Time-Variability in Higher Moments Is Important for Asset Allocation

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

October 2008

Abstract
It is well known that strategies that allow investors to allocate their wealth using return and volatility forecasts, the use of which are termed market and volatility timing, are of significant value. In this paper, we show that distribution timing, i.e., the ability to use distribution forecasts for deciding asset allocation, also yields significant economic value. By considering the weekly asset allocation among the three largest international stock markets, we find that distribution timing yields around 100 basis points per year. To control for the parameter uncertainty of the model, we cast the model into a Bayesian setting. We also consider alternative preference structures and datasets. In all cases, the value of distribution timing remains highly significant.

Keywords: Portfolio allocation, distribution 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 support from the Swiss National Science Foundation through NCCR FINRISK (Financial Valuation and Risk Management). This paper replaces ‘The Economic Value of Distributional Timing.’ We have benefited from comments of and discussions with Tim Bollerslev, Peter Bossaerts, Ales Cerny, Darrell Duffie, Bernard Dumas, René Garcia, Campbell Harvey, Steward Hodges, Eric Jacquier, Nour Meddahi, and Raman Uppal, and the participants at the European Finance Association meeting in Ljubljana (2007). We remain responsible for any remaining errors.
1 Introduction

A rational investor, if asked to choose between two assets with the same mean and variance, is very likely to invest in the asset with the highest skewness and the lowest kurtosis (Scott and Horvath, 1980; Dittmar, 2002). In this vein, research on optimal asset allocation has provided evidence that heterogeneity in higher moments is influential in explaining the cross-section of stock returns and that skewness should be priced (Kraus and Litzenberger, 1976; Friend and Westerfield, 1980; Barone-Adesi, 1985). Harvey and Siddique (2000) show that there exists a significant premium for systematic skewness (i.e., co-skewness with the market return). Barberis and Huang (2005) have shown that idiosyncratic skewness does also matter in cross-section. Kumar (2005), Mitton and Vorkink (2007), and Goetzmann and Kumar (2008) have empirically confirmed that institutional and retail investors who have preference for skewness voluntarily decide to hold under-diversified portfolios.

However, early empirical evidence reveals that, in a static setting, the mean-variance criterion results in allocations that are very similar to those obtained from a direct optimization of the expected utility, suggesting that higher moments do not play a significant role in practice (Levy and Markowitz, 1979; Kroll, Levy, and Markowitz, 1984). Recent papers have proposed alternative approaches to deal with portfolio allocation under non-normality. Most of this work provides evidence that the mean-variance criterion correctly approximates the expected utility except in situations departing significantly from normality or in cases of highly levered portfolios. For instance, Das and Uppal (2004) have shown that, in the presence of unexpected jumps occurring at the same time in multiple countries, loss from a reduction in diversification is not substantial and the cost of ignoring common jumps is large only for highly levered positions. Jondeau and Rockinger (2006) have reported that the mean-variance criterion fails to approximate the expected utility when assets are characterized by highly asymmetric and fat-tailed distributions. In such a case, optimization strategies based on higher moments provide better approximations of the expected utility. On the whole, therefore, these studies conclude that the mean-variance framework may fail, but only in extreme cases.

This conclusion has been reached under the assumption that the distribution of the opportunity set remains constant through time, while recent work demonstrates that the distribution may vary through varying and be at least partially predictable (Hansen, 1994; Harvey and Siddique, 1999; Jondeau and Rockinger, 2003; Patton, 2004). Evaluating the effect of dynamic higher moments on portfolio allocation requires the design of a data generating process that accounts for time variability in the higher moments. The first attempts towards this goal were made by Ang and Bekaert (2002) and Guidolin and Timmermann (2007) using a switching-regime approach. In this framework, mean and variance of asset returns change depending on the regime. Although the dynamics of higher moments are not explicitly modeled, Guidolin and Timmermann (2007) show that taking into account such changes in regime improves the performance of the portfolio allocation. Despite all of this
work, researchers have not yet conclusively established that improvements due to adding the complexity of dynamic higher moments to the data generating process are sufficient to yield economic value to investors. The main objective of this paper is to establish the economic value of distribution timing, that is the investor’s ability to forecast the subsequent distribution of asset returns and to invest accordingly.

The concept of distribution timing essentially echoes the concepts of market timing and volatility timing that have already been explored in the literature. While market timing, which involves the expected return predictability, has been extensively studied (Kandel and Stambaugh, 1996; Balduzzi and Lynch, 1999; Barberis, 2000; Cremers, 2002, and many others), studies with a similar focus on volatility timing have come only relatively recently. Graham and Harvey (1996) and Busse (1999) have shown that investors design strategies that exploit predictability in volatility. Several papers have shown that volatility timing is of significant economic value for investing on daily to monthly horizons (e.g., Fleming, Kirby, and Ostdiek, 2001, 2003, Marquering and Verbeek, 2001, Johannes, Polson, and Stroud, 2002). These authors have constructed strategies based on volatility forecasts and have shown that such strategies are valuable. In this paper, we study the incremental value of taking skewness and kurtosis into account, and we compare the magnitude of distribution timing relative to that of volatility timing. For this purpose, we consider a dynamic mean-variance strategy in which investors try to benefit from their ability to predict subsequent volatility, and a dynamic higher-moment strategy, in which investors try to benefit from their ability to predict not only volatility, but also the distribution of returns. Our setting allows us to demonstrate, both statistically and economically, the gain of the higher-moment strategy over the mean-variance strategy.

Evaluating the economic importance of distribution timing requires a relatively elaborate statistical model. We extend the dynamic conditional correlations (DCC) model of Engle (2002) and Engle and Sheppard (2001) to the case of a joint distribution that allows for both asymmetry and fat tails. This extension incorporates several statistical features that are known to characterize the dynamics of asset returns. We demonstrate that it provides a good fit of the temporal evolution of the various assets considered in this paper. We then derive closed-form solutions for the moments of the distribution of the portfolio. These moments can be used directly as inputs to a fourth-order approximation of the expected utility. Within this framework, we show that time variability in higher moments does matter for effective portfolio allocation.\footnote{Several authors have proposed portfolio criteria based on an extension of the mean-variance criterion. Examples are the higher-dimensional efficient frontier (Athayde and Flores, 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 such as the hyperbolic absolute risk aversion function.}

One major problem in measuring the economic value of distribution timing is
the parameter uncertainty encountered in estimating the model. Intuitively, one may expect more uncertainty in the estimation of parameters controlling the conditional distribution than in the estimation of first and second moment parameters, the intuition being that there are fewer observations that describe the more extreme parts of the distribution. To account for parameter uncertainty, we perform the estimation of the model in a Bayesian setting, which has several additional advantages in our context. First, it is a framework that naturally handles the estimation of nonlinear models with a large number of parameter constraints. Second, we can take advantage of the large number of parameter draws to test economic hypotheses. In particular, we can directly test the statistical significance of distribution timing. Finally, the Bayesian setting provides a very efficient way to investigate the consequences of the investor’s aversion to model uncertainty on the optimal asset allocation. For all of these reasons, the use of Bayesian analysis appears relevant in the context of distribution timing.

We apply our approach to several datasets. We report, however, only the weekly allocation of wealth among the three largest stock markets, that is those of the US, Japan, and the UK. 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 excessive risk taking and significant opportunity cost, as compared to a strategy based on higher moments. The performance fee that an investor would be willing to pay to benefit from the higher-moment dynamic strategy (distribution 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 distribution timing is approximately 100 basis points per year, much greater than the economically insignificant value of volatility timing. When we investigate alternative sample periods, preference structures or datasets, we find that the economic value of distribution timing remains sizeable and comparable to volatility timing, even after taking transaction costs into account.

The outline of this paper is as follows. In Section 2, we formulate our approach for modeling returns with a non-normal multivariate distribution and for measuring distribution timing. Section 3 presents the empirical results. We discuss the estimation of the model and the main characteristics of the portfolios obtained assuming mean-variance or higher-moments strategies. Finally, we measure the economic value of these strategies, under alternative preference structures and we provide some robustness checks of our main results. Section 4 concludes the paper.\textsuperscript{2}

\textsuperscript{2}The appendix contains the details of the statistical model describing the evolution of returns. A Technical Appendix, available upon request from the authors, provides further details on the estimation techniques used in the paper, as well as additional empirical results.
2 Methodology

This section describes our methodology for solving the conditional asset allocation problem with non-normal returns. We first present the data generating process that we adopt for asset returns. The process consists of a DCC model, with a multivariate skewed $t$ (Sk-$t$) distribution, which allows for both asymmetry and fat tails. The parameters driving the shape of the conditional distribution are allowed to vary over time as a function of past shocks. For a complete description of the model, see Jondeau and Rockinger (2008). We then describe how including forecasts of higher moments 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 in order to obtain the optimal asset allocation. Finally, we describe how to measure the gain of distribution timing and how to test its significance.

2.1 The Multivariate Return Process

Given our interest in the effect of the 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:

\begin{align}
  r_t & = \mu + \varepsilon_t, \quad (1) \\
  \varepsilon_t & = \Sigma_t^{1/2} z_t, \quad (2) \\
  z_t & \sim g(z_t|\eta_t). \quad (3)
\end{align}

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, $\Sigma_t = E_{t-1}[\varepsilon_t (\varepsilon_t - \mu)^\prime]$. The vector of independent innovations, $z_t$, has zero mean and identity covariance matrix. We denote by $\Sigma_t^{1/2}$ a matrix such that $\Sigma_t = \Sigma_t^{1/2} \Sigma_t^{1/2\prime}$. In this paper, we use the Choleski decomposition. Finally, equation (3) specifies that innovations, $z_t$, follow a conditional distribution with (possibly time-varying) shape parameters, $\eta_t$. When the conditional distribution is normal, there is no shape parameter, since the normal distribution is characterized entirely by its mean and variance. In more general cases, the shape parameters typically involve parameters that capture the asymmetry and fat-tailedness of the distribution.

In Appendix 1, we provide additional details on the 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). The Sk-$t$ distribution is able to capture both fat-tailedness and asymmetry, which are often found in actual data through two shape parameters for each asset $i$:
the degree of freedom of the $t$ distribution, $\nu_{i,t}$, and the asymmetry parameter, $\xi_{i,t}$.

In order to model time variability in both shape parameters, we adopt the following specification:

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

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

where $N_{i,t} = 1_{\{z_{i,t} < 0\}}$ and $L$ is the lag operator. The parameter $\nu$ is the lower bound for the degree of freedom. These specifications generate asymmetry in the return distribution subsequent to shocks.

Unreported goodness-of-fit tests show that the model fits the data very well. Most importantly, it is possible to analytically derive the moments of a portfolio in which asset returns are driven by this distribution. This insight, combined with a Taylor approximation of the expected utility, allows us to perform allocation in very efficient manner.

The model that we use to describe the evolution of returns is relatively complex. Given this complexity, the correct convergence of the maximum likelihood optimizer is not guaranteed and the measurement of parameter precision may be prone to numerical inaccuracies. For this reason, we estimate the model in a Bayesian setting, where only the computation of the likelihood is required. This setting also allows us to demonstrate that the reported economic values are not due to chance but, rather, are robust to parameter uncertainty. In other words, this setting allows us to deal directly with estimation risk. This issue has already been addressed by Kandel and Stambaugh (1996), Barberis (2000), Johannes, Polson, and Stroud (2002), and Harvey et al. (2004). Using our draws in the empirical distribution of the parameters, it is easy 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 for an investor with aversion to model uncertainty. A complete description of the estimation technique used in this paper can be found in the Technical Appendix.

\[\text{For a modeling of these parameters in a univariate setting, see Hansen (1994) and Jondeau and Rockinger (2003). Alternative distributions allowing for asymmetry and fat tails have been proposed in an allocation perspective. See for instance Patton (2004) or Mencia and Sentana (2005).}\]

\[\text{Other work, see Patton (2004), obtains the expected utility via Monte Carlo integration of a bivariate distribution. This setting would be very time-consuming in a higher-dimensional context. Our model allows for model estimation and portfolio allocations, even in cases with many assets.}\]

\[\text{Several authors have investigated how model uncertainty may be incorporated into the portfolio optimization problem. See Barberis (2000), Polson and Tew (2000), Johannes, Polson, and Stroud (2002), Harvey et al. (2004). In most papers, estimation risk is addressed in a Bayesian framework, which provides a direct way to maximize the expected utility given the distribution of parameters.}\]
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, $W_{t+1}$. 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 $r_{t+1} = (r_{1,t+1}, \cdots, r_{n,t+1})'$ and a risk-free asset with return $r_{f,t}$ from time $t$ to time $t+1$. End-of-period wealth is $W_{t+1} = 1 + r_{f,t+1}$, where $r_{p,t+1} = r_{f,t} + \alpha_t' (r_{t+1} - r_{f,t}e)$ denotes the portfolio return, with $\alpha_t = (\alpha_{1,t}, \cdots, \alpha_{n,t})'$ the vector of weights allocated to the various risky assets at time $t$. With $e$ we denote the $n \times 1$ vector of ones. 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 also assume 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.

Optimal portfolio weights are obtained by maximizing the expected utility

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

(6)

In general, this problem does not have a closed-form solution and must be solved numerically. 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 distributions would require a number of quadrature points that increases exponentially with the number of assets. Hence solving the optimization problem with numerical integration is practically intractable for more than two or three assets.

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

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

Wang (2005) and Garlappi, Uppal, and Wang (2007) adopt alternative approaches for incorporating estimation risk into the optimization procedure. In Section 3.5, we discuss the case of investors with aversion to model uncertainty.

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 Barberis (2000) 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 defer this extension to a multi-period investment to future research.

This framework can easily be extended to allow for portfolio rebalancing when changes in the optimal weights exceed a certain threshold.
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

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

(7)

such that it depends on all of the moments of the distribution of the portfolio return.\(^8\) The investor’s preference (or aversion) towards the $k$-th moment is directly given by the $k$-th derivative of the utility function. Since our econometric model is designed to capture the dynamics of the third and fourth moments, we focus on a Taylor series expansion up to the fourth order.\(^9\) The expected utility (7) can therefore be rewritten as

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

(8)

where $m_{p,t+1}^{(i)} = E_t[r_{p,t+1}^i]$ are 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 (8) becomes

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

(9)

The effect of the third and fourth moments on the approximated expected utility is unambiguous. This finding is consistent with the theoretical arguments developed by

---

\(^8\)Such an approach has been adopted in a number of contributions. See Rubinstein (1973), Kraus and Litzeneberger (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.

\(^9\)In fact, 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.

\(^{10}\)Dittmar (2002), Jondeau and Rockinger (2006), and Guidolin and Timmermann (2007) also maximize this four-moment expected utility. In Appendix 2, we explain how to obtain the moments of the portfolio return from those of the asset returns.
Scott and Horvath (1980). Expected utility decreases with large negative skewness (i.e., left-skewed distributions) and large kurtosis (i.e., fat-tailed distributions).

Maximizing expression (??) for each date \( t \) defines a dynamically rebalanced portfolio that maximizes the expected utility of the investor. This expression clearly shows how forecasts of the higher moments of the portfolio return distribution will affect the optimal weights at date \( t + 1 \). At this point, we have a model for asset returns and a criterion for asset allocation. We now turn to a discussion of how to evaluate the economic relevance of distribution timing.

### 2.3 Allocation Strategies

In our model, we have constant expected returns, as it is also assumed by Fleming, Kirby, and Ostdiek (2001, 2003). Some recent papers have shown that macroeconomic variables can have predictive power for monthly returns (Kandel and Stambaugh, 1996, Brandt, 1999, Marquering and Verbeek, 2001), although this is still a controversial topic (see Cremers, 2002). To our knowledge, published literature regarding the short-term predictability of returns is very scarce. Since our focus is on the evaluation of volatility and distribution timing, we will defer the investigation of the predictability of returns on short horizons to future 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 moments over the estimation period and then holds these parameter estimates constant over the allocation period. Thus, for the benchmark strategy, optimal portfolio weights vary only when the risk-free rate varies.\(^{11}\)

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 by using our general model that allows for time-varying parameters. With this strategy, the investor is allowed to take full advantage of volatility and distribution timing. This dynamic higher-moment strategy is denoted \( HM^d \).

\(^{11}\)Given the relative performance of the various assets over time, there are also some transaction costs incurred when the investor has to rebalance her portfolio to maintain constant weights. It turns out that this rebalancing effect is negligible.
2.4 Measuring the Gains of Distribution Timing

We compare the performance of the various strategies using many different measures. A first measure of performance is the standard Sharpe ratio, which is computed using the ex-post average return $m_p$ and the 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 with different levels of risk, we also consider the modified Sharpe ratio $mSR$ introduced by Graham and Harvey (1997)

$$mSR = \frac{\sigma_0}{\sigma_p} (m_p - r_f) - (m_0 - r_f), \quad (10)$$

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

These measures have however an obvious drawback in our context, since they do not capture the effect of non-normality. Given this, we consider another tool for evaluating the economic value of volatility and distribution 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 to both strategies

$$E_t [U (1 + \hat{r}_{p,t+1})] = E_t [U (1 + r^*_{p,t+1} - \vartheta)], \quad (11)$$

where $r^*_{p,t+1}$ is the optimal portfolio return obtained under the dynamic strategy, and $\hat{r}_{p,t+1}$ is the optimal portfolio return obtained under the static strategy.\footnote{We also considered the certainty equivalent, previously adopted 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) that 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 provide the same measure of the economic gain (up to a few basis points), we only report the former in our empirical evidence.}

The performance fee is obtained by solving equation (11) numerically.

We also compute the success rate, i.e., the percentage of allocation periods for which the dynamic strategy has a greater realized return than the static strategy. This measure, denoted $Z$, is useful to check if the out-performance of the dynamic strategy is due to some very specific events or, rather, to an enhanced ability to capture changes in investment opportunities. For instance, very small values of $Z$ suggest that the out-performance is essentially due to a few favorable events.

Finally, since we compare static and dynamic strategies, we must control for the possible effect of transaction costs. Indeed, given that 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 of costs may be incurred depending on the asset and the type of customer relation. For this reason, we follow the approach of Han (2006) and compute the breakeven transaction cost, denoted $\tau^{bc}$. This parameter measures the level of transaction costs required to make the investor indifferent when choosing 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 breakeven transaction cost between the dynamic strategy $d$ and the static strategy $s$ is defined as $\tau^{bc} = (\bar{r}^{d} - \bar{r}^{s}) / (tv^{d} - tv^{s})$. If the actual transaction cost is lower than $\tau^{bc}$, the investor will prefer the dynamic strategy over the static strategy.\(^{13}\)

There are several issues in testing the statistical significance of the gains due to distribution timing. First, while the static strategy requires the estimation of only the sample mean vector and the covariance matrix, the dynamic strategies rely on estimation of the dynamics of the covariance and higher co-moments matrices. To avoid any overfitting of the data or data snooping, we will use two non-overlapping subsamples for the estimation and allocation stages.\(^{14}\)

Another important issue in the evaluation of the economic value of a strategy is estimation risk. Our results suggest that distribution timing has an economically sizeable value, but this value may be statistically insignificant if the uncertainty surrounding parameter estimates is too large. To address this issue, we use Bayesian estimation to generate draws from the finite-sample distribution of the parameters and to evaluate the significance of the performance measures reported in Section 3.3.

---

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 extensive discussion in Han (2006).

14 Overfitting may arise from the introduction of too many parameters in a model. Some parameters may be significant only because they help in capturing very specific episodes. They would be helpful to improve in-sample allocation, but useless (at best) for out-of-sample allocation. Data snooping occurs if the same sample is used for both estimation and allocation.
3 Empirical Investigation

3.1 Data Description

In order to demonstrate that our results are general, we used several datasets to evaluate the economic significance of distribution timing. In this section, we report results from our first dataset, consisting of the 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.\(^\text{15}\) The data are weekly and cover the period from January 1977 through June 2007, for a total of 1588 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 June 2007, 292 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 and well-developed markets. As a consequence, it is less likely to be characterized by extreme behaviors that may drive the results. In the Technical Appendix, we report evidence based on alternative subsamples or datasets. It turns out that our main results are not significantly altered.

Table 1 reports several summary statistics for the market returns under investigation for both the estimation and the allocation periods (Panel A). Over the estimation period 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 slightly smaller return than the UK market, with a relatively low volatility, while the Japanese market is characterized by low return and 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 measures are between 4.1 and 5.3, a range that is inconsistent with the assumption of normality. 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

\(^{15}\)The data consist of 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 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 in the model.
returns are strongly correlated which suggests temporal variation in second moments. We observe some changes in the average returns over the allocation period. In particular, the US market has the lowest expected return, while the Japanese market ranks first. The UK market is characterized by a large negative skewness, while still having a low kurtosis relative to the other markets.

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 correlation between the US and the Japanese markets is smallest (0.218). Given the well-known time variability of correlations, these sample correlations may be misleading for allocation purposes. For instance, the sample correlation between the US and the UK is as high as 0.722 over the (out-of-sample) allocation period. Hence, the static strategy is likely to overstate the diversification ability of the UK market.

3.2 Model Estimation

Table 2 reports the parameter estimates of the multivariate model with Sk-t distribution and time-varying shape parameters. In all cases, as expected, the asymmetry-in-volatility parameter, \( \psi_i \), is significantly positive, suggesting that bad news has a stronger effect on volatility than good news. In addition, the persistence of volatilities is rather large in US and Japanese markets, but much less so for the UK market. Turning to the dynamics of correlations, the persistence parameter, \( \delta_2 \), takes a value of 0.978, translating the fact that correlation dynamics are highly persistent.

Regarding the shape parameters, the table reveals that the degree-of-freedom parameter varies over time for all markets under study. It is strongly related to the absolute value of the past innovation. Indeed, most of the parameters \( c_{i1}^- \) and \( c_{i1}^+ \) are significantly positive. This result indicates that, after a large (positive or negative) shock, the degree of freedom of the Sk-t distribution increases, so that kurtosis decreases. This suggests that large shocks do not cluster at weekly frequency. Regarding the asymmetry parameter, we observe that the parameter \( d_{i1}^- \) is positive for the US and Japan. This suggests that, after a large negative shock, the subsequent distribution is more negatively skewed and the probability of another large negative shock increases. Similarly, for Japan and the UK, large positive shocks are more likely to be followed by large shocks with the same sign.

Figures 1 to 4 display the dynamics of volatilities \( \sigma_{i,t} \), correlations \( \rho_{ij,t} \) and individual conditional higher moments \( sk_{i,t} \) and \( ku_{i,t} \), respectively. Inspection of these figures reveals several interesting features from a portfolio perspective. First, the volatility of the UK market is relatively low over the entire allocation period. In particular, it is less than 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. For instance, the conditional correlation between the US and UK markets varies around 0.5 over the allocation period, while it is
below 0.4 over the estimation period. Third, the conditional skewness measures are in the range of the sample measures of asymmetry for the US and Japan. However, the UK market return turns out to be more negatively skewed over the allocation period. Finally, the most striking result is the variability in the conditional kurtosis: For the US and the UK, the conditional estimates are below the sample estimates (around 4 and 3.5 for the US and the UK, while the sample measures in Table 1 are 5.3 and 4.1 respectively). In contrast, the conditional kurtosis of the Japanese market is found to be much higher (around 7) than the sample estimate.\footnote{Estimation of conditional higher moments is based on the complete structural description of the return’s dynamics, whereas sample moments may be strongly affected by a few extreme events.} Once the temporal evolution of the higher moments is taken into consideration in the portfolio decision problem, the investor is likely to put more weight on the UK market, at the expense of the Japanese market.

Unreported results show that all the co-skewness measures between the US and Japanese markets, and between the US and UK markets are negative. This implies that the US market provides a bad hedge against adverse changes in volatility in the Japanese and UK markets, and vice-versa. Conversely, the co-skewness measures between the Japanese and UK markets are positive, suggesting that the return in one market is likely to be larger than expected when the volatility in the other market is high, thus providing a good hedge against high volatility. In addition, most co-kurtosis measures are positive and well above the value predicted by a normal distribution. Interestingly, there is a positive trend in the co-kurtosis between the US and UK markets, suggesting that the ability of these two markets to hedge each other is worsening over time.

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

### 3.3 Portfolio and Performance Analysis

We now turn to the analysis of the performance of the various dynamic trading strategies described above. Each strategy will provide some insight on the economic value of the volatility and distribution timing. We therefore compare the out-of-sample performance of the dynamic strategies to that of the static strategy. The evolution of the portfolio weights over the out-of-sample period is displayed in Figure 5 (for a risk aversion of $\gamma = 5$). The statistics on the realized portfolio returns of the various strategies are reported in Table 3, while the statistics on the relative performances of the strategies are reported in Table 4. We begin our discussion with the gain that results from switching from the static to the $MV^d$ strategy (volatility timing). We will then discuss the gain that results from switching to the $HM^d$ strategy (distribution timing).
3.3.1 Economic Value of Volatility Timing

The estimation of the optimal portfolio weights implied by the various strategies under study is performed as follows. 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 to produce portfolio weights. The ex-ante optimal static portfolio comprises mainly the US and UK markets (50% and 44% respectively, on average), while the weight of the Japanese market is only 9% (3% of the wealth gets borrowed at the risk-free rate). Although these optimal weights do not fully reflect the capitalization of these markets, they are consistent with the means and variances reported in Table 1.

When the investor accounts for the time-variability in the covariance matrix (\(MV^d\) strategy), the weight of the US market significantly decreases at the beginning of the allocation period, to below 25% over the first two years, 2002-03, mainly because the volatility is higher than it is over the estimation period. In subsequent years, the US market is allocated a higher weight on average, while the Japanese market receives a lower weight reflecting the changes in the volatility forecasts (as shown in Figure 1). The UK market has a higher weight over the allocation period (47%), corresponding to a slightly lower volatility than that seen during the estimation period.

As the tables reveal, however, the \(MV^d\) strategy cannot benefit from volatility timing. Indeed, the static and \(MV^d\) strategies show similar performance. First, the ex-post return and volatility of the portfolio are close for both strategies: For a risk aversion of \(\gamma = 5\), the annualized return is equal to \(\mu = 8.7\%\) for the \(MV^d\) strategy against 8.8% for the static strategy, while realized volatilities are equal to \(\sigma = 13.5\%\). As a result, the strategies have the same Sharpe ratio (\(SR = 0.45\)). Accordingly, the risk-adjusted excess return of the dynamic strategy, measured by \(mSR\), is negative but insignificantly different from zero.

The table also reveals that the performance fee a static investor is willing to pay to switch to the dynamic \(MV^d\) strategy is negative. This means that capturing volatility timing does not increase utility to the investor. The same result holds when one considers alternative levels of risk aversion: for all values of \(\gamma\), the performance fee is insignificantly different from 0. Our estimate of the value of volatility timing is lower than the estimates reported, for instance, by Fleming, Kirby, and Ostdiek (2001, 2003), and by Han (2006). This result suggests that, over our dataset and allocation period, the ability to forecast the subsequent covariance matrix of asset returns is essentially worthless for a power investor.

\[\text{As already mentioned, the weights of the static strategy are not necessarily held constant, since the risk-free rate varies over time.}\]

\[\text{In a previous version of the paper, the out-of-sample period was 2002-2005, and the reported performance fee was significantly positive. The reason for this difference is that, over the end of the allocation period, the static strategy performs better than the dynamic one.}\]
3.3.2 Economic Value of Distribution Timing

When the investor takes the temporal evolution of the conditional distribution into account \((HM^d)\) strategy, the new feature that she has to consider is the trade-off between skewness and kurtosis in asset returns. Over the allocation period, between the end of 2003 and the beginning of 2006, we observe dramatic changes in portfolio weights relative to the \(MV^d\) strategy. The weight of the UK market increases to more than 100% on average, whereas the US weight decreases from 100% to less than 90%. The explanation of this result can be found in Figures 3 and 4. On the one hand, we notice that the skewness of the UK market is only slightly negative, while its kurtosis is very low, close to 3, i.e., the value of the kurtosis of a Gaussian variable. On the other hand, the skewness of the US return remains significantly lower (below \(-0.2\)) and its kurtosis reaches higher levels during the 2004-05 period (above 4). Finally, we do not observe significant changes in the Japanese weight with respect to the \(MV^d\) strategy. Indeed, although its skewness is highly positive (around 0.2), the kurtosis is extremely high over the allocation period and reaches peak values up to 9.

It turns out that, over this period, the UK market has outperformed the other markets (about 33% vs. 11% for the US and vs. 25% for Japan), meaning that the \(HM^d\) strategy yields a significant increase in the realized returns (from 8.7% for the \(MV^d\) strategy to 11.5% per year).\(^{19}\) This result shows that the \(HM^d\) strategy has been able to benefit from the better perspectives of the UK market and from the relatively worse perspectives for the US and Japanese markets. Indeed, for some dates, the higher expected skewness of the UK market produced higher ex-post returns, leading to a higher ex-post portfolio return.\(^{20}\)

The economic gain due to distribution 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 102 bp for \(\gamma = 5\). The ability to benefit from distribution timing therefore generates an additional performance fee of about 107 bp (102.1 + 4.7). Depending on the level of risk aversion, the performance fee ranges between 40 to 150 bp.\(^{21}\)

\(^{19}\)Some contributions (Mitton and Vorkink, 2007, or Goetzmann and Kumar, 2008) have investigated the equilibrium implications of investors having preferences for skewness. The main result is that, at equilibrium, such investors are willing to pay a premium for holding positively skewed assets. Thus, all other things being equal, the expected returns of these assets should be lower than the ones without skewness. In our framework with preference for skewness and aversion for kurtosis, the combined effect is likely to be ambiguous and the equilibrium implications are unclear. We wish to defer this investigation to future research.

\(^{20}\)This does not necessarily imply an increase in the ex-post portfolio skewness (see Table 3). This is because the higher ex-post portfolio return is obtained through a succession of small gains on the UK market relative to the US and Japanese markets.

\(^{21}\)We have also measured the gain of distributional timing when only the asymmetry of the
The evidence presented above has been obtained in the absence of transaction costs and it is clear that the gain of the dynamic strategy becomes smaller once such costs are included. The last column of the table reports the breakeven transaction cost, $\tau_{be}$, of the $HM^d$ strategy. It is about 20 bp for all levels of risk aversion. These values are in the lower 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, the distribution of the breakeven transaction cost across simulations clearly takes on a range of positive values. This suggests that a large part of the gain resulting from the $HM^d$ strategy is given back when transaction costs are accounted for. With a breakeven cost around 20 bp, which is clearly above the range of actual transaction costs, the gain of implementing the $HM^d$ strategy will be only slightly reduced in practice.

To assess the statistical significance of the gains to distribution timing, Figure 7 depicts the empirical distribution of the performance fee for $MV^d$ and $HM^d$ strategies with respect to the static strategy. One can clearly see that the performance fee of the $HM^d$ strategy is significantly positive, confirming the economic value of distribution timing. At the same time, we note that the distribution of the performance fee for the $MV^d$ strategy is concentrated around 0, while that of the $HM^d$ strategy is concentrated around 100 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 all significance levels. We conclude that the performance of the $HM^d$ strategy is superior to that of the $MV^d$ one.

### 3.4 Direct Preferences for Higher Moments

As several contributions to the recent literature demonstrate, 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 ones that she keeps are expected to show more extreme positive outcomes. The resulting deterioration in the mean-variance measures, by construction ex-ante optimal for a mean-variance investor, may be compensated by the higher skewness that increases utility to an investor who also cares about higher moments. Such behavior has also been described in the early contribution of Conine and Tamarkin (1981) and more recently by Mitton and Vorkink (2007) as well as Kumar (2005) and Goetzmann and Kumar (2008), who provide empirical evidence 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, with a similar magnitude, to distributional timing. We have also performed the same experiment without short sales. The magnitude of distributional timing was slightly reduced, but it remained statistically and economically significant.
that such investors under-diversify their portfolios. Similar theoretical conclusions are reached by Barberis and Huang (2005), who endow the investor with prospect theoretic preferences.\(^{22}\) In this last work, the preference for right-skewed assets results from a misrepresentation of the probability distribution of returns.

Our framework is designed to investigate direct preferences for skewness and kurtosis. Formally, we consider the preference structure

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

where \(\gamma_{Va}, \gamma_{Sk},\) and \(\gamma_{Ku}\) are positive parameters, which measure the direct preferences for variance, skewness, and kurtosis, respectively.\(^{23}\)

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

**Table 5** presents the performance measures for the various preference structures that have just been described, as well as an under-diversification measure introduced by Mitton and Vorkink (2007). This measure, \(D_2\), is defined as the sum of the squared weights of the risky portfolio, excluding the risk-free asset.\(^{24}\) Thus, less-diversified portfolios have a higher value of \(D_2\).

We first consider the case of utilities #1 and #2, already 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 though the (unreported) ex-ante Sharpe ratio decreases by definition, we notice that, for the data at hand, the ex-post Sharpe ratio may increase. In addition, a static investor would be willing to pay a larger performance fee to switch to the \(HM^d\) strategy (utility #2) than to switch to the \(MV^d\) strategy (utility #1). If we compare the mean-variance investor with the investor who only dislikes kurtosis (utility #3), we again find less diversification and a higher performance fee. Intuitively, dislike of kurtosis may lead to greater diversification just as an increase of risk aversion in a

\(^{22}\)Brunnermeier and Parker (2005) show how prospect theoretic preferences may arise as investors choose optimal beliefs.

\(^{23}\)Mitton and Vorkink (2007) consider an investor with \(\gamma_{Ku} = 0\) and focus on the equilibrium implications.

\(^{24}\)We also computed their diversification measure that takes correlations into account. This measure moves very much like \(D_2\), so that we only present one measure. We also computed a measure including the risk-free asset. Again, we obtained a measure with the same ranking as the reported measure.
mean-variance setting would lead to a more diversified portfolio. We do not observe
this feature at this stage (but will find it later) because the assets chosen by the
mean-variance allocation (utility #1) tend to have a relatively large kurtosis. Hence
an investor who cares about kurtosis will select stocks that have the smallest kur-
tosis, though the exact choice obviously also depends on issues such as co-kurtosis.
This finding indicates that preferences for both skewness and kurtosis may play an
important role in asset allocation, leading to under-diversified portfolios.

This leads us to the power-utility investor, who has \( \gamma = 5 \) (utility #4) and has
already been considered in Tables 3 and 4. This investor is still willing to pay a
relatively high performance fee to benefit from distribution timing. Her portfolio
will be less diversified than in the mean-variance case but more diversified than that
of the investor who likes skewness.

We now consider the cases in which \( \gamma_{Va} = 5 \) but where \( \gamma_{Sk} \) and \( \gamma_{Ku} \) successively
take values that correspond to an investor with \( \gamma = 10 \) (utilities #5 to #7). The
benchmark case is (#4). On the one hand, the increase in the skewness preference
results in less diversified portfolios, a result consistent with the evidence put forward
by Mitton and Vorkink (2007) in a similar context. Interestingly, if one increases
only the dislike of kurtosis (utility #6), the portfolio becomes more diversified than
that of utility #4. The reason for this is that kurtosis now plays a role similar to
that of volatility. This result suggests that the investor is willing to avoid large
risks (whether positive or negative), in addition to the diversification effect already
brought on by aversion to volatility.

The last case corresponds to an investor who has the same preferences for skew-
ness and kurtosis as with \( \gamma = 5 \) in equation (??), but who strongly dislikes volatility
(utility #8). Such an investor would diversify more strongly than in the case of
\( \gamma = 5 \).

As this investigation demonstrates, explicit preferences for higher moments may
have an important impact on the degree of diversification of a portfolio. A preference
for skewness will lead to portfolios with more positively skewed assets. Increased
kurtosis aversion may lead to more or less diversified portfolios depending on the
initial higher moment preferences.

3.5 Aversion to Model Uncertainty

At this point, it may be argued that incorporating parameter uncertainty in the
evaluation of the significance of distribution timing is not enough, since it may
directly affect the behavior of investors. As already mentioned in footnote 5, this
issue has been addressed in a set of contributions, which use Bayesian techniques
to evaluate how investors with aversion 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 minimum expected utility
actually reflects the preferences of an investor who is averse to uncertainty about
the probability distribution. The corresponding max-min optimization program is

$$\max_{\alpha_t} \min_{\theta \in \Theta} E_t[U(W_{t+1})],$$  \hspace{1cm} (12)$$

for each period of time, where $\Theta$ is the domain characterizing the range of the parameters required to compute the expectation. Given the complexity of the model, it is not possible to follow the approach adopted, for instance, by Garlappi, Uppal, and Wang (2007) in order 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 Bayesian estimation. The parameter vector $\theta$ has to obey a set of constraints such as those ensuring stationarity and positivity of the covariance matrix. In addition to this, the range of plausible values of $\theta$ is delimited by the Bayesian prior and the likelihood of the model. We use draws from the MCMC estimation, which is the posterior distribution, in order to describe the possible domain, $\Theta$, to which $\theta$ can belong.

We solve the optimization problem (12) as follows. For a given date $t$, we consider all the possible sets of parameters in $\Theta$ and maximize the corresponding expected utility over portfolio weights $\alpha_t$. This yields a solution, say $\alpha^*_t(\theta)$. We then seek the allocation that solves equation (12). It is this solution that will ensure that the investor will do best in the event of a worst-case scenario.

As Table 6 shows, switching from the static strategy to the $HM^d$ improves expected returns at the cost of higher volatility. The Sharpe ratio improves as one considers volatility timing and then again as one considers distribution timing. Comparison with Table 3 reveals that the conservative investor accepts a decrease in expected return 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 is uncertain of her parameter estimates.

Table 7 documents that the economic value of volatility timing under aversion to model uncertainty amounts to 27 bp, whereas the economic value of distribution timing is 111 bp (138.5—27.1). These estimates suggest that even when the strategies

\footnote{Several recent papers also demonstrate the importance of 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.}

\footnote{In the Bayesian allocations performed so far, we computed, for each draw of $\theta$, the allocations $\alpha_t$ for each date $t$, which we then averaged across simulations. Hence we also had the posterior distribution of $\alpha_t$ and it was possible to estimate its dispersion. In the case of aversion to model uncertainty, there is only one optimal allocation among all of the draws. As a consequence, although we are still able to evaluate the performance fee for switching from one strategy to another, it is no longer possible to report statistics on its distribution.}
are constrained to be more conservative in order to take worst-case scenarios into account, the gain of distribution timing is still much larger than that of volatility timing. As the breakeven transaction costs show, this performance gain can 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, the choice of conservative allocations to avoid risk due to erroneous parameter estimates appears to provide additional utility to the investor.

3.6 Robustness Analysis

As already discussed, we did our best to control statistical issues. We accounted for overfitting by using two separate subperiods for the parameter estimation and asset allocation, and we accounted for parameter estimation risk by computing the finite-sample distribution of all of our performance measures. To further evaluate the robustness of the gain due to distribution timing, we have performed an additional set of analyses, the main results of which are reported in this section. A complete description of the additional work can be found in a Technical Appendix.

In a first experiment, we used the same dataset on international market returns, but considered a change in the estimation and allocation periods. We re-estimated the model between 1977 and 1998 and performed the allocation over the longer period from 1999 to June 2007, which includes both bearish and bullish markets. As expected, the performances of the various strategies are markedly different from those obtained over the 2002-June 2007 allocation period. Annualized returns are much lower, while annualized volatilities are much higher, resulting in negative Sharpe ratios for the static and $MV^d$ strategies. When the dynamic strategies are compared to the static one, the $HM^d$ strategy is found to perform 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 for this 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. Once again the gain due to distribution timing is sizeable (around 200 bp) and larger than over the 2002-June 2007 period.

In a second experiment, we use data already considered in the literature. We used stock, bond, and gold returns, as already studied by Fleming, Kirby, and Ostdiek (2001, 2003) in their evaluation of volatility timing. Again, we find that

---

27The data consist of weekly returns for the US stock and bond markets and for gold. The estimation sample is from January 1983 to May 2000 (909 weekly observations) and the allocation
the $MV^d$ and $HM^d$ strategies perform much better than the static strategy. The Sharpe ratios of dynamic strategies are about twice that of the static strategy. More importantly, the modified Sharpe ratio and 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 extremely high, mainly because the exposure to risky markets has been reduced during turbulent periods. Although it is smaller in magnitude, the value of distribution timing is sizeable and significant, around 100 bp per year.

4 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 remain constant through time, several recent papers have explored the consequences of ignoring the time variability of some aspects of the distribution of returns: 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. The present study contributes to this literature with several insights. 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 when having to account for many assets.

We show that, for all levels of risk aversion, the performance fee an investor is willing to pay to benefit from distribution timing is of the same order of magnitude as the performance fee to pay to benefit from volatility timing. We perform several alternative experiments designed to assess the robustness of our findings. We cast the model in a Bayesian setting, which 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, with all of our findings suggesting 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 sample is from June 2000 to February 2006 (300 observations). As for international markets, the investor is allowed to short her position since futures are available for these assets.
rather demanding, in light of the way that multi-period investment problems are usually solved. From this point of view, it may be more convenient to express the problem in continuous time.

Another natural extension would be to analyze the effect of skewed and fat-tailed assets in an equilibrium setting. Such an analysis has recently been provided by Mitton and Vorkink (2007) for skewed assets: they have shown that preference for skewness results in portfolio under-diversification. As we have seen in Section 3.3, increasing the weight of kurtosis in the expected utility has an ambiguous effect on 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

The DCC Model. The dynamics of the returns vector \( r_t \) is

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

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, and is 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 \( D_t = \{\sigma_{i,i,t}\}_{i=1,\ldots,n} \) is a diagonal matrix with 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)

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

where all parameters are positive. Equivalently, we have

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

where \( \gamma_i \) denotes the variance persistence. The constraint \( \gamma_i < 1 \) guarantees the 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)

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

where \( u_t = D_t^{-1} \varepsilon_t \) denotes the vector of normalized unexpected returns, and \( \text{diag}(Q_t) \) denotes the \( n \times n \) matrix whose entries are zeros, except for the diagonal that contains the diagonal of \( Q_t \). The matrix \( \tilde{Q} \) 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 are drawn from \( n \) independent Sk-t distributions. As shown in equation (??), the 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

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

24
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:

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

Shape parameters \( \nu_i \) and \( \xi_i \) correspond to the individual degree of freedom and the asymmetry parameter respectively.\(^{28}\) The marginal distribution of \( z_{i,t} \) is a 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 \( t(\cdot | \nu_1) \). If the \( r \)-th moment of the \( t(\cdot | \nu_1) \) 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

\[
M_{i,r} = m_{i,r} - \frac{\xi_i^{r+1}}{\xi_i + 1},
\]

where \( m_{i,r} = 2E[Z_i^r | Z_i > 0] = \frac{\Gamma \left( \frac{\nu_i - r}{2} \right) \Gamma \left( \frac{r + 1}{2} \right) (\nu_i - 2)^{\frac{r+1}{2}}}{\sqrt{\pi}\Gamma \left( \frac{\nu_i}{2} \right)} \) is the \( r \)-th moment of \( t(\cdot | \nu_i) \) truncated to the positive real values. Provided that they exist, the third and fourth central moments of \( z_{i,t} \) are

\[
\mu_i^{(3)} = E[Z_i^3] = M_{i,3} - 3M_{i,1}M_{i,2} + 2M_{i,1}^3, \quad \text{(21)}
\]

\[
\mu_i^{(4)} = E[Z_i^4] = M_{i,4} - 4M_{i,1}M_{i,3} + 6M_{i,2}M_{i,1}^2 - 3M_{i,1}^4. \quad \text{(22)}
\]

The skewness and kurtosis are therefore non-linear functions of the degree of freedom, \( \nu_1 \), and the asymmetry parameter, \( \xi_i \).

**The 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 \( t \) in order for the distribution to be well defined. We choose the specification

\[
(1 - c_{i,2}L) \log (\nu_{i,t} - \nu) = c_{i,0} + c_{i,1} |z_{i,t-1}| N_{i,t-1} + c_{i,1}^+ |z_{i,t-1}| (1 - N_{i,t-1}), \quad \text{(23)}
\]

\[
(1 - d_{i,2}L) \log (\xi_{i,t}) = d_{i,0} + d_{i,1} z_{i,t-1} N_{i,t-1} + d_{i,1}^+ z_{i,t-1} (1 - N_{i,t-1}), \quad \text{(24)}
\]

\(^{28}\)Parameters \( a_i \) and \( b_i \) are required in order 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.\(^{29}\) 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 the Moments of Portfolio Returns**

Analytical expressions for the portfolio conditional moments can be easily obtained for a multivariate Sk-\( t \) distribution. The third and fourth central moments of a Sk-\( t \) distributed random variable are given by equations \((3)\) and \((4)\). Next, since unexpected returns are defined as \( \varepsilon_{t+1} = \sum_{r=1}^{1/2} \omega_{r,t+1} z_{r,t+1} \), we have \( E_t[\varepsilon_{t+1}] = 0 \) and \( V_t[\varepsilon_{t+1}] = \Sigma_{t+1} \). We denote \( \Sigma_{t+1/2} = (\omega_{i,j,t+1})_{i,j=1,...,n} \) the Choleski decomposition of the covariance matrix of returns, so that \( r_{i,t+1} = \mu_i + \sum_{r=1}^{n} \omega_{ir,t+1} z_{r,t+1} \). In addition, denoting by \( \otimes \) the Kronecker product, the \( n \times n^2 \) third central co-moment matrix is defined as\(^{30}\)

\[
S_{t+1} = E_t \left[ (r_{t+1} - \mu) (r_{t+1} - \mu) \otimes (r_{t+1} - \mu) \right] = \{s_{ijk,t+1}\},
\]

with component \((i,j,k)\)

\[
s_{ijk,t+1} = \sum_{r=1}^{n} \omega_{ir,t+1} \omega_{jr,t+1} \omega_{kr,t+1} \mu_{r,t+1}^{(3)},
\]

and the \( n \times n^3 \) fourth central co-moment matrix is defined as

\[
K_{t+1} = E_t \left[ (r_{t+1} - \mu) (r_{t+1} - \mu) \otimes (r_{t+1} - \mu) \otimes (r_{t+1} - \mu) \right] = \{k_{ijkl,t+1}\},
\]

with component \((i,j,k,l)\)

\[
k_{ijkl,t+1} = \sum_{r=1}^{n} \omega_{ir,t+1} \omega_{jr,t+1} \omega_{kr,t+1} \omega_{lr,t+1} \mu_{r,t+1}^{(4)} + \sum_{r=1}^{n} \sum_{s \neq r} \psi_{rs,t+1},
\]

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

The last step consists of the computation of portfolio moments. For a given portfolio weight vector \( \alpha_t \), the conditional expected return, the conditional variance,
and the conditional third and fourth moments of the portfolio return are defined as:

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

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.\(^{31}\)

The relations between the central and non-central moments, which are required in the evaluation of the Taylor approximation of the expected utility, are:

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

\(^{31}\)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 the standardized central moments, \( E_t \left[ (r_{p,t+1} - m_{p,t+1}) / \sigma_{p,t+1} \right]^i \), for \( i = 3, 4. \)
References


Captions

Table 1: This table reports summary statistics on international market returns for the estimation period (Panel A) and the allocation period (Panel B): the average, the standard deviation, the skewness, the kurtosis, the Jarque-Bera normality test statistic (JB), the Ljung-Box test statistic for no serial correlation [LB(4)], the Lee-King test statistic for no serial correlation in squared returns [LK(4)], the first-order serial correlation of returns \([\rho(r)]\) and of squared returns \([\rho(r^2)]\), and the correlation matrix. The exponent \(a\) indicates that a statistic is significant at the 1% level.

Table 2: This table reports Bayesian parameter estimates of the model with a joint Sk-t distribution with time-varying shape parameters. The first two columns contain the mean and standard deviation of the posterior distribution of the parameter estimates. The last columns contain the 5%, median, and 95% quantiles of the distribution.

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, obtained with 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, \(mSR\), is defined by equation (??). The performance fee, \(\vartheta\), is estimated from the sample counterpart of the relation (??). 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 in the text. For each level of risk aversion, the first row corresponds to the median statistics and 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 \(\gamma_{Va}\) corresponds to aversion towards volatility. \(\gamma_{Sk}\) and \(\gamma_{Ku}\) are the preference parameters for skewness and kurtosis respectively. The statistic \(D_2\), defined as the sum of squared portfolio weights, measures the degree of diversification of a portfolio. Other measures are as in Table 4. For a general level of risk aversion of 5, a CRRA Taylor expansion yields \(\gamma_{Va} = 5\), \(\gamma_{Sk} = 30\) and \(\gamma_{Ku} = 210\). For a risk aversion of 10, these values become \(\gamma_{Va} = 10\), \(\gamma_{Sk} = 110\) and \(\gamma_{Ku} = 1320\).

Table 6: This table reports summary statistics on the optimal portfolio return in the case of aversion to model uncertainty. The level of risk aversion is \(\gamma = 5\). The statistics are the same as in Table 3.
Table 7: This table reports statistics on the performance of the optimal portfolios in the case of aversion to model uncertainty. The level of risk aversion is $\gamma = 5$. The statistics are the same as in Table 4.

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 January 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 January 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 January 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 January 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 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 one of the dynamic strategies. 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>Mean</td>
<td>0.215</td>
<td>0.179</td>
<td>0.251</td>
</tr>
<tr>
<td>Std dev.</td>
<td>2.099</td>
<td>2.985</td>
<td>2.505</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.249</td>
<td>0.226</td>
<td>0.049</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.257</td>
<td>4.956</td>
<td>4.133</td>
</tr>
<tr>
<td>JB</td>
<td>289.259</td>
<td>69.997</td>
<td>218.221</td>
</tr>
<tr>
<td>LB(4)</td>
<td>63.630</td>
<td>60.400</td>
<td>156.831</td>
</tr>
<tr>
<td>LK(4)</td>
<td>6.222</td>
<td>6.273</td>
<td>9.908</td>
</tr>
<tr>
<td>$\rho(r)$</td>
<td>-0.092</td>
<td>-0.032</td>
<td>-0.004</td>
</tr>
<tr>
<td>$\rho(r^2)$</td>
<td>0.180</td>
<td>0.234</td>
<td>0.116</td>
</tr>
<tr>
<td>Correlation matrix</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.218</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UK</td>
<td>0.397</td>
<td>0.309</td>
<td>-</td>
</tr>
<tr>
<td><strong>Panel B: Allocation period (2002-June 2007)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.118</td>
<td>0.231</td>
<td>0.223</td>
</tr>
<tr>
<td>Std dev.</td>
<td>1.884</td>
<td>2.912</td>
<td>1.964</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.257</td>
<td>0.194</td>
<td>-0.306</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.116</td>
<td>4.415</td>
<td>3.510</td>
</tr>
<tr>
<td>JB</td>
<td>289.259</td>
<td>218.221</td>
<td>69.997</td>
</tr>
<tr>
<td>LB(4)</td>
<td>63.630</td>
<td>156.831</td>
<td>60.400</td>
</tr>
<tr>
<td>LK(4)</td>
<td>6.222</td>
<td>9.908</td>
<td>6.273</td>
</tr>
<tr>
<td>$\rho(r)$</td>
<td>0.049</td>
<td>-0.014</td>
<td>-0.054</td>
</tr>
<tr>
<td>$\rho(r^2)$</td>
<td>0.226</td>
<td>0.051</td>
<td>-0.047</td>
</tr>
<tr>
<td>Correlation matrix</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.385</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UK</td>
<td>0.722</td>
<td>0.414</td>
<td>-</td>
</tr>
</tbody>
</table>


Table 2: Estimation of the model with Sk-t distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std dev.</th>
<th>5%</th>
<th>median</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expected return</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.217</td>
<td>0.009</td>
<td>0.203</td>
<td>0.217</td>
<td>0.231</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.239</td>
<td>0.010</td>
<td>0.222</td>
<td>0.239</td>
<td>0.255</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>0.236</td>
<td>0.009</td>
<td>0.220</td>
<td>0.236</td>
<td>0.251</td>
</tr>
<tr>
<td><strong>Volatility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.161</td>
<td>0.044</td>
<td>0.092</td>
<td>0.159</td>
<td>0.232</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.068</td>
<td>0.025</td>
<td>0.032</td>
<td>0.066</td>
<td>0.113</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>0.078</td>
<td>0.022</td>
<td>0.046</td>
<td>0.076</td>
<td>0.117</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.877</td>
<td>0.028</td>
<td>0.829</td>
<td>0.879</td>
<td>0.918</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.235</td>
<td>0.039</td>
<td>0.178</td>
<td>0.235</td>
<td>0.300</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.045</td>
<td>0.013</td>
<td>0.026</td>
<td>0.043</td>
<td>0.070</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>0.095</td>
<td>0.016</td>
<td>0.071</td>
<td>0.092</td>
<td>0.124</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.866</td>
<td>0.020</td>
<td>0.832</td>
<td>0.867</td>
<td>0.896</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.206</td>
<td>0.074</td>
<td>0.100</td>
<td>0.202</td>
<td>0.329</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.071</td>
<td>0.013</td>
<td>0.052</td>
<td>0.070</td>
<td>0.094</td>
</tr>
<tr>
<td>$\psi_3$</td>
<td>0.026</td>
<td>0.012</td>
<td>0.008</td>
<td>0.025</td>
<td>0.049</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.886</td>
<td>0.021</td>
<td>0.851</td>
<td>0.887</td>
<td>0.920</td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.007</td>
<td>0.002</td>
<td>0.003</td>
<td>0.007</td>
<td>0.011</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.978</td>
<td>0.008</td>
<td>0.965</td>
<td>0.978</td>
<td>0.992</td>
</tr>
<tr>
<td><strong>Degree of freedom</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_{1,0}$</td>
<td>0.031</td>
<td>0.006</td>
<td>0.023</td>
<td>0.031</td>
<td>0.042</td>
</tr>
<tr>
<td>$c_{1,1}$</td>
<td>2.635</td>
<td>0.343</td>
<td>2.103</td>
<td>2.607</td>
<td>3.133</td>
</tr>
<tr>
<td>$c_{1,1}$</td>
<td>0.054</td>
<td>0.280</td>
<td>-0.576</td>
<td>-0.021</td>
<td>0.358</td>
</tr>
<tr>
<td>$c_{1,2}$</td>
<td>-0.107</td>
<td>0.167</td>
<td>-0.359</td>
<td>-0.117</td>
<td>0.182</td>
</tr>
<tr>
<td>$c_{2,0}$</td>
<td>0.031</td>
<td>0.006</td>
<td>0.022</td>
<td>0.031</td>
<td>0.041</td>
</tr>
<tr>
<td>$c_{2,1}$</td>
<td>11.904</td>
<td>0.172</td>
<td>11.638</td>
<td>11.894</td>
<td>12.189</td>
</tr>
<tr>
<td>$c_{2,1}$</td>
<td>1.518</td>
<td>0.142</td>
<td>1.266</td>
<td>1.535</td>
<td>1.714</td>
</tr>
<tr>
<td>$c_{2,2}$</td>
<td>-0.368</td>
<td>0.107</td>
<td>-0.544</td>
<td>-0.366</td>
<td>-0.200</td>
</tr>
<tr>
<td>$c_{3,0}$</td>
<td>0.031</td>
<td>0.006</td>
<td>0.022</td>
<td>0.030</td>
<td>0.042</td>
</tr>
<tr>
<td>$c_{3,1}$</td>
<td>0.778</td>
<td>0.640</td>
<td>0.053</td>
<td>0.517</td>
<td>2.135</td>
</tr>
<tr>
<td>$c_{3,1}$</td>
<td>6.304</td>
<td>0.755</td>
<td>5.246</td>
<td>6.058</td>
<td>7.642</td>
</tr>
<tr>
<td>$c_{3,2}$</td>
<td>-0.104</td>
<td>0.126</td>
<td>-0.319</td>
<td>-0.101</td>
<td>0.098</td>
</tr>
<tr>
<td><strong>Asymmetry parameter</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{1,0}$</td>
<td>1.005</td>
<td>0.065</td>
<td>0.889</td>
<td>1.010</td>
<td>1.101</td>
</tr>
<tr>
<td>$d_{1,1}$</td>
<td>1.916</td>
<td>0.639</td>
<td>1.029</td>
<td>1.762</td>
<td>2.809</td>
</tr>
<tr>
<td>$d_{1,1}$</td>
<td>-1.726</td>
<td>0.466</td>
<td>-2.365</td>
<td>-1.868</td>
<td>-0.806</td>
</tr>
<tr>
<td>$d_{1,2}$</td>
<td>-0.170</td>
<td>0.253</td>
<td>-0.665</td>
<td>-0.097</td>
<td>0.136</td>
</tr>
<tr>
<td>$d_{2,0}$</td>
<td>1.097</td>
<td>0.039</td>
<td>1.036</td>
<td>1.095</td>
<td>1.166</td>
</tr>
<tr>
<td>$d_{2,1}$</td>
<td>1.124</td>
<td>0.238</td>
<td>0.732</td>
<td>1.196</td>
<td>1.431</td>
</tr>
<tr>
<td>$d_{2,1}$</td>
<td>0.492</td>
<td>0.267</td>
<td>0.018</td>
<td>0.508</td>
<td>0.949</td>
</tr>
<tr>
<td>$d_{2,2}$</td>
<td>0.159</td>
<td>0.089</td>
<td>0.002</td>
<td>0.170</td>
<td>0.288</td>
</tr>
<tr>
<td>$d_{3,0}$</td>
<td>0.873</td>
<td>0.036</td>
<td>0.815</td>
<td>0.872</td>
<td>0.931</td>
</tr>
<tr>
<td>$d_{3,1}$</td>
<td>-1.056</td>
<td>0.250</td>
<td>-1.462</td>
<td>-1.049</td>
<td>-0.671</td>
</tr>
<tr>
<td>$d_{3,1}$</td>
<td>2.751</td>
<td>0.160</td>
<td>2.466</td>
<td>2.757</td>
<td>2.989</td>
</tr>
<tr>
<td>$d_{3,2}$</td>
<td>-0.057</td>
<td>0.128</td>
<td>-0.250</td>
<td>-0.064</td>
<td>0.169</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>18.096</td>
<td>33.936</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>17.870</td>
<td>33.905</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>24.618</td>
<td>38.076</td>
</tr>
<tr>
<td>$\gamma = 5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>8.816</td>
<td>13.561</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>8.733</td>
<td>13.545</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>11.482</td>
<td>15.322</td>
</tr>
<tr>
<td>$\gamma = 10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>5.723</td>
<td>6.770</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>5.687</td>
<td>6.761</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>7.080</td>
<td>7.676</td>
</tr>
<tr>
<td>$\gamma = 15$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>4.692</td>
<td>4.508</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>4.669</td>
<td>4.500</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>5.602</td>
<td>5.121</td>
</tr>
<tr>
<td>Strategy</td>
<td>mSR $(bp)$</td>
<td>Performance fee $\vartheta$ $(bp)$</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>$\gamma = 2$</td>
<td>$MV^d$</td>
<td>$-19.4$</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>$403.7$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[238.1; 579.2]$</td>
</tr>
<tr>
<td>$\gamma = 5$</td>
<td>$MV^d$</td>
<td>$-7.6$</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>$161.0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[93.4; 232.6]$</td>
</tr>
<tr>
<td>$\gamma = 10$</td>
<td>$MV^d$</td>
<td>$-3.2$</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>$81.0$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[47.8; 117.2]$</td>
</tr>
<tr>
<td>$\gamma = 15$</td>
<td>$MV^d$</td>
<td>$-1.8$</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>$54.4$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[32.1; 78.2]$</td>
</tr>
</tbody>
</table>
Table 5: Direct preferences for skewness and kurtosis

<table>
<thead>
<tr>
<th>Utility #</th>
<th>$\gamma_{Va}$</th>
<th>$\gamma_{Sk}$</th>
<th>$\gamma_{Ku}$</th>
<th>Sharpe ratio</th>
<th>$mSR$ (bp)</th>
<th>Performance fee $\vartheta$ (bp)</th>
<th>$D_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0.449</td>
<td>−7.6</td>
<td>−4.7</td>
<td>0.550</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>30</td>
<td>0</td>
<td>0.576</td>
<td>164.5</td>
<td>96.9</td>
<td>0.823</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>210</td>
<td>0.572</td>
<td>159.7</td>
<td>112.5</td>
<td>0.756</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>30</td>
<td>210</td>
<td>0.573</td>
<td>161.0</td>
<td>102.1</td>
<td>0.792</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>110</td>
<td>210</td>
<td>0.586</td>
<td>178.9</td>
<td>74.7</td>
<td>0.915</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>30</td>
<td>1320</td>
<td>0.575</td>
<td>163.4</td>
<td>130.4</td>
<td>0.682</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>110</td>
<td>1320</td>
<td>0.582</td>
<td>173.4</td>
<td>109.7</td>
<td>0.758</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>30</td>
<td>210</td>
<td>0.575</td>
<td>82.6</td>
<td>59.5</td>
<td>0.323</td>
</tr>
</tbody>
</table>
Table 6: Moments of realized portfolio return over the allocation period  
(Model with aversion to model uncertainty)

<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>Static</td>
<td>8.007</td>
<td>11.839</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>8.460</td>
<td>12.079</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>10.813</td>
<td>13.635</td>
</tr>
</tbody>
</table>
Table 7: Measures of portfolio performance over the allocation period
(Model with aversion to model uncertainty)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>$mSR$ (bp)</th>
<th>Performance fee $\vartheta$ (bp)</th>
<th>$\mathcal{Z}$ (%)</th>
<th>$\tau^b_c$ (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MV^d$</td>
<td>0.338</td>
<td>27.1</td>
<td>0.505</td>
<td>3.345</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>1.732</td>
<td>138.5</td>
<td>0.560</td>
<td>28.220</td>
</tr>
</tbody>
</table>
Figure 1: Evolution of the conditional volatilities

<table>
<thead>
<tr>
<th>Year</th>
<th>US</th>
<th>Japan</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1980</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1985</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1990</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>1995</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2010</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2: Evolution of the conditional correlations

<table>
<thead>
<tr>
<th>Year</th>
<th>US-Japan</th>
<th>US-UK</th>
<th>Japan-UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1980</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>1985</td>
<td>0.4</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>1990</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>1995</td>
<td>0.8</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td>2000</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>2005</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>2010</td>
<td>0.6</td>
<td>0.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>
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 the performance fee for $MV^d$ and $HM^d$ strategies