The Economic Value of Distributional Timing

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

November 2006

Abstract

We evaluate how non-normality of asset returns and the temporal evolution of volatility and higher moments affects the conditional allocation of wealth. We show that if one neglects these aspects, as would be the case in a mean-variance allocation, a significant cost would arise. The performance fee the investor is willing to pay to benefit from our allocation is as high as the fee she is willing to pay to benefit from volatility timing. Many tests of robustness are performed, yet, the economic value of taking the non-normality and the temporal evolution of the distribution into account remains.

Keywords: Non-normality, volatility timing, distributional timing, GARCH, portfolio allocation.

JEL classification: G11, F37, C22, C51.

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

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

Both authors acknowledge help from the Swiss National Science Foundation through NCCR FINRISK (Financial Valuation and Risk Management). We benefited from comments of and discussions with Tim Bollerslev, Peter Bossaert, Ales Cerny, Bernard Dumas, René Garcia, Steward Hodges, Eric Jacquier, Nour Meddahi, and Raman Uppal. The usual disclaimer applies.
1 Introduction

Several recent papers have highlighted that volatility timing –i.e. the gain for the investor to capture the dynamics of volatility when allocating her portfolio– has a significant economic value for short horizons, e.g., Fleming, Kirby, and Ostdiek (2001, 2003), Marquering and Verbeek (2001), Johannes, Polson, and Stroud (2002). All these papers acknowledge the predictability of volatilities and show that strategies based on volatility-timing are valuable. Fleming, Kirby, and Ostdiek (2001) show that volatility timing strategies outperform strategies based on constant expected returns and volatilities. Fleming, Kirby, and Ostdiek (2003) investigate alternative volatility models and obtain that the use of an appropriate measure of conditional volatilities may significantly improve the gain of volatility timing strategies. Marquering and Verbeek (2001) show that the economic value of strategies based on the timing of expected returns and volatility exceeds that of strategies based on market timing only. The relevance of volatility timing with respect to ‘expected return’ timing is demonstrated by Johannes, Polson, and Stroud (2002) who argue that in fact strategies based on time-varying expected returns perform rather poorly, due to estimation risk, while volatility timing strategies do not suffer from this problem.

All these contributions assume a mean-variance investor, who allocates her portfolio among some risky assets and the risk-free asset. The mean-variance criterion implicitly assumes that returns are normal, or at least that higher moments (beyond mean and variance) are not relevant for asset allocation. In this paper, we address a closely related issue, namely distributional timing, i.e. the ability to invest according to changes in the distribution of asset returns. Academic interest in the effect of higher moments on asset allocation goes back to Arditti (1967), Samuelson (1970), Rubinstein (1973), and Kraus and Litzenberger (1976). Early empirical evidence however suggests that the mean-variance criterion results in allocations that are very similar to those obtained from a direct optimization of the expected utility (Levy and Markowitz, 1979, and Kroll, Levy, and Markowitz, 1984). More recently, Chunachinda et al. (1997), Harvey et al. (2002), Das and Uppal (2004) and Jondeau and Rockinger (2006) propose alternative approaches to deal with portfolio allocation under non-normality. Most of this work provides evidence that the mean-variance
criterion approximates correctly the expected utility correctly except for large departures from normality or for highly levered portfolios. Indeed, Das and Uppal (2004) show that, in presence of jumps occurring at the same time across countries, the loss from the reduction in diversification is not substantial and that the cost of ignoring common jumps is large only for highly levered positions. Jondeau and Rockinger (2006) report that the mean-variance criterion fails to approximate the expected utility if assets are characterized by highly asymmetric and fat-tailed distributions. In such a case, three- and four-moment optimization strategies provide better approximations of the expected utility. These studies, therefore, conclude that the mean-variance framework may fail, but only in extreme cases.

This conclusion is, however, reached under the assumption that the distribution of the opportunity set is constant through time, while a recent literature demonstrates that the distribution of asset returns is time varying and at least partially predictable (Hansen, 1994, or Harvey and Siddique, 1999, Jondeau and Rockinger, 2003, Brooks, Burke, and Persand, 2005). Extension to a conditional set-up, however, is a difficult task due to the fact that returns need to be modeled by a multivariate density with complex shapes. A first avenue was followed by Ang and Bekaert (2002) and Guidolin and Timmermann (2005) using a switching-regime approach. In this framework, returns' mean and variance change depending on the regime. Although the returns' distribution is conditionally normal, the model provides a measure of the opportunity cost of assuming iid normal returns rather than time changing first two moments. Ang and Bekaert (2002) report that the cost of ignoring regime shifts can be large even in the presence of a risk-free asset. It should be noticed that although this approach is appropriate to investigate the consequences of time-varying first and second moments, it is less convenient for addressing the consequences of time-varying higher moments. The reason is that the conditional higher moments are implied by the dynamics of the first two moments, so that they cannot be modeled separately. Guidolin and Timmermann (2005) introduce several approximations of the utility function using up to the fourth moment of returns, and show that taking higher moments into account improves the allocation of assets significantly.
In this paper, we propose a solution to the asset-allocation problem when the joint conditional distribution of returns is non-normal and varies over time. Modeling asset returns in such a general setting requires some distributional assumption that capture the volatility clustering, asymmetry and fat-tailedness features found in the data. The building blocks of our model to describe asset returns are the following: volatility clustering is captured using a multivariate GARCH model with dynamic conditional correlation (DCC) (Engle and Sheppard, 2001). We propose an extension to this model by modeling innovations as a multivariate skewed Student $t$ (Sk-$t$) distribution (see also Sahu, Dey, and Branco, 2003). Such a model allows us to also incorporate time variation of higher moments. The proposed framework is sufficiently general to reproduce the statistical features of returns and to remain tractable even when several assets are included. A clear advantage of this model is that moments of returns’ distribution can be computed analytically. The asset allocation builds on a high-order Taylor expansion of the utility function, so that the optimization program only depends on returns’ moment forecasts. Consequently, once the multivariate model has been estimated, rather general asset allocation problems can be addressed, in real time, even with a large number of assets.

We apply our approach to the weekly allocation of wealth between the three largest stock markets, i.e. the US, Japanese, and UK markets. Weekly market returns are known to be non-normally distributed, thus requiring an extension of the standard mean-variance setting. We show that our model is able to capture the main statistical characteristics of these market returns. In addition, we find that the mean-variance criterion results in an excessive risk taking and a significant opportunity cost, as compared to a strategy based on higher moments. The performance fee an investor would be willing to pay to switch from a static to a mean-variance dynamic strategy (volatility timing) is about the same as the fee she would be willing to pay to switch from the mean-variance dynamic strategy to the fully dynamic strategy (distributional timing). On average for various levels of risk aversion, the economic value of the volatility and distributional timing is around 60 basis points per year (to be compared with an expected return of about 5% over the allocation period).
The outline of the paper is as follows. In Section 2, we present the methodology adopted to measure the economic value of non-normality. We describe how the conditional asset-allocation problem can be solved when returns are non-normal. We also demonstrate how portfolio moments can be computed in a very efficient way. In Section 3, we formulate our approach for modeling returns with a non-normal multivariate distribution. In Section 4, we present the data and discuss the results of the estimation of the model. In Section 5, we consider the case of time-varying investment opportunities and examine the consequences of using the mean-variance criterion under strong departure from normality. In Section 6, we provide several robustness checks of our main results. Section 7 concludes. Several appendices contain the details of the statistical model describing the evolution of returns as well as preliminary statistical estimations.

2 Methodology

This section describes the conditional asset allocation problem with non-normal returns. In this context, the standard mean-variance criterion is likely to be inappropriate to select the optimal portfolio. Incorporating the effect of higher moments on the expected utility of investors is likely to improve the allocation of wealth (Harvey and Siddique, 2000, and Dittmar, 2002). As in Guidolin and Timmermann (2005), we approximate the expected utility up to the fourth moment and obtain the optimal asset allocation. Given that our econometric model provides us with a time varying conditional distribution of returns, combining it with an optimal portfolio allocation model yields time-changing allocations. This gives rise to what we call distributional timing. Restrictions of the most general econometric model, shutting down temporal variation or skewness and kurtosis allows us to gauge the relative importance of the various components.

2.1 The multivariate return process

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

\[ r_t = \mu_t(\theta) + \varepsilon_t, \quad (1) \]
\[ \varepsilon_t = \Sigma_t(\theta)^{1/2} z_t, \quad (2) \]
\[ z_t \sim g(z_t|\eta_t). \quad (3) \]

Equation (1) decomposes the return at time \( t \) into two \( n \times 1 \) vectors, i.e. the conditional mean, \( \mu_t \equiv \mu_t(\theta) = E[r_t|I_{t-1}] \), and the unexpected return, \( \varepsilon_t \). Equation (2) indicates that the unexpected return \( \varepsilon_t \) is defined as the product of independent innovations \( z_t \) and the conditional covariance matrix of returns, \( \Sigma_t \equiv \Sigma_t(\theta) = E[(r_t - \mu_t)(r_t - \mu_t)'|I_{t-1}] \). The \( n \times 1 \) vector of independent innovations, \( z_t = \Sigma_t^{-1/2} (r_t - \mu_t) \), has zero mean and identity covariance matrix. We denote by \( \Sigma_t^{1/2} \) the Choleski decomposition of \( \Sigma_t \). The vector \( \theta \) contains all the parameters associated with the conditional mean and the conditional variance equations. Last, equation (3) specifies that innovations \( z_t \) follow a conditional distribution \( g \) with (possibly) time-varying shape parameters, \( \eta_t \). When the conditional distribution is normal, there is no need for a shape parameter since the normal distribution is entirely characterized by its mean and variance. In more general cases, shape parameters typically involve parameters capturing asymmetry and fat-tailedness of the distribution.

In Appendix 1, we describe a general model, in which expected returns, volatilities and the joint conditional distribution itself are allowed to vary over time. This setting supports most statistical features of stock market returns and appears to be well-suited for solving the asset allocation problem presented below. The model we use accounts for the well-known properties of volatility clustering (Engle, 1982) and time-varying correlations (Engle and Sheppard, 2001). As a matter of fact, our model is an extension of this model, where innovations are assumed to be drawn from a multivariate Sk-\( t \) distribution. Thereby, we capture both asymmetry and fat-tailedness, often found in actual data. This distribution has been developed, in a univariate context, by Hansen (1994) and studied by Jondeau and Rockinger (2003). It has been extended to the multivariate setting by Sahu, Dey, and Branco (2003). A nice feature of the distribution is that it is a straightforward extension.
of the normal and $t$ distributions and that the associated parameters have a rather natural interpretation. Last but not least, this distribution appears to fit the data very well.

### 2.2 Distributional timing

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 $W_{t+1}$. We do not consider a multi-period investment problem since Brandt (1999) as well as Ang and Bekaert (2002) show that even if portfolio weights may be slightly affected by the investment horizon, the opportunity cost of a myopic strategy is negligible, at least for a relative limited time horizon. This result suggests that hedging against unfavorable changes in the investment set does not result in any significant gain. The initial wealth $W_t$ is arbitrarily set equal to one. There are $n$ risky assets with return vector $r_{t+1} = (r_{1,t+1}, \cdots, r_{n,t+1})'$ and a risk-free asset with return $r_{f,t}$ for the period between $t$ and $t + 1$. Return is defined as the simple rate of return of asset $i$ from time $t - 1$ to time $t$. End-of-period wealth is $W_{t+1} = 1 + r_{p,t+1}$, where $r_{p,t+1} = r_{f,t} + \alpha_t' (r_{t+1} - r_{f,t})$ denotes the portfolio return and $\alpha_t = (\alpha_{1,t}, \cdots, \alpha_{n,t})'$ the vector of fractions of wealth allocated to the various risky assets. Portfolio weights are constrained to be positive, so that short-selling is not allowed. We assume also that the investor has forecasts for the expected mean vector $\mu_{t+1}$, the covariance matrix $\Sigma_{t+1}$, and possibly the higher order co-moment matrices.

Formally, optimal portfolio weights at time $t$ are obtained by maximizing the expected utility

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

s.t. $0 \leq \alpha_{i,t} \leq 1, \forall i, \sum_{i=1}^{n} \alpha_{i,t} \leq 1$,

where $E_t [\cdot]$ is the expectation conditional on information available at time $t$ and $e = (1, \cdots, 1)'$ denotes the $n \times 1$ vector of ones. For non-normal returns, equation (4) generally does not have a closed-form solution and numerical techniques must be used. Gallant and Tauchen (1989) and Tauchen and Hussey (1991) suggest using
quadrature rules to solve this problem. This approach has been applied to normal iid returns (see Campbell and Viceira, 1999) or to regime-switching conditionally normal returns (Ang and Bekaert, 2002). Non-normal distributions may require a number of quadrature points that increases exponentially with the number of assets, so that solving the optimization problem using numerical integration is often intractable for more than two or three assets.

Since we are primarily interested in measuring the effect of higher moments on the asset allocation, we follow an alternative approach that approximates the expected utility as a function of the moments of the portfolio return. Such an approach has been adopted in a large number of contributions, see Kraus and Litzenberger (1976), Pratt and Zeckhauser (1987), Harvey and Siddique (2000), Dittmar (2002), Guidolin and Timmermann (2005), and Jondeau and Rockinger (2006). The utility function can be written as an infinite-order Taylor series expansion of the form

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

where $W_{t+1} - W_t = r_{f,t} + \alpha_t'(r_{t+1} - r_{f,t}e) = r_{p,t+1}$ denotes the portfolio return vector at date $t + 1$. Under rather mild conditions, the expected utility is given by

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

Therefore, the expected utility depends on all the moments of the distribution of the end-of-period wealth.

Necessary conditions for the infinite Taylor series expansion to converge to the expected utility have been explored by Loistl (1976) and Lhabitant (1998). The region of convergence of the series depends on the utility function considered. In particular, the exponential or polynomial utility functions do not put any restriction on the wealth range, while the power utility function converges 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 when short-selling is not authorized. In contrast, it may be too small for options, due to their leverage effect. These results hold for arbitrary return distributions.

This expression of the expected utility by a Taylor series expansion is directly
related to the investor’s preference (or aversion) towards all moments of the distribution, that are directly given by derivatives of the utility function. Scott and Horvath (1980) show that, under certain conditions, the following inequalities hold:\(^1\)

\[
U^{(k)}(W_{t+1}) > 0 \quad \forall W_{t+1} \quad \text{if } k \text{ is odd and}
\]

\[
U^{(k)}(W_{t+1}) < 0 \quad \forall W_{t+1} \quad \text{if } k \text{ is even.}
\]

Further discussion on the conditions that yield such moment preferences or aversions may be found in Pratt and Zeckhauser (1987), Kimball (1993), and Dittmar (2002).

Now, since the infinite Taylor series expansion is not suitable for numerical implementation, the expected utility can be approximated by truncating the infinite expansion at a given value. Although it is not clear a priori whether there exists an optimal order, we can give a general condition such that the inclusion of an additional moment will improve the quality of the approximation. Indeed, under rather mild assumptions, a general condition for the smoothness of the convergence of the Taylor series expansion is that preference-weighted odd central moments are not dominated by their consecutive preference-weighted even central moments, so that

\[
\frac{U^{(2n+1)}(W_t)}{2n + 1} E[r_{p,t+1}]^{2n+1} < -\frac{U^{(2n+2)}(W_t)}{2n + 2} E[r_{p,t+1}]^{2n+2},
\]

with \(n\) integer. In this case, including skewness and kurtosis always leads to a better approximation of the expected utility.

In the following, we focus on an approximation up to the fourth order. The main reason is that our econometric model is designed to capture the dynamics of the third and fourth moments through the modeling of the asymmetry and degree-of-freedom parameters. The dynamics of higher-order moments would not be independent from the one of the third and fourth moments. It should be noticed that a fourth-order approximation of the expected utility should be enough to highlight the economic value of distributional timing. A modeling of the dynamics of the higher-order

\(^1\)These conditions are positive marginal utility, decreasing absolute risk aversion at all wealth levels, and strict consistency for moment preference. An investor exhibits strict consistency for moment preference if a given moment is always associated with the same preference direction regardless of the wealth level.
moments would probably only reinforce the extent of distributional timing. Focusing on terms up to the fourth order, we obtain the following expected utility

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

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

where \(m_{p,t+1}^{(i)} = E_t[r_{p,t+1}]\) denotes the non-central moments of order \(i\).

Let us consider an investor with a CRRA utility function:

\[
U(W_{t+1}) = \begin{cases} 
W_{t+1}^{1-\gamma}/(1-\gamma) & \text{if } \gamma > 1, \\
\log(W_{t+1}) & \text{if } \gamma = 1,
\end{cases}
\]

where \(\gamma\) measures the investor’s constant relative risk aversion. For this utility function, expression (5) becomes

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

We obtain an unambiguous effect of the third and fourth moments on the approximated expected utility function. Expected utility decreases with large negative skewness (i.e. left-skewed distributions) and large kurtosis (i.e. fat-tailed distributions). These effects are consistent with the theoretical arguments developed by Scott and Horvath (1980). In Appendix 2, we explain how to maximize the fourth-moment expected utility (6) in a very efficient way, when we are given expected returns and co-moment matrices.

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

\[
m_{p,t+1} = \mu_{p,t+1}, \\
m_{p,t+1}^{(2)} = \sigma_{p,t+1}^2 + \mu_{p,t+1}^2, \\
m_{p,t+1}^{(3)} = s_{p,t+1}^3 + 3\sigma_{p,t+1}^2 \mu_{p,t+1} + \mu_{p,t+1}^3, \\
m_{p,t+1}^{(4)} = \kappa_{p,t+1}^4 + 4s_{p,t+1}^3 \mu_{p,t+1} + 6\sigma_{p,t+1}^2 \mu_{p,t+1}^2 + \mu_{p,t+1}^4,
\]

and where \(\sigma_{p,t+1}^2, s_{p,t+1}^3\) and \(\kappa_{p,t+1}^4\) stand for central moments \(E_t[r_{p,t+1} - \mu_{p,t+1}]^i\) for \(i = 2, 3\) and 4 respectively. Central moments \(s_{p,t+1}^3\) and \(\kappa_{p,t+1}^4\) should not be confused with skewness \(sk_{p,t+1}\) and kurtosis \(ku_{p,t+1}\) defined as standardized central moments \(E\left[\left((r_{p,t+1} - \mu_{p,t+1})/\sigma_{p,t+1}\right)^i\right]\) for \(i = 3, 4\).
Maximizing expression (6) for each date $t$ clearly defines a dynamically rebalanced portfolio that maximizes the expected utility of the investor. Optimal weights at date $t$ will therefore be functions of the higher central moments of the portfolio for date $t + 1$ and consequently of the higher co-moment matrices of the risky assets for date $t + 1$. Presently, we are endowed with a model for asset returns, a model for asset allocation, thus, we may develop a protocol for investigating the relevance of various features of the most general model.

2.3 Allocation strategies

To evaluate the economic gain of the volatility and distributional timing, we describe strategies that reflect such timing ability. The models we consider have constant conditional mean, assuming no ability to forecast changes in expected returns. The reason for this choice is that there is very little information content in past variables regarding future returns. Some recent papers have shown that macroeconomic variables have predictive power for monthly returns (see, e.g., Pesaran and Timmermann, 1995, Marquering and Verbeek, 2001), but no clear evidence has been provided up to now regarding weekly returns. Han (2006) also fails to find an economic value of return timing for weekly or monthly allocation horizons, while he obtains a very sizeable effect for a daily horizon. We therefore follow Fleming, Kirby, and Ostdiek (2001, 2003) and assume constant expected returns to concentrate on the changes in the covariance matrix and the conditional distribution.

To avoid the problem with in-sample overfitting as well as spurious findings, we decided to gauge all performance measures out of sample. Thus, we will estimate the model over a first subsample, referred to as the estimation period, then we will take the second subsample, referred to as the allocation period, to measure the portfolio performance.

For convenience, the main characteristics of the various strategies are summarized in Table 1. The first strategy we consider is the static (or buy-and-hold) strategy. The allocation holding for this strategy is obtained by computing constant first and second moments over the estimation period. Then, using the mean-variance criterion, the optimal allocation is obtained. This allocation is maintained and the
performances of the resulting portfolio are investigated. The investor assumes constant investment opportunities and adopts a mean-variance criterion. Thus, the expected returns and the covariance matrix are estimated using sample moments over the estimation period. The conditional distribution is normal. In the second strategy, the investor still adopts a mean-variance criterion, but she assumes a time-varying conditional covariance matrix. Thus a DCC model is estimated under the assumption of a joint normal distribution. This strategy will be denoted by $MV^d$. The performance of the $MV^d$ strategy relative to the static strategy provides a measure of the economic value of volatility timing.

To investigate the effect of the conditional distribution on the asset allocation, we consider two strategies where innovations are distributed as a Sk-$t$ distribution. In the first strategy, we keep shape parameters constant over time. This would mean that the investor does not pay attention to distributional timing and his focus would therefore only be the effect of non-normality, thus higher moments, on the asset allocation. For these reasons, we label this strategy with $HM^c$. In the last strategy, shape parameters are themselves time varying, so that the investor takes full advantage of distributional timing as well as of non-normality. We name the strategy based on this latter model $HM^d$ (for dynamic higher moments).

### 2.4 Performance measures

To compare the performance of the various strategies, we consider a set of performance measures. Most of these measures are well known and have been widely used in the previous literature. For a description of these measures, we refer to Fleming, Kirby, and Ostdiek (2001) and Han (2006).

A first measure of performance is the standard Sharpe ratio, which is computed using the ex-post average return $\mu_p$ and volatility $\sigma_p$, as $SR_p = (\mu_p - r_f) / \sigma_p$. Since the Sharpe ratio does not provide a measure of out-performance over alternative strategies, we also consider the performance measure $M2$ introduced by Modigliani and Modigliani (1997)

$$M2 = \frac{\sigma_0 \left( \mu_p - r_f \right) - (\mu_0 - r_f)}{\sigma_p \left( \mu_p - r_f \right)}.$$
where \( \mu_0 \) and \( \sigma_0 \) are the average return and volatility of the benchmark portfolio (in our set-up, the static strategy). This measure corresponds to a scaled difference of the prices of risk of the two allocations under comparison.

These measures have however some obvious drawbacks in our context. First, they assume that portfolio volatility is constant. Second, they do not take the effect of non-normality into account. We therefore consider a second tool to evaluate the economic value of volatility and distributional timing, namely the performance fee measure proposed by Fleming, Kirby, and Ostdiek (2001). It measures the management fee an investor would be willing to pay to switch from the static strategy to one of the dynamic strategies. If we denote \( r^*_p,t+1 \) the optimal portfolio return obtained under a given dynamic strategy, and \( \hat{r}_p,t+1 \) the optimal portfolio return obtained using the static strategy, then the performance fee (or opportunity cost), denoted \( \theta \) and obtained numerically, is defined as the average return that has to be subtracted from the return of the dynamic strategy, so that the investor becomes indifferent between both strategies\(^3\)

\[
E_t \left[ U \left( 1 + \hat{r}_p,t+1 \right) \right] = E_t \left[ U \left( 1 + r^*_p,t+1 - \theta \right) \right].
\]  

We also compute the success rate, i.e. the percentage of times in which the return of the dynamic strategy exceeds the static strategy. This measure, denoted \( Z \), is useful to check whether the out-performance of the dynamic strategy is due to some very specific events or to a better ability to capture changes in investment opportunities. For instance, if this statistic is very small, it means that the out-performance is essentially due to a few happy instances.\(^4\)

Finally, since we compare static and dynamic strategies, it is important to cap-

\(^3\)We also considered the certainty equivalent, previously adopted among others by Kandel and Stambaugh (1996), Campbell and Viceira (1999), Ang and Bekaert (2002), and Das and Uppal (2004). It is defined as the compensation (in percentage of initial wealth) an investor must receive so that she becomes willing to put one dollar in the sub-optimal strategy rather than in the optimal one. Since the performance fee and the certainty equivalent provided the same measure of the economic gain (up to a few basis points), we only report the former in our empirical evidence.

\(^4\)We also investigated several alternative performance measures. Since they all provided the same patterns as reported in the following, we do not report them to save space. They are however available upon request from the authors.
ture the possible effect of transaction costs. Indeed, since by construction, the static strategy does not generate large transaction costs, the gain of a dynamic strategy may be partly offset by transaction costs. In practice, it is however rather difficult to estimate these transaction costs since a wide range applies depending on the asset and the type of customer relation. For this reason, we follow the approach of Han (2006) and measure the breakeven transaction cost. This is the level of transaction costs that makes the investor indifferent between the dynamic strategy and the static strategy. If we assume that transaction costs are equal to a fixed fraction $\tau$ of the value traded for all stocks in the portfolio, the average weekly transaction cost of this strategy is $\tau tc$, where

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

where $\alpha_{i,t}$ denotes the weight of asset $i$ at date $t$ and $r_{p,t}$ is the portfolio return at date $t$. Finally, the breakeven transaction cost between the dynamic strategy $d$ and the static strategy $s$ is defined as

$$\tau^{be} = \frac{\bar{r}_p^d - \bar{r}_p^s}{tC^d - tC^s}.$$  \hspace{1cm} (8)

In other words, if the investor has a transaction cost lower than $\tau^{be}$, she will prefer the dynamic strategy to the static strategy.

3 Data and preliminary analysis

3.1 Data

We consider the three largest international markets: the United States, Japan, and the United Kingdom.\textsuperscript{5} The asset allocation problem is viewed from the perspective of an unhedged US investor, so that returns are expressed in US dollars. The risk-free rate is the 7-day US Treasury bill rate. The data are weekly and cover the period from January 1977 through December 2005, for a total of 1510 observations.\textsuperscript{5}

\footnotesize{At the end of 2005, the US, Japanese, and UK markets represent respectively 41.3%, 18.4%, and 7.5% of the world market capitalization. Market returns are measured by the return on the indices S&P 500, Nikkei, and FTSE 100 respectively.}
This sample period is broken in two subsamples: the first sample (from 1977 to 2001, 1304 observations) is used for the estimation of the model, while the second sample (from 2002 to 2005, 206 observations) is used for an out-of-sample investigation.

Table 2 reports several summary statistics for the market returns under study (Panel A). Average returns are all positive and significant, ranging between 0.164% and 0.237% per week. Volatilities range between 2.126% and 3% per week. The US market yields a moderate return, with a low volatility. In contrast, the Japanese market is characterized by a low return and a high volatility. For a mean-variance investor, the former would be probably preferred to the latter.

Concerning higher moments, we notice a rather large dispersion in the magnitude of skewness across markets. US and UK returns are strongly negatively skewed (about $-0.4$), suggesting that crashes occur more often than booms. The Japanese market, on the other side, has a large positive skewness (0.182), suggesting that booms are more likely to occur than crashes.\(^6\) A high level of kurtosis is found for all markets, a result that is not consistent with the normality assumption: it ranges between 5 and 7.4, while normality would give a kurtosis of 3. It may be argued that the large skewness and kurtosis observed over the sample are affected by the October 1987 crash. If we remove the entry corresponding to the week of the crash (Panel B), we obtain a significant increase of the skewness for the US market (to $-0.25$) and for the UK (to 0.036). In addition, the kurtosis of the UK market return decreases to 4.116 (instead of 7.447). The shape of the distribution is not affected for Japanese returns. For this market, removing the observation of the October 1987 crash barely decreases the sample kurtosis. We concluded from these statistics, that the crash may have a dramatic effect on the shape of the distribution and on the dynamics of the higher moments. For this reason, we removed the week containing the 1987 crash from our sample.\(^7\)

\(^6\)The positive skewness of the Japanese market return suggests that the multivariate model with jumps occurring at the same time across countries, as proposed by Das and Uppal (2004), is probably at odds with our data. In their dataset on developed countries, the skewness of the Japanese market is very close to 0.

\(^7\)We also investigated the consequences of keeping this observation in the estimation sample. The main effect was to increase the standard error of parameter estimates.
We also tested for normality and were able to reject this assumption with great confidence in all cases and this with or without the 1987 crash. Similarly, we tested the null of no autocorrelation and of no heteroscedasticity. We could not reject, at the weekly frequency, the assumption that the data is not autocorrelated but, as would be expected, we found evidence for heteroscedasticity.

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

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

\[
s_{ijk} = \frac{E[(r_{i,t} - \mu_i)(r_{j,t} - \mu_j)(r_{k,t} - \mu_k)]}{\sigma_i \sigma_j \sigma_k},
\]

(9)

\[
k_{ijkl} = \frac{E[(r_{i,t} - \mu_i)(r_{j,t} - \mu_j)(r_{k,t} - \mu_k)(r_{l,t} - \mu_l)]}{\sigma_i \sigma_j \sigma_k \sigma_l},
\]

(10)

where \( \mu_i \) and \( \sigma_i \) denote the sample mean and standard deviation of return \( i \) respectively. Expressions \( s_{iii} \) and \( k_{iiii} \) are the standard measures of individual skewness and kurtosis (reported in Panel A). Co-skewness \( s_{iji} \) measures the strength of the link between the squared deviation of asset \( i \) and the deviation of asset \( j \) from their respective means. A negative value indicates that asset \( j \) will yield a low return when the volatility of asset \( i \) is high, therefore, asset \( j \) does not provide a good hedge against an increase in the volatility of asset \( i \). The table shows that this co-skewness cannot be distinguished from 0. This suggests that it is not possible for the investor to hedge her portfolio against increases in the volatility of any of the three markets under study. Such a result was expected, given that bear markets generally combine low returns and high volatilities in all markets.

Co-kurtosis \( k_{iiij} \) measures the strength of the link between the squared deviations of the two assets \( i \) and \( j \) from their means. A large value would mean that
the volatility of asset $i$ increases as the volatility of asset $j$ increases (or vice-versa). Both assets would provide bad hedges against high volatility in the other market. Co-kurtosis $ku_{iij}$ measures the strength of the link between the third power of the deviation of asset $i$ and the deviation of asset $j$ from their means. A large value would mean that the distribution of asset $i$ becomes more negatively skewed when the return of asset $j$ is lower than expected, therefore providing a bad hedge against a fall in asset $j$. Similarly, a large co-kurtosis $ku_{iijk}$ indicates that the covariance between assets $j$ and $k$ increases as the volatility of asset $i$ increases. Since this co-kurtosis is positive, it means that in general, volatility increases in one country go together with increases in dependency elsewhere. From the point of view portfolio allocation, this is obviously a bad news. Portfolios where this measure will be downweighted ought to be preferred.

As for the standard individual kurtosis, it is customary to compare co-kurtosis with the level a multivariate normal distribution would imply. The table indicates that most co-kurtosis are significantly larger than what would be expected from normal returns. In particular, the US and UK markets provide bad hedges against adverse return or volatility in the other market. A similar result holds for the UK and Japanese markets.

### 3.2 Estimation of the model

The parameter estimates for the general model are relegated to Appendix 1. In this section we only wish to discuss those results that are not so well documented in the literature.

**Figures 1 and 2** display the dynamics of conditional higher moments. One of the most striking results is that the variability in the individual skewness and kurtosis is rather large for Japan but moderate for the US and the UK. From the Japanese sample moments, one could have concluded that the distribution of Japanese returns is not so far from the normal one. The large variability of skewness and kurtosis shows however that this finding is due to averaging and that allocating

---

8In the normal case, all co-skewness are equal to zero, while co-kurtosis are given by $ku_{iii} = 3$, $ku_{iij} = 1$ for $i \neq j$, and 0 for all other cases, see Kotz, Balakrishnan, and Johnson (2000).
wealth on the basis of the sample moments would be misleading. It follows that the temporal variability of moments may play an important role.

We notice that all the co-skewness between the US and Japanese markets and between the US and UK markets are negative. This implies that the US market provides a bad hedge against adverse changes in volatility in the Japanese and UK markets (and vice-versa). On the opposite, the co-skewness between the Japanese and UK markets are positive, suggesting that the return in one market is likely to be larger than expected when the volatility in the other market is high, thus providing a good hedge against high volatility.

The figures also show that most co-kurtosis are positive and well above the value predicted by a normal distribution. In addition, we notice a positive trend in the co-kurtosis between the US and UK markets, suggesting that the ability of these two markets to hedge each other is worsening through time.

4 Portfolio analysis

We begin our analysis of the allocation results with some characteristics of the optimal portfolio weights implied by the various strategies under study. For each week of the sample, we forecast the first four moments and co-moments of market returns using the various models described above and maximize the approximated expected utility function (6) to produce portfolio weights. Summary statistics on portfolio weights are reported in Table 3. If one focuses, for instance, on the weights obtained for a risk aversion of $\gamma = 5$, the optimal ex-ante static portfolio is mainly composed of the US (49.4%) and UK (42.5%) market returns, while the weight of the Japanese market is 7.3% only, the remaining weight being invested in the risk-free asset. The main reason for the small weight attributed to Japan is that over the sample it has a small average return and a high volatility.\footnote{Notice that weights of the static strategy are not necessarily held constant, because the risk-free rate varies over time.}

When the covariance matrix is allowed to vary over time ($MV^d$ strategy), the weight of the UK market decreases significantly (to 34.3% on average), because the
increase in the correlation with the US market over the allocation period makes the UK market a poor hedge against adverse return shocks in the US market. At the same time, the weight of the Japanese market importantly increases (to 28.1%), because the time-varying volatility of the Japanese market is on average much lower over the allocation period than over the estimation period.

When (constant) non-normality of returns is taken into account (HM$^c$ strategy), we only observe little changes as compared to the MV$^d$ strategy. As reported in Table 2, the US return is negatively skewed, while the Japanese return displays positive skewness. This strategy therefore puts more weight on the Japanese market (relative to the US market) than the MV$^d$ strategy, in order to benefit from a more positively skewed portfolio return distribution. However, the tails of the distribution of the Japanese market return are also fatter than those of the US market return. As a consequence, the strategy should not put too much a weight on the Japanese market, because turbulent periods are more likely to occur. All in all, the final effect is rather limited.

In contrast, when higher moments are allowed to vary over time (HM$^d$ strategy), we obtain sizeable differences in portfolio weights. The reason is once again related to the properties of Japanese returns. As shown in Figure 2, at the end of the allocation period, the kurtosis of this series increases significantly. At the same time, the kurtosis of the UK market decreases below its average level. Consequently, the investor significantly reduces the weight of the Japanese market and increases accordingly the weight of the UK market. Last the optimal weight of the US market is rather small, given the persistently negative skewness of this market and the increase of its kurtosis at the end of the allocation period.

As expected, when the risk aversion increases, the weight of the risk-free asset increases significantly. While the static portfolio is fully invested in equity for $\gamma = 2$, the share of the US T-bill is as high as 62.3% for $\gamma = 15$, so that the risk of the optimal portfolio is much lower. Interestingly, among all strategies, HM$^d$ requires the smallest increase of the US T-bill weight (to 56%), suggesting for this strategy a better control of the overall risk of the portfolio without requiring a reduction in the share of risky assets.
Considering the average weights obtained for the three dynamic strategies, one may argue that the differences are all in all rather small. The evolution of the portfolio weights over the out-of-sample period is displayed in Figure 3 (for $\gamma = 5$). It shows that quite large differences in the portfolio weights are obtained between the dynamic strategies, in particular over the most recent period. It should also be emphasized that the actual differences are mitigated by averaging over the allocation period.

To obtain a quantitative tool to measure the difference in the allocations, we presently define two measures. A first measure of the distance between two series of weights can be defined as $\text{Dist} = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t}^{*} - \hat{\alpha}_{i,t}| / 2$, where $\hat{\alpha}$ denotes weights of the static strategy, while $\alpha^{*}$ denotes weights of a dynamic strategy. Similarly, we measure turnover from one week to the next using the measure $\text{Turn} = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t} - \alpha_{i,t-1}| / 2$.\(^{10}\)

For $\gamma = 5$, Table 3 shows that the average distance between the weights of the static and the $HM^d$ strategies is 0.137, i.e. the difference between the weights of the two strategies amounts to 14 percentage points on average for the whole portfolio. We also observe that the higher the risk aversion, the smaller the discrepancy between the static strategy and the dynamic strategies. The reason is that the investor is more reluctant to take risks and therefore invests more in the risk-free asset. For instance, the distance between the static strategy and the $HM^d$ strategy decreases from 15.7 percentage points for $\gamma = 2$ to 8.7 percentage points for $\gamma = 15$.

Besides, the turnover ratio is at most 0.04 whatever the strategy and the risk aversion, indicating that each week the investor re-allocates around 4% of her portfolio. This suggests that transaction costs should not be an issue for comparing the performance of the dynamic strategies with respect to the static one. Similarly to above, we notice that as the risk aversion increases, the turnover ratio decreases, because the investor increases her diversification to reduce her exposition to risk and therefore becomes closer and closer to the $1/n$ rule.\(^{11}\) Consequently, the composition of the portfolio is less affected by changes in investment opportunities.

\(^{10}\)The division by 2 corrects for double counting.

\(^{11}\)The $1/n$ rule consists in allocating an equal fraction of wealth to each asset. It has been recently advocated by DeMiguel, Garlappi, and Uppal (2006).
5 Performance analysis

We now analyze the performance of the various dynamic trading strategies described above. Each strategy shall provide some insight on the economic value of the volatility and distributional timing. We therefore compare the out-of-sample performance of the dynamic strategies to that of the static strategy, which corresponds to the ex-ante optimal static strategy based on the sample mean and covariance over the estimation period. We also compute the performance fee an investor is willing to pay to switch from a suboptimal strategy A to an optimal strategy B, assuming that strategy B corresponds to a correct approximation of the true DGP of asset returns. In the following, we will discuss the results displayed in Table 4. That table reports several performance measures for the various strategies and levels of risk aversion $\gamma$. In the following section we will discuss the gain from switching from the static to the $MV^d$ strategy. Subsequent sections will be devoted to the passage to the $HM^e$ and $HM^d$ strategies.

**Economic value of volatility timing.** The focus of this section is to investigate the economic gain of volatility timing on our data. It is measured by comparing the performance of the static and $MV^d$ strategies. Both strategies assume a mean-variance criterion. But while the former assumes a constant covariance matrix (expected returns are always held constant), in the latter the investor forecasts the covariance matrix at each date. As the table reveals, all the reported statistics point in favor of the dynamic strategy. First, $MV^d$ provides a higher realized return than the static strategy: for a risk aversion of $\gamma = 5$, the annualized return is equal to $\mu = 6.63\%$ for the $MV^d$ strategy against $5.69\%$ for the static strategy. Since realized volatility is also lower ($\sigma = 12.53\%$ against $12.87\%$), this results in a higher Sharpe ratio ($SR = 0.39$ against $0.30$).\(^{12}\) Accordingly, the risk-adjusted excess return of the dynamic strategy measured by $M2$ is as high as 108 basis point per year.

A static investor is willing to pay to switch to the dynamic strategy. For $\gamma = 5$, it is equal to $\theta = 126$ basis points per year. The value of this fee can be broken down

\(^{12}\)Notice that in addition realized skewness is also larger and kurtosis is lower, suggesting an improvement of all the characteristics of the realized returns’ distribution.
into two components. First, the $MV^d$ strategy provides the investor an excess return of 94 bp with respect to the static strategy. Second, the $MV^d$ strategy reduces the risk exposure of the portfolio, as measured by volatility and the higher moments: it reduces portfolio volatility and also improves the characteristics of the portfolio return’s distribution, since skewness increases and kurtosis decreases. The extra fee the investor is willing to pay for this improvement is equal to 32 bp. To sum up, the 126 bp fee the investor is willing to pay corresponds to 94 bp for the increase in expected return and 32 bp for the improvement in the volatility and distribution.\footnote{The last component should not be confused with volatility or distributional timing. By adopting the $MV^d$ strategy, the investor is at the same time able to improve the distributional properties of her portfolio, even if she does not take explicitly care for it.}

For alternative levels of risk aversion, the dynamic strategy still out-performs the static strategy. The Sharpe ratios of the static and dynamic strategies are equal to 0.33 and 0.35 for $\gamma = 2$ and to 0.32 and 0.41 for $\gamma = 10$. The performance fees are found to be equal to 21 bp for $\gamma = 2$ and 51 basis points for $\gamma = 10$. It is slightly lower than for $\gamma = 5$. An interpretation of this result is that the dynamic strategy is not able to reduce the realized volatility for low and large risk aversions, therefore reducing the performance fee the static investor is willing to pay for switching her strategy.

Another interesting result is that the success rate markedly increases when the level of risk aversion increases. For $\gamma = 2$, it is as low as $Z = 0.47$, suggesting that the gain of the dynamic strategy may be due to some specific dates. For $\gamma = 10$, in contrast, we have $Z = 0.56$, indicating that the outperformance of the dynamic strategy is not due to just a few lucky guesses but to an improvement in the allocation method.

The evidence presented above has been obtained in the absence of transaction costs. Clearly, the gain of the dynamic strategy would be smaller in presence of such costs. The last column of the table reports the breakeven transaction cost, $\tau^{be}$, of the $MV^d$ strategy. It ranges between 5 bp for low risk aversion to 12 bp for large risk aversion. These values are larger than the transaction costs typically charged for an individual portfolio.
Our empirical evidence regarding volatility timing is similar to that reported for instance by Fleming, Kirby, and Ostdiek (2001) and Han (2006), although using different assets, sample periods, and forecasting models.

**Economic value of distributional timing.** Returns’ non-normality may affect the optimal asset allocation along two directions. First, a rational investor should take non-normality into account, because it is likely to affect her expected utility. For instance, everything else being equal, the investor dislikes assets with a negative skewness and a large kurtosis, because they are more prone to crashes than positively skewed assets. This first effect is not related to the conditional distribution being time dependent. If in addition the conditional distribution is time varying, the investor should forecast changes in the conditional distribution and allocate her wealth accordingly. This second effect, which we call distributional timing, reflects the ability of the investor to anticipate changes in the returns’ distribution.

Table 4 further reveals that the two effects of non-normality play a role in the performance of dynamic strategies. The $HM^c$ strategy measures the economic gain of an investor’s ability to take non-normality into account. The performance fee a static investor is willing to pay to switch to the $HM^c$ strategy is equal to 171 basis points for $\gamma = 5$. Taking non-normality into account therefore implies an additional increase of the fee by 45 basis points as compared to the $MV^d$ strategy. In the same vein, as compared to the $MV^d$ strategy, the Sharpe ratio increases to 0.42 from 0.39, while the $M2$ measure increases from 1.08 to 1.52. The main change in the portfolio allocation that explains the increase in the performance measures is the significant increase in the realized return (from 6.63% to 7.03% per year). This suggests that the $HM^c$ strategy is able to benefit from the positive skewness of the ex-ante portfolio return distribution. Since for some dates some unexpected positive events occur, we do not observe such a positive skewness in the ex-post distribution but rather we obtain a higher realized return.

This phenomenon is confirmed by the decomposition of the performance fee between the $MV^d$ and $HM^c$ strategies. The annual fee is 45 bp, but the main contribution comes from the increase in realized portfolio return (40 bp). The
distributional properties of realized return are not affected by the change of strategy.

Changing the level of risk aversion does not alter these results significantly. The performance fee ranges around 30 bp. This measure has to be compared to an average of 60 bp for the performance fee an investor is willing to pay to switch from the static to the $MV^d$ strategy. Therefore, the economic gain to capture non-normality is broadly equal to half the economic gain of the volatility timing. This is quite a sizeable effect.

The global effect of non-normality also includes distributional timing. This corresponds to the ability of the investor to correctly forecast the changes in the returns’ conditional distribution. In our context, it consists in forecasting the skewness and the kurtosis of market returns. The $HM^d$ strategy is based on this approach. Table 4 shows that the economic gain of distributional timing is significant and comparable to that of volatility timing. For risk aversion $\gamma = 5$, the performance fee the investor with a $MV^d$ strategy is willing to pay to switch to the $HM^d$ strategy is equal to 57 bp. It is the same order of magnitude as the performance fee to switch from the static to the $MV^d$ strategy. For other levels of risk aversion, this economic gain is even higher. It is equal to 79 bp for $\gamma = 2$ and 78 bp for $\gamma = 10$. The $HM^d$ strategy also displays some gain with respect to the $HM^c$ strategy, indicating that distributional timing represents an additional effect to the direct effect of (constant) non-normality. Each of these two effects represents 30 bp on average for the various levels of risk aversion.

To summarize our empirical evidence, we observe that the average economic gains of volatility timing and of taking changes in the conditional distribution into account are both around 60 bp. Regarding the latter effect, the average economic gains of using a non-normal distribution and distributional timing are both around 30 basis points. It is worth emphasizing that whatever the level of risk aversion, the $HM^d$ strategy is the dynamic strategy that incurs the smallest turnover. It results in high breakeven transaction costs (between 15 and 36 bp). This result indicates that the large economic gains of the $HM^d$ are not likely to be offset by transaction costs. This strategy could be easily implemented.
**Statistical significance of distributional timing.** While the static strategy only requires the estimation of the sample mean vector and the covariance matrix, the dynamic strategies rely on the estimation of the dynamics of the covariance and higher co-moments matrices. To avoid any overfitting of the data or data snooping, we used two non-overlapping subsamples for the estimation and allocation stages.\(^{14}\)

Another important issue in the evaluation of economic value of a strategy is estimation risk. Our previous results suggest that distributional timing has an economically sizeable value. However, this value may be statistically insignificant if the uncertainty surrounding parameter estimates is too large. To address this issue, we use Monte-Carlo simulations to evaluate the significance of the performance measures reported in Table 4. We proceed as follows. For a given dynamic strategy, we draw a set of parameters in the asymptotic distribution of the corresponding econometric model and then simulate a sample of returns. Using these simulated returns, we perform the optimal allocation over 206 periods (the size of the allocation period). This gives the characteristics of the optimal portfolio return for this draw. We repeat the same exercise for 1'000 draws in the parameter distribution and for each dynamic strategy. Eventually, we evaluate the significance of the various performance measures reported above.

**Figure 4** depicts the distribution of the performance fee for \(MV^d\) and \(HM^d\) strategies with respect to the static strategy. As it appears clearly, the two measures are significantly positive, confirming the economic value of volatility and distributional timing. In addition, while the distribution of the performance fee for the \(MV^d\) strategy is concentrated around 120 bp, that of the \(HM^d\) strategy is around 180 bp. From this graphical analysis, we conclude that the performance of the \(HM^d\) strategy dominates the one of \(MV^d\). In particular, the most sophisticated strategy \(HM^d\) does not suffer from any additional risk, since the evaluation of its performance fee is more precise than for the \(MV^d\) strategy.

\(^{14}\)Over-fitting may arise by the introduction of too many parameters in a model. If this is the case, some parameters may be significant only because they help capturing very specific episodes. They would be helpful to improve the in-sample allocation, but useless, at best, for the out-of-sample allocation. Data snooping would occur if the same sample was used for the estimation and the allocation.
6 Robustness analysis

Our general result, that optimal allocation can be improved by taking the properties of the conditional distribution into account, may also be sensitive to some of our modeling assumptions. We already discussed above the effect of risk aversion and showed that the economic value of distributional timing is positive and sizeable for all levels of risk aversion considered. To further evaluate the robustness of our results, we presently investigate if our findings are sensitive to a set of assumptions: (1) the assumption regarding short-sales and borrowing constraints; (2) the choice of the allocation period; (3) the estimates of expected returns and of the parameters driving the dynamics of higher moments and co-moments.

6.1 No short-sales and borrowing constraints

In the literature on the economic value of volatility timing, it is common practice, following Fleming, Kirby, and Ostdiek (2001), to allow for short selling and unrestricted borrowing. So far, since many financial institutions practice long-only strategies, thereby discarding shortselling or leveraged portfolios, we concentrated on an investor who invests a positive weight in all available assets including the risk-free asset. In this section, as a robustness check, we investigate the consequence of releasing the no-shortsale constraint. In practice, since futures contracts are available for indices, such a strategy could also be implemented.

The set of new portfolio weights and measures of performance are reported in Tables 5 and 6 respectively. Figure 5 displays the evolution of the optimal weights over time. We first notice that portfolio allocations are affected by the borrowing constraint only for low levels of risk aversion. For $\gamma = 10$ or 15, the optimal allocation and the portfolio performances are not altered as compared to the previous case. We also notice that, even in the absence of short-sales and borrowing constraints, the investor does not short sell any risky asset, but rather borrows to increase her exposure to risky assets. As Table 5 shows, the investor typically increases the fraction of wealth allocated to the US and Japanese markets.

As expected, the table also shows that the distance between the weights of the
dynamic strategies and the static one increase significantly. The turnover of dynamic strategies is also larger than with short-sales and borrowing constraints. For $\gamma = 5$, the turnover is equal to 5.8% per week for the $MV^d$ strategy and 4.8% for the $HM^d$ strategy.

Table 6 reveals that short selling and unrestricted borrowing exacerbate the economic value of distributional timing, while the effect of volatility timing becomes ambiguous depending on the level of risk aversion. If we begin with $\gamma = 5$, the $MV^d$ strategy yields a larger realized return than the static strategy (8.05% against 6.53%), but also a larger volatility (15.48% against 14.88%). Consequently, the Sharpe ratio does not improve as compared to the case with short-sales and borrowing constraints. Furthermore, the performance fee to switch from the static strategy to the $MV^d$ strategy decreases to 97 bp per year.

Turning to the measure of distributional timing, we observe a much higher realized return (10.63%) that is only partly offset by a higher risk. In terms of Sharpe ratio, the gain with respect to the static strategy is very sizeable (0.52 against 0.32). It also results that the investor is willing to pay a performance fee as high as 235 bp to switch from the $MV^d$ strategy to the $HM^d$ strategy (against 161 bp in the case with short-sales and borrowing constraints). This fee is attributable for 64 bp to the ability to take the (constant) non-normality into account and for 75 bp to distribution timing.

For low risk aversion ($\gamma = 2$), the gain of volatility timing improves as compared to the case with short-sales and borrowing constraints. The corresponding fee is equal to 209 bp (against 21 bp). In the meanwhile, the fee to switch from the $MV^d$ strategy to the $HM^d$ strategy is as high as 318 bp: 175 bp come from the ability to take non-normality into account and 143 bp from distribution timing. Finally, allowing for short sales and borrowing reinforces our initial result that the ability of investor to take non-normality into account is economically valuable.

6.2 Sensitivity analysis

In this section, we investigate the effect the specific value of some parameter estimates may have on the optimal allocations and more particularly on their perfor-
mances.

6.2.1 Sensitivity to expected returns

We begin our sensitivity analysis by investigating the stability of allocations to changes in expected returns. This exercise is of relevance since expected returns are arguably difficult to estimate and moreover they may strongly affect the weights of the optimal allocations (see Jorion, 1985, Dumas and Jacquillat, 1990).

To neutralize the effect of expected returns on our allocation, we follow Das and Uppal (2004) and simply take the average of the expected return estimates across markets as a proxy for the expected returns on all the markets. Since the three markets have exactly the same expected returns, the allocation is no longer affected by differences in expected returns.

Table 7 reports the portfolio performance of assuming the same expected return for all markets. Interestingly, the main consequence of neutralizing the effect of expected returns is to reduce volatility timing. Indeed, while the realized return of the static strategy slightly increases when expected returns are equal (from 5.7% per year to 6.1%), the return on the $MV^d$ strategy dramatically decreases (from 6.6% to 5.6%). Although the volatility of the $MV^d$ strategy also reduces, we notice a fall in the Sharpe ratio of this strategy, while the ratio of the static strategy slightly increases. Finally, the performance fee a static investor is willing to pay to switch to the $MV^d$ strategy decreases from 126 bp with our baseline model to only 17 bp under equal expected returns.

In contrast, the higher-moment strategies are almost unaltered. The fall in realized returns is mainly offset by the fall in volatility, so that the Sharpe ratios are essentially the same as under our baseline model. In terms of performance fee, we notice a decrease in the fee with respect to the static strategy (from 183 bp to 125 bp for the $HM^d$ strategy), but an increase in the fee with respect to the $MV^d$ strategy (from 57 bp to 108 bp).
6.2.2 Sensitivity to shape parameters

Another issue of interest is the sensitivity of the optimal allocation with respect to changes of the shape parameters’ dynamics. As argued before, the economic value of distribution timing comes from the ability to forecast changes in the conditional distribution. It may be argued that these changes may be quite difficult to anticipate and therefore that too a large sensitivity to the shape parameters’ estimates would be a drawback of the approach, since it would be also very sensitive to estimation error.

For obvious reasons we focus on the $HM^d$ strategy which involves time-varying shape parameters. To investigate the sensitivity to shape parameters, we adopt the same approach as for the expected returns. We simply take the average across markets of the parameters pertaining to the dynamics of shape parameters. As a consequence, although the level of the higher moments differs from one market to the other, the parameters driving their dynamics are now assumed to be the same.

Table 7 reveals that the optimal allocation is not sensitive to the specific values taken by the parameters driving the dynamics of higher moments. Indeed, we observe only a slight decrease in the performance fee the investor is willing to pay to benefit from the $HM^d$ strategy, while the shape parameters of this model are no longer optimally estimated. It suggests that it is the general tendency of time variability of the shape parameters that drives the result. Their precise evolution does not need to be known as long as skewness and kurtosis demonstrate a variability of consistent direction.

7 Conclusion

In this paper, we have investigated the consequences of non-normality of returns on the optimal asset allocation when the distribution of asset returns changes over time. While most previous work has been devoted to the case where the characteristics of investment opportunities are constant through time, several recent papers have explored the consequences of ignoring the time variability of some aspects of the returns’ distribution: Fleming, Kirby, and Ostdiek (1999, 2001) and Han (2006)
evaluate the value of volatility timing, while Ang and Bekaert (2002) as well as
Guidolin and Timmermann (2005) measure the cost of ignoring the presence of
regime shifts. Das and Uppal (2004) consider the consequences of jumps in the joint
distribution of market returns. We contribute to this literature along two dimen-
sions. First, from the point of view return dynamic, we propose an econometric
model that captures most statistical features of market returns, such as volatility
clustering, correlation persistence, asymmetry and fat-tailedness of the distribution.
The estimation of this model remains tractable, even in the case of several assets.

Second, we compute the optimal asset allocation for several strategies that allow
us to evaluate the economic value of volatility and distributional timing. We show
that, even for moderate levels of risk aversion, the performance fee an investor is
willing to pay to benefit from distributional timing is of the same order of magnitude
as the performance fee to benefit from volatility timing.

Thus, our evaluation of the economic value of taking non-normality into account,
reveals a sizeable benefit. The economic value is similar in magnitude to volatility
timing, highlighted by Fleming, Kirby, and Ostdiek (1999, 2001). Importantly, it re-
results from two cumulative effects: first, the non-normality of market returns should
be taken into account in the allocation process by using a criterion that depends on
the higher moments of the portfolio return’s distribution; second, the ability to fore-
cast changes in the conditional distribution through time (the distributional timing)
also contributes significantly to the performance of the higher-moment strategies.
8 Appendices

8.1 Appendix 1: A multivariate model for returns

In this Appendix, we describe the conditional multivariate model for returns. This model allows for distributions of returns with a rich pattern.

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

\[
r_t = \mu_t(\theta) + \varepsilon_t, \tag{11}
\]

\[
\varepsilon_t = \Sigma_t(\theta)^{1/2} z_t, \tag{12}
\]

where \( \mu_t(\theta) = E[r_t|I_{t-1}] \) denotes the \( n \times 1 \) conditional mean vector, given the information set \( I_{t-1} \) available at date \( t-1 \), \( \varepsilon_t = r_t - \mu_t \) is the \( n \times 1 \) vector of unexpected returns, \( \Sigma_t(\theta) = E[\varepsilon_t\varepsilon_t'|I_{t-1}] = \{\sigma_{ij,t}\}_{i,j=1,\ldots,n} \) is the \( n \times n \) conditional covariance matrix. The \( z_t \) is the \( n \times 1 \) vector of innovations, such that \( E[z_t] = 0 \) and \( V[z_t] = I_n \), where \( I_n \) is the identity matrix of order \( n \). The conditional covariance matrix of returns \( \Sigma_t \) is defined as

\[
\Sigma_t = D_t \Gamma_t D_t, \tag{13}
\]

where \( D_t = \{\sigma_{i,t}\}_{i=1,\ldots,n} \) is the \( n \times n \) diagonal matrix with standard deviations on the diagonal, and \( \Gamma_t = \{\rho_{ij,t}\}_{i,j=1,\ldots,n} \) is the \( n \times n \) 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)

\[
\sigma_{i,t}^2 = \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i \varepsilon_{i,t-1}^2 + \gamma_i \varepsilon_{i,t-1}^2 1\{\varepsilon_{i,t-1} < 0\}, \quad i = 1, \ldots, n, \tag{14}
\]

where the constraint \( \alpha_i + \beta_i + \gamma_i/2 < 1 \) guarantees stationarity of the variance process. The conditional correlation matrix, \( \Gamma_t \), is time-varying, following the DCC specification of Engle (2002) and Engle and Sheppard (2001)

\[
\Gamma_t = (\text{diag}(Q_t))^{-1/2} \cdot Q_t \cdot (\text{diag}(Q_t))^{-1/2}, \tag{15}
\]

\[
Q_t = (1 - \delta_1 - \delta_2) \bar{Q} + \delta_1 (u_{t-1} u_{t-1}') + \delta_2 Q_{t-1}, \tag{16}
\]
where \( u_t = D_t^{-1} \varepsilon_t \) denotes the vector of normalized unexpected returns, and \( \text{diag}(Q_t) \) denotes the \( n \times n \) matrix with zeros, except for the diagonal that contains the diagonal of \( Q_t \). The matrix \( \bar{Q} \) is the \( n \times n \) unconditional covariance matrix of \( u_t \). We impose the restrictions \( 0 \leq \delta_1, \delta_2 \leq 1 \) and \( \delta_1 + \delta_2 \leq 1 \), so that the conditional correlation matrix is guaranteed to be positive definite.

The multivariate Sk-t distribution. Innovations \( z_t \) are drawn from \( n \) independent Sk-t distributions. As shown in equation (12), correlation among returns is introduced via a Cholesky decomposition. The main advantage of this approach is that it explicitly separates the modeling of the univariate conditional distributions (through the parameters of the Sk-t distribution) and the modeling of the multivariate dependence (through the parameters of the covariance matrix).\(^{15}\) We assume that the \( n \times 1 \) vector of innovations \( z_i \) is drawn from the following multivariate Sk-t distribution

\[
g(z_i|\eta) = \prod_{i=1}^{n} \frac{2b_i}{\xi_i} \frac{\Gamma \left( \frac{\nu_i+1}{2} \right)}{\Gamma \left( \frac{\nu_i}{2} \right)} \left( 1 + \frac{\kappa_{i,t}^2}{\nu_i} \right) \left( 1 + \frac{\kappa_{i,t}^2}{\nu_i} \right)^{-\frac{\nu_i+1}{2}}, \tag{17}
\]

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

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

and

\[
a_i = \frac{\Gamma \left( \frac{\nu_i-1}{2} \right)}{\sqrt{\pi} \Gamma \left( \frac{\nu_i}{2} \right)} \left( \frac{1}{\xi_i} - 1 \right),
\]

\[
b_i^2 = \xi_i^2 + \frac{1}{\xi_i^2} - 1 - a_i^2.
\]

Focusing on a given component \( i \), shape parameters \( \nu_i \) and \( \xi_i \) correspond to the individual degree of freedom and the asymmetry parameter respectively. The marginal distribution of \( z_{i,t} \) is a univariate Sk-t distribution \( g(\nu_i, \xi_i) \). It is defined for \( 2 < \nu_i < \infty \) and \( \xi_i > 0 \). Parameters \( a_i \) and \( b_i \) are required to center and scale the asymmetric distribution so that \( z_{i,t} \) has a zero mean and unit variance.

\(^{15}\)A similar multivariate distribution has been analyzed by Brooks, Burke, and Persand (2005) yet in its symmetric version.
Higher moments of \( z_{i,t} \) are easily deduced from those of the symmetric  \( t(\cdot|\nu_i) \). If the \( r \)-th moment of a random variable with distribution \( t(\cdot|\nu_i) \) exists, then the associated variable \( z_{i,t} \) with distribution \( g(\cdot|\nu_i,\xi_i) \) has a finite \( r \)-th moment, defined as

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

where

\[
m_{i,r} = 2E[Z_{i}^r|Z_i > 0] = \frac{\Gamma \left( \frac{\nu_i-r}{2} \right) \Gamma \left( \frac{r+1}{2} \right) (\nu_i - 2)^{r+1}}{\sqrt{\pi(\nu_i - 2)} \Gamma \left( \frac{\nu_i}{2} \right)}, \tag{18}
\]

is the \( r \)-th moment of \( t(\cdot|\nu_i) \) truncated to the positive real values. Note that we have \( E[Z_{i}^3|\xi_i] = -E[Z_{i}^3|1/\xi_i] \) and \( E[Z_{i}^3] = 0 \) when \( \xi_i = 1 \). Provided that they exist, third and fourth central moments of \( z_{i,t} \) are

\[
s_k^Z_i = \mu^{(3)}_i = E[Z_{i}^3] = M_{i,3} - 3M_{i,1}M_{i,2} + 2M_{i,1}^3, \tag{19}
\]

\[
k_u^Z_i = \mu^{(4)}_i = E[Z_{i}^4] = M_{i,4} - 4M_{i,1}M_{i,3} + 6M_{i,2}M_{i,1}^2 - 3M_{i,1}^4. \tag{20}
\]

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

**Dynamics of the conditional higher moments.** Finally, we allow the asymmetry parameters \( \xi_i \) and the degree-of-freedom parameters \( \nu_i \) to vary over time. Such time variability in the higher moments has already been analyzed in Hansen (1994), Harvey and Siddique (1999), and Jondeau and Rockinger (2003). The dynamics of these parameters cannot be chosen arbitrarily because of the constraints imposed on their dynamics. The degree-of-freedom parameter \( \nu_{i,t} \) has to be larger than 2 and the asymmetry parameter \( \xi_{i,t} \) has to be positive at each date \( t \) for the distribution to be well defined. We adopt the following model, that is, in its structure, very close to an asymmetric GARCH model:

\[
\log(\nu_{i,t} - \nu) = c_{i,0} + c_{i,1}^+ |z_{i,t-1}|N_{i,t-1} + c_{i,1}^- |z_{i,t-1}| P_{i,t-1} + c_{i,2} \log(\nu_{i,t-1} - \nu), \tag{21}
\]

\[
\log(\xi_{i,t}) = d_{i,0} + d_{i,1}^- z_{i,t-1}N_{i,t-1} + d_{i,1}^+ z_{i,t-1} P_{i,t-1} + d_{i,2} \log(\xi_{i,t-1}), \tag{22}
\]

33
where $N_{i,t} = 1_{\{z_{i,t} < 0\}}$ and $P_{i,t} = 1 - N_{i,t}$. The parameter $\nu_1$ is the lower bound for the degree of freedom.\textsuperscript{16} Three main features of these specifications are worth emphasizing. First, the degree-of-freedom parameter $\nu_{i,t}$ is related to the absolute value of lagged standardized innovations, since $z_{i,t-1}$ is expected to affect the heaviness of the distribution’s tails regardless of its sign. In contrast, the asymmetry parameter naturally depends on signed residuals, since $\xi_{i,t}$ is likely to reflect the sign and size of the recent shocks. Second, instead of assuming that positive and negative shocks have the same impact on the shape of the distribution, we allow an asymmetry in the reaction of the shape parameters to recent shocks. Finally, equations (21) and (22) include a lag of the dependent variable, in order to capture the possible persistence in the dynamics of the higher moments. Given the dynamics of shape parameters (equations (21) and (22)), higher moments are deduced using equations (19) and (20).

**Estimation.** One of the convenient properties of the DCC model is that, under normality, the estimation of the joint model can be performed in several steps. The reason is that the log-likelihood can be broken down into components that involve the parameters of the univariate GARCH models and a component that involves the parameters of the correlation matrix (Engle, 2002). As a consequence, estimating this model is generally rather fast, at least for small-dimensional models. In case of a Sk-$t$ distribution, such a decomposition of the log-likelihood is not available, and the estimation of the DCC model requires estimating all the parameters simultaneously. Considering the model described above, the set of location and dispersion parameters is denoted $\theta = (\mu_i, \omega_i, \alpha_i, \beta_i, \gamma_i (i = 1, \ldots, n), \delta_1, \delta_2)'$. When innovations are drawn from a multivariate Sk-$t$ distribution with constant shape parameters, we have $\eta = (\nu_1, \ldots, \nu_n, \xi_1, \ldots, \xi_n)'$. When shape parameters are time varying as in equations (21) and (22), we have $\eta = (c_{i,0}, c_{i,-1}^+, c_{i,-1}^-, c_{i,2}, d_{i,0}, d_{i,-1}^+, d_{i,-1}^-, d_{i,2} (i = 1, \ldots, n))'$.

The sample log-likelihood function of the multivariate DCC model with Sk-$t$

\textsuperscript{16}Since we are actually interested in an asset allocation based on the four first moments, we impose that moments up to the fourth one are defined, i.e. $\nu = 4$. 

34
distribution is therefore
\[
\log L(r_1, \cdots, r_T|\theta, \eta) = \sum_{t=1}^{T} \left[ \log \left( g \left( \Sigma_t(\theta)^{-1/2} (r_t - \mu_t(\theta)) | \eta \right) \right) - \frac{1}{2} \log |\Sigma_t(\theta)| \right],
\]
where \( g (\cdot | \eta) \) is defined in equation (17). Maximizing the log-likelihood with respect to parameter vectors \( \theta \) and \( \eta \) yields maximum-likelihood (ML) estimates. Since all parameters are estimated jointly, it is quite time-consuming, but it remains reasonable for moderate-size systems.\(^{17}\)

**Empirical results.** Table A reports the results of the estimation of the multivariate model with Sk-t distribution and time-varying shape parameters (Model 4).\(^{18}\) First of all, conditional volatilities are modeled through univariate asymmetric GARCH models. In all cases, as expected, the asymmetry parameter \( \gamma_i \) is significantly positive, suggesting that a bad news has a stronger effect on volatility than a good news. In addition, the persistence of volatilities is rather large for the US and Japan, but much less so for the UK. The dynamics of variances are reported in Figure A1. Turning to the dynamics of correlations, the parameter \( \delta_2 \) takes a value of 0.96 translating the fact that correlation movements are persistent. Figure A2 displays the temporal evolution of correlations. We notice that the correlations of the US and UK with Japan have increased over the first half of the period with a peak around 1992, then they steadily decreased till 2000. Since 2001, they tended again upwards.

Regarding the dynamics of shape parameters, the table reveals that the degree-of-freedom parameter varies over time for all markets under study. It is strongly related to the absolute value of the past innovation. Indeed, parameters \( c_{i,1}^- \) and \( c_{i,1}^+ \) are systematically found to be positive, and significantly different from zero. This result indicates that, after an extreme (positive or negative) event, the degree of

---

\(^{17}\)Estimation is performed using MATLAB. Gradient and Hessian are computed numerically. Since we do not claim that the Sk-t distribution is necessarily the correct distribution, we systematically report robust standard errors, based on the sandwich estimator of the covariance matrix.

\(^{18}\)The parameter estimates of Models 1 to 3 are not reported in the paper, to save space, but they are available upon request from the authors. They are very close to those obtained with Model 4. In addition, with only 41 estimated parameters, Model 4 provides a parsimonious description of the constant conditional higher moments.
freedom of the Sk-t distribution increases, so that kurtosis decreases. This suggests that extreme events do not cluster at the weekly frequency. Regarding the asymmetry parameter, we observe that for the US and Japan, parameter $d_{i,1}^+$ is positive while $d_{i,1}^-$ is negative. This suggests that, after an extreme event, the probability of another extreme event with the same sign decreases, so that once again extreme events do not cluster. This result is in accordance with the evidence already reported by Das and Uppal (2004) using a multivariate model with jumps.

To test the ability of this model to fit the data, we performed goodness-of-fit (GoF) tests as suggested by Diebold, Gunther, and Tay (1998). The test for iidness of the margins indicates that the model is able to filter out all the dynamics found in the first four moments of the distribution for the UK and Japan. (For the US, there is still some residual correlation in the first and third moments.) In addition, the GoF test statistic does not reject the hypothesis that the assumed Sk-t distribution correctly fits the empirical distribution.

8.2 Appendix 2: Computing portfolio returns’ moments

Analytical expressions for the portfolio conditional moments can be easily obtained for a multivariate Sk-t distribution. First, third and fourth central moments of a univariate Sk-t distributed random variable $z_{t+1}$ are given by equations (19) and (20). Second, since unexpected returns are defined as $\varepsilon_{t+1} = \Sigma_{t+1}^{1/2} z_{t+1}$, their first four moments can be computed using matrix calculus, instead of numerical integration. We obviously have $E_t [\varepsilon_{t+1}] = 0$ and $V_t [\varepsilon_{t+1}] = \Sigma_{t+1}$. We denote $\Sigma_{t+1}^{1/2} = (\omega_{ij,t+1})_{i,j=1,\ldots,n}$ the Choleski decomposition of the covariance matrix of

---

19 These authors show that, if the marginal distributions are correctly specified, the margins of the distribution should be iid Uniform(0,1). The test is performed in two steps. First, we question if the model correctly captures the dynamics of returns and test if the margins $u_{i,t} = \tilde{F} (z_{i,t})$ are serially correlated, where $\tilde{F} (z_{i,t})$ denotes the marginal cumulative distribution function of $z_{i,t}$. Second, we question if the model correctly reproduces the conditional distribution of returns and test the null hypothesis that $u_{i,t}$ is distributed as an Uniform(0,1). For this purpose, we cut the empirical and theoretical distributions into $N$ bins and test whether the two distributions significantly differ on each bin. For this test, the statistic is distributed as a $\chi^2$ with $(N - 1)$ degrees of freedom. We use $N = 20$. 

36
returns, so that \( r_{i,t+1} = \mu_{i,t+1} + \sum_{r=1}^n \omega_{ir,t+1} z_{r,t+1} \). In addition, using tensor notations and denoting by \( \otimes \) the Kronecker product, the \( n \times n^2 \) third central co-moment matrix is defined as\(^{20}\)

\[
S_{t+1} = E_t \left[ (r_{t+1} - \mu_{t+1}) (r_{t+1} - \mu_{t+1})' \otimes (r_{t+1} - \mu_{t+1}) \right] = \{ s_{ijk,t+1} \},
\]

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

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

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

\[
K_{t+1} = E_t \left[ (r_{t+1} - \mu_{t+1}) (r_{t+1} - \mu_{t+1})' \otimes (r_{t+1} - \mu_{t+1}) \right] \otimes (r_{t+1} - \mu_{t+1})' \]

\[
= \{ \kappa_{ijkl,t+1} \},
\]

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 \( \psi_{rs} = \omega_{ir} \omega_{jr} \omega_{ks} \omega_{ls} + \omega_{is} \omega_{jr} \omega_{ks} \omega_{ls} + \omega_{is} \omega_{jr} \omega_{kr} \omega_{ls} \).

The numerical computation of these expressions is very fast.

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

\[
\mu_{p,t+1} = r_{ft} + \alpha_t' (\mu_{t+1} - r_{f,t} e),
\]

\[
\sigma_{p,t+1}^2 = \alpha_t' \Sigma_{t+1} \alpha_t,
\]

\[
s_{p,t+1}^3 = \alpha_t' S_{t+1} (\alpha_t \otimes \alpha_t),
\]

\[
\kappa_{p,t+1}^4 = \alpha_t' K_{t+1} (\alpha_t \otimes \alpha_t \otimes \alpha_t).
\]

Conditional skewness and kurtosis of the portfolio return are deduced using the relations \( s k_{p,t+1} = s_{p,t+1}^3 / \sigma_{p,t+1}^3 \) and \( k u_{p,t+1} = \kappa_{p,t+1}^4 / \sigma_{p,t+1}^4 \).

\(^{20}\)Using these notations, central co-moment matrices can be conveniently represented as bi-dimensional matrices.

\(^{21}\)For simplicity, we suppressed the index \( t \) in the expression for \( \psi_{rs,t+1} \).
References


Chosen by a Bayesian Technique: 1967-85, *Journal of Banking and Finance*,
14(2/3), 539–558.

timates of the Variance of United Kingdom Inflation, *Econometrica*, 50(4),
987–1007.

tivariate Generalized Autoregressive Conditional Heteroskedasticity Models.

Dynamic Conditional Correlation Multivariate GARCH, NBER Working Paper
8554.

473–509.


[17] Gallant, R., and G. Tauchen (1989), Seminonparametric Estimation of Con-
ditionally Constrained Heterogeneous Processes: Asset Pricing Applications,
*Econometrica*, 57(5), 1091–1120.

[18] Glosten, R. T., R. Jagannathan, and D. Runkle (1993), On the Relation be-
tween the Expected Value and the Volatility of the Nominal Excess Return on

Regime Switching, Skew and Kurtosis Preferences, Federal Reserve Bank of St.


### Captions

**Table 1:** This table provides a description of the various models and strategies under study. For each model/strategy, we describe the main property of the conditional mean, covariance matrix, distribution and dynamics of higher moments.

**Table 2:** This table reports summary statistics on market returns. They include the average, the standard deviation, the skewness, and the kurtosis. We also report the Lee and King (1993) test for serial correlation of returns, $\rho_1$, and of squared returns, $\rho_2$. Panel A corresponds to the full sample including the week of the October 1987 crash. Panel B corresponds to the sample without the week of the October 1987 crash. Panel C reports correlation, co-skewness and co-kurtosis matrices. Exponents $a$ and $b$ indicate that a statistic is significant at the 1%, 5% level respectively.

**Table 3:** This table reports statistics on the optimal portfolio weights, for the various strategies and for values of the risk-aversion parameter $\gamma$ ranging from 2 to 15. We report the optimal weights, the distance between two portfolios and the turnover ratio. The distance between two series of weights is measured by $Dist = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t}^* - \hat{\alpha}_{i,t}|/2$, while the turnover ratio from one week to the other is measured by $Turn = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t} - \alpha_{i,t-1}|/2$.

**Table 4:** This table reports statistics on the performance of the optimal portfolios, for the various strategies and for values of the risk-aversion parameter $\gamma$ ranging from 2 to 15. We report the first four realized moments of the portfolio return, the Sharpe ratio, the cumulative return (CR), and several measures of performance of the strategies. The $M^2$ measure is defined by equation (2.4). The performance fee, $\theta$, is estimated from the sample counterpart of the relation $E[U(1 + \hat{r}_{p,t+1} + \theta)] = E[U(1 + r_{p,t+1}^*)]$, where $r_{p,t+1}^*$ denotes the portfolio return obtained with the optimal strategy, and $\hat{r}_{p,t+1}$ the portfolio return obtained with a sub-optimal strategy. The sub-optimal strategy is the static one for $\theta_s$, the $MV^d$ strategy for $\theta_d$ and the $HM^c$ strategy for $\theta_c$. The success rate $Z$ is the percentage of times in which the dynamic strategy out-performed the static strategy. The breakeven transaction cost $\tau^{be}$ is defined by equation (8).

**Table 5:** This table reports statistics on the optimal portfolio weights, for the various strategies and for values of the risk-aversion parameter $\gamma$ ranging from 2 to 15, in the case where short sales and borrowing are allowed. We report the optimal weights, the distance between two portfolios and the turnover ratio. The distance between two series of weights is measured by $Dist = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t}^* - \hat{\alpha}_{i,t}|/2$, while the turnover ratio from one week to the other is measured by $Turn = \frac{1}{Tn} \sum_{t=1}^{T} \sum_{i=1}^{n} |\alpha_{i,t} - \alpha_{i,t-1}|/2$.

**Table 6:** This table reports statistics on the performance of the optimal portfolios, for the various strategies and for values of the risk-aversion parameter $\gamma$ ranging from
2 to 15, in the case where short sales and borrowing are allowed. For a description of the reported statistics, refer to Table 4.

**Table 7:** This table reports statistics on the performance of the optimal portfolios, for the various strategies and for $\gamma = 5$, under various scenarios regarding the model’s parameters. For a description of the reported statistics, refer to Table 4.

**Table A:** This table reports parameter estimates, summary statistics, and specification tests for the model with a Sk-$t$ distribution with time-varying higher moments (Model 4). Conditional correlations are not reported, since they are estimated at their sample values. $\ln L$ denotes the sample log-likelihood. Following Diebold, Gunther, and Tay (1998), the lower part of the table reports the goodness-of-fit (GoF) test statistics for the null hypothesis that the margins are Uniform(0,1) and the LM test statistics for the null hypothesis that the margins are iid processes. All numbers in parenthesis are $p$-values for the tests. Under the null, the statistic GoF is computed using 20 bins and is distributed as a $\chi^2(20)$. The LM test statistics are obtained by regressing $(u_{it} - \bar{u})^k$, for $k = 1, \cdots, 4$, on $q$ lags of the variable. The statistics are defined as $(T - q) R^2$, where $R^2$ is the coefficient of determination of the regression, and is distributed as a $\chi^2(q)$ under the null. We used $q = 20$.

**Figure 1:** This figure displays the evolution of the conditional skewness, as estimated by the model with a Sk-$t$ distribution with time-varying shape parameters (Model 4). The period from 2002 to 2005 is used for the out-of-sample performance analysis. This remark also applies to subsequent figures. The indices 1, 2 and 3 correspond to the US, Japan, and the UK respectively.

**Figure 2:** This figure displays the evolution of the conditional kurtosis over the estimation period, as estimated by the model with a Sk-$t$ distribution with time-varying shape parameters (Model 4). The straight line corresponds to the value expected from a multivariate normal distribution.

**Figure 3:** This figure displays the optimal portfolio weights, for the various strategies and for $\gamma = 5$. The time variation of the various weights for the static strategy in the out-of-sample part is due to time variation in the risk free rate.

**Figure 4:** This figure displays the distribution of the performance fee for $MV^d$ and $HM^d$ strategies. It is a kernel density approximation of the empirical distribution of the performance fee obtained using Monte-Carlo simulations as described in Section 5.3.

**Figure 5:** This figure displays the optimal portfolio weights, for the various strategies and for $\gamma = 5$, in the case where short sales and borrowing are allowed.

**Figure A1:** This figure displays the evolution of the conditional variances estimated by the model with a Sk-$t$ distribution with time-varying shape parameters (Model
4). We kept the period from 2002 to 2005 for the out-of-sample performance analysis and therefore did not use it in the estimation. This remark applies to this figure as well as to the following one.

**Figure A2:** This figure displays the evolution of the conditional correlations, as estimated by the model with a Sk-t distribution with time-varying shape parameters (Model 4).
Table 1: Description of the various models and strategies under study

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Variance</th>
<th>Distribution</th>
<th>Higher moments of innovations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>constant</td>
<td>constant</td>
<td>Normal</td>
<td>$sk = 0, ku = 3$</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>constant</td>
<td>DCC</td>
<td>Normal</td>
<td>$sk = 0, ku = 3$</td>
</tr>
<tr>
<td>$HM^c$</td>
<td>constant</td>
<td>DCC</td>
<td>Sk-$t$</td>
<td>constant</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>constant</td>
<td>DCC</td>
<td>Sk-$t$</td>
<td>time-varying</td>
</tr>
</tbody>
</table>
Table 2: Summary statistics on market returns

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>Japan</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Sample including the 1987’s crash</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.205</td>
<td>0.164</td>
<td>0.237</td>
</tr>
<tr>
<td>Std dev.</td>
<td>2.126</td>
<td>3.003</td>
<td>2.561</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.380</td>
<td>0.182</td>
<td>-0.398</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.864</td>
<td>5.034</td>
<td>7.447</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.072</td>
<td>-0.032</td>
<td>0.008</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.239</td>
<td>0.219</td>
<td>0.049</td>
</tr>
<tr>
<td><strong>Panel B: Sample excluding the 1987’s crash</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.215</td>
<td>0.173</td>
<td>0.253</td>
</tr>
<tr>
<td>Std dev.</td>
<td>2.099</td>
<td>2.985</td>
<td>2.492</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.250</td>
<td>0.228</td>
<td>0.036</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.246</td>
<td>4.947</td>
<td>4.116</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.093</td>
<td>-0.030</td>
<td>-0.001</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.180</td>
<td>0.234</td>
<td>0.127</td>
</tr>
<tr>
<td><strong>Panel C: Multivariate Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation matrix</td>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r_3$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_2$</td>
<td>0.219</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$r_3$</td>
<td>0.397</td>
<td>0.312</td>
<td>1.000</td>
</tr>
<tr>
<td>Co-skewness matrix</td>
<td>$r_1^2$</td>
<td>$r_2^2$</td>
<td>$r_3^2$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>-0.250</td>
<td>0.010</td>
<td>-0.014</td>
</tr>
<tr>
<td>$r_2$</td>
<td>0.037</td>
<td>0.227</td>
<td>0.002</td>
</tr>
<tr>
<td>$r_3$</td>
<td>-0.106</td>
<td>0.000</td>
<td>0.036</td>
</tr>
<tr>
<td>Co-kurtosis matrix</td>
<td>$r_1^3$</td>
<td>$r_2^3$</td>
<td>$r_3^3$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>5.238</td>
<td>0.668</td>
<td>1.752</td>
</tr>
<tr>
<td>$r_2$</td>
<td>0.974</td>
<td>4.940</td>
<td>1.305</td>
</tr>
<tr>
<td>$r_3$</td>
<td>2.481</td>
<td>1.154</td>
<td>4.109</td>
</tr>
<tr>
<td></td>
<td>$r_1 r_2^2$</td>
<td>$r_1 r_3^2$</td>
<td>$r_2 r_3^2$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>1.323</td>
<td>2.107</td>
<td></td>
</tr>
<tr>
<td>$r_2$</td>
<td></td>
<td>0.667</td>
<td>1.341</td>
</tr>
<tr>
<td>$r_3$</td>
<td>0.596</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3: Optimal portfolio weights

<table>
<thead>
<tr>
<th>Strategy</th>
<th>US</th>
<th>Japan</th>
<th>UK</th>
<th>US T-bill</th>
<th>Dist</th>
<th>Turn</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = 2$</td>
<td>Static</td>
<td>0.377</td>
<td>0.000</td>
<td>0.624</td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>0.174</td>
<td>0.384</td>
<td>0.443</td>
<td>0.000</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>0.224</td>
<td>0.407</td>
<td>0.369</td>
<td>0.000</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>0.120</td>
<td>0.258</td>
<td>0.622</td>
<td>0.000</td>
<td>0.157</td>
</tr>
<tr>
<td>$\gamma = 5$</td>
<td>Static</td>
<td>0.494</td>
<td>0.073</td>
<td>0.425</td>
<td>0.008</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>0.331</td>
<td>0.281</td>
<td>0.343</td>
<td>0.045</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>0.339</td>
<td>0.290</td>
<td>0.314</td>
<td>0.057</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>0.267</td>
<td>0.220</td>
<td>0.453</td>
<td>0.060</td>
<td>0.137</td>
</tr>
<tr>
<td>$\gamma = 10$</td>
<td>Static</td>
<td>0.285</td>
<td>0.051</td>
<td>0.230</td>
<td>0.434</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>0.239</td>
<td>0.169</td>
<td>0.197</td>
<td>0.395</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>0.246</td>
<td>0.176</td>
<td>0.186</td>
<td>0.392</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>0.222</td>
<td>0.144</td>
<td>0.288</td>
<td>0.347</td>
<td>0.130</td>
</tr>
<tr>
<td>$\gamma = 15$</td>
<td>Static</td>
<td>0.190</td>
<td>0.034</td>
<td>0.153</td>
<td>0.623</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>0.159</td>
<td>0.112</td>
<td>0.131</td>
<td>0.597</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>0.164</td>
<td>0.117</td>
<td>0.124</td>
<td>0.594</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>0.148</td>
<td>0.096</td>
<td>0.192</td>
<td>0.564</td>
<td>0.087</td>
</tr>
</tbody>
</table>
Table 4: Measures of portfolio performance

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Moments of realized ptf return</th>
<th>$M_2$</th>
<th>Performance fee</th>
<th>$Z$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$sk$</td>
<td>$ku$</td>
<td>$SR$</td>
</tr>
<tr>
<td>$\gamma = 2$</td>
<td>Static</td>
<td>6.05</td>
<td>13.02</td>
<td>-0.21</td>
<td>3.67</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>6.70</td>
<td>14.16</td>
<td>-0.02</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>7.03</td>
<td>14.34</td>
<td>0.01</td>
<td>3.38</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>7.33</td>
<td>13.76</td>
<td>-0.03</td>
<td>3.23</td>
</tr>
<tr>
<td>$\gamma = 5$</td>
<td>Static</td>
<td>5.69</td>
<td>12.87</td>
<td>-0.15</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>6.63</td>
<td>12.53</td>
<td>0.05</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>7.03</td>
<td>12.45</td>
<td>0.06</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>7.07</td>
<td>12.33</td>
<td>0.00</td>
<td>3.08</td>
</tr>
<tr>
<td>$\gamma = 10$</td>
<td>Static</td>
<td>4.14</td>
<td>7.43</td>
<td>-0.12</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>4.90</td>
<td>7.73</td>
<td>-0.14</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>5.28</td>
<td>7.81</td>
<td>-0.12</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>6.11</td>
<td>8.20</td>
<td>-0.14</td>
<td>3.09</td>
</tr>
<tr>
<td>$\gamma = 15$</td>
<td>Static</td>
<td>3.34</td>
<td>4.94</td>
<td>-0.13</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>$MV^d$</td>
<td>3.85</td>
<td>5.14</td>
<td>-0.14</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>$HM^c$</td>
<td>4.11</td>
<td>5.21</td>
<td>-0.12</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>$HM^d$</td>
<td>4.63</td>
<td>5.48</td>
<td>-0.16</td>
<td>3.18</td>
</tr>
</tbody>
</table>
Table 5: Optimal portfolio weights  
(short-sales and borrowing allowed)

<table>
<thead>
<tr>
<th>γ</th>
<th>Strategy</th>
<th>US</th>
<th>Japan</th>
<th>UK</th>
<th>US T-bill</th>
<th>Dist</th>
<th>Turn</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Static</td>
<td>1.429</td>
<td>1.153</td>
<td>0.257</td>
<td>-1.839</td>
<td>-</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>MVd</td>
<td>1.197</td>
<td>0.990</td>
<td>0.847</td>
<td>-2.033</td>
<td>0.551</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>HMc</td>
<td>1.218</td>
<td>0.924</td>
<td>0.879</td>
<td>-2.021</td>
<td>0.605</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>HMd</td>
<td>1.090</td>
<td>1.407</td>
<td>0.747</td>
<td>-2.244</td>
<td>0.643</td>
<td>0.120</td>
</tr>
<tr>
<td>5</td>
<td>Static</td>
<td>0.571</td>
<td>0.461</td>
<td>0.103</td>
<td>-0.135</td>
<td>-</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>MVd</td>
<td>0.478</td>
<td>0.396</td>
<td>0.338</td>
<td>-0.212</td>
<td>0.220</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>HMc</td>
<td>0.492</td>
<td>0.371</td>
<td>0.352</td>
<td>-0.215</td>
<td>0.244</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>HMd</td>
<td>0.442</td>
<td>0.566</td>
<td>0.300</td>
<td>-0.307</td>
<td>0.260</td>
<td>0.048</td>
</tr>
<tr>
<td>10</td>
<td>Static</td>
<td>0.285</td>
<td>0.230</td>
<td>0.051</td>
<td>0.434</td>
<td>-</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>MVd</td>
<td>0.239</td>
<td>0.197</td>
<td>0.169</td>
<td>0.395</td>
<td>0.110</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>HMc</td>
<td>0.246</td>
<td>0.186</td>
<td>0.176</td>
<td>0.392</td>
<td>0.122</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>HMd</td>
<td>0.222</td>
<td>0.283</td>
<td>0.150</td>
<td>0.345</td>
<td>0.131</td>
<td>0.024</td>
</tr>
<tr>
<td>15</td>
<td>Static</td>
<td>0.190</td>
<td>0.153</td>
<td>0.034</td>
<td>0.623</td>
<td>-</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>MVd</td>
<td>0.159</td>
<td>0.131</td>
<td>0.112</td>
<td>0.597</td>
<td>0.073</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>HMc</td>
<td>0.164</td>
<td>0.124</td>
<td>0.117</td>
<td>0.594</td>
<td>0.081</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>HMd</td>
<td>0.148</td>
<td>0.189</td>
<td>0.100</td>
<td>0.563</td>
<td>0.087</td>
<td>0.016</td>
</tr>
</tbody>
</table>
Table 6: Measures of portfolio performance
(short-sales borrowing allowed)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Moments of realized ptf return</th>
<th>$M^2$</th>
<th>Performance fee</th>
<th>$\tau^{be}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$sk$</td>
<td>$ku$</td>
</tr>
<tr>
<td>$\gamma = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>13.71</td>
<td>37.24</td>
<td>-0.11</td>
<td>3.61</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>17.52</td>
<td>38.74</td>
<td>-0.13</td>
<td>3.06</td>
</tr>
<tr>
<td>$HM^c$</td>
<td>19.30</td>
<td>38.81</td>
<td>-0.11</td>
<td>3.06</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>23.16</td>
<td>40.74</td>
<td>-0.15</td>
<td>3.15</td>
</tr>
<tr>
<td>$\gamma = 5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>6.53</td>
<td>14.88</td>
<td>-0.12</td>
<td>3.61</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>8.05</td>
<td>15.48</td>
<td>-0.14</td>
<td>3.06</td>
</tr>
<tr>
<td>$HM^c$</td>
<td>8.80</td>
<td>15.60</td>
<td>-0.12</td>
<td>3.07</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>10.36</td>
<td>16.42</td>
<td>-0.16</td>
<td>3.17</td>
</tr>
<tr>
<td>$\gamma = 10$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>4.14</td>
<td>7.43</td>
<td>-0.12</td>
<td>3.61</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>4.90</td>
<td>7.73</td>
<td>-0.14</td>
<td>3.07</td>
</tr>
<tr>
<td>$HM^c$</td>
<td>5.28</td>
<td>7.81</td>
<td>-0.12</td>
<td>3.07</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>6.06</td>
<td>8.22</td>
<td>-0.16</td>
<td>3.18</td>
</tr>
<tr>
<td>$\gamma = 15$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>3.34</td>
<td>4.94</td>
<td>-0.13</td>
<td>3.61</td>
</tr>
<tr>
<td>$MV^d$</td>
<td>3.85</td>
<td>5.14</td>
<td>-0.14</td>
<td>3.07</td>
</tr>
<tr>
<td>$HM^c$</td>
<td>4.11</td>
<td>5.21</td>
<td>-0.12</td>
<td>3.07</td>
</tr>
<tr>
<td>$HM^d$</td>
<td>4.63</td>
<td>5.48</td>
<td>-0.16</td>
<td>3.18</td>
</tr>
</tbody>
</table>
Table 7: Measures of portfolio performance under various scenarios (for $\gamma = 5$)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Moments of realized ptf return</th>
<th>$M^2$</th>
<th>Performance fee</th>
<th>$Z$</th>
<th>$\theta^b$s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$ $\sigma$ $sk$ $ku$ $SR$</td>
<td></td>
<td>$\theta_s$ $\theta_d$ $\theta_c$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: Short-sales and borrowing constraints

- Baseline model
  - Static
    - $MV^d$: 5.69 12.87 -0.15 3.61 0.30 - - - - - -
    - $HM^c$: 7.03 12.45 0.06 2.91 0.42 1.52 1.71 0.45 - 0.52 13.53
    - $HM^d$: 7.07 12.33 0.00 3.08 0.43 1.61 1.83 0.57 0.11 0.54 18.61

- Same expected returns for all markets
  - Static
    - $MV^d$: 6.10 13.13 -0.06 3.27 0.33 - - - - - 0.51 0.00
    - $HM^c$: 6.21 12.06 0.01 2.85 0.37 0.50 0.94 0.76 - 0.52 1.16
    - $HM^d$: 6.47 11.97 -0.01 3.03 0.39 0.82 1.26 1.08 0.32 0.54 5.20

- Same dynamics for shape parameters for all markets
  - $HM^d$: 7.00 12.32 -0.01 3.08 0.42 1.55 1.77 0.51 0.05 0.54 18.12

Panel B: Short-sales and borrowing allowed

- Baseline model
  - Static
    - $MV^d$: 6.53 14.88 -0.12 3.61 0.32 - - - - - -
    - $HM^c$: 8.80 15.60 -0.12 3.07 0.45 1.94 1.61 0.64 - 0.55 17.63
    - $HM^d$: 10.36 16.42 -0.16 3.17 0.52 3.02 2.35 1.38 0.74 0.58 33.34

- Same expected returns for all markets
  - Static
    - $MV^d$: 7.04 15.61 -0.04 3.31 0.34 - - - - - -
    - $HM^c$: 8.21 15.43 -0.19 3.05 0.42 1.25 1.29 0.99 - 0.53 9.39
    - $HM^d$: 9.51 15.88 -0.20 3.17 0.49 2.34 2.14 1.85 0.86 0.56 22.57

- Same dynamics for shape parameters for all markets
  - $HM^d$: 10.21 16.41 -0.17 3.21 0.51 2.89 2.20 1.23 0.59 0.58 33.46
Table A: Estimation of the model with Sk-t distribution and time-varying higher moments

<table>
<thead>
<tr>
<th></th>
<th>Param. est.</th>
<th>Std dev.</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expected returns $\mu_i$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.213</td>
<td>0.072</td>
<td>2.974</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.235</td>
<td>0.049</td>
<td>4.792</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>0.235</td>
<td>0.072</td>
<td>3.283</td>
</tr>
<tr>
<td><strong>Dynamics of the variances $\sigma_{i,t}^2$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.087</td>
<td>0.021</td>
<td>4.161</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.059</td>
<td>0.002</td>
<td>28.982</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.036</td>
<td>0.003</td>
<td>10.846</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.918</td>
<td>0.040</td>
<td>22.941</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.273</td>
<td>0.030</td>
<td>9.010</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.034</td>
<td>0.009</td>
<td>3.705</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.104</td>
<td>0.018</td>
<td>5.646</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.868</td>
<td>0.051</td>
<td>17.013</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>0.157</td>
<td>0.044</td>
<td>3.560</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.059</td>
<td>0.014</td>
<td>4.255</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.022</td>
<td>0.017</td>
<td>1.287</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.907</td>
<td>0.040</td>
<td>22.681</td>
</tr>
<tr>
<td><strong>Dynamics of the correlations $\rho_{ij,t}$</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.008</td>
<td>0.002</td>
<td>4.489</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.977</td>
<td>0.034</td>
<td>28.724</td>
</tr>
</tbody>
</table>
Table A (cont’d): Estimation of the model with Sk-t distribution and time-varying higher moments

<table>
<thead>
<tr>
<th>Dynamics of the degrees of freedom $\nu_{i,t}$</th>
<th>Param. est.</th>
<th>Std dev.</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_{1,0}$</td>
<td>0.015</td>
<td>0.001</td>
<td>13.143</td>
</tr>
<tr>
<td>$c_{1,1}$</td>
<td>3.013</td>
<td>0.089</td>
<td>33.778</td>
</tr>
<tr>
<td>$c_{1,2}$</td>
<td>0.174</td>
<td>0.674</td>
<td>0.258</td>
</tr>
<tr>
<td>$c_{1,1}$ $c_{1,0}$</td>
<td>0.044</td>
<td>0.017</td>
<td>2.520</td>
</tr>
<tr>
<td>$c_{2,0}$</td>
<td>0.012</td>
<td>0.001</td>
<td>22.039</td>
</tr>
<tr>
<td>$c_{2,1}$</td>
<td>2.910</td>
<td>0.089</td>
<td>32.631</td>
</tr>
<tr>
<td>$c_{2,2}$</td>
<td>1.720</td>
<td>0.051</td>
<td>33.503</td>
</tr>
<tr>
<td>$c_{2,2}$ $c_{2,0}$</td>
<td>-0.321</td>
<td>0.008</td>
<td>-38.888</td>
</tr>
<tr>
<td>$c_{3,0}$</td>
<td>0.006</td>
<td>0.001</td>
<td>7.472</td>
</tr>
<tr>
<td>$c_{3,1}$</td>
<td>14.655</td>
<td>0.508</td>
<td>28.853</td>
</tr>
<tr>
<td>$c_{3,2}$</td>
<td>8.629</td>
<td>0.707</td>
<td>12.212</td>
</tr>
<tr>
<td>$c_{3,2}$ $c_{3,0}$</td>
<td>-0.115</td>
<td>0.037</td>
<td>-3.089</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamics of the asymmetry parameters $\xi_{i,t}$</th>
<th>Param. est.</th>
<th>Std dev.</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{1,0}$</td>
<td>0.861</td>
<td>0.072</td>
<td>12.014</td>
</tr>
<tr>
<td>$d_{1,1}$</td>
<td>0.780</td>
<td>0.512</td>
<td>1.524</td>
</tr>
<tr>
<td>$d_{1,2}$</td>
<td>-0.542</td>
<td>0.433</td>
<td>-1.251</td>
</tr>
<tr>
<td>$d_{1,1}$ $d_{1,0}$</td>
<td>-0.821</td>
<td>0.182</td>
<td>-4.498</td>
</tr>
<tr>
<td>$d_{2,0}$</td>
<td>1.086</td>
<td>0.041</td>
<td>26.791</td>
</tr>
<tr>
<td>$d_{2,1}$</td>
<td>1.164</td>
<td>0.618</td>
<td>1.884</td>
</tr>
<tr>
<td>$d_{2,2}$</td>
<td>-0.335</td>
<td>0.451</td>
<td>-0.743</td>
</tr>
<tr>
<td>$d_{2,2}$ $d_{2,0}$</td>
<td>0.695</td>
<td>0.178</td>
<td>3.899</td>
</tr>
<tr>
<td>$d_{3,0}$</td>
<td>0.920</td>
<td>0.053</td>
<td>17.199</td>
</tr>
<tr>
<td>$d_{3,1}$</td>
<td>-0.214</td>
<td>0.424</td>
<td>-0.506</td>
</tr>
<tr>
<td>$d_{3,1}$ $d_{3,0}$</td>
<td>2.335</td>
<td>0.905</td>
<td>2.581</td>
</tr>
<tr>
<td>$d_{3,2}$</td>
<td>-0.191</td>
<td>0.189</td>
<td>-1.009</td>
</tr>
</tbody>
</table>

| lnL                                             | 8639.156    |

<table>
<thead>
<tr>
<th>GoF</th>
<th>US</th>
<th>Japan</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u - \bar{u}$</td>
<td>27.138</td>
<td>22.000</td>
<td>13.077</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>(0.101)</td>
<td>(0.284)</td>
<td>(0.835)</td>
</tr>
<tr>
<td>$u - \bar{u}$</td>
<td>42.428</td>
<td>29.681</td>
<td>24.699</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>(0.002)</td>
<td>(0.075)</td>
<td>(0.213)</td>
</tr>
<tr>
<td>$(u - \bar{u})^2$</td>
<td>15.262</td>
<td>30.280</td>
<td>18.057</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>(0.761)</td>
<td>(0.065)</td>
<td>(0.584)</td>
</tr>
<tr>
<td>$(u - \bar{u})^3$</td>
<td>36.810</td>
<td>29.559</td>
<td>17.787</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>(0.012)</td>
<td>(0.077)</td>
<td>(0.601)</td>
</tr>
<tr>
<td>$(u - \bar{u})^4$</td>
<td>15.021</td>
<td>28.728</td>
<td>17.965</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>(0.775)</td>
<td>(0.093)</td>
<td>(0.590)</td>
</tr>
</tbody>
</table>
Figure 1: Evolution of the conditional skewness

$sk_{1,1,1}$ (US)

$sk_{2,2,2}$ (Japan)

$sk_{3,3,3}$ (UK)

$sk_{1,1,2}$

$sk_{1,2,2}$

$sk_{1,1,3}$

$sk_{1,3,3}$

$sk_{2,2,3}$

$sk_{1,2,3}$

$sk_{2,3,3}$

$sk_{1,2,3}$
Figure 2: Evolution of the conditional kurtosis

- $k_{u_{1,1,1}}$ (US)
- $k_{u_{2,2,2}}$ (Japan)
- $k_{u_{3,3,3}}$ (UK)

Allocation period

$k_{u_{1,1,2}}$ and $k_{u_{2,2,2}}$ (. .)

$k_{u_{1,1,3}}$ and $k_{u_{1,3,3}}$ (. .)

$k_{u_{2,2,3}}$ and $k_{u_{2,3,3}}$ (. .)

$K_{u_{2,2,3}}$

$K_{u_{1,2,3}}$

$K_{u_{1,2,2}}$

$K_{u_{1,3,3}}$
Figure 3: Evolution of the optimal (out-of-sample) portfolio weights with short-sales and borrowing constraints ($\gamma = 5$).
Figure 4: Distribution of the performance fee for $MV^d$ and $HM^d$ strategies
Figure 5: Evolution of the optimal (out-of-sample) portfolio weights without short-sales and borrowing constraints ($\gamma = 5$)
Figure A1: Evolution of the conditional variances
Figure A2: Evolution of the conditional correlations

US – Japan

US – UK

Japan – UK