

On the Importance of Time Variability in Higher Moments for Asset Allocation

ERIC JONDEAU

Swiss Finance Institute and University of Lausanne

MICHAEL ROCKINGER

Swiss Finance Institute and University of Lausanne

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*, defined here as the ability to use forecasts for moments up to the fourth one, yields significant incremental economic value. By considering the weekly asset allocation among the five largest international stock markets, we find that distribution timing yields a gain of around 140 basis points per year over the last decade. To control for the parameter uncertainty of the model, we cast the model into a Bayesian setting. We also consider alternative preference structures and model specifications. In all cases, the value of distribution timing remains economically significant. (JEL: G11, F37, C22, C51)

KEYWORDS: Bayesian estimation, distribution timing, GARCH model, nonnormality, parameter uncertainty, Portfolio allocation, volatility timing

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, Steward Hodges, Eric Jacquier, Nour Meddahi, and Raman Uppal, and the participants at the European Finance Association meeting in Ljubljana (2007). We are grateful to two referees for their detailed comments. We remain responsible for any remaining errors. This work was supported by the Swiss National Science Foundation through NCCR FINRISK (Financial Valuation and Risk Management). Address correspondence to Michael Rockinger, Faculty of Business and Economics, University of Lausanne, CH 1015 Lausanne, Switzerland, or e-mail: michael.rockinger@unil.ch.

doi: 10.1093/jfifnec/nbr006

© The Author 2011. Published by Oxford University Press. All rights reserved.

For permissions, please e-mail: journals.permissions@oup.com.

A rational investor, if asked to choose between two assets with the same mean and variance, is likely to invest in the asset with the highest skewness and the lowest kurtosis (Scott and Horvath 1980; Dittmar 2002). 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 (2008) show that idiosyncratic skewness also matters in the cross section.

However, early empirical evidence on asset allocation in a static setting reveals that 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 provided evidence that indeed 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 (2009) have reported empirical evidence that the mean–variance criterion may fail to approximate the constant relative risk aversion 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, whereas recent work demonstrates that higher moments may vary through time and may be at least partially predictable (Hansen 1994; Harvey and Siddique 1999; Jondeau and Rockinger 2003; Patton 2004). This result suggests that the role of higher moments found in the cross section come from the ability of the investor to predict the evolution of the higher moments. Evaluating the effect of dynamic higher moments on asset allocation requires the design of a data-generating process that accounts for time variability in the higher moments. The first attempts toward this goal were made by Ang and Bekaert (2002), Guidolin and Timmermann (2007, 2008), and Guidolin and Nicodano (2009), using a switching-regime approach. In this framework, returns' mean and variance change depending on the regime. Although the dynamics of higher moments are not explicitly modeled, (Guidolin and Timmermann 2008) as well as Tu (2010) show that taking into account such changes in regime improves the performance of the asset allocation. Despite all 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 was to establish the economic value of *distribution timing*, that is, the investor's ability to forecast the subsequent characteristics of the 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; Barberis 2000, among many others), research on volatility timing has appeared 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 daily to monthly investment horizons (e.g., Fleming, Kirby, and Ostdiek 2001, 2003; Marquering and Verbeek 2004; 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. We decided to name allocations that use information on up to the fourth moment as distribution timing to avoid possible confusion with higher-moment timing that could mean timing of skewness and kurtosis only.¹ For this purpose, we consider two strategies: first, a dynamic mean–variance strategy, in which investors try to benefit from their ability to predict subsequent volatility; then, 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. This 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 with asymmetry and fat tails. This extension incorporates several statistical features that characterize the dynamics of asset returns. This econometric model has been extensively analyzed in Jondeau and Rockinger (2009).² 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

¹We recognize the limitations of this naming convention since, as shown by Cenesizoglu and Timmermann (2008), predicting moments is not exactly the same thing as predicting an entire distribution.

²While we adopt the same econometric model for the dynamic and the joint distribution of the asset returns, we clearly depart from that paper by describing how to use this model in a portfolio allocation context. We also investigate a different issue, that is, whether or not an investor can benefit from her ability to forecast the subsequent joint distribution of asset returns instead of investigating the predictability of higher moments. Finally, we show that this approach can be used in different contexts, at different frequencies, and with more assets.

the expected utility. Within this framework, we show that time variability in higher moments does matter for effective portfolio allocation.³

One major problem in measuring the economic value of distribution timing is the parameter uncertainty encountered in estimating the model. We therefore 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 highly 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, we perform resampling asset allocation à la Michaud (1998), which allows us to take care of the parameter uncertainty as well as robust asset allocation, which aims at maximizing the minimum expected utility. For all of these reasons, the use of Bayesian analysis appears relevant in the context of distribution timing.

We applied our approach to the weekly allocation of wealth among the five largest stock markets, that is, those of the United States, Japan, the UK, Germany, and France, which represent, at the end of 1999, 60% of world market capitalization. We first demonstrate that our model captures the main statistical characteristics of these market returns. Then, 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 an investor would be willing to pay to benefit from the higher-moment dynamic strategy (distribution timing) is similar in magnitude to the fee she would be willing to pay to benefit from the mean–variance dynamic strategy (volatility timing). For common levels of risk aversion, the economic value of distribution timing is about 140 basis points (bp) per year, while the economic value of volatility timing is about 300 bp per year. When we investigate alternative preference structures, sample periods, or model specifications, we find that distribution timing remains sizeable and comparable to volatility timing.

The outline of this paper is as follows. In Section 1, we formulate our approach for modeling returns with a nonnormal multivariate distribution and for measuring distribution timing. Section 2 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. Then, we measure the economic value of these strategies, under alternative preference structures, and we provide some robustness checks of our main results. Section 3 concludes the paper.⁴

³Several authors have proposed portfolio criteria based on an extension of the mean–variance criterion. Examples are the higher-dimensional efficient frontier (Athayde and Flóres (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 (2008)), ambiguity aversion (Ait-Sahalia and Brandt (2001)), or an approximation of a general utility function based on higher moments (Jondeau and Rockinger (2006); Guidolin and Timmermann (2008); Harvey et al. (2010)).

⁴The appendix contains the details of the statistical model, describing the evolution of returns.

1 METHODOLOGY

This section describes our methodology for solving the conditional asset allocation problem with nonnormal returns. We first present the data-generating process 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 \(2009\)](#). We then describe how including higher-moments forecasts in the expected utility of investors is likely to improve the allocation of wealth. As in [Jondeau and Rockinger \(2006\)](#) and [Guidolin and Timmermann \(2008\)](#), 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.

1.1 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 returns, r_t , in excess of the risk-free rate, $r_{f,t-1}$. The return's dynamic is written as follows:

$$\tilde{r}_t = r_t - r_{f,t-1} = \mu_t + \varepsilon_t, \quad (1)$$

$$\varepsilon_t = \Sigma_t^{1/2} z_t, \quad (2)$$

$$z_t \sim g(z_t | \eta). \quad (3)$$

Equation (1) decomposes the excess return at time t , \tilde{r}_t , into two $n \times 1$ vectors, the expected excess returns, μ_t , and the unexpected excess returns, ε_t . Equation (2) describes the unexpected returns ε_t , where z_t denotes the $n \times 1$ vector of independent innovations, with zero mean and an identity covariance matrix, and $\Sigma_t^{1/2}$ denotes the Choleski decomposition of the conditional covariance matrix of returns, $\Sigma_t = E_{t-1}[(\tilde{r}_t - \mu_t)(\tilde{r}_t - \mu_t)']$. Equation (3) specifies that z_t follows a conditional distribution $g(\cdot)$ with shape parameters η .

The specification we adopt for Equations (1)–(3) is described in detail in [Jondeau and Rockinger \(2009\)](#) and summarized in Appendix A. The main characteristics of this model are the following. First, we follow [Fleming, Kirby, and Ostdiek \(2001, 2003\)](#) and assume constant expected returns ($\mu_t = \mu, \forall t$). Some recent papers have shown that macroeconomic variables can have predictive power for monthly returns ([Kandel and Stambaugh 1996](#); [Campbell and Thompson 2008](#); [Chochrane 2008](#)) in particular if forecasts get pooled ([Rapach, Strauss, and Zhou 2010](#)). Predictability is, however, a controversial topic (see [Cremers 2002](#); [Goyal and Welch 2008](#)). Because our focus is on the evaluation of volatility and distribution timing, we mainly focus on the case with constant expected returns.⁵ Second, the covariance matrix Σ_t is described as a dynamic conditional correlation model

(Engle 2002, and Engle and Sheppard 2001), with asymmetric GARCH dynamics for the conditional variances. This specification is designed to account for the well-known properties of volatility clustering and dynamic correlations. Third, innovations are drawn from the following multivariate Sk- t distribution (Sahu, Dey, and Branco 2003):

$$g(z_t|\eta) = \prod_{i=1}^n \frac{2b_i}{\xi_i + \frac{1}{\xi_i}} \frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\pi}(\nu_i-2)\Gamma\left(\frac{\nu_i}{2}\right)} \left(1 + \frac{\kappa_{i,t}^2}{\nu_i-2}\right)^{-\frac{\nu_i+1}{2}}, \tag{4}$$

where $\eta = (\nu_1, \dots, \nu_n, \xi_1, \dots, \xi_n)'$ denotes the vector of shape parameters and

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

with

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

The location and dispersion parameters a_i and b_i , respectively, ensure that $z_{i,t}$ has a zero mean and unit variance. The Sk- t distribution is able to capture both the asymmetry and the fat-tailedness through two shape parameters for each asset i : the degree of freedom, ν_i , and the asymmetry parameter, ξ_i . Finally, the shape parameters are rendered time varying, as described in the following specification, proposed in Jondeau and Rockinger (2009):⁶

$$(1 - c_{i,2}L) \log(\nu_{i,t} - \underline{\nu}) = \log c_{i,0} + c_{i,1}^- |z_{i,t-1}| N_{i,t-1} + c_{i,1}^+ |z_{i,t-1}| (1 - N_{i,t-1}) \tag{5}$$

$$(1 - d_{i,2}L) \log(\xi_{i,t}) = \log d_{i,0} + d_{i,1}^- z_{i,t-1} N_{i,t-1} + d_{i,1}^+ z_{i,t-1} (1 - N_{i,t-1}), \tag{6}$$

where $N_{i,t} = 1_{\{z_{i,t} < 0\}}$ and L is the lag operator. The degree of freedom and the asymmetry parameter are driven by past shocks, as in an exponential GARCH process, allowing some asymmetry in the reaction of the higher moments to negative and positive shocks. The parameter $\underline{\nu}$ is the lower bound for the degree of freedom.⁷ Two main features of equations (5) and (6) are worth emphasizing. First, $\nu_{i,t}$ is related to the absolute value of lagged standardized innovations because $z_{i,t-1}$ 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

⁵In the robustness section, we show that the distribution timing still holds in a GARCH-in-mean setup, when conditional volatility drives expected returns.

⁶For 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 may be found in Patton (2004) or Mencia and Sentana (2005).

⁷We impose $\underline{\nu} = 4$ to make sure that the conditional moments up to the fourth order are well defined.

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.

This model has several appealing properties. First, it nests the standard normal and t distributions with constant or time-varying degree of freedom. Second, it allows us to analytically compute the moments of a portfolio composed of assets driven by this distribution. This key insight, combined with a Taylor approximation of the expected utility, allows us to perform allocation in a very efficient manner. Third, the model is able to capture most of the features usually observed in actual financial returns. These features include the time variability in the covariance matrix, and the presence of asymmetry and fat tails in the distribution. This model therefore provides an appropriate setting to empirically investigate the magnitude of the volatility and distribution timing.

The model used to describe the evolution of returns involves a large number of parameters in a nonlinear manner. Given this complexity, the convergence of the maximum likelihood (ML) 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, which allows us to take account for estimation risk. This setting also allows us to demonstrate that the reported economic values are not due to chance but are robust to parameter uncertainty. More details on Bayesian estimation can be found in Appendix B.

It is worth emphasizing that there are different ways to incorporate parameter uncertainty into the allocation problem in a Bayesian setup. In the first approach, which we adopt, *Bayesian estimation* is used to produce the multivariate posterior distribution of the parameter set. Then, we use resampling from this posterior distribution to estimate the optimal weights for all the dates of the sample, from which we evaluate the economic significance of the portfolio weights and performance measures reported in the paper. This approach is similar in principle to [Michaud \(1998\)](#), although we use the posterior distribution of the parameters instead of their asymptotic distribution. As a by-product, this posterior distribution also provides a solution to the allocation problem for an investor with aversion to parameter uncertainty.

Alternatively, one could incorporate parameter uncertainty directly into the portfolio optimization problem. This *Bayesian allocation* approach has been followed, for instance, by [Barberis \(2000\)](#), [Polson and Tew \(2000\)](#), [Johannes, Polson, and Stroud \(2002\)](#), [Tu and Zhou \(2004\)](#), and [Harvey et al. \(2010\)](#). The idea is to estimate the so-called predictive distribution of asset returns. This distribution incorporates the uncertainty about the parameters and yields a unique set of optimal portfolio weights. Our purpose in this paper is not to claim that one approach performs better than the other does. Our objective is to provide the empirical distribution of the performance measures of the various allocation strategies we consider.⁸

⁸The relative performance of both approaches has been studied by [Markowitz and Usmen \(2003\)](#) and [Harvey, Liechty, and Liechty \(2008\)](#).

This objective is naturally addressed using the resampling approach. A description of the estimation technique used in this paper can be found in Appendix B.

1.2 The Asset Allocation Strategies

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} .⁹ 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 $r_{t+1} = (r_{1,t+1}, \dots, r_{n,t+1})'$ and a risk-free asset with return $r_{f,t}$ from time t to time $t + 1$. Excess returns are denoted by $\tilde{r}_{t+1} = r_{t+1} - r_{f,t}$. End-of-period wealth is $W_{t+1} = 1 + r_{p,t+1}$, where $r_{p,t+1} = r_{f,t} + \alpha_t' \tilde{r}_{t+1}$ denotes the portfolio return, with $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{n,t})'$ as 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) as well as positive (lending). The investor uses our econometric model (1)–(3) to forecast the expected mean vector $\mu_{t+1} = E_t(\tilde{r}_{t+1})$, the covariance matrix $\Sigma_{t+1} = E_t[(\tilde{r}_{t+1} - \mu_{t+1})(\tilde{r}_{t+1} - \mu_{t+1})']$, and possibly the third and fourth co-moment matrices:

$$\begin{aligned} \mathcal{S}_{t+1} &= E_t[(\tilde{r}_{t+1} - \mu_{t+1})(\tilde{r}_{t+1} - \mu_{t+1})' \otimes (\tilde{r}_{t+1} - \mu_{t+1})'], \\ \mathcal{K}_{t+1} &= E_t[(\tilde{r}_{t+1} - \mu_{t+1})(\tilde{r}_{t+1} - \mu_{t+1})' \otimes (\tilde{r}_{t+1} - \mu_{t+1})' \otimes (\tilde{r}_{t+1} - \mu_{t+1})']. \end{aligned}$$

In Appendix A, we indicate how to compute the components \mathcal{S}_{t+1} and \mathcal{K}_{t+1} as well as how to compute the moments of the portfolio return distribution for a given vector of portfolio weights.

Optimal portfolio weights are obtained by maximizing the expected utility

$$\max_{\{\alpha_t\}} E_t[U(W_{t+1}(\alpha_t))] = E_t[U(1 + r_{f,t} + \alpha_t' \tilde{r}_{t+1})]. \tag{7}$$

In general, this problem does not have a closed-form solution and must be solved numerically. It can be done using quadrature rules (Balduzzi and Lynch 1999; Ang and Bekaert 2002) or using Monte Carlo integration (Detemple, Garcia, and Rindisbacher 2003; Patton 2004; Guidolin and Timmermann 2007). However, these approaches are practically intractable in a higher-dimensional context.

Because 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.

⁹We do not consider a multiperiod 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 multiperiod investment to future research.

This approach allows handling the case of many assets. The utility function can be written as an infinite-order Taylor series expansion around the wealth at date t :

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

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

$$E_t[U(W_{t+1})] = E_t\left[\sum_{k=0}^{\infty} \frac{U^{(k)}(W_t) r_{p,t+1}^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{U^{(k)}(W_t)}{k!} m_{p,t+1}^{(k)}, \quad (8)$$

with $m_{p,t+1}^{(k)} = E_t[r_{p,t+1}^k]$ denoting the noncentral moments of order k .¹⁰ Therefore, the expected utility depends on all of the moments of the distribution of the portfolio return. The investor's preference (or aversion) toward the k th moment is directly given by the k th derivative of the utility function.

Because our aim is to evaluate the gain of forecasting higher moments in addition to the variance, we consider a Taylor series expansion up to the fourth order yielding

$$\hat{E}_t[U_{[4]}(W_{t+1})] = \varphi_0 + \varphi_1 m_{p,t+1}^{(1)} + \varphi_2 m_{p,t+1}^{(2)} + \varphi_3 m_{p,t+1}^{(3)} + \varphi_4 m_{p,t+1}^{(4)}, \quad (9)$$

where $\varphi_k = \frac{1}{k!} U^{(k)}(W_t)$. Following [Jondeau and Rockinger \(2006\)](#) and [Guidolin and Timmermann \(2008\)](#), we calibrate the parameters φ_k using the power utility function $U(W_{t+1}) = W_{t+1}^{1-\gamma} / (1-\gamma)$, where $\gamma > 0$ ($\gamma \neq 1$) measures the investor's constant relative risk aversion. In this case, we obtain: $\varphi_0 = 1/(1-\gamma)$, $\varphi_1 = 1$, $\varphi_2 = -\gamma/2$, $\varphi_3 = \gamma(\gamma+1)/3!$, and $\varphi_4 = \gamma(\gamma+1)(\gamma+2)/4!$.

The effects of the third and fourth moments on the approximated expected utility are unambiguously positive and negative, respectively. This finding is consistent with the theoretical arguments developed by [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 (9) for each date t defines a dynamically rebalanced portfolio that maximizes the expected utility of the investor. This expression clearly demonstrates how forecasts of the higher moments of the portfolio return distribution will affect the optimal weights at date $t + 1$.

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

We now describe the strategies used for comparing the magnitude of distribution timing and volatility timing. In the first strategy, the investor estimates a DCC model with a joint Sk-*t* distribution, as described in Equations (1)–(6). She therefore forecasts the time-varying covariance matrix as well as the conditional distribution of asset returns. With this strategy, the investor is allowed to take full advantage of volatility and distribution timing. The allocation criterion is based on the fourth-order Taylor series expansion, $\hat{E}_t[U_{[4]}(W_{t+1})]$ given by Equation (9). This *dynamic higher-moment strategy* is denoted by HM^d.

In the second strategy, the investor still estimates a DCC model but assumes a joint normal distribution. She therefore forecasts the time-varying conditional covariance matrix. The utility function is approximated by a Taylor series expansion up to the second order, giving

$$\hat{E}_t[U_{[2]}(W_{t+1})] = \varphi_0 + \varphi_1 m_{p,t+1}^{(1)} + \varphi_2 m_{p,t+1}^{(2)}. \tag{10}$$

Its performance relative to the static strategy provides a measure of the economic value of volatility timing. This *dynamic mean-variance strategy* is denoted by MV^d.

To evaluate the distribution and volatility timing, these dynamic strategies are compared to a benchmark strategy, the *naive (mean-variance) strategy*. It corresponds to 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. Optimal portfolio weights are thus constant over time. For purposes of comparison, we follow the approach recommended by DeMiguel, Garlappi, and Uppal (2009) and consider two basic strategies, the 1/*N* and the minimum-variance strategies.

1.3 Measuring the Gains of Distribution Timing

We compare the performance of the various strategies using 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 σ_p , as $SR_p = (m_p - r_f) / \sigma_p$. Because 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), defined as

$$mSR = \frac{\sigma_0}{\sigma_p} (m_p - r_f) - (m_0 - r_f), \tag{11}$$

where m_0 and σ_0 are the average return and volatility of the naive strategy. This measure corresponds to a scaled difference in the prices of risk for the two allocations being compared.

However, these measures have an obvious drawback in our context, as they do not capture the effect of nonnormality. Therefore, 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 by ϑ , is defined as the average return that has to be subtracted from the return of the dynamic strategy, such that the investor becomes indifferent to both strategies

$$E_t[U(1 + \hat{r}_{p,t+1})] = E_t[U(1 + r_{p,t+1}^* - \vartheta)], \quad (12)$$

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 naive strategy. The performance fee is obtained by solving equation (12) numerically.¹¹

Finally, we report two widely used measures that reflect the ability of a strategy to account for non-normality features: the value-at-risk (VaR) and the expected shortfall (ES) of the portfolio return, with a threshold of 1%. These measures are expected to improve when we use the HM^d strategy because this strategy explicitly takes the tail behavior of market returns into account.¹²

There are several issues in testing the statistical significance of the gains due to distribution timing. First, while the naive 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 use two nonoverlapping subsamples for the estimation and allocation stages.¹³

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.

¹¹We 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 in order to be willing to put 1 dollar in the suboptimal strategy rather than in the optimal one. Because 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.

¹²In a previous draft of the paper, we also considered several alternative measures, such as the success rate or the break-even transaction cost used by Han (2006) among others. Because all these measures provide the same evidence in favor of the HM^d strategy, we will only report some of these measures in the following.

¹³Overfitting may arise from the introduction of too many parameters in a model. Some parameters may be significant only because they help to capture very specific episodes. They would be helpful for improving in-sample allocation, but useless (at best) for out-of-sample allocation. Data snooping occur if the same sample is used for both estimation and allocation.

2 EMPIRICAL INVESTIGATION

2.1 Data Description

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 five largest international markets (the United States, Japan, the UK, Germany, and France). The asset allocation problem is viewed from the perspective of an unhedged U.S. investor, thus, returns are expressed in U.S. dollars. Excess returns are defined over the risk-free 7-day U.S. Federal funds rate.¹⁴ The data are weekly and cover the period from January 1973 through December 2009, for a total of 1931 observations. To avoid in-sample overfitting as well as spurious findings, this sample period is broken in two subsamples: the first sample (from 1973 to 1999, 1409 observations) is used for the estimation of the model, while the second sample (from 2000 to December 2009, 522 observations) is used for the out-of-sample investigation. This dataset has been selected as the benchmark to establish our results because it covers indices corresponding to large markets. Consequently, it is less likely to be characterized by extreme behavior that may drive the results. Therefore, for more “exotic” data, we would expect even stronger findings. In Section 2.5, we report additional evidence based on alternative subsamples or model specifications. It turns out that our main results are not significantly altered.¹⁵

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, annualized average returns are all positive and significant, ranging between 13.6% and 16.8%. Annualized volatilities range between 15.3% and 20.6%. UK and France have high expected returns with high risk, whereas the United States and Germany have low expected returns and low risk. Japan has low average returns and high risk over the sample.

Skewness measures are dispersed across markets. United States and French returns are negatively skewed, suggesting that crashes occur more often than booms, while the Japanese and UK markets have a large positive skewness. Kurtosis measures are between 4.4 for Germany and 9.3 for the UK, a range that is

¹⁴The data consist of Friday-to-Friday weekly returns based on closing prices from Datastream International. At the end of 1999, the United States, Japanese, and UK markets represent 32.9%, 10.2%, 8.3%, 4.1%, and 4.3% of the world market capitalization, respectively. Market returns are measured by the weekly return on the S&P 500, Nikkei 225, FTSE 100, DAX 30, and CAC 40 indices, respectively. Returns are converted into U.S. dollars using the exchange rate of the same day. Nonsynchronicity of the markets is expected to be softened by the use of the weekly frequency. These market indices are easy to trade because they all have tradable futures. Because the transaction costs on these futures are very low, transaction costs will not be a key issue in the model.

¹⁵A technical appendix, available from the authors, contains an analysis of a portfolio containing weekly returns of stocks, bonds, and gold. Using daily data, we also analyzed a portfolio containing size portfolios sampled at daily frequency. In all these cases, distribution timing was found to add economic value.

Table 1 Summary statistics on market returns

	United States	Japan	UK	Germany	France
Panel A: Estimation period (1973–1999)					
Mean	14.321	13.585	16.196	13.936	16.756
SD	15.326	20.364	20.479	16.906	20.592
Skewness	-0.478	0.282	0.440	-0.179	-0.319
Kurtosis	5.755	5.190	9.263	4.420	4.934
JB	498.815	299.918	2346.610	125.841	243.318
LB(4)	0.920	8.471	5.442	7.660	11.733
LK(4)	7.865	9.523	10.582	8.749	9.750
$\rho(r)$	-0.016	0.050	0.052	0.062	0.105
$\rho(r^2)$	0.266	0.167	0.166	0.176	0.219
Correlation matrix					
Japan	0.241	-	0.294	0.365	0.337
UK	0.391	0.294	-	0.384	0.443
Germany	0.318	0.365	0.384	-	0.535
France	0.352	0.337	0.443	0.535	-
Panel B: Allocation period (2000–2009)					
Mean	1.160	-1.787	3.754	4.751	6.178
SD	19.130	21.493	21.312	24.420	23.803
Skewness	-0.318	0.018	-0.412	-0.554	-0.450
Kurtosis	6.849	4.550	5.715	5.180	5.363
JB	336.010	53.078	177.719	132.059	141.234
LB(4)	3.829	4.697	3.917	4.213	9.776
LK(4)	5.424	5.459	8.938	7.285	6.822
$\rho(r)$	-0.054	-0.010	-0.082	-0.060	-0.112
$\rho(r^2)$	0.200	0.146	0.372	0.209	0.265
Correlation matrix					
Japan	0.432	-	0.460	0.482	0.503
UK	0.733	0.460	-	0.833	0.878
Germany	0.734	0.482	0.833	-	0.931
France	0.732	0.503	0.878	0.931	-

This table reports summary statistics on international market returns for the estimation period (Panel A) and the allocation period (Panel B): the annualized average return and standard deviation, the standardized skewness and 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 eventually the correlation matrix. The critical values at 5% are 5.99 for JB and 9.488 for LB(4) and LK(4).

inconsistent with the assumption of normality. We reject normality with great confidence for all markets and, interestingly, these statistics demonstrate that international stock markets are characterized by very different distribution patterns. Regarding temporal dependence, we find no systematic evidence for serial

correlation in market returns, but squared returns are strongly correlated, which suggests temporal variation in second moments.

We observe some changes in the average returns over the allocation period (2000–2009). Germany and France now have high expected returns and high risk, whereas the United States and Japan have low expected returns and low risk. European returns have a highly negatively skewed distribution. The United States and UK return distributions display the heaviest tails.

Turning to the multivariate characteristics of market returns over the 1973–1999 period, we notice that the correlation is the largest between the French and German markets (0.535), while correlation is the lowest between Japan and the United States (0.241). Given the well-known time variability of correlations, these sample correlations may be misleading for allocation purposes. Indeed, the correlations between these markets have been much higher over the last decade. For instance, as may be gleaned from the correlation matrix in Panel B, the sample correlation between the French and German markets is as high as 0.93. On average, correlations increased by 15% to 40% between the two subperiods. Hence, the naive strategy is likely to overstate the diversification ability of the stock markets.

2.2 Model Estimation

Table 2 reports the Bayesian 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, ψ_i , is significantly positive, suggesting that bad news has a stronger effect on volatility than good news. In addition, the volatility persistence, calculated as $\alpha_i + \psi_i/2 + \beta_i$, is rather large in Japan, but much less so in France. Turning to the dynamics of correlations, the persistence parameter, δ_2 , takes a value of 0.93, translating the fact that correlation dynamics are also highly persistent.

The second part of the table presents the parameters of the higher-moment dynamics. Regarding the degree of freedom, we notice that large negative shocks are generally followed by an increase in the subsequent degree of freedom and therefore by a decrease in the subsequent kurtosis ($c_{i,1}^- > 0$). This suggests that large returns are less likely to occur. In contrast, large positive shocks are often followed by a large kurtosis, indicating that the probability of large subsequent shocks increases ($c_{i,1}^+ < 0$).

Regarding the asymmetry of the distribution, we observe that after a large negative shock, the subsequent skewness tends to be negative in all countries but France ($d_{i,1}^- < 0$), suggesting that another negative shock is more likely. After a large positive shock, the subsequent skewness is negative in the United States and Japan, indicating that a negative shock is more likely ($d_{i,1}^+ < 0$). In European markets, a positive shock is more likely to be followed by another positive shock ($d_{i,1}^+ > 0$).

To sum up, in the United States and Japan, large shocks (both negative and positive) tend to be followed by negative shocks, suggesting that negative shocks are persistent while positive shocks are short-lived. For European countries, the patterns are less clear, probably because these markets are also contaminated by the evolution of the U.S. market. Broadly speaking, the evidence provided by these weekly estimates is similar to that found by [Jondeau and Rockinger \(2009\)](#) for the daily frequency.

Table 3 summarizes goodness-of-fit tests. It reports the log-likelihood, the Akaike and Schwarz information criteria, and the Pearson's goodness-of-fit statistic proposed by [Diebold, Gunther, and Tay \(1998\)](#) (using 20 cells). The

Table 2 Estimation of the GARCH-DCC parameters

Country	Parameter	Mean	SD	5%	Median	95%
United States	μ	0.062	0.013	0.040	0.063	0.085
	ω	0.514	0.031	0.448	0.520	0.551
	α	0.088	0.004	0.078	0.088	0.093
	ψ	0.142	0.008	0.126	0.143	0.153
	β	0.794	0.009	0.785	0.792	0.815
Japan	μ	0.076	0.004	0.069	0.075	0.082
	ω	0.985	0.015	0.953	0.989	0.999
	α	0.128	0.019	0.095	0.131	0.156
	ψ	0.285	0.044	0.197	0.294	0.343
	β	0.711	0.013	0.692	0.708	0.729
UK	μ	0.082	0.010	0.065	0.083	0.099
	ω	0.353	0.001	0.352	0.353	0.355
	α	0.097	0.000	0.096	0.097	0.097
	ψ	0.018	0.006	0.007	0.019	0.026
	β	0.863	0.001	0.861	0.863	0.864
Germany	μ	0.148	0.010	0.133	0.148	0.165
	ω	0.544	0.014	0.515	0.545	0.564
	α	0.118	0.014	0.100	0.115	0.143
	ψ	0.021	0.010	0.007	0.020	0.039
	β	0.821	0.011	0.802	0.822	0.836
France	μ	0.173	0.010	0.156	0.173	0.189
	ω	0.463	0.057	0.362	0.482	0.539
	α	0.072	0.012	0.045	0.076	0.088
	ψ	0.062	0.012	0.041	0.059	0.085
	β	0.756	0.027	0.720	0.749	0.806
Correlation	δ_1	0.007	0.001	0.006	0.007	0.008
	δ_2	0.930	0.006	0.921	0.931	0.940

(continued)

Table 2 (continued). Estimation of the Sk-*t* distribution parameters

Country	Parameter	Mean	SD	5%	Median	95%
United States	c_0	3.186	0.642	2.280	3.130	4.280
	c_1^-	17.384	3.096	12.681	17.740	23.079
	c_1^+	-7.302	1.401	-9.633	-7.163	-5.251
	c_2	0.618	0.074	0.515	0.608	0.752
	d_0	1.020	0.029	0.960	1.022	1.065
	d_1^-	0.146	0.035	0.088	0.147	0.202
	d_1^+	-0.084	0.052	-0.174	-0.082	0.008
	d_2	0.518	0.114	0.332	0.527	0.710
Japan	c_0	3.172	0.636	2.230	3.110	4.400
	c_1^-	44.455	8.895	29.291	45.084	57.760
	c_1^+	-9.768	1.049	-11.422	-9.488	-8.252
	c_2	-0.494	0.176	-0.712	-0.536	-0.192
	d_0	1.198	0.052	1.135	1.180	1.300
	d_1^-	0.093	0.052	0.008	0.090	0.181
	d_1^+	-0.234	0.065	-0.346	-0.229	-0.134
	d_2	-0.094	0.197	-0.459	-0.084	0.215
UK	c_0	3.114	0.628	2.280	3.025	4.360
	c_1^-	-10.301	1.798	-12.743	-10.927	-7.525
	c_1^+	-6.943	4.516	-15.857	-5.972	-1.347
	c_2	-0.060	0.038	-0.126	-0.056	-0.005
	d_0	1.032	0.053	0.926	1.041	1.103
	d_1^-	0.174	0.049	0.091	0.177	0.257
	d_1^+	0.033	0.075	-0.062	0.011	0.179
	d_2	-0.010	0.194	-0.323	-0.009	0.292
Germany	c_0	3.114	0.492	2.380	3.250	3.810
	c_1^-	29.884	4.165	21.950	30.710	35.619
	c_1^+	8.432	1.254	6.614	8.265	10.636
	c_2	-0.573	0.064	-0.684	-0.572	-0.459
	d_0	1.019	0.058	0.926	1.019	1.115
	d_1^-	0.307	0.035	0.239	0.309	0.361
	d_1^+	0.212	0.041	0.151	0.208	0.289
	d_2	-0.397	0.111	-0.589	-0.393	-0.217
France	c_0	2.499	0.275	1.920	2.550	2.860
	c_1^-	-12.083	2.869	-16.353	-11.893	-7.307
	c_1^+	22.698	4.458	15.116	23.138	29.383
	c_2	0.149	0.102	0.001	0.142	0.304
	d_0	0.846	0.069	0.731	0.846	0.955
	d_1^-	-0.158	0.089	-0.300	-0.169	-0.020
	d_1^+	0.279	0.076	0.146	0.288	0.385
	d_2	-0.141	0.202	-0.413	-0.175	0.237

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

Table 3 Goodness-of-fit test statistics

	Normal distribution		Sk- <i>t</i> distribution	
	Statistics	p-value	Statistics	p-value
lnL	-15526.156	-	-15491.485	-
AIC	19.2861	-	19.2937	-
BIC	19.5261	-	19.8892	-
Pearson's goodness-of-fit statistics				
United States	36.039	0.010	21.947	0.287
Japan	30.402	0.047	19.555	0.422
UK	43.641	0.001	26.388	0.120
Germany	39.456	0.004	21.548	0.307
France	9.306	0.968	19.470	0.427

This table reports the log-likelihood (lnL), Akaike (AIC) and Schwarz information criteria, and goodness-of-fit test statistics for the model with normal distribution and with Sk-*t* distribution. Information criteria are the Akaike and Schwarz criteria. In the lower part of the table, we report, for each market, Pearson's goodness-of-fit statistic proposed by [Diebold, Gunther, and Tay \(1998\)](#), with 20 cells.

log-likelihood test statistic (equal to 69.34 with a p-value of 0.3%) and the information criteria indicate that the complete GARCH-DCC model with joint Sk-*t* distribution clearly dominates the model with joint normal distribution. In addition, goodness-of-fit tests indicate that the normal distribution fails to fit four of the five market returns at hand. On the other hand, the Sk-*t* distribution fits the data very well.¹⁶

Figures 1 and 2 display the dynamics of volatilities $\sigma_{i,t}$ and correlations $\rho_{ij,t}$, respectively. Inspection of these figures reveals several interesting features from a portfolio perspective. First, volatilities of the UK, German, and French markets are relatively low over the allocation period. In particular, they are lower than the U.S. volatility, whereas sample estimates were ranking the United States as the safest market. Second, we observe some trends in the correlations across markets. For instance, the conditional correlation between the U.S. and UK markets is anchored above 0.5 over the allocation period, while it is below 0.5 over the estimation period. Similarly, the correlation between the German and French markets increases from about 0.6 over the estimation period to about 0.7 over the allocation period.

Figures 3 and 4 display the dynamics of the conditional skewness $sk_{i,t}$ and kurtosis $ku_{i,t}$ for the five markets under study. Several comments are of interest. First, the conditional skewness and kurtosis differ at times significantly from the

¹⁶In Table 3, the column labeled 'Normal distribution' corresponds to [Engle's \(2002\)](#) standard DCC model.

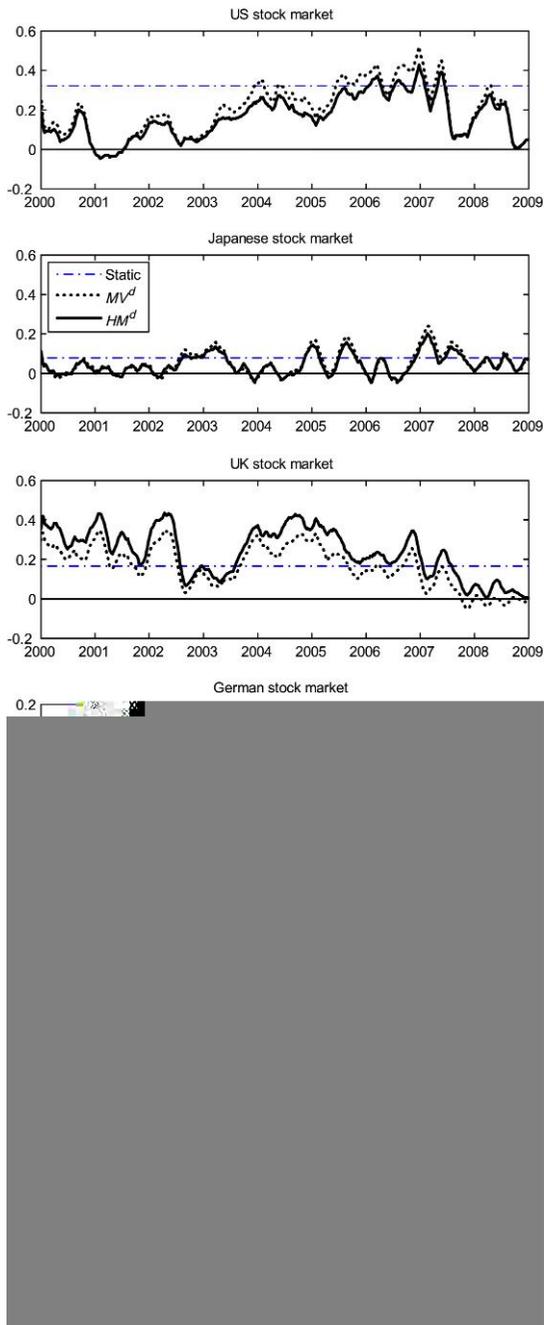


Figure 5 The optimal portfolio weights, for the naive, MV^d , and HM^d strategies with a risk aversion of $\gamma = 5$, over the allocation period are displayed.

Table 5 Measures of portfolio performance over the allocation period

Strategy	mSR (%)	Performance fee ϑ (%)	VaR(1%) (%)	ES(1%) (%)
$\gamma = 5$				
MV ^d	0.935 [0.32; 1.81]	3.020 [1.96; 4.01]	6.883 [6.43; 7.42]	7.922 [7.36; 8.49]
HM ^d	0.965 [0.74; 1.36]	4.395 [4.03; 4.76]	5.722 [5.56; 6.29]	6.810 [6.58; 7.32]
$\gamma = 10$				
MV ^d	0.482 [0.17; 0.91]	1.423 [0.94; 1.89]	3.386 [3.16; 3.64]	3.915 [3.64; 4.19]
HM ^d	0.493 [0.39; 0.68]	1.993 [1.84; 2.15]	2.846 [2.76; 3.10]	3.386 [3.27; 3.63]
$\gamma = 15$				
MV ^d	0.331 [0.12; 0.62]	0.949 [0.64; 1.25]	2.219 [2.07; 2.39]	2.581 [2.39; 2.76]
HM ^d	0.322 [0.25; 0.45]	1.284 [1.18; 1.38]	1.879 [1.83; 2.06]	2.238 [2.16; 2.41]

This table reports statistics on the performance of the optimal portfolios for the MV^d and HM^d strategies and for values of the risk aversion γ ranging from 5 to 15. We report several measures of performance of the strategies: the modified Sharpe ratio, mSR, defined by equation (11); the annualized performance fee, ϑ , estimated from the sample counterpart of relation (12); and the VaR and ES, estimated for a threshold of 1%. For each allocation strategy, the first row corresponds to the median statistics and the second row reports the 5% and 95% quantiles of the finite-sample distribution.

2.3.2 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, we observe some sizeable changes in portfolio weights relative to the MV^d strategy. In particular, the weights of the UK and German markets increase by about 6% on average, whereas the French weight decreases by 14% compared to the MV^d strategy. The resulting average weights are $\alpha_{HM^d}^{\gamma=5} = (16\%, 6\%, 22\%, -24\%, 75\%)$.

The explanation of these changes in portfolio weights can be found in Figures 1 through 4. On the one hand, there is a clear trade-off between the German and French markets, which are highly correlated over the allocation period. While the volatility of the French market is lower than the German counterpart, we notice that the skewness is more negative and that the kurtosis is slightly higher in the French market than that of the German market. Hence, the HM^d investor puts more emphasis on Germany and less on France. There is also another less pronounced trade-off between the U.S. and UK markets, which are also highly

correlated over the allocation period. The skewness is clearly less negative in the UK than in the United States. We should emphasize that at the beginning and at the end of the period (2000–2003 and 2006–2007), the kurtosis is lower in the United States than in the UK. Over these subperiods, we do not observe any decrease in the U.S. portfolio weight.

We now turn to the relative performance of the MV^d and HM^d strategies. Table 4 indicates that the realized return of the latter is slightly lower than the realized return of the lower (for instance, 4.3% vs. 4.4% per year, for $\gamma = 5$), while the volatility is lower for the HM^d , such that both strategies yield similar Sharpe ratios. We wish to emphasize that *ex ante* the MV^d strategy should yield a higher Sharpe ratio than the HM^d strategy by construction. To objectively compare both strategies, it is necessary to use a criterion that incorporates the nonnormal character of returns. Observation of the higher moments indicates that skewness is less negative and kurtosis lower for the HM^d strategy. This evidence suggests that the HM^d strategy is able to generate a portfolio return dynamic that is less prone to extreme events.

The economic gain due to distribution timing is measured by comparing the performance of the MV^d and HM^d strategies. Table 5 reveals that a naive investor is willing to pay 3% per year to switch to the MV^d strategy and 440 bp per year to switch to the HM^d strategy, for $\gamma = 5$. Therefore, ability to benefit from distribution timing generates an additional performance fee of about 140 bp (440 – 300). The performance fee decreases to 60 bp for $\gamma = 10$ and 35 bp for $\gamma = 15$. As the investor reduces her exposure to risky assets, the strategies are less likely to produce large differences in terms of performance measures.

To assess the statistical significance of the gains to distribution timing, Figure 6 depicts the empirical distribution of the performance fee for MV^d and HM^d strategies with respect to the naive strategy. One can clearly see that the performance fee of the HM^d strategy is significantly above the fee of the MV^d strategy, confirming the economic value of distribution timing. 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 economically and statistically superior to that of the MV^d one.

Finally, we investigate the distribution properties of realized portfolio returns implied by both dynamic strategies. Table 5 reports the VaR and the ES at the 1% threshold. As is apparent, the gain of adopting the HM^d strategy is both statistically and economically significant, as it decreases the probability and the size of extreme negative events. For $\gamma = 5$, the average returns below the 1% quantile are –8.9%, –7.9%, and –6.8% for the naive, MV^d , and HM^d strategies, respectively.

2.4 Aversion to Parameter Uncertainty

At this point, it may be argued that incorporating parameter uncertainty in the evaluation of the significance of distribution timing is insufficient because it may

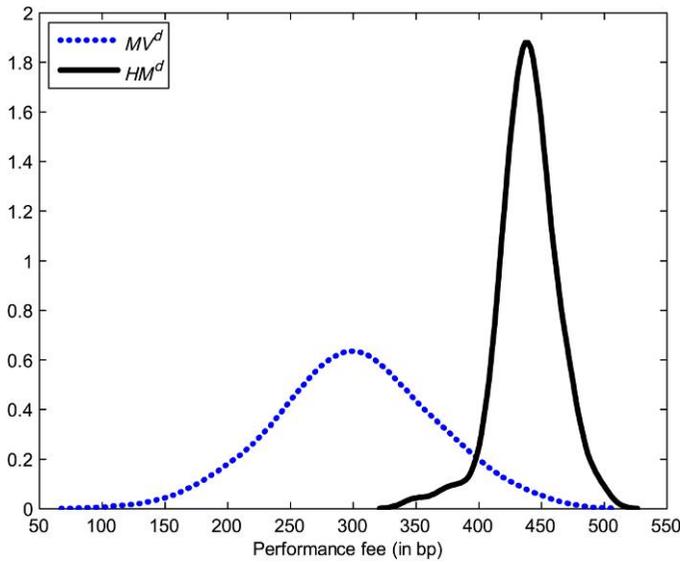


Figure 6 The distribution of the performance fee for MV^d and HM^d strategies are given. The two curves are obtained by drawing parameters from the Markov chain generated during the parameter estimation. The performance fee is the premium the investor is willing to pay to switch from a naive strategy to either of the dynamic strategies. The dynamic mean-variance strategy is tantamount to volatility timing, whereas the dynamic higher-moment strategy corresponds to distribution timing.

directly affect the behavior of investors. As already mentioned in Section 1.1, this issue has been addressed in a series of contributions that use Bayesian techniques to evaluate how investors with aversion to parameter uncertainty choose portfolios that maximize the minimum expected utility. This research follows the approach of Gilboa and Schmeidler (1989), which demonstrates that the minimum expected utility actually reflects the preferences of an investor who is averse to uncertainty about the probability distribution.¹⁷ The corresponding max–min optimization program is

$$\max_{\alpha_t} \min_{\theta \in \Theta} E_t[U(W_{t+1})], \tag{13}$$

for each period of time, where Θ is the domain characterizing the range of the parameters required to compute the expectation. Given the complexity of the model,

¹⁷Several 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 (2008). 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.

it is not possible to follow the approach adopted, for instance, by [Garlappi, Uppal, and Wang \(2007\)](#), who 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 θ 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 θ is delimited by the Bayesian prior and the likelihood of the model. We use draws from the posterior distribution of the parameters to describe the possible domain, Θ , to which θ can belong.

We solve the optimization problem (13) as follows. For a given date t , we consider all the possible sets of parameters in Θ and maximize the corresponding expected utility over portfolio weights α_t . This yields a solution, say $\alpha_t^*(\theta)$. We then seek the allocation that solves equation (13). This portfolio weight vector is expected to produce the best outcome to the investor in the event of a worst-case scenario.

As expected, the optimal portfolio weights found under ambiguity aversion are more conservative than those found using the previous resampling approach. In particular, we observe a reduction in the weight of the French market. For instance, for $\gamma = 5$, it decreases from 89% to 82% for the MV^d strategy and from 75% to 72% for the HM^d strategy. In Tables 6 and 7, we present moments and characteristics of the resulting allocations. As a comparison of Table 6 with Table 4 indicates, a conservative investor who is uncertain of her parameter estimates accepts a

Table 6 Moments of realized portfolio return over the allocation period (Model with aversion to parameter uncertainty)

Strategy	Moments of realized portfolio return				Sharpe ratio
	μ	σ	sk	ku	
$\gamma = 5$					
Naive	3.202	13.798	-0.621	5.961	0.020
MV ^d	4.027	16.331	-0.438	4.635	0.067
HM ^d	4.126	15.427	-0.386	4.158	0.077
$\gamma = 10$					
Naive	3.219	7.273	-0.626	5.901	0.039
MV ^d	3.516	8.139	-0.433	4.637	0.072
HM ^d	3.543	7.732	-0.380	4.160	0.079
$\gamma = 15$					
Naive	3.065	4.757	-0.633	6.071	0.028
MV ^d	3.330	5.410	-0.426	4.641	0.074
HM ^d	3.342	5.160	-0.376	4.165	0.080

This table reports summary statistics on the optimal portfolio return in the case of aversion to parameter uncertainty. The level of risk aversion γ ranges from 5 to 15. The statistics are the same as in Table 4. Moments of realized portfolio return over the allocation period.

Table 7 Measures of portfolio performance over the allocation period (Model with aversion to parameter uncertainty)

Strategy	mSR (%)	Performance fee ϑ (%)	VaR(1%) (%)	ES(1%) (%)
$\gamma = 5$				
MV ^d	0.656	0.936	6.660	7.775
HM ^d	0.799	1.980	5.572	6.775
$\gamma = 10$				
MV ^d	0.235	0.655	3.249	3.841
HM ^d	0.288	1.062	2.769	3.363
$\gamma = 15$				
MV ^d	0.217	0.435	2.140	2.536
HM ^d	0.245	0.672	1.843	2.221

This table reports statistics on the performance of the optimal portfolios in the case of aversion to parameter uncertainty. The level of risk aversion γ ranges from 5 to 15. The statistics are the same as in Table 5. Measures of portfolio performance over the allocation period.

decrease in expected return in order to reduce volatility. The table also demonstrates that switching from the naive strategy to the dynamic strategies 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.

Table 7 documents that the economic value of volatility timing under aversion to parameter uncertainty amounts to 95 bp, whereas the economic value of distribution timing is 105 bp (200 – 95). These estimates suggest that, even when the strategies are constrained to be more conservative in order to take worst-case scenarios into account, the gain of distribution timing is large compared to that of volatility timing. We also observe that, in this context, the HM^d strategy also provides lower VaR and lower ES compared to the MV^d strategy. Comparison with Table 5 also reveals that the VaR and the ES are systematically lower for the conservative investor than for the standard power utility investor. Thus, the choice of conservative allocations to avoid risk due to erroneous parameter estimates appears to provide additional felicity to the investor.

2.5 Robustness Analysis

As already discussed, we did our best to control for statistical issues. We accounted for overfitting by using two separate subperiods for the parameter estimation and asset allocation, and we accounted for parameter estimation risk by using the finite-sample distribution for all 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 described in this section.

All these experiments are based on the same data, yet with different specifications of the model. We start with a subperiod analysis. We split the allocation period into two subperiods of equal length. The first subsample covers the period 2000–2004, which includes the dotcom bubble burst. The second period (2005–2009) covers the subprime crisis. Then, we consider the relative performance of both dynamic strategies for each of the subperiods. Table 8 (Panel A) reports the moments for realized portfolio returns for $\gamma = 5$. As is apparent, compared to the naive strategy, the dynamic strategies performed well during the first period. The HM^d strategy also managed to generate a lower volatility. Over the second period, the HM^d strategy outperformed the naive and MV^d approaches, according to all criteria, with a higher expected return, a lower volatility, a higher skewness, and a lower kurtosis. Observing Table 9 (Panel A) reveals that the performance fee is much higher for the HM^d strategy than for the MV^d strategy over both periods.

Table 8 Moments of realized portfolio return over the allocation period (robustness analysis)

Strategy	Moments of realized portfolio return				Sharpe ratio
	μ	σ	sk	ku	
Panel A: Subperiod analysis					
First subperiod (2000–2004)					
Naive	1.983	15.506	−0.060	4.654	−0.058
MV ^d	2.384	17.206	−0.213	5.116	−0.002
HM ^d	2.583	15.923	−0.227	4.653	−0.019
Second subperiod (2005–2009)					
Naive	4.858	17.348	−1.007	6.811	0.107
MV ^d	5.990	16.836	−0.643	3.854	0.179
HM ^d	6.040	15.193	−0.561	3.558	0.200
Panel B: Constant higher moments					
Naive	3.418	16.453	−0.610	6.015	0.030
MV ^d	4.409	17.030	−0.421	4.510	0.086
HM ^d	4.299	15.577	−0.386	4.157	0.088
Cst kurtosis	4.447	17.696	−0.415	4.373	0.085
Cst skewness	4.471	17.602	−0.418	4.467	0.087
Panel C: GARCH-in-Mean model					
Naive	3.418	16.453	−0.610	6.015	0.030
MV ^d	3.140	16.348	−0.500	3.779	0.013
HM ^d	8.341	16.028	−0.412	5.144	0.337

This table reports summary statistics on the optimal portfolio return for alternative specifications of the allocation period and the econometric model. The level of risk aversion is $\gamma = 5$. The statistics are the same as in Table 4.

Table 9 Measures of portfolio performance over the allocation period (robustness analysis)

Strategy	mSR (%)	Performance fee θ (%)	VaR(1%) (%)	ES(1%) (%)
Panel A: Subperiod analysis				
First subperiod (2000–2004)				
MV ^d	−0.912	−0.800	6.742	8.741
HM ^d	−2.307	−0.162	6.630	7.914
Second subperiod (2005–2009)				
MV ^d	1.259	2.528	7.329	7.470
HM ^d	1.615	4.200	5.726	6.090
Panel B: Constant higher moments				
MV ^d	0.935	3.020	6.883	7.921
HM ^d	0.965	4.395	5.722	6.810
Cst kurtosis	0.912	2.343	7.133	8.159
Cst skewness	0.940	2.408	7.139	8.231
Panel C: GARCH-in-Mean model				
MV ^d	−0.274	2.461	6.547	7.313
HM ^d	5.058	7.720	6.353	8.119

This table reports statistics on the performance of the optimal portfolios for alternative specifications of the allocation period and the econometric model. The level of risk aversion is $\gamma = 5$. The statistics are the same as in Table 5.

The values of distribution timing are equal to 65 and 170 bp per year over these two periods, respectively. In addition, the VaR of the HM^d strategy is always much lower than the VaR of the MV^d strategy, especially over the most recent period.

We then turn to two special cases of our general model: in the first one, the skewness is time varying, but the kurtosis is constrained to be constant over time; in the second model, the skewness is constant, and the kurtosis is time varying. The idea is to identify the source of the gains found in distribution timing. Tables 8 and 9 (Panel B) demonstrate that the models with constant kurtosis or with constant skewness do not perform very well. In particular, we notice that the performance fee is lower than the performance of the MV^d strategy, around 2.3% per year. These findings suggest that it is the joint dynamic of the skewness and kurtosis that causes the gains of distribution timing.

In the last experiment, we introduce a conditional mean in the return’s dynamic. Specifically, we allow the expected excess return in market i ($\mu_{i,t}$) to depend on the conditional volatility ($\sigma_{i,t}$), corresponding to a so-called GARCH-in-mean effect. In this model, the return process is given by

$$\tilde{r}_{i,t} = \mu_i + \lambda_i \sigma_{i,t} + \varepsilon_t.$$

We experimented with several specifications for the risk measure ($\sigma_{i,t}$, $\sigma_{i,t}^2$, and $\log \sigma_{i,t}$, as recommended by Engle, Lilien, and Robbins 1987). The volatility specification that we adopt is the one with the highest log likelihood. The estimates of λ_i for the United States and the UK are highly significant, but not so for the other countries. Concerning the performances of the strategies, we find that capturing GARCH-in-mean effects does not improve the MV^d allocation. Instead, the performances of this strategy are worsened compared to the benchmark because the associated performance fee decreases from 300 to 246 bp per year. On the other hand, the HM^d strategy performs much better in the case of the GARCH-in-mean model. The realized return is much higher, at the cost of a small increase in the volatility and kurtosis. All in all, the performance fee of the HM^d strategy increases from 140 bp, obtained in the benchmark case, to 530 bp.

3 CONCLUSION

In this paper, we investigate the consequences of nonnormality of returns on the optimal asset allocation when the distribution of asset returns changes over time. Whereas most previous work has been devoted to the case in which 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 (2001, 2003) and Han (2006) evaluate the value of volatility timing, while Ang and Bekaert (2002) and Guidolin and Timmermann (2008) measure the cost of ignoring the presence of regime shifts. Patton (2004) considers a bivariate model with predictability in the asymmetric behavior of asset returns. The present study contributes to this literature by providing several additional 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 we account for several assets. This setting allows us to integrate out parameter uncertainty as we consider the performance measures.

We demonstrate that, for all levels of risk aversion, the performance fee an investor is willing to pay to benefit from distribution timing is of a similar magnitude to the performance fee she would be willing to pay to benefit from volatility timing. We perform several alternative experiments designed to assess the robustness of our findings. We consider conservative investors who take parameter uncertainty into account in their allocation process. We measure the economic value of distribution timing for several specifications and subperiods and confirmed in all cases the relevance of taking into account the temporal variation of the conditional distribution of asset returns.

Several extensions to this research may be considered. It would be interesting, for instance, to have multiperiod investments in order to evaluate the consequences of nonnormality on hedging demands. As already mentioned, this

extension would be rather demanding, in light of the way that multiperiod investment problems are usually solved.

APPENDIX A: A MULTIVARIATE MODEL FOR RETURNS

A.1 The DCC Model

The dynamic of the excess return vector, \tilde{r}_t , is

$$\tilde{r}_t = \mu_t + \varepsilon_t, \tag{A1}$$

$$\varepsilon_t = \Sigma_t^{1/2} z_t, \tag{A2}$$

where μ_t denotes the vector of expected excess returns and ε_t the vector of unexpected excess returns, $\Sigma_t = \{\sigma_{ij,t}\}_{i,j=1,\dots,n}$ is the conditional covariance matrix, z_t is the vector of innovations, such that $E[z_t] = 0$ and $V[z_t] = I_n$, where I_n is the identity matrix. The conditional covariance matrix of returns Σ_t is defined as $\Sigma_t = D_t \Gamma_t D_t$, where $D_t = \{\sigma_{i,t}\}_{i=1,\dots,n}$ is a diagonal matrix with standard deviations on the diagonal, and $\Gamma_t = \{\rho_{ij,t}\}_{i,j=1,\dots,n}$ is the symmetric, positive, definite correlation matrix. Each conditional variance, $\sigma_{i,t}^2$, is described by an asymmetric GARCH model as in [Glosten, Jagannathan, and Runkle \(1993\)](#):

$$\sigma_{i,t}^2 = \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i \varepsilon_{i,t-1}^2 + \psi_i \varepsilon_{i,t-1}^2 1_{\{\varepsilon_{i,t-1} < 0\}}, \quad i = 1, \dots, n, \tag{A3}$$

where all parameters are positive. Equivalently, we have

$$\sigma_{i,t}^2 - \bar{\sigma}_i^2 = \tilde{\omega}_i + \gamma_i (\sigma_{i,t-1}^2 - \bar{\sigma}_i^2) + (\alpha_i + \psi_i 1_{\{\varepsilon_{i,t-1} < 0\}}) (\varepsilon_{i,t-1}^2 - \sigma_{i,t-1}^2), \tag{A4}$$

where γ_i denotes the variance persistence. The constraint $\gamma_i < 1$ guarantees the stationarity of the variance process. The conditional correlation matrix, Γ_t , is time varying, following the DCC specification of [Engle \(2002\)](#) and [Engle and Sheppard \(2001\)](#):

$$\Gamma_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \tag{A5}$$

$$Q_t = (1 - \delta_1 - \delta_2) \bar{Q} + \delta_1 (u_{t-1} u_{t-1}') + \delta_2 Q_{t-1}, \tag{A6}$$

where $u_t = D_t^{-1} \varepsilon_t$ denotes the vector of normalized unexpected returns, and Q_t^* denotes the $n \times n$ diagonal matrix composed of the diagonal elements of Q_t . The matrix \bar{Q} is the unconditional covariance matrix of u_t . We impose the restrictions $0 \leq \delta_1, \delta_2 \leq 1$ and $\delta_1 + \delta_2 \leq 1$ so that the conditional correlation matrix is guaranteed to be positive definite.

A.2 Moments of the Sk- t Distribution

The $n \times 1$ vector of innovations, z_t , is drawn from the multivariate Sk- t distribution defined in Equation (4), with time-varying shape parameters $\nu_{i,t}$ and $\zeta_{i,t}$

corresponding to the individual degree of freedom and the asymmetry parameter, respectively. The marginal distribution of $z_{i,t}$ is a univariate Sk- t distribution $g(z_{i,t}|v_{i,t}, \xi_{i,t})$. It is defined for $2 < v_{i,t} < \infty$ and $\xi_{i,t} > 0$ for all t . As shown in Equation (A2), dependence across returns is introduced via the covariance matrix Σ_t .

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

$$M_{i,t}^{(r)} = m_{i,t}^{(r)} \frac{\xi_{i,t}^{r+1} + \frac{(-1)^r}{\xi_{i,t}^{r+1}}}{\xi_{i,t} + \frac{1}{\xi_{i,t}}},$$

where

$$m_{i,t}^{(r)} = 2E[Z_{i,t}^r | Z_{i,t} > 0] = \frac{\Gamma\left(\frac{v_{i,t}-r}{2}\right) \Gamma\left(\frac{r+1}{2}\right) (v_{i,t} - 2)^{\frac{r+1}{2}}}{\sqrt{\pi}(v_{i,t} - 2) \Gamma\left(\frac{v_{i,t}}{2}\right)} \tag{A7}$$

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

$$\mu_{i,t}^{(3)} = E[Z_{i,t}^3] = M_{i,t}^{(3)} - 3M_{i,t}^{(1)}M_{i,t}^{(2)} + 2(M_{i,t}^{(1)})^3, \tag{A8}$$

$$\mu_{i,t}^{(4)} = E[Z_{i,t}^4] = M_{i,t}^{(4)} - 4M_{i,t}^{(1)}M_{i,t}^{(3)} + 6M_{i,t}^{(2)}(M_{i,t}^{(1)})^2 - 3(M_{i,t}^{(1)})^4. \tag{A9}$$

The skewness and kurtosis are therefore nonlinear functions of the degree of freedom, $v_{i,t}$, and the asymmetry parameter, $\xi_{i,t}$.

A.3 Moments of the Portfolio Return

Analytical expressions for the portfolio's 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 (A8) and (A9). Next, because unexpected excess returns are defined as $\varepsilon_{t+1} = \sum_{i=1}^{1/2} z_{t+1}$, we have $E_t[\varepsilon_{t+1}] = 0$ and $V_t[\varepsilon_{t+1}] = \Sigma_{t+1}$. We denote $\Sigma_{t+1}^{1/2} = (\omega_{ij,t+1})_{i,j=1,\dots,n}$ as the Choleski decomposition of the covariance matrix of excess returns, such that $\tilde{r}_{i,t+1} = \mu_{i,t} + \sum_{r=1}^n \omega_{ir,t+1} z_{r,t+1}$. In addition, denoting by \otimes the Kronecker product, the $n \times n^2$ third central co-moment matrix is defined as¹⁸

$$S_{t+1} = E_t[(\tilde{r}_{t+1} - \mu_{t+1})(\tilde{r}_{t+1} - \mu_{t+1})' \otimes (\tilde{r}_{t+1} - \mu_{t+1})'] = \{s_{ijk,t+1}\}$$

¹⁸Using these notations, central co-moment matrices can be conveniently represented as bidimensional matrices. The formulae presented are convenient for theoretical purposes. In our programs, we exploit, for numerical efficiency, the symmetric structure of these matrices, thereby reducing existing redundancies.

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

$$\begin{aligned} \mathcal{K}_{t+1} &= E_t[(\tilde{r}_{t+1} - \mu_{t+1})(\tilde{r}_{t+1} - \mu_{t+1})' \otimes (\tilde{r}_{t+1} - \mu_{t+1})' \otimes (\tilde{r}_{t+1} - \mu_{t+1})'] \\ &= \{\kappa_{ijkl,t+1}\} \end{aligned}$$

with component (i, j, k, l)

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

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 α_t , the conditional expected return, the conditional variance, and the conditional third and fourth moments of the portfolio return are defined as

$$\begin{aligned} m_{p,t+1} &= r_{f,t} + \alpha_t' \mu_{t+1}, \\ \sigma_{p,t+1}^2 &= \alpha_t' \Sigma_{t+1} \alpha_t, \\ s_{p,t+1}^3 &= \alpha_t' \mathcal{S}_{t+1} (\alpha_t \otimes \alpha_t), \\ \kappa_{p,t+1}^4 &= \alpha_t' \mathcal{K}_{t+1} (\alpha_t \otimes \alpha_t \otimes \alpha_t), \end{aligned}$$

where $\sigma_{p,t+1}^2$, $s_{p,t+1}^3$, and $\kappa_{p,t+1}^4$ stand for central moments $E_t[(r_{p,t+1} - m_{p,t+1})^i]$ for $i = 2, 3$, and 4, respectively.¹⁹

The relationships between the central and noncentral moments, which are required in the evaluation of the Taylor approximation of the expected utility, are

$$\begin{aligned} m_{p,t+1}^{(2)} &= \sigma_{p,t+1}^2 + m_{p,t+1}^2, \\ m_{p,t+1}^{(3)} &= s_{p,t+1}^3 + 3\sigma_{p,t+1}^2 m_{p,t+1} + m_{p,t+1}^3, \\ m_{p,t+1}^{(4)} &= \kappa_{p,t+1}^4 + 4s_{p,t+1}^3 m_{p,t+1} + 6\sigma_{p,t+1}^2 m_{p,t+1}^2 + m_{p,t+1}^4. \end{aligned}$$

¹⁹Central moments $s_{p,t+1}^3$ and $\kappa_{p,t+1}^4$ should not be confused with skewness and kurtosis, defined as the standardized central moments, $E_t[((r_{p,t+1} - m_{p,t+1})/\sigma_{p,t+1})^i]$, for $i = 3, 4$.

APPENDIX B: BAYESIAN ESTIMATION

Given the complexity of the model, we use Bayesian estimation to obtain parameter estimates. This setting also allows us to assess statistically the economic gain of our strategy. Bayesian estimation is performed using a Markov chain Monte Carlo algorithm. We obtain this chain using the Metropolis–Hastings algorithm. For completeness, we provide a short description of this technique, following the algorithm A.24 of [Robert and Casella \(1999\)](#). We denote by $\theta^{(t)}$ the vector of parameters obtained at step t . In each step, we generate a new guess, X , for a vector of parameters. This guess should not be too distant from the previous vector of parameters. For the first step, $\theta^{(1)}$ is drawn from the asymptotic (normal) distribution of the ML estimation. Subsequent steps $t > 1$ are given by:

1. Generate $X \sim q(x|\theta^{(t)})$.
2. Take

$$\theta^{(t+1)} = \begin{cases} X & \text{with probability } \rho(\theta^{(t)}, X), \\ \theta^{(t)} & \text{with probability } 1 - \rho(\theta^{(t)}, X), \end{cases}$$

where

$$\rho(\theta^{(t)}, X) = \min \left\{ \frac{p(X|y)q(\theta^{(t)}|X)}{p(\theta^{(t)}|y)q(X|\theta^{(t)})}, 1 \right\}.$$

The meanings of the various elements entering the algorithm are as follows: 1) $q(x|\theta^{(t)})$ is the so-called proposal or instrumental density. In our case, we choose for q the asymptotic multivariate normal distribution resulting from the ML estimation. In the particular case of a symmetric q function, the ratio $q(\theta^{(t)}|X)/q(X|\theta^{(t)}) = 1$. 2) $p(\theta|y)$ is the so-called objective or target density, with $y = \{\tilde{r}_1, \dots, \tilde{r}_T\}$. In our case, the target density is the posterior distribution:

$$p(\theta|y) = L(y|\theta)f(\theta),$$

where $L(y|\theta)$ is the data density or likelihood of the model, and $f(\theta)$ is the prior density of the parameter set. The likelihood is obtained by taking the exponential of

$$\log L(r_1, \dots, r_T | \zeta, \eta) = \sum_{t=1}^T \left[\log(g(\Sigma_t(\zeta)^{-1/2}(r_t - \mu)|\eta)) - \frac{1}{2} \log |\Sigma_t(\zeta)| \right], \quad (\text{A10})$$

where $g(\cdot|\eta)$ is defined in Equation (4). In words, at step t , once a guess X of the parameter vector θ has been obtained, we obtain the likelihood of the actual data y and measure the probability of occurrence of the vector θ , given the prior assumptions. The Metropolis–Hastings algorithm accepts the candidate vector X , setting $\theta^{(t+1)} = X$ if its likelihood $p(X|y)$ is larger than $p(\theta^{(t)}|y)$, which is the likelihood of the previously retained vector. With a certain probability, given by ρ , the algorithm accepts parameter vectors that decrease the likelihood. In the case of

ill-defined likelihoods with local maxima, the Metropolis–Hastings algorithm allows the parameters to eventually reach regions with globally higher likelihood. In the case in which an investor is uncertain about the true parameters, the chain provides a sequence of possible values. Thus, the chain implicitly captures parameter uncertainty.

Before presenting the priors put on the parameters, it should be emphasized that for each model we ran the chain over 1.5 million estimations and, after a burn-in period of one million observations, we performed allocations using independent draws from the Markov chain. These independent draws are typically obtained by keeping one out of 100 draws. We checked for independence of draws and diagnosed convergence of the chains as discussed in Robert and Casella 1999, chapter 8). As usual in such an exercise, the stability of the chain is investigated using several chains, each obtained with a different starting value for $\theta^{(1)}$.

The prior distributions of the parameters have been chosen to ensure that the model is stationary. Expected returns, μ , are drawn from a normal distribution $f(\mu) \propto N(\hat{\mu}, \hat{\sigma}_\mu^2)$, where $\hat{\mu}$ and $\hat{\sigma}_\mu^2$ correspond to the ML estimators of the expected return and its variance, respectively. Parameters of the GARCH processes (15) are drawn from Beta(p, q) distributions to ensure positivity:

$$\begin{aligned} f(\omega) &\propto B(p_\omega, q_\omega), & f(\alpha) &\propto B(p_\alpha, q_\alpha), \\ f(\psi) &\propto B(p_\psi, q_\psi), & f(\gamma) &\propto B(p_\gamma, q_\gamma). \end{aligned}$$

Parameters p and q are selected to ensure that the parameters ω , α , ψ , and γ are in the range usually obtained for GARCH models on weekly returns.²⁰ In a similar way, parameters of the DCC model are drawn from Beta distributions²¹:

$$f(\delta_1) \propto B(p_{\delta_1}, q_{\delta_1}), \quad f(\delta_1 + \delta_2) \propto B(p_{\delta_2}, q_{\delta_2}).$$

The parameters driving the degree of freedom, ν_t , and the asymmetry parameter, ξ_t , of the innovation distribution have the following prior distributions:

$$\begin{aligned} f(\log(c_0)) &\propto N(\mu_{c_0}, \sigma_{c_0}^2) & f(\log(d_0)) &\propto N(\mu_{d_0}, \sigma_{d_0}^2) \\ f(c_1^-) &\propto N(\mu_{c_1}, \sigma_{c_1}^2), & f(d_1^-) &\propto N(\mu_{d_1}, \sigma_{d_1}^2), \\ f(c_1^+) &\propto N(\mu_{c_1}, \sigma_{c_1}^2), & f(d_1^+) &\propto N(\mu_{d_1}, \sigma_{d_1}^2), \\ f(c_2) &\propto N(\mu_{c_2}, \sigma_{c_2}^2) \mathcal{I}_{c_2 \in [-1;1]}, & f(d_2) &\propto N(\mu_{d_2}, \sigma_{d_2}^2) \mathcal{I}_{d_2 \in [-1;1]}. \end{aligned}$$

Our priors for these parameters are consistent with the null hypothesis that there are no dynamics in the conditional distribution. Indeed, we assume a mean value equal to 0 for c_1^- , c_1^+ , d_1^- , and d_1^+ with a large standard deviation ($\sigma_{c_1} = \sigma_{d_1} = 2$).

²⁰In practice, we take $p_\omega = 2, q_\omega = 10, p_\alpha = 2, q_\alpha = 20, p_\psi = 2, q_\psi = 20, p_\gamma = 25$, and $q_\gamma = 1.5$. We then deduce the prior distribution of β .

²¹For the estimation, we take $p_{\delta_1} = 1.5, q_{\delta_1} = 100, p_{\delta_2} = 100$, and $q_{\delta_2} = 1.5$.

For c_2 and d_2 , we also set a central value equal to 0 with a standard deviation $\sigma_{c_2} = \sigma_{d_2} = 0.3$. These distributions are truncated to ensure that the lagged parameter is between -1 and 1 . Finally, the constant terms μ_{c_0} and μ_{d_0} have a central value deduced from the unconditional moments reported in Table 1, with large standard deviations ($\sigma_{c_0} = \sigma_{d_0} = 5$). Therefore, in the case that there are no dynamics in the degree of freedom and the asymmetry parameter, the central values are such that the degree of freedom and the asymmetry parameter are equal to their sample value.

Received February 15, 2009; revised July 26, 2011; accepted July 26, 2011.

REFERENCES

- Ait-Sahalia, Y., and M. W. Brandt. 2001. Variable Selection for Portfolio Choice. *Journal of Finance* 56: 1297–1351.
- Ang, A., and G. Bekaert. 2002. International Asset Allocation with Regime Shifts. *Review of Financial Studies* 15: 1137–1187.
- Ang, A., J. Chen, and Y. Xing. 2006. Downside Risk. *Review of Financial Studies* 19: 1191–1239.
- de Athayde, G. M., and R. G. Flôres Jr. 2004. Finding a Maximum Skewness Portfolio—A General Solution to Three Moments Portfolio Choice. *Journal of Economic Dynamics and Control* 28: 1335–1352.
- Balduzzi, P., and A. Lynch. 1999. Transaction Costs and Predictability: Some Utility Costs Calculations. *Journal of Financial Economics* 52: 47–78.
- Barberis, N. 2000. Investing in the Long-Run When Returns Are Predictable. *Journal of Finance* 55: 225–264.
- Barberis, N., and M. Huang. 2008. Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review* 98: 2066–2100.
- Barone-Adesi, G. 1985. Arbitrage Equilibrium with Skewed Asset Returns. *Journal of Financial and Quantitative Analysis* 20: 299–313.
- Busse, J. A. 1999. Volatility Timing in Mutual Funds: Evidence from Daily Returns. *Review of Financial Studies* 12: 1009–1041.
- Campbell, J. Y., and S. B. Thompson. 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies* 21: 1509–1531.
- Campbell, J. Y., and L. M. Viceira. 1999. Consumption and Portfolio Decisions When Expected Returns Are Time Varying. *Quarterly Journal of Economics* 114: 433–495.
- Cenesizoglu, T., and Timmermann. 2008. “Is the Distribution of Stock Returns Predictable?” Working paper. UCSD.
- Cochrane, J. H. 2008. The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies* 21: 1533–1575.
- Cremers, K. J. M. 2002. Stock Return Predictability: A Bayesian Model Selection Perspective. *Review of Financial Studies* 15: 1223–1249.

- Das, S. R., and R. Uppal. 2004. Systematic Risk and International Portfolio Choice. *Journal of Finance* 59: 2809–2834.
- DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *Review of Financial Studies* 22: 1915–1953.
- Detemple, J., R. Garcia, and M. Rindisbacher. 2003. A Monte Carlo Method for Optimal Portfolios. *Journal of Finance* 58: 401–446.
- Diebold, F. X., T. A. Gunther, and A. S. Tay. 1998. Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review* 39: 863–883.
- Dittmar, R. F. 2002. Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns. *Journal of Finance* 57: 369–403.
- Engle, R. F. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics* 20: 339–350.
- Engle, R. F., D. M. Lilien, and R. P. Robbins. 1987. Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model. *Econometrica* 55: 391–407.
- Engle, R. F., and K. Sheppard. 2001. "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." NBER Working paper 8554.
- Fleming, J., C. Kirby, and B. Ostdiek. 2001. The Economic Value of Volatility Timing. *Journal of Finance* 56: 329–352.
- Fleming, J., C. Kirby, and B. Ostdiek. 2003. The Economic Value of Volatility Timing Using "Realized" Volatility. *Journal of Financial Economics* 67: 473–509.
- Friend, I., and R. Westerfield. 1980. Co-Skewness and Capital Asset Pricing. *Journal of Finance* 35: 897–913.
- Garlappi, L., R. Uppal, and T. Wang. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies* 20: 41–81.
- Gilboa, I., and D. Schmeidler. 1989. Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics* 18: 141–153.
- Glosten, R. T., R. Jagannathan, and D. Runkle. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* 48: 1779–1801.
- Goyal, A., and I. Welch. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* 21: 1455–1508.
- Graham, J. R., and C. R. Harvey. 1997. Grading the Performance of Market-Timing Newsletters. *Financial Analysts Journal* 54: 54–66.
- Graham, J. R., and C. R. Harvey. 1996. Market Timing Ability and Volatility Implied in Investment Newsletters' Asset Allocation Recommendations. *Journal of Financial Economics* 42: 397–421.
- Guidolin, M., and G. Nicodano. 2009. Small Caps in International Equity Portfolios: The Effects of Variance Risk. *Annals of Finance* 5: 15–48.

- Guidolin, M., and A. Timmermann. 2007. Asset Allocation under Multivariate Regime Switching. *Journal of Economic Dynamics and Control* 31: 3503–3544.
- Han, Y. 2006. Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *Review of Financial Studies* 19: 237–271.
- Guidolin, M., and A. Timmermann. 2008. International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences. *Review of Financial Studies* 21: 889–935.
- Hansen, B. E. 1994. Autoregressive Conditional Density Estimation. *International Economic Review* 35: 705–730.
- Hansen, L. P., T. J. Sargent, and D. Tallarini. 1999. Robust Permanent Income and Pricing. *Review of Economic Studies* 66: 873–907.
- Harvey, C. R., J. C. Liechty, and M. W. Liechty. 2008. Bayes vs. Resampling: A Re-match. *Journal of Investment Management* 6: 29–45.
- Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller. 2010. Portfolio Selection with Higher Moments. *Quantitative Finance* 10: 469–485.
- Harvey, C. R., and A. Siddique. 2000. Conditional Skewness in Asset Pricing Tests. *Journal of Finance* 55: 1263–1295.
- Harvey, C. R., and A. Siddique. 1999. Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis* 34: 465–487.
- Johannes, M., N. Polson, and J. Stroud. 2002. “Sequential Optimal Portfolio Performance Market and Volatility Timing.” Working paper.
- Jondeau, E., and M. Rockinger. 2009. The Impact of Shocks on Higher Moments. *Journal of Financial Econometrics* 7: 77–105.
- Jondeau, E., and M. Rockinger. 2006. Optimal Portfolio Allocation Under Higher Moments. *European Financial Management* 12: 29–55.
- Jondeau, E., and M. Rockinger. 2003. Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements. *Journal of Economic Dynamics and Control* 27: 1699–1737.
- Kandel, S., and R. Stambaugh. 1996. On the Predictability of Stock Returns: An Asset Allocation Perspective. *Journal of Finance* 51: 385–424.
- Kraus, A., and R. H. Litzenberger. 1976. Skewness Preference and the Valuation of Risk Assets. *Journal of Finance* 31: 1085–1100.
- Kroll, Y., H. Levy, and H. M. Markowitz. 1984. Mean-Variance Versus Direct Utility Maximization. *Journal of Finance* 36: 47–61.
- Leippold, M., F. Trojani, and P. Vanini. 2008. Learning and Asset Prices Under Ambiguous Information. *Review of Financial Studies* 21: 2565–2597.
- Levy, H., and H. M. Markowitz. 1979. Approximating Expected Utility by a Function of Mean and Variance. *American Economic Review* 69: 308–317.
- Loistl, O. 1976. The Erroneous Approximation of Expected Utility by Means of a Taylor’s Series Expansion: Analytic and Computational Results. *American Economic Review* 66: 904–910.
- Maenhout, P. J. 2004. Robust Portfolio Rules and Asset Pricing. *Review of Financial Studies* 17: 951–983.
- Markowitz, H., and N. Usmen. 2003. Resampled Frontiers vs Diffuse Bayes: An Experiment. *Journal of Investment Management* 1: 9–25.

- Marquering, W., and M. Verbeek. 2004. The Economic Value of Predicting Stock Index Returns and Volatility. *Journal of Financial and Quantitative Analysis* 39: 407–429.
- Mencia, J., and E. Sentana. 2005. “Estimation and Testing of Dynamic Models with Generalized Hyperbolic Innovations.” CEPR Discussion paper 5177.
- Michaud, R. 1998. *Efficient Asset Management*. New York: Oxford University Press.
- Patton, A. J. 2004. On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation. *Journal of Financial Econometrics* 2: 130–168.
- Polson, N. G., and B. V. Tew. 2000. Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970–1996. *Journal of Business and Economic Statistics* 18: 164–173.
- Rapach, D., J. Strauss, and G. Zhou. 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies* 23: 821–862.
- Robert, C. P., and G. Casella. 1999. *Monte Carlo Statistical Methods*. New York: Springer Verlag.
- Rubinstein, M. 1973. The Fundamental Theorem of Parameter-Preference Security Valuation. *Journal of Financial and Quantitative Analysis* 8: 61–69.
- Sahu, S. K., D. K. Dey, and M. D. Branco. 2003. A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models. *Canadian Journal of Statistics* 31: 129–150.
- Scott, R. C., and P. A. Horvath. 1980. On the Direction of Preference for Moments of Higher Order than the Variance. *Journal of Finance* 35: 915–919.
- Tu, J. 2010. Is Regime Switching in Stock Returns Important in Portfolio Decisions? *Management Science* 56: 1198–1215.
- Tu, J., and G. Zhou. 2004. Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions? *Journal of Financial Economics* 72: 385–421.
- West, K. D., H. J. Edison, and D. Cho. 1993. A Utility-Based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics* 35: 23–45.