the correlation is the largest between the US and the UK markets (0.397), while the smallest correlation is between the US and Japan (0.219). Given the well-known time variability of correlations, these sample correlations may be misleading for allocation purposes. While the sample correlation between the US and the UK is 0.397 over the estimation period, it is as high as 0.68 over the (out-of-sample) allocation period. Hence, the static strategy is likely to overstate the diversification ability of the UK market.

Finally, the table also reports co-skewness and co-kurtosis defined as the empirical analogues to

$$sk_{ijk} = \frac{E[(r_{i,t} - m_i)(r_{j,t} - m_j)(r_{k,t} - m_k)]}{\sigma_i\sigma_j\sigma_k}, \quad (11)$$

$$ku_{ijkl} = \frac{E[(r_{i,t} - m_i)(r_{j,t} - m_j)(r_{k,t} - m_k)(r_{l,t} - m_l)]}{\sigma_i\sigma_j\sigma_k\sigma_l}, \quad (12)$$

where $m_i$ and $\sigma_i$ denote the sample mean and standard deviation of return $i$ respectively (Panel B). Expressions $sk_{iii}$ and $ku_{iii}$ are the standard measures of individual skewness and kurtosis (reported in Panel A). On the one hand, the table shows that the co-skewness cannot be distinguished from 0. On the other hand, co-kurtosis are highly significant. As for the standard individual kurtosis, it is customary to compare co-kurtosis with the level that a multivariate normal distribution would imply. As the table indicates, most co-kurtosis are significantly larger than the level expected from normal returns. These results suggest that the US and UK markets provide bad hedges against adverse volatility in the other market. A similar result holds for the US and Japanese markets.

### 3.2 Estimation of the Model

Measuring the economic value of distributional timing requires a rather sophisticated model since the temporal evolution of the conditional distribution has to be specified. Given the complexity of the model, it is of great importance to show that the reported significance of the economic value is not due to chance and is robust to parameter uncertainty. A natural concern with such a general data generating process is estimation risk, when it is used for asset allocation purpose. This issue has been already addressed among others by Kandel and Stambaugh (1996), Barberis (2000), Johannes, Polson, and Stroud (2002), and Harvey et al. (2004). A natural way to accommodate parameter uncertainty is Bayesian estimation. This technique provides a very efficient way of estimating our general model and it also allows us to evaluate the statistical significance of the portfolio weights and performance measures reported in the paper. One additional benefit of Bayesian estimation is that it produces the empirical distribution of the parameters. As a by-product, this
empirical distribution provides a solution to the allocation problem of an investor with aversion to model uncertainty.

In the Technical Appendix, we describe the Bayesian estimation procedure used in this paper and comment the parameter estimates. In this section, we discuss some results on higher moments that are not so well documented in the literature.

Figures 1 to 4 display the dynamics of volatilities $\sigma_{i,t}$, correlations $\rho_{ij,t}$ and individual conditional higher moments $sk_{iii,t}$ and $ku_{iiii,t}$ respectively. Inspection of these figures reveals several interesting features from a portfolio perspective. First, over the allocation period, the volatility of the UK market is relatively low. In particular, it is below the US volatility, while sample estimates were ranking the US as the safest market. Second, at the end of the period, correlations across markets are higher than the sample estimates. As already mentioned, the correlation between the US and UK markets is around 0.7, while the sample correlation is only 0.4. Third, the conditional skewness estimated for the three markets are in the range of the sample measures of asymmetry for the US and Japan. However, the UK market return turns out to be negatively skewed once temporal variation of skewness is allowed. Finally, the most striking result is the variability in the conditional kurtosis: for the US and the UK, the conditional estimates are far below the sample estimates (around 4 and 3.5 for the US and the UK, while the sample measures in Table 1 are 5.25 and 4.1 respectively). On the opposite, the conditional kurtosis of the Japanese market is found to be much higher (around 7) than the sample estimate. Once the temporal evolution of the higher moments is taken into consideration in the portfolio decision problem, the investor is likely to put more weights on the UK market at the expense of the Japanese market.

These results suggest that allocating wealth on the basis of the sample moments only is likely to be misleading and that the temporal variability of moments, including higher moments, may play an important role.

4 Portfolio Analysis

We begin our analysis of the allocation results with some characteristics of the optimal portfolio weights implied by the various strategies under study. For each week of the sample, we forecast the first four moments and co-moments of market returns using the model described above and maximize the approximated expected utility (7) to produce portfolio weights. Summary statistics on portfolio weights are reported in Table 2. For a risk aversion of $\gamma = 5$, the ex-ante optimal static portfolio is mainly composed of the US (56%) and UK (47%) markets, while the weight of the Japanese market is 12% only (15% of the wealth is borrowed at the risk-free rate). Although they do not fully reflect the market capitalization, these
optimal weights are consistent with the means and variances reported in Table 1.\footnote{As already mentioned, the weights of the static strategy are not necessarily held constant, as the risk-free rate varies over time.}

When the covariance matrix is allowed to vary over time (\( MV^d \) strategy), the weight of the UK market increases significantly (to 56\% on average). As argued before, this change is mainly due to the low volatility of the UK market over the allocation period. It is noteworthy that since the allocation period is out of sample, the sample volatility can be a poor estimate of the volatility that prevails over the allocation period. The optimal weights put on the US and Japanese markets by the \( MV^d \) strategy are much closer to the static allocation.

When the investor takes the temporal evolution of the conditional distribution into account (\( HM^d \) strategy), the new feature she has to consider is the trade-off between the UK return (negative skewness and low kurtosis) and the Japanese return (positive skewness and high kurtosis). Over the allocation period, the kurtosis of the Japanese market reaches very high levels in 2003-04 (above 9). Accordingly, during this period, the investor significantly increases the weight of the UK market in her portfolio (to 69\% on average) while reducing the weight of the Japanese market (below 10\%). The US market also receives a smaller fraction of wealth (44\%).

Considering the average weights obtained for the two dynamic strategies, one may argue that the differences are, all in all, rather small. This observation is actually due to averaging over the allocation period. A picture is needed to reveal the pattern of allocations over time. The evolution of the portfolio weights over the out-of-sample period is displayed in Figures 5 (for \( \gamma = 5 \)). It shows that relatively large differences in the portfolio weights are obtained between the dynamic strategies, in particular over the second half of the sample.

To measure the difference between the allocations, we consider the distance
\[
Dist = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha^*_i,t - \hat{\alpha}_i,t|/2,
\]
where \( \hat{\alpha} \) and \( \alpha^* \) denote the weights of the static strategy and the dynamic strategy respectively.\footnote{The division by 2 corrects for double counting.} Table 2 shows that the average distance between the weights of the static and the \( HM^d \) strategies is 0.15, i.e. the difference between the weights of the two strategies amounts to 15 percentage points on average for the whole portfolio. We also observe that the higher the risk aversion, the smaller the discrepancy between the static strategy and the dynamic strategies. The reason is that the investor is more reluctant to take risk and therefore invests more in the risk-free asset. For instance, the distance between the static and \( HM^d \) strategies decreases from 34 percentage points for \( \gamma = 2 \) to 4 percentage points for \( \gamma = 15 \).

So far, the reported weights represent averages of portfolio weights obtained for a given parameter vector generated by the Markov chain. This setting represents a natural framework within which the investor can measure the accuracy of a given weight. It is also possible to test if there are statistically significant changes in the weights over time.
As a first exercise, we trace kernel-density estimates for the UK weights resulting from the $MV^d$ and the $HM^d$ allocation for January 2002 and January 2005. As Figure 6 displays, the allocations differ across dates. Furthermore, we also notice that there is a strong difference between the allocation involving the distributional timing and the one involving the volatility timing only. For instance, focusing on January 2002 (densities represented with thick lines), the $HM^d$ strategy is anchored at around 0.15, whereas the $MV^d$ strategy is anchored at around 0.55.

In our setting, it is also possible to perform statistical tests on the relevance of certain strategies. For instance, we can test if the strategy consisting in equal weights across markets is supported by the data. We reject the constant equally-weighted portfolio for all the dates in the allocation sample.

5 Performance Analysis

Presently, we analyze the performance of the various dynamic trading strategies described above. Each strategy shall provide some insight on the economic value of the volatility and distributional timing. We therefore compare the out-of-sample performance of the dynamic strategies to that of the static strategy based on the sample mean and covariance matrix over the estimation period. We also compute the performance fee an investor is willing to pay to switch from the sub-optimal static strategy to the optimal dynamic strategy. These computations are reported in Tables 3 and 4. Table 3 displays statistics on realized portfolio returns, whereas Table 4 reports several performance measures for the various strategies and levels of risk aversion $\gamma$. We begin our discussion with the gain that results from switching from the static to the $MV^d$ strategy (volatility timing) and then discuss the gain from switching to the $HM^d$ strategy (distributional timing).

5.1 Economic Value of Volatility Timing

As the tables reveal, all the reported statistics point in favor of the dynamic strategy. First, the $MV^d$ strategy provides a higher realized return than the static strategy: for a risk aversion of $\gamma = 5$, the annualized return is equal to $\mu = 7.1\%$ for the $MV^d$ strategy against 6.9\% for the static strategy. Since realized volatility is also lower ($\sigma = 14.5\%$ against 14.9\%), the Sharpe ratio is higher ($SR = 0.37$ against 0.34). Accordingly, the risk-adjusted excess return of the dynamic strategy, measured by $mSR$, is as high as 191 bp per year.

For $\gamma = 5$, a static investor is willing to pay a performance fee $\vartheta$ of 55 bp per year to switch to the dynamic strategy. This fee includes the increase in the excess return.

---

\[18\] From our Bayesian estimation, one may easily construct a covariance matrix, say $J$, and set up a Wald test of the type $g'J^{-1}g$, where $g$ is the vector of distances of the average portfolio weights to 1/3, the weight required here for an equally-weighted portfolio.
and the decrease in the risk exposure of the portfolio as measured by volatility and the higher moments.\textsuperscript{19}

The $MV^d$ strategy still outperforms the static strategy for other levels of risk aversion. For low levels, the investor takes more extreme positions in order to improve the return on her portfolio. The performance fee is the highest for $\gamma = 2$ (at 136 bp), while it decreases to 19 bp for $\gamma = 15$. The performance fee found for volatility timing is in the same range as the one reported for instance by Fleming, Kirby, and Ostdiek (2001, 2003) and Han (2006).

The evidence presented above has been obtained in the absence of transaction costs. Clearly, the gain of the dynamic strategy would be smaller once such costs are included. The last column of the table reports the break-even transaction cost, $\tau_{be}$, of the $MV^d$ strategy. It is about 2 bp for all levels of risk aversion. These values are in the low part of the range reported by Han (2006), although they are larger than the transaction costs typically charged for a portfolio of futures. In addition, a significant fraction of the distribution of the break-even transaction cost over the simulations is negative. This suggests that a large part of the gain resulting of the $MV^d$ strategy is given back when transaction costs are accounted for.

\section{5.2 Economic Value of Distributional Timing}

We now turn to the extra performance an investor can expect from her ability to forecast changes in the return’s distribution. The economic gain of distributional timing is measured by comparing the performance of the $MV^d$ and $HM^d$ strategies. Table 4 reveals that the performance fee a static investor is willing to pay to switch to the $HM^d$ strategy is equal to 193 bp for $\gamma = 5$. The ability to benefit from distributional timing therefore generates an additional performance fee by about 140 bp (192.5 – 54.6). In the same line, as compared to the $MV^d$ strategy, the Sharpe ratio increases to 0.5 from 0.37, while the $mSR$ measure increases to 234 from 42. The main change in the portfolio allocation that explains the increase in the performance measures is the increase in the UK weight and the decrease in the US weight. Given the over-performance of the UK market in 2004-05 over the US market (about 24% vs. 13%), this resulted in a significant increase of the realized return (from 7.2% to 9.8% per year). This suggests that the $HM^d$ strategy is able to benefit from the positive skewness of the ex-ante portfolio return distribution. Since for some dates some positive events occur, we do not observe such a positive skewness in the ex-post distribution but rather a higher realized return.

A last noteworthy result of the table is that the breakeven transaction cost is much higher for the $HM^d$ strategy than for the $MV^d$ strategy. With a breakeven

\textsuperscript{19}The last component should not be confused with distributional timing. By adopting the $MV^d$ strategy, the investor improves the distributional properties of her portfolio, even if she does not explicitly take care of non-normality issues.
cost in the range 18 to 23 bp, clearly above the range of actual transaction costs, the
gain of implementing the $HM^d$ strategy should be only slightly reduced in practice.

As it appears clearly from the tables, all performance measures of the $HM^d$
strategy are significantly positive and above their counterparts for the $MV^d$ strategy.
For instance, for $\gamma = 5$, the economic value of volatility timing is about 55 bp while
the economic value of distribution timing is about 140 bp about three times higher.
A similar order of magnitude is obtained for the other levels of risk aversion.\textsuperscript{20}

**Statistical significance.** While the static strategy only requires the estimation
of the sample mean vector and the covariance matrix, the dynamic strategies rely on
the estimation of the dynamics of the covariance and higher co-moments matrices.
To avoid any overfitting of the data or data snooping, we used two non-overlapping
subsamples for the estimation and allocation stages.\textsuperscript{21} Another important issue in
the evaluation of economic value of a strategy is estimation risk. Our previous results
suggest that distributional timing has an economically sizeable value. However, this
value may be statistically insignificant if the uncertainty surrounding parameter
estimates is too large. To address this issue, we used a Bayesian estimation to
generate draws from the finite-sample distribution of the parameters and to evaluate
the significance of the performance measures reported in Table 4.

**Figure 7** depicts the empirical distribution of the performance fee for $MV^d$ and
$HM^d$ strategies with respect to the static strategy. As it appears clearly, the two
measures are significantly positive, confirming the economic value of volatility and
distributional timing. In addition, while the distribution of the performance fee
for the $MV^d$ strategy is concentrated around 50 bp, that of the $HM^d$ strategy is
around 190 bp. We performed a Kolmogorov-Smirnov test for the null hypothesis
that the two distributions are the same and the null was overwhelmingly rejected
at any significance level. We conclude that the performance of the $HM^d$ strategy
dominates over the $MV^d$ one.

\textsuperscript{20}We also measured the gain of distributional timing when only the asymmetry of the conditional
distribution is assumed to be time varying and when only the degree of freedom is assumed to be
time varying. We found that both contribute significantly to distributional timing, with a similar
magnitude. We also performed the same experiment without short sales. The magnitude of
distributional timing was slightly reduced, but remained statistically and economically significant.

\textsuperscript{21}Overfitting may arise by the introduction of too many parameters in a model. Some parameters
may be significant only because they help capturing very specific episodes. They would be helpful
to improve in-sample allocation, but useless, at best, for out-of-sample allocation. Data snooping
would occur if the same sample were used for the estimation and the allocation.
5.3 Direct Preferences for Higher Moments

Several contributions to the recent literature demonstrate that there is evidence that investors have direct preferences for skewness. For instance, Mitton and Vorkink (2007) demonstrate that, in an equilibrium model, preferences for more skewness will yield less-diversified portfolios. By taking skewness into account, an investor will pick fewer assets but the one’s that she keeps are expected to show more extreme positive outcomes. The resulting deterioration of the mean-variance measures, by construction ex-ante optimal for a mean-variance investor, may be compensated by the increased skewness that provides felicity to an investor who also cares about higher moments. Mitton and Vorkink (2007) as well as Goetzmann and Kumar (2005) and Kumar (2005) provide empirical evidence that such investors under-diversify their portfolios. Similar theoretical conclusions are reached by Barberis and Huang (2005) who endow the investor with prospect theoretical preferences.

Our framework lends itself naturally to investigate direct preferences for skewness and kurtosis. Formally, we consider as preference structure

\[
E_t [U (W_{t+1})] = \frac{1}{1 - \gamma_2} + m_{p,t+1} - \frac{\gamma_2}{2} m_{p,t+1}^{(2)} + \frac{\gamma_{Sk}}{3!} m_{p,t+1}^{(3)} - \frac{\gamma_{Ku}}{4!} m_{p,t+1}^{(4)},
\]

where the parameters \(\gamma_{Sk} \geq 0\) and \(\gamma_{Ku} \geq 0\) measure the direct preferences for skewness and kurtosis respectively.

Adapting such a preference structure brings about the choice of \(\gamma_2\), \(\gamma_{Sk}\), and \(\gamma_{Ku}\). In the following, we will consider various preferences, ranging from a mean-variance investor with \(\gamma_2 = 5\), \(\gamma_{Sk} = \gamma_{Ku} = 0\) to a power-utility maximizer with parameter \(\gamma = 5\) (corresponding to \(\gamma_2 = 5\), \(\gamma_{Sk} = 30\) and \(\gamma_{Ku} = 210\)) and ending with an investor whose volatility preference is \(\gamma_2 = 5\) but whose preferences for skewness and kurtosis correspond to a power-utility investor with \(\gamma = 10\) (i.e., \(\gamma_{Sk} = 110\) and \(\gamma_{Ku} = 1320\)).

Table 5 presents, for the various preference structures, the performance measures as well as an under-diversification measure introduced by Mitton and Vorkink (2007). This measure, \(D_2\), is defined as the sum of squared weights. Thus, less-diversified portfolios have a higher value of \(D_2\).

In rows 1 and 2 of the table, we consider the case discussed in Mitton and Vorkink (2007). We find that, indeed, once an investor cares also about skewness rather than just mean and variance, her portfolio becomes less diversified. Even

\[\text{Brunnermeier and Parker (2005) show how prospect theoretic preferences may arise as investors choose optimal beliefs.}\]

\[\text{Mitton and Vorkink (2007) consider an investor with } \gamma_{Ku} = 0 \text{ and focus on the equilibrium implications.}\]

\[\text{We also computed their diversification measure that takes correlations into account. This measure moves very much like } D_2, \text{ so that we only present one measure.}\]
though the ex-ante Sharpe ratio decreases by definition, we notice that, for the data at hand, the ex-post Sharpe ratio may increase. Also the investor would be willing to pay a larger performance fee than the mean-variance investor to switch from the static strategy to the $HM^d$ strategy. If we compare the mean-variance investor of row 1 with the investor of row 3 who only dislikes kurtosis, we again find less diversification and a higher performance fee. This finding indicates that preferences for both skewness and kurtosis may play a role for asset allocation.

This leads us to the power-utility investor in row 4 with $\gamma = 5$ already considered in Tables 3 and 4. This investor is still willing to pay a relatively high performance fee to benefit from distributional timing. Her portfolio will be less diversified than in the mean-variance case. As it appears clearly from Figure 5, the investor puts more weights on the UK market to benefit from its distributional properties, and therefore reduces her diversification.

Rows 5 to 7 consider the case where $\gamma_2 = 5$ but where $\gamma_{Sk}$ and $\gamma_{Ku}$ successively take values that correspond to an investor with $\gamma = 10$. On the one hand, the increase in the skewness preference results in less diversified portfolios, consistently with the results put forward by Mitton and Workink (2007) in a similar context. On the other hand, an increase in the weight put on kurtosis results in more diversified portfolios. This result suggests that the investor is willing to avoid large (positive or negative) risks, in addition to the diversification effect already brought by the standard risk (volatility) aversion.

Last, row 8 considers an investor who has the same preferences for skewness and kurtosis as with $\gamma = 5$ in equation (7) but who strongly dislikes volatility. Such an investor would strongly diversify with respect to the case of $\gamma = 5$ and select a portfolio that is close to the $1/n$ strategy.

In an international finance context, investors have been proved to under-diversify, a phenomenon known as home bias, e.g. French and Poterba (1991). Our findings suggest that, if one had to explain this phenomenon with preferences, a model where international investors care more about positive skewness and, up to some degree, dislikes kurtosis more than power-utility investors, may be an avenue for future research.\footnote{A similar remark may be found in Brunnermeier and Parker (2005).} In all cases, investors who have direct preferences for higher moments are willing to pay an additional performance fee to capture the dynamics of skewness and kurtosis over mean-variance investors or over power-utility investors.\footnote{We are grateful to a referee who advised us to investigate the role played by direct higher-moment preferences.}

### 5.4 Aversion to Model Uncertainty

At this point, it may be argued that incorporating parameter uncertainty in the evaluation of the significance of distributional timing is not enough, because it may directly affect the behavior of the investors. As already mentioned in footnote 4,
this issue has been already addressed in a set of contributions that use Bayesian techniques to evaluate how investors averse to model uncertainty choose a portfolio that maximizes the minimum expected utility. This research follows the approach of Gilboa and Schmeidler (1989), who show that the max-min criterion actually reflects the preferences of an investor who is averse to uncertainty about the probability distribution.\footnote{Several recent papers also demonstrate the importance played by ambiguity aversion in asset allocation. Recent contributions in this domain are Hansen, Sargent, and Tallarini (1999), Maenhout (2004), Garlappi, Uppal, and Wang (2007) as well as Leippold, Trojani, and Vanini (2007). In these papers, the utility is also modeled by introducing a max-min criterion, hence, the investor seeks an allocation that will be optimal under the worst case scenario.}

The corresponding optimization program is

\[
\max_{\alpha_t} \min_{\theta \in \Theta} E_t[U(W_{t+1})],
\]

for each period of time, where \(\Theta\) is the domain characterizing the range of the parameters of the distribution required to compute the expectation. Given the complexity of the model, it is not possible to follow the approach adopted by Garlappi, Uppal, and Wang (2007) and to infer for a given parameter which part of its domain is more likely to produce the worst case scenario. To solve this problem, we take advantage of the Bayesian estimation. The parameter vector \(\theta\) has to obey originally certain constraints such as ensuring stationarity and positivity of the covariance matrix. In addition, the range of plausible values of \(\theta\) is delimited by the Bayesian prior and the likelihood of the model. This implies that the domain to which \(\theta\) belongs is relatively restricted. We use the draws from the MCMC estimation, i.e. the posterior distribution, to describe the possible domain, \(\Theta\), to which \(\theta\) can belong.

We solve the optimization problem (13) as follows. For a given date \(t\), we consider all the possible sets of parameters in \(\Theta\) and maximize over portfolio weights \(\alpha_t\) the corresponding expected utility. This yields a solution, say \(\alpha^*_t(\theta)\). Then we seek the allocation that solves equation (13). It is this solution that will ensure that the investor will do best in the worst case.

As the first panel of Table 6 shows, switching from the static strategy to the \(HM^d\) improves expected returns at the cost of a higher volatility.\footnote{In the Bayesian allocations performed so far, we computed, for each draw of \(\theta\), allocations \(\alpha_t\) which we averaged across simulations. In that approach, it is possible to obtain a confidence interval for each allocation. In the case at hand, it is no longer possible to compute such a confidence interval, because there is only one optimal allocation among all the draws. As a consequence, we are still able to evaluate the performance fee to switching from one strategy to another, but it is no longer possible to obtain its distribution.} The Sharpe ratios improve as one considers volatility timing and then distributional timing. Comparison with Table 3 reveals that the conservative investor accepts a decrease in expected returns to obtain less volatility. At the same time, this investor would also seek more positively skewed portfolios with less kurtosis. Such an allocation
would be expected from a conservative investor who does not trust her parameter estimates.

As Table 7 documents, the economic value of volatility timing with conservativeness amounts to 120 bp, whereas the economic value of distributional timing is 85 bp. These estimates suggest that even when the strategies are constrained to be more conservative (to take worst case scenarios into account), the gain of distributional timing is still very sizeable relatively to volatility timing. As the break-even transaction costs show, this performance gain could be implemented in practice even for an investor who chooses conservative allocations. Comparison with Table 4 reveals systematically higher performance fees for the conservative investor. Thus, by choosing conservative allocations to avoid risk due to erroneous parameter estimates appears to provide additional felicity to the investor.

6 Robustness Analysis

As already discussed, we did our best to control for statistical issues. We accounted for overfitting by using two separate subperiods for the parameter estimation and the asset allocation, and we accounted for parameter estimation risk by computing the finite-sample distribution of all our performance measures. To further evaluate the robustness of our result that optimal allocation can be improved by forecasting the conditional distribution, we investigate if it is sensitive to a set of assumptions: First, using the same dataset on international market returns, we consider a change in the estimation and allocation periods as well as a change in the priors of expected returns. Second, we use alternative datasets already considered in the literature and estimate the performance fee of the $MV^d$ and $HM^d$ strategies.

6.1 International Market Returns

Sensitivity to Allocation Period. Changing the allocation period may affect the significance of our results and downweigh the magnitude of distributional timing. For our dataset on international market returns, one reason may be that the allocation period is between 2002 and 2005, a relatively quiet period following the Internet bubble crash. To control for this possible bias, we re-estimated the model between 1977 and 1999 and performed the allocation over the longer period from 2000 to 2005, which includes both bearish and bullish markets.

The results of the asset allocation for all the robustness checks are reported in Tables 6 and 7, for the case $\gamma = 5$. As expected, the performances of the various strategies are very different from the 2002-05 allocation period. Annualized returns are much lower, while annualized volatilities are much higher, resulting in negative Sharpe ratios for the three strategies. However, once the dynamic strategies are compared to the static one, Table 7 reveals that the $HM^d$ strategy still performs
much better than the $MV^d$ one. Indeed, the modified Sharpe ratio is negative for the latter strategy but positive for the former one. This result indicates that in terms of mean-variance trade-off, the $MV^d$ strategy slightly under-performs while the $HM^d$ strategy slightly over-performs the static strategy. When all the characteristics of the distribution are taken into account, the two dynamic strategies perform better than the static one. The reason is that the investor is able to reduce her exposure to stock markets during the turbulent period, resulting in a lower variance and kurtosis in the portfolio return distribution. As a consequence, the estimated performance fee is clearly positive for both dynamic strategies. Yet once again the gain of the $HM^d$ strategy relative to the $MV^d$ one remains sizeable and of the same order of magnitude (around 100 bp) as over the 2002-05 period.

**Sensitivity to Expected Returns.** We also investigate the sensitivity of the distributional timing to a change in expected returns. It is well known that expected returns are difficult to estimate and that small changes in their relative magnitude may strongly affect the weights of the optimal allocations (see Jorion, 1985). To neutralize the effect of expected returns on our allocation, we follow Das and Uppal (2004) and take the average of the sample expected returns across markets as a proxy for the expected returns on all the markets.

As the tables reveal, neutralizing the effect of expected returns barely affects the measures of volatility and distributional timing. The Sharpe ratios remain at the same level as in the baseline case. The performance fee for the $MV^d$ strategy decreases from 55 bp with our baseline model to 38 bp under equal expected returns. We also obtain a decrease in the performance fee for the $HM^d$ strategy (from 193 bp to 139 bp), so that the gain of the $HM^d$ strategy relative to the $MV^d$ one is again about 100 bp. These results suggest that the assumptions regarding expected returns are not a key factor affecting the relative magnitude of volatility and distribution timing.

### 6.2 Alternative Asset Classes

The last analysis we perform to assess the robustness of our main results relies on alternative assets that have been investigated in previous research.29

**Stocks, Bonds, and Gold.** Our first alternative dataset is based on stocks, bonds, and gold. These asset classes were already studied by Fleming, Kirby, and Ostdiek (2001, 2003) in their evaluation of the volatility timing. The data consists of weekly returns for the US stock and bond markets and for gold (source: Datastream International). The estimation sample is from January 1983 to May 2000 (909

---

29A more complete analysis of these results based on alternative datasets is reported in a Technical Appendix.
weekly observations) and the allocation sample is from June 2000 to February 2006 (300 observations). As for international markets, since futures are available for these assets, the investor is allowed to short her position.

Tables 6 and 7 show that again the $MV^d$ and $HM^d$ strategies perform much better than the static strategy. The Sharpe ratios of dynamic strategies are about twice the ratio of the static strategy. More importantly, the modified Sharpe ratio and the performance fee are positive for both strategies, even when parameter uncertainty is taken into account. With this dataset, the economic value of volatility timing is very high, once again because the exposure to risky markets can be reduced during turbulent periods. Although it is smaller in magnitude, the value of distributional timing is sizeable and significant, around 100 bp per year.

**Size Portfolios.** Finally we consider the allocation of wealth between US size portfolios. More precisely, the investment set consists of portfolios on small, intermediate, and large-sized firms. Size portfolios have been previously used by Moskowitz (2003) and Patton (2004) in a similar context. To illustrate the robustness of our result, we perform the same allocation experiment as we did with international markets, but at a daily horizon. The estimation sample is from 1995 to 2005 (2750 observations) and the allocation sample is from January to December 2006 (250 observations). Contrary to the previous data, these assets are highly correlated and their co-skewness and co-kurtosis are very high. This pattern is due to the fact that extreme realizations on a given stock market occur in general at the same dates.

Again the $MV^d$ and $HM^d$ strategies perform better than the static strategy. The realized portfolio return over the allocation period is very high for both dynamic strategies. Even if the realized portfolio volatility is also higher, it results in a much higher Sharpe ratio than for the static strategy. The modified Sharpe ratio suggests that both dynamic strategies generate similar risk-adjusted returns, when risk is measured by the portfolio volatility. However, once all the characteristics of the return’s distribution are taken into account, through the performance fee, the economic gain of the $MV^d$ strategy is barely significant. In contrast, the performance fee of the $HM^d$ strategy is highly significant. The result can be explained by the fact that the return on the big-size portfolio is much more positively skewed than the return of other portfolios. The investor adopting the $HM^d$ strategy therefore had the opportunity to overweight the large-size portfolio.

Inspection of the allocations through time (displayed in the Technical Appendix) confirms the findings of Patton (2004) that the optimal allocation implies a short sale of the small-sized firms. Short-sale restrictions would have prevented investors to take advantage of the characteristics of these stocks during that period.

---

30Size portfolios consist in three portfolios composed of the 30% smallest, 40% intermediate, and 30% largest firms in the CRSP database. Details on the construction of the portfolios can be found in Fama and French (1993). We obtained the data from Professor French’s website.
The interesting study by Tu and Zhou (2004) considers static allocations involving the small-minus-big portfolio return sampled at a monthly frequency. Even though that study found that fat-tailed returns may have important consequences for portfolio, it reports only a small utility gain for an investor who would switch from an erroneous normal distribution to a symmetric Student $t$. Our evidence suggests that as one switches to data sampled at higher frequency, not only are allocations different, but also the gain to take non-normality into account becomes economically relevant.\footnote{We also assume a power-utility function rather than a mean-variance criterion as in Tu and Zhou (2004).}

### 7 Conclusion

In this paper, we have investigated the consequences of non-normality of returns on the optimal asset allocation when the distribution of asset returns changes over time. While most previous work has been devoted to the case where the characteristics of investment opportunities are constant through time, several recent papers have explored the consequences of ignoring the time variability of some aspects of the returns’ distribution: Fleming, Kirby, and Ostdiek (1999, 2001) and Han (2006) evaluate the value of volatility timing, while Ang and Bekaert (2002) as well as Guidolin and Timmermann (2007) measure the cost of ignoring the presence of regime shifts. Patton (2004) investigates a bivariate model with predictability in the asymmetric behavior of asset returns. We contribute to this literature along several dimensions. From the point of view of return dynamics, we propose a model that captures most statistical features of market returns, such as volatility clustering, correlation persistence, asymmetry and fat-tailedness of the distribution. The Bayesian estimation of this model remains tractable, even in the case of several assets.

We show that, for all levels of risk aversion, the performance fee an investor is willing to pay to benefit from distributional timing is of the same order of magnitude as the performance fee to pay to benefit from volatility timing. We perform several alternative experiments to assess the robustness of our findings. We cast the model in a Bayesian setting that allows us to integrate out parameter uncertainty as we consider the performance measures. We also consider conservative investors who take model uncertainty into account in their allocation process. We measure the economic value of distribution timing over several datasets, all pointing towards the relevance of taking into account the temporal variation of the conditional distribution of asset returns.

Several extensions to this research may be considered. First, it would be interesting to have multi-period investments, in order to evaluate the consequences of non-normality on hedging demands. As already mentioned, this extension would
be rather demanding, given the way multi-period investment problems are usually solved. From this point of view, it may be easier to write the problem in continuous time.

Another natural extension would be to analyze the effect of skewed and fat-tailed assets at the equilibrium. Such analysis has been recently provided for skewed assets by Mitton and Vorkink (2007), who showed that preference for skewness should result in portfolio under-diversification. We have seen in Section 5 that increasing the weight of kurtosis in the expected utility implies a higher diversification. Clearly the effect of the interaction between skewness and kurtosis on the optimal allocation remains an open question.
Appendices

Appendix 1: A Multivariate Model for Returns

An extension to the DCC model. The dynamics of returns' vector \( r_t \) is

\[
\begin{align*}
r_t &= \mu + \varepsilon_t, \\
\varepsilon_t &= \Sigma_t^{1/2} z_t,
\end{align*}
\]  

(14) (15)

where \( \mu \) denotes the \( n \times 1 \) mean vector, \( \varepsilon_t \) is the vector of unexpected returns, \( \Sigma_t = \{\sigma_{ij,t}\}_{i,j=1,\ldots,n} \) is the conditional covariance matrix, \( z_t \) is the vector of innovations, such that \( E[z_t] = 0 \) and \( V[z_t] = I_n \), where \( I_n \) is the identity matrix. The conditional covariance matrix of returns \( \Sigma_t \) is defined as \( \Sigma_t = D_t \Gamma_t D_t \), where \( D_t = \{\sigma_{i,t}\}_{i=1,\ldots,n} \) is a diagonal matrix with the standard deviations on the diagonal, and \( \Gamma_t = \{\rho_{ij,t}\}_{i,j=1,\ldots,n} \) is the symmetric positive definite correlation matrix.

Each conditional variance, \( \sigma^2_{i,t} \), is described by an asymmetric GARCH model as in Glosten, Jagannathan, and Runkle (1993)

\[
\sigma^2_{i,t} = \omega_i + \beta_i \sigma^2_{i,t-1} + \alpha_i \varepsilon^2_{i,t-1} + \psi_i \varepsilon^2_{i,t-1} \mathbb{1}_{\{\varepsilon_{i,t-1} < 0\}},
\]

(16)
or, equivalently,

\[
\sigma^2_{i,t} - \bar{\sigma}^2_i = \tilde{\omega}_i + \gamma_i \left( \sigma^2_{i,t-1} - \bar{\sigma}^2_i \right) + (\alpha_i + \psi_i \mathbb{1}_{\{\varepsilon_{i,t-1} < 0\}}) \left( \varepsilon^2_{i,t-1} - \sigma^2_{i,t-1} \right),
\]

(17)

where \( \gamma_i \) denotes the variance persistence. The constraint \( \gamma_i < 1 \) guarantees stationarity of the variance process. The conditional correlation matrix, \( \Gamma_t \), is time-varying, following the DCC specification of Engle (2002) and Engle and Sheppard (2001)

\[
\begin{align*}
\Gamma_t &= (\text{diag} (Q_t))^{-1/2} \cdot Q_t \cdot (\text{diag} (Q_t))^{-1/2}, \\
Q_t &= (1 - \delta_1 - \delta_2) \bar{Q} + \delta_1 \left( u_{t-1} u'_{t-1} \right) + \delta_2 Q_{t-1},
\end{align*}
\]

(18) (19)

where \( u_t = D_{t-1}^{-1} \varepsilon_t \) denotes the vector of normalized unexpected returns, and \( \text{diag}(Q_t) \) denotes the \( n \times n \) matrix with zeros, except for the diagonal that contains the diagonal of \( Q_t \). The matrix \( \bar{Q} \) is the unconditional covariance matrix of \( u_t \). We impose the restrictions \( 0 \leq \delta_1, \delta_2 \leq 1 \) and \( \delta_1 + \delta_2 \leq 1 \), so that the conditional correlation matrix is guaranteed to be positive definite.

The multivariate Sk-t distribution. Innovations \( z_t \) are drawn from \( n \) independent Sk-t distributions. As shown in equation (15), correlation among returns is introduced via a Cholesky decomposition. The \( n \times 1 \) vector of innovations \( z_t \) is drawn from the following multivariate Sk-t distribution

\[
g(z_t|\eta) = \prod_{i=1}^{n} \left( \frac{2 \beta_i}{\xi_i + \frac{1}{\xi_i}} \right) \frac{\Gamma \left( \frac{\nu_i + 1}{2} \right)}{\sqrt{\pi (\nu_i - 2)} \Gamma \left( \frac{\nu_i}{2} \right)} \left( 1 + \frac{\kappa^2_{i,t}}{\nu_i - 2} \right)^{-\frac{\nu_i + 1}{2}},
\]

(20)
where \( \eta = (\nu_1, \ldots, \nu_n, \xi_1, \ldots, \xi_n)' \) denotes the vector of shape parameters,

\[
\kappa_{i,t} = \begin{cases} 
(b_i z_{i,t} + a_i) \xi_i, & \text{if } z_{i,t} \leq -a_i/b_i, \\
(b_i z_{i,t} + a_i) / \xi_i, & \text{if } z_{i,t} > -a_i/b_i,
\end{cases}
\]

with

\[
a_i = \frac{\Gamma \left( \frac{\nu_i - 1}{2} \right) \sqrt{\nu_i - 2}}{\sqrt{\pi} \Gamma \left( \frac{\nu_i}{2} \right)} \left( \xi_i - \frac{1}{\xi_i} \right), \text{ and } b_i^2 = \xi_i^2 + \frac{1}{\xi_i^2} - 1 - a_i^2.
\]

Shape parameters \( \nu_i \) and \( \xi_i \) correspond to the individual degree of freedom and the asymmetry parameter respectively.\(^{32}\) The marginal distribution of \( z_{i,t} \) is an univariate Sk-t distribution \( g(z_{i,t} | \nu_i, \xi_i) \). It is defined for \( 2 < \nu_i < \infty \) and \( \xi_i > 0 \).

Higher moments of \( z_{i,t} \) are easily deduced from those of the symmetric \( t \) distribution \( t(\cdot | \nu_i) \). If the \( r \)-th moment of the \( t(\cdot | \nu_i) \) distribution exists, then the associated variable \( z_{i,t} \) with distribution \( g(\cdot | \nu_i, \xi_i) \) has a finite \( r \)-th moment, defined as

\[
M_{i,r} = \frac{\xi_{i,r}^{1 + \frac{(r-1)\nu_i}{\nu_i - 2}}}{\xi_i + \frac{1}{\xi_i}},
\]

where

\[
m_{i,r} = 2E[Z_i^r | Z_i > 0] = \frac{\Gamma \left( \frac{\nu_i - r}{2} \right) \Gamma \left( \frac{\nu_i + 1}{2} \right) (\nu_i - 2)^{\frac{r+2}{2}}}{\sqrt{\pi} (\nu_i - 2) \Gamma \left( \frac{\nu_i}{2} \right)}
\]

is the \( r \)-th moment of \( t(\cdot | \nu_i) \) truncated to the positive real values. Provided that they exist, third and fourth central moments of \( z_{i,t} \) are

\[
\mu_i^{(3)} = E[Z_i^3] = M_{i,3} - 3M_{i,1}M_{i,2} + 2M_{i,1}^3, \quad (22)
\]

\[
\mu_i^{(4)} = E[Z_i^4] = M_{i,4} - 4M_{i,1}M_{i,3} + 6M_{i,1}M_{i,2}^2 - 3M_{i,1}^2. \quad (23)
\]

Therefore, skewness and kurtosis are non-linear functions of the degree of freedom \( \nu_i \) and the asymmetry parameter \( \xi_i \).

**Dynamics of the conditional higher moments.** Finally, we allow the degree of freedom and the asymmetry parameter to vary over time. The dynamics of these parameters cannot be chosen arbitrarily because of the constraints imposed on their dynamics. The degree of freedom \( \nu_{i,t} \) has to be larger than 2 and the asymmetry parameter \( \xi_{i,t} \) has to be positive at each date \( t \) for the distribution to be well defined. We chose as specification

\[
(1 - c_{i,2} L) \log (\nu_{i,t} - \nu) = c_{i,0} + c_{i,1}^+ |z_{i,t-1}| N_{i,t-1} + c_{i,1}^- |z_{i,t-1}| (1 - N_{i,t-1}), \quad (24)
\]

\[
(1 - d_{i,2} L) \log (\xi_{i,t}) = d_{i,0} + d_{i,1}^+ z_{i,t-1} N_{i,t-1} + d_{i,1}^- z_{i,t-1} (1 - N_{i,t-1}), \quad (25)
\]

\(^{32}\)Parameters \( a_i \) and \( b_i \) are required to center and scale the asymmetric distribution so that \( z_{i,t} \) has a zero mean and unit variance.
where \( N_{i,t} = 1 \{ z_{i,t} < 0 \} \). The parameter \( \nu \) is the lower bound for the degree of freedom.\(^{33}\) Two main features of these specifications are worth emphasizing. First, \( \nu_{i,t} \) is related to the absolute value of lagged standardized innovations, since \( z_{i,t-1} \) is expected to affect the heaviness of the distribution’s tails regardless of its sign. In contrast, \( \xi_{i,t} \) is expected to depend on signed residuals. Second, instead of assuming that positive and negative shocks have the same impact on the shape of the distribution, we allow an asymmetric reaction of the shape parameters to recent shocks.

### Appendix 2: Computing portfolio returns’ moments

Analytical expressions for the portfolio conditional moments can be easily obtained for a multivariate Sk-t distribution. Third and fourth central moments of a Sk-t distributed random variable are given by equations (22) and (23). Next, since unexpected returns are defined as

\[
\varepsilon_{t+1} = \frac{1}{2} r_{t+1} - \mu + \sum_{s=1}^{n} \psi_{rs,t+1},
\]

where \( \psi_{rs} = \omega_{is} \omega_{jr} \omega_{kr} \omega_{ls} + \omega_{ir} \omega_{js} \omega_{kr} \omega_{ls} + \omega_{is} \omega_{jr} \omega_{ks} \omega_{ls} + \omega_{ir} \omega_{js} \omega_{ks} \omega_{ls} \).\(^{35}\) The numerical computation of these expressions is very fast.

The last step consists in the computation of portfolio moments. For a given portfolio weight vector \( \alpha_t \), the conditional expected return, the conditional variance, \(^{34}\) imposing that conditional moments up to the fourth one are defined, i.e. \( \nu = 4 \). Using these notations, central co-moment matrices can be conveniently represented as bi-dimensional matrices.

\(^{33}\)Since we are actually interested in an asset allocation based on the four first moments, we impose that conditional moments up to the fourth one are defined, i.e. \( \nu = 4 \).

\(^{34}\)For simplicity, we suppressed the temporal index in the expression for \( \psi_{rs,t+1} \).
and the conditional third and fourth moments of the portfolio return are defined as:

\[
\begin{align*}
    m_{p,t+1} &= r_{f,t} + \alpha_t' (\mu - r_{f,t} e), \\
    \sigma^2_{p,t+1} &= \alpha_t' \Sigma_{t+1} \alpha_t, \\
    s^3_{p,t+1} &= \alpha_t' S_{t+1} (\alpha_t \otimes \alpha_t), \\
    \kappa^4_{p,t+1} &= \alpha_t' \mathcal{K}_{t+1} (\alpha_t \otimes \alpha_t \otimes \alpha_t),
\end{align*}
\]

where \( \sigma^2_{p,t+1}, \ s^3_{p,t+1}, \) and \( \kappa^4_{p,t+1} \) stand for central moments \( E_t [r_{p,t+1} - m_{p,t+1}]^i \) for \( i = 2, 3 \) and \( 4 \) respectively.\(^{36}\)

The relations between these non-central moments and the usual moments are:

\[
\begin{align*}
    m^{(2)}_{p,t+1} &= \sigma^2_{p,t+1} + m^2_{p,t+1}, \\
    m^{(3)}_{p,t+1} &= s^3_{p,t+1} + 3\sigma^2_{p,t+1} m_{p,t+1} + m^3_{p,t+1}, \\
    m^{(4)}_{p,t+1} &= \kappa^4_{p,t+1} + 4s^3_{p,t+1} m_{p,t+1} + 6\sigma^2_{p,t+1} m^2_{p,t+1} + m^4_{p,t+1},
\end{align*}
\]

\(^{36}\)Central moments \( s^3_{p,t+1} \) and \( \kappa^4_{p,t+1} \) should not be confused with skewness \( sk_{p,t+1} \) and kurtosis \( ku_{p,t+1} \) defined as standardized central moments, \( E_t [((r_{p,t+1} - m_{p,t+1})/\sigma_{p,t+1})^i] \), for \( i = 3, 4 \).
References


Captions

Table 1: This table reports summary statistics on international market returns. Panel A reports the average, the standard deviation, the skewness, the kurtosis, and the first-order serial correlation of returns, $\rho(r)$, and of squared returns, $\rho(r^2)$. Panel B reports correlation, co-skewness, and co-kurtosis matrices. The exponent $a$ indicates that a statistics is significant at the 1% level. For a multivariate normal distribution, all co-skewness are equal to zero, while co-kurtosis are equal to $ku_{iiii} = 3$, $ku_{iijj} = 1$ for $i \neq j$, and 0 for all other cases. The sample does not include the week of the October 1987 crash, because this observation may have a dramatic effect on the shape of the distribution and on the dynamics of the higher moments.

Table 2: This table reports statistics on the optimal portfolio weights, for the various strategies and for values of the risk aversion $\gamma$ ranging from 2 to 15. We also report the distance between the weights of a given dynamic strategy and the weights of the static strategy. It is measured by $Dist = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t}^* - \hat{\alpha}_{i,t}| / 2$. In brackets we report the [5%,95%] interval of the sample weights.

Table 3: This table reports summary statistics on the optimal portfolio return, for the various strategies and for values of the risk aversion $\gamma$ ranging from 2 to 15. We report the first four realized moments of the portfolio return and the Sharpe ratio. First and second moments are annualized.

Table 4: This table reports statistics on the performance of the optimal portfolios, for the dynamic strategies and for values of the risk aversion $\gamma$ ranging from 2 to 15. We report several measures of performance of the strategies. The modified Sharpe ratio, $mSR$, measure is defined by equation (8). The performance fee, $\vartheta$, is estimated from the sample counterpart of the relation (9). The success rate $Z$ is the percentage of times in which the dynamic strategy out-performed the static strategy. The breakeven transaction cost $\tau^{be}$ is defined by equation (10). For each level of risk aversion, the first row corresponds to the median statistics, while the second row reports the 5% and 95% quantiles, resulting from our Bayesian sampling.

Table 5: This table presents various statistics for portfolios obtained for investors with direct preferences for skewness and kurtosis. The parameter $\gamma_2$ corresponds to aversion towards volatility. $\gamma_{Sk}$ and $\gamma_{Ku}$ are the direct preference parameters for skewness and kurtosis respectively. The statistics $D_2$, defined as the sum of squared portfolio weights, measures the degree of diversification of a portfolio. Other measures are as in Table 4.

Table 6: This table reports summary statistics on the optimal portfolio return, for various alternative experiments. The statistics are the same as in Table 3. Panel A considers the allocations obtained for an investor averse to model uncertainty. In Panel B, we consider a longer time span, namely from January 2000 to December 34.
2005 for a power-utility investor (with $\gamma = 5$). Panel C considers again the period from January 2002 to December 2005 as allocation period, but all assets have the same expected return. Panel D considers the weekly allocation obtained for a portfolio including stocks, bonds, and gold, from January 2000 to December 2006. Panel E considers the daily allocation among the size-based portfolios of Fama and French (1993), from January to December 2006.

Table 7: This table reports statistics on the performance of the optimal portfolios, for various alternative experiments. The statistics are the same as in Table 4. The experiments are the same as in Table 6. Given the way the construction of the allocation under aversion to model uncertainty, there are no confidence intervals on the statistics.

Figure 1: This figure displays the evolution of the conditional volatility, as estimated by the model with a Sk-t distribution with time-varying shape parameters. The allocation subperiod begins in 2002.

Figure 2: This figure displays the evolution of the conditional correlation, as estimated by the model with a Sk-t distribution with time-varying shape parameters. The allocation subperiod begins in 2002.

Figure 3: This figure displays the evolution of the conditional skewness, as estimated by the model with a Sk-t distribution with time-varying shape parameters. The allocation subperiod begins in 2002.

Figure 4: This figure displays the evolution of the conditional kurtosis, as estimated by the model with a Sk-t distribution with time-varying shape parameters. The allocation subperiod begins in 2002.

Figure 5: This figure displays the optimal portfolio weights, for the various strategies and for risk aversion $\gamma = 5$. The time variation of the various weights for the static strategy is due to time variation in the risk free rate.

Figure 6: This figure displays the distribution of the UK weights for two dates (the first week of January 2002 and the first week of January 2005) for the $MV^d$ and $HM^d$ strategies.

Figure 7: This figure displays the distribution of the performance fee for $MV^d$ and $HM^d$ strategies. The two curves are obtained by drawing parameters from the Markov chain generated during the parameter estimation. For each set of parameters drawn, the dynamics of the state variables are reconstructed, and for the out-of-sample period we obtain the portfolio allocation. The performance fee is the premium the investor is willing to pay to switch from a static strategy to either the dynamic strategy. The dynamic mean-variance strategy is tantamount to volatility timing, whereas the dynamic higher-moment strategy corresponds to distribution timing.
Table 1: Summary statistics on market returns

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>Japan</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Univariate statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.215</td>
<td>0.179</td>
<td>0.253</td>
</tr>
<tr>
<td>Std dev.</td>
<td>2.099</td>
<td>2.985</td>
<td>2.492</td>
</tr>
<tr>
<td>Skewness</td>
<td>$-0.250^a$</td>
<td>0.228$^a$</td>
<td>0.036$^a$</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.246$^a$</td>
<td>4.947$^a$</td>
<td>4.116$^a$</td>
</tr>
<tr>
<td>$\rho(r)$</td>
<td>$-0.093$</td>
<td>$-0.030$</td>
<td>$-0.001$</td>
</tr>
<tr>
<td>$\rho(r^2)$</td>
<td>0.180</td>
<td>0.234</td>
<td>0.127</td>
</tr>
</tbody>
</table>

|                  |      |       |      |
| **Panel B: Multivariate statistics** |      |       |      |
| Correlation matrix | $r_1$ | $r_2$ | $r_3$ |
| $r_1$             | 1.000 |       |       |
| $r_2$             | 0.219$^a$ | 1.000 |       |
| $r_3$             | 0.397$^a$ | 0.312$^a$ | 1.000 |

|                  |      |       |      |
| Co-skewness matrix | $r_1$ | $r_2$ | $r_3$ |
| $r_1^2$           | $-0.250^a$ | 0.037 | $-0.106$ |
| $r_2^2$           | 0.010 | 0.227$^a$ | 0.000 |
| $r_3^2$           | $-0.014$ | 0.002 | 0.036 |
| $r_2r_3$          | $-0.023$ |       |       |

|                  |      |       |      |
| Co-kurtosis matrix | $r_1$ | $r_2$ | $r_3$ |
| $r_1^3$           | 5.238$^a$ | 0.974$^a$ | 2.481$^a$ |
| $r_2^3$           | 0.668$^a$ | 4.940$^a$ | 1.154$^a$ |
| $r_3^3$           | 1.752$^a$ | 1.305$^a$ | 4.109$^a$ |
| $r_1r_2^2$        | 1.323$^a$ |       | 0.596$^a$ |
| $r_1r_3^2$        | 2.107$^a$ |       |       |
| $r_2^3$           | 0.667$^a$ | 1.341$^a$ | 0.736$^a$ |
| $r_1^2r_2$        |       |       |       |
Table 2: Optimal portfolio weights
(average over the allocation period)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>$\gamma = 2$</th>
<th>$\gamma = 5$</th>
<th>$\gamma = 10$</th>
<th>$\gamma = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Static$</td>
<td>$Static$</td>
<td>$Static$</td>
<td>$Static$</td>
</tr>
<tr>
<td>US</td>
<td>1.406</td>
<td>0.562</td>
<td>0.280</td>
<td>0.187</td>
</tr>
<tr>
<td>Japan</td>
<td>0.290</td>
<td>0.116</td>
<td>0.058</td>
<td>0.039</td>
</tr>
<tr>
<td>UK</td>
<td>1.181</td>
<td>0.472</td>
<td>0.236</td>
<td>0.157</td>
</tr>
<tr>
<td>US T-bill</td>
<td>$-1.876$</td>
<td>$-0.149$</td>
<td>$0.426$</td>
<td>$0.618$</td>
</tr>
<tr>
<td>$Dist$</td>
<td>$[-1.516; 1.107]$</td>
<td>$[0.606; 0.442]$</td>
<td>$[0.534; 0.385]$</td>
<td>$[0.690; 0.590]$</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>1.372</td>
<td>0.548</td>
<td>0.273</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>$[0.333; 0.173]$</td>
<td>$[0.133; 0.069]$</td>
<td>$[0.246; 0.207]$</td>
<td>$[0.624; 0.207]$</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>1.026</td>
<td>0.417</td>
<td>0.209</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>$[0.866; 0.242]$</td>
<td>$[0.366; 0.045]$</td>
<td>$[0.377; 0.173]$</td>
<td>$[0.128; 0.014]$</td>
</tr>
<tr>
<td>$Dist$</td>
<td>$[-2.106; -2.082]$</td>
<td>$[-0.242; -0.232]$</td>
<td>$[0.312; 0.171]$</td>
<td>$[0.661; 0.014]$</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>1.094</td>
<td>0.091</td>
<td>0.045</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>$[0.666; 0.033]$</td>
<td>$[0.729; 0.091]$</td>
<td>$[0.088; 0.009]$</td>
<td>$[0.086; 0.023]$</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>1.026</td>
<td>0.417</td>
<td>0.209</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>$[0.866; 0.242]$</td>
<td>$[0.366; 0.045]$</td>
<td>$[0.377; 0.173]$</td>
<td>$[0.128; 0.014]$</td>
</tr>
<tr>
<td>$Dist$</td>
<td>$[-0.850; -3.153]$</td>
<td>$[-0.372; -0.660]$</td>
<td>$[0.312; 0.171]$</td>
<td>$[0.661; 0.014]$</td>
</tr>
</tbody>
</table>
Table 3: Moments of realized portfolio return over the allocation period

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Moments of realized portfolio return</th>
<th>Sharpe ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>$\gamma = 2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>14.563</td>
<td>37.222</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>15.239</td>
<td>36.340</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>21.780</td>
<td>40.018</td>
</tr>
<tr>
<td>$\gamma = 5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>6.876</td>
<td>14.874</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>7.148</td>
<td>14.521</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>9.819</td>
<td>16.114</td>
</tr>
<tr>
<td>$\gamma = 10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>4.314</td>
<td>7.424</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>4.455</td>
<td>7.250</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>5.798</td>
<td>8.073</td>
</tr>
<tr>
<td>$\gamma = 15$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>3.459</td>
<td>4.942</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>3.554</td>
<td>4.824</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>4.236</td>
<td>5.284</td>
</tr>
</tbody>
</table>
Table 4: Measures of portfolio performance over the allocation period

<table>
<thead>
<tr>
<th>Strategy</th>
<th>mSR (bp)</th>
<th>Performance fee ϑ (bp)</th>
<th>Z (%)</th>
<th>τ_{bei} (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MV\textsuperscript{d}</td>
<td>104.4</td>
<td>136.3</td>
<td>0.507</td>
<td>2.0</td>
</tr>
<tr>
<td>HM\textsuperscript{d}</td>
<td>583.6</td>
<td>487.4</td>
<td>0.502</td>
<td>21.9</td>
</tr>
<tr>
<td>γ = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MV\textsuperscript{d}</td>
<td>41.8</td>
<td>54.6</td>
<td>0.507</td>
<td>1.9</td>
</tr>
<tr>
<td>HM\textsuperscript{d}</td>
<td>233.7</td>
<td>192.5</td>
<td>0.502</td>
<td>21.6</td>
</tr>
<tr>
<td>γ = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MV\textsuperscript{d}</td>
<td>21.3</td>
<td>27.8</td>
<td>0.507</td>
<td>2.0</td>
</tr>
<tr>
<td>HM\textsuperscript{d}</td>
<td>116.5</td>
<td>95.3</td>
<td>0.502</td>
<td>22.7</td>
</tr>
<tr>
<td>γ = 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MV\textsuperscript{d}</td>
<td>14.5</td>
<td>18.7</td>
<td>0.507</td>
<td>2.1</td>
</tr>
<tr>
<td>HM\textsuperscript{d}</td>
<td>62.7</td>
<td>51.9</td>
<td>0.502</td>
<td>18.0</td>
</tr>
</tbody>
</table>
Table 5: Direct preferences for skewness and kurtosis.

<table>
<thead>
<tr>
<th>Utility</th>
<th>$\gamma_2$</th>
<th>$\gamma_{Sk}$</th>
<th>$\gamma_{Ku}$</th>
<th>Sharpe ratio</th>
<th>$mSR$ (bp)</th>
<th>Perf. fee (bp)</th>
<th>$D_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.343</td>
<td>41.8</td>
<td>54.6</td>
<td>0.632</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>30</td>
<td>0</td>
<td>0.501</td>
<td>234.6</td>
<td>188.6</td>
<td>1.047</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>210</td>
<td>0.496</td>
<td>226.5</td>
<td>200.2</td>
<td>0.970</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>30</td>
<td>210</td>
<td>0.501</td>
<td>233.7</td>
<td>192.5</td>
<td>1.011</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>110</td>
<td>210</td>
<td>0.515</td>
<td>255.2</td>
<td>179.7</td>
<td>1.161</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>30</td>
<td>1320</td>
<td>0.493</td>
<td>222.7</td>
<td>215.0</td>
<td>0.881</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>110</td>
<td>1320</td>
<td>0.503</td>
<td>237.6</td>
<td>202.7</td>
<td>0.973</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>30</td>
<td>210</td>
<td>0.498</td>
<td>114.9</td>
<td>97.2</td>
<td>0.251</td>
</tr>
</tbody>
</table>
Table 6: Moments of realized portfolio return over the allocation period
(Robustness analysis)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Moments of realized portfolio return</th>
<th>Sharpe ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>Panel A: Aversion to model uncertainty (2002-05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>6.172</td>
<td>12.632</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>7.316</td>
<td>12.899</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>9.051</td>
<td>14.175</td>
</tr>
<tr>
<td>Panel B: Longer allocation period (1999-2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>0.038</td>
<td>20.457</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>−0.786</td>
<td>17.253</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>0.847</td>
<td>18.074</td>
</tr>
<tr>
<td>Panel C: Equal expected returns (2002-05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>7.019</td>
<td>15.059</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>7.424</td>
<td>15.038</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>9.030</td>
<td>15.859</td>
</tr>
<tr>
<td>Panel D: Stocks, bonds, and gold (2000-06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>7.276</td>
<td>23.018</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>10.779</td>
<td>21.227</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>10.660</td>
<td>20.310</td>
</tr>
<tr>
<td>Panel E: Size portfolios (2006, daily)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>1.917</td>
<td>2.144</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>5.901</td>
<td>6.918</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>5.018</td>
<td>5.323</td>
</tr>
</tbody>
</table>
Table 7: Measures of portfolio performance over the allocation period (Robustness analysis)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>mSR (bp)</th>
<th>Performance fee (bp)</th>
<th>Z (%)</th>
<th>Z^{be} (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Aversion to model uncertainty (2002-05)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( MV^d )</td>
<td>119.4</td>
<td>120.4</td>
<td>0.541</td>
<td>8.2</td>
</tr>
<tr>
<td>( HM^d )</td>
<td>228.0</td>
<td>205.8</td>
<td>0.555</td>
<td>30.5</td>
</tr>
</tbody>
</table>

| **Panel B: Longer allocation period (1999-2005)** |
| \( MV^d \) | -147.1 | 229.9 | 0.507 | -5.1 |
| \( HM^d \) | 55.0 | 312.9 | 0.500 | 5.5 |

| **Panel C: Equal expected returns (2002-05)** |
| \( MV^d \) | 42.5 | 37.5 | 0.517 | 2.9 |
| \( HM^d \) | 168.4 | 138.7 | 0.517 | 14.2 |

| **Panel D: Stocks, bonds, and gold (2000-06)** |
| \( MV^d \) | 415.1 | 623.1 | 0.473 | 13.1 |
| \( HM^d \) | 445.6 | 720.8 | 0.483 | 12.9 |

| **Panel E: Size portfolios (2006, daily)** |
| \( MV^d \) | 598.2 | 252.4 | 0.507 | 18.9 |
| \( HM^d \) | 700.0 | 597.2 | 0.512 | 25.5 |
Figure 1: Evolution of the conditional volatilities

![US Volatility Chart](chart1)

![Japan Volatility Chart](chart2)

![UK Volatility Chart](chart3)

Figure 2: Evolution of the conditional correlations

![US-Japan Correlation Chart](chart4)

![US-UK Correlation Chart](chart5)

![Japan-UK Correlation Chart](chart6)
Figure 3: Evolution of the conditional skewness

Figure 4: Evolution of the conditional kurtosis
Figure 5: Evolution of the optimal (out-of-sample) portfolio weights ($\gamma = 5$)
Figure 6: Distribution of UK weights for two dates

Figure 7: Distribution of the performance fee for $MV^d$ and $HM^d$ strategies