Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Small Caps in International Equity Portfolios: The Effects of Variance Risk

Massimo Guidolin
and
Giovanna Nicodano

Working Paper 2005-075B
http://research.stlouisfed.org/wp/2005/2005-075.pdf

December 2005
Revised August 2007

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

Small Caps in International Equity Portfolios: The Effects of Variance Risk*

Massimo GUIDOLIN†
Federal Reserve Bank of St. Louis and Manchester Business School

Giovanna NICODANO‡
CeRP Collegio Carlo Alberto and University of Turin

July 2007

Abstract

We show that predictable covariances between means and variances of stock returns may have a first order effect on portfolio composition. In an international asset menu that includes both European and North American small capitalization equity indices, we find that a three-state, heteroskedastic regime switching VAR model is required to provide a good fit to weekly return data and to accurately predict the dynamics in the joint density of returns. As a result of the non-linear dynamic features revealed by the data, small cap portfolios become riskier in bear markets, i.e. display negative co-skewness with other stock indices. Because of this property, a power utility investor ought to hold a well-diversified portfolio, despite the high risk premium and Sharpe ratios offered by small capitalization stocks. On the contrary small caps command large optimal weights when the investor ignores variance risk, by incorrectly assuming joint normality of returns. These results provide the missing partial equilibrium rationale for the presence of co-skewness in the empirical asset pricing models that have been proposed to explain the cross-section of stock returns.

Keywords: intertemporal portfolio choice; return predictability; co-skewness and co-kurtosis; international portfolio diversification.

JEL code: G11, G15, F30, C32.

.

*We thank one anonymous referee, Michael Brandt (a discussant), Andrew Ang, Giovanni Barone Adesi, Giulio Casuccio, Enrico Perotti, Ludo Phalippou, and participants at the American Finance Association 2006 Boston meetings, the Citigroup Quantitative Finance conference in Lisbon (June 2006), EFMA 2006 Madrid meetings, Ente Einaudi Anniversary Conference in Rome (December 2005), European Summer Finance Symposium at Gerzensee (July 2006), III Global Finance Conference in Dublin (June 2005), the C6 CSEF-IGIER Symposium on Economics and Institutions in Capri (July 2005), Royal Economic Society 2006 Nottingham meetings, University of Amsterdam Finance Group, University of Salerno, University of Venice, University of Virginia, and the University of Turin (CeRP) for useful comments. Financial support from CeRP and the Italian Research Ministry is gratefully acknowledged. All errors remain our own.

†Research Division, St. Louis, MO 63166, USA. E-mail: Massimo.Guidolin@stls.frb.org; phone: 314-444-8550.

‡University of Turin, Faculty of Economics, Corso Unione Sovietica, 218bis - 10134 Turin, ITALY. E-mail: giovanna.nicodano@unito.it; phone: +39-011.6706073.

Abstract

We show that predictable covariances between means and variances of stock returns may have a first order effect on portfolio composition. In an international asset menu that includes both European and North American small capitalization equity indices, we find that a three-state, heteroskedastic Markov switching VAR model is required to provide a good fit to weekly return data and to accurately predict the dynamics in the joint density of returns. As a result of the non-linear dynamic features revealed by the data, small cap portfolios become riskier in bear markets, i.e. display negative co-skewness with other stock indices. Because of this property, a power utility investor ought to hold a well-diversified portfolio, despite the high risk premium and Sharpe ratios offered by small capitalization stocks. On the contrary small caps command large optimal weights when the investor ignores variance risk, by incorrectly assuming joint normality of returns. These results provide the missing partial equilibrium rationale for the presence of co-skewness in the empirical asset pricing models that have been proposed to explain the cross-section of stock returns.

1. **Introduction**

Small capitalization stocks have become important to international investors with the development of new technologies and venture capital. They are however known to be rather peculiar assets in that their returns display – along with high average risk premia – asymmetric risk across bull and bear markets (Ang and Chen, 2002). For instance, small caps generally imply high risk in cyclical downturns due to tighter credit constraints associated to lower firm collateral (Perez-Quiros and Timmermann, 2000). Several papers focus on international portfolio choice under a variety of assumptions concerning the asset menu and the process generating asset returns, e.g. Ang and Bekaert (2002) and De Santis and Gerard (1997). However, no specific attention has been given to small capitalization firms. Our paper studies the contribution of small caps to the international diversification of stock portfolios under realistic specifications for the stochastic process driving asset returns, that allow for asymmetric risk.

Developing such a perspective on small capitalization firms appears to be warranted in the light of recent asset pricing research showing that the cross-sectional distribution of the equity risk premium is related to "variance risk" (Harvey and Siddique, 2000; Dittmar, 2002; Barone-Adesi, Gagliardini, and Urga, 2004), i.e. the correlation between returns and aggregate volatility as well as between individual stock volatility and aggregate volatility. Acharya and Pedersen (2005) have provided the economic rationale for why co-skewness (what Black, 1976, referred to as "leverage") should impact expected returns: An investor may dislike small caps – whose illiquidity/volatility increase when the market is bear – because they fail to provide liquidity when she may want to trade. Similarly, an investor may prefer large to small caps because the returns on the latter fall when market volatility/illiquidity is high, and therefore fail to provide insurance against market volatility. Hence small caps command higher expected returns than large caps, which happen to have low variance risk, i.e. positive co-skewness. Additionally, also co-kurtosis could be priced, if an investor dislikes assets whose risk increases in volatile markets.

While predictable variance risk plays a role in these pricing models, it is absent in the literature on dynamic portfolio choice that has mostly focused on the ability to forecast expected returns (Campbell and Viceira, 1999), even in multivariate asset menus (Campbell, Chan, and Viceira, 2003). However, it is clear that the partial equilibrium, portfolio choice counterparts of the findings in the asset pricing literature ought to show that optimal portfolio weights respond to predictable changes in the covariances between returns and volatility as well as among cross-sectional volatilities. Our paper contributes to the literature on intertemporal portfolio choice with predictable returns by showing that the interaction of predictability in mean returns and volatilities has first order effects on portfolio composition in a multivariate asset menu. In particular, we show that the portfolio share of small caps is significantly reduced by their high variance risk, which is captured by their negative (positive) co-skewness. This provides the missing partial equilibrium rationale for the role of variance risk in explaining differences in expected returns associated to firm size.

Our paper makes two choices in order to tease out the effects of variance risk on asset allocation. First, we focus on an international equity diversification problem in which both U.S. and European small cap portfolios figure prominently. The case for studying a non-U.S. portfolio of small caps one is based on two observations. First, the European size effect has been basically neglected by the asset pricing literature that has instead focused primarily on U.S. data. Since such a focus poses data-snooping problems, it is important to prevent our estimates of the small caps share in optimal portfolios to depend entirely on well-known but possibly random features of U.S. data. Second, U.S. small caps have experienced an unprecedented performance in the first part of our sample, from January 1999 to June 2001. Since a concern has been expressed that the size premium may contain long and persistent swings (see Pástor, 2000 and Guidolin and Timmermann, 2005), it is useful to obtain evidence involving at least one additional, major market for small caps.

Second, measuring the effect of predictable co-skewness on portfolio choice requires abandoning the traditional mean-variance approach. On the one hand, we assume that the investor has a power utility function, implying a preference for positively skewed wealth as well as aversion to kurtosis of final wealth. On the other hand, we allow the return process to generate non-normal and/or predictable returns. In particular, we examine the fit of competing models of asset returns, including multivariate ARCH models, linear VARs as well as Markov switching VAR processes. It turns out that the latter are able to account for both non-normality, asymmetric correlations, and predictability. Finally, a parametric Markov switching framework allows us to obtain precise estimates of the high order (co-)moments that characterize variance risk.

Using a 1999-2007 weekly international data set, we find that the joint distribution of equity returns is well captured by a three-state model. The states can be ordered by increasing risk premia. In two of the regimes – that we label *bear* and *bull* because of the implied levels of expected returns – European small caps returns exhibit both a relatively low variance and a high Sharpe ratio. Thus a risk averse investor, who is assumed to start from this regime, would invest in excess of 60% her equity portfolio in European small caps for horizons up to two years. On the other hand, the change in regime-specific variance is the highest just for European small caps: in particular, variance almost triples when the regime shifts from bear to the third, *crash* state. The high variance 'excursion' across regimes is compounded by the presence of high and negative co-skewness with other asset returns, which means that the European small variance is high when other excess returns are negative, and European small returns are small when the 'market' is highly volatile. Similarly, the co-kurtosis of European small excess returns with other excess returns series is high – i.e., the variance of the European small class tends to correlate with the variance of other assets. Both these features suggest a tendency of European small caps to suffer from a disproportionate variance risk. The striking implication is that a rational investor ought to give European small caps a limited weight (as low as 0% for short horizons) when she is ignorant about the nature of the regime, which is a realistic situation. Further experiments reveal that the dominant factor in inducing such shifts in optimal weights is represented by the co-skewness, the predictable, time-varying covariance between returns and volatilities. This shows that higher moments of the return distribution can considerably reduce the desirability of an asset. We quantify such an effect in about 100 basis points per year under the steady-state distribution for returns. These results provide a demand-side justification for the dependence of asset prices on co-skewness — as uncovered by Harvey and Siddique (2000).

Importantly, in this paper we estimate a variety of non-linear models – in particular from the multivariate, asymmetric ARCH-in-mean family, besides regime switching models – for the dynamics of international equity returns. Consistently with prior evidence by Ang and Chen (2002), who report that regime switching models may replicate the asymmetries in correlations observed in stock returns data better than GARCH-M and processes, we find that both in-sample and out-of-sample, Markov switching models with time-varying covariance matrix fare as well as (or better than) multivariate ARCH models. However, our robustness checks also confirm that our main portfolio implications would be qualitatively intact were we to adopt a dynamic conditional correlations EGARCH(1,1) VAR(1) model as our baseline specification. Additionally, our results prove qualitatively robust when *both* European and North American small caps are introduced in the analysis.

Our work is closely related to the research on the effects of predictability on intertemporal portfolio choice. This strand of research has often concluded that predictable variance does not exert large effects. We extend and qualify this observation by showing that the interaction of predictable variance with predictable mean returns has first order effects on investors' choices provided that assets with non-symmetric return distributions are included in the investment set. Our application also bears similarities with Ang and Bekaert (2002), and Guidolin and Timmermann (2006) who investigate the effects on international diversification of time-varying moments when regime shifts are accounted for. Similarly to these papers, we overlook the analysis of inflation risk, informational differences, and currency hedging costs that – while generally important – may not radically affect the rational choices of a large investor who can hedge currency risk. Differently from these papers, we focus on issues of international diversification across small and large capitalization firms.

Cvitanić, Polimenis, and Zapatero (2007) characterize optimal portfolio weights for the case of the choice between one risky and a riskless asset when the dynamics of the risky asset is subject to pure-jump risk and the jump arrival rates are stochastic. Differently from Liu et al. (2003), in which jumps arrive at a finite Poisson rate, jumps may arrive at an infinite rate. The paper shows that under pure-jump dynamics all the moments of risky returns are affected by the presence of jumps generating skewness and kurtosis and departures from normality. When the process is specified to be a Variance-Gamma type, numerical examples are offered that imply that welfare costs of misspecifying the dynamics of the return process may be substantial. Although our econometric framework is different, our paper provides a specific, multivariate case study in which first-order effects are derived from non-Gaussian dynamics. Similarly to Cvitanić, Polimenis, and Zapatero (2007) we uncover substantial utility costs from misspecifying the process of returns.[1]

Finally, Harvey and Siddique (2000) show that conditional skewness contributes to the explanation of cross sectional U.S. expected returns. They highlight that small cap portfolios have high expected returns together with negative co-skewness while low expected returns, large cap portfolios have positive co-skewness. Therefore they suggest analyzing portfolio choice in a richer conditional mean-variance-skewness framework. Dittmar (2002) allows for expected returns to be related also to co-kurtosis between returns and aggregate wealth, finding however a modest impact. In our framework, all moments above the second may be responsible for departures of portfolio shares from the mean variance ones. We are able to show, however, that a large fraction of such departures come from the (co-) skewness of the multivariate return distribution.

The paper is organized as follows. Section 2 presents a range of econometric models that have the capability of generating variance risk. Section 3 introduces the portfolio choice problem and illustrates the methodology employed to compute welfare costs from imposing restrictions on either the asset allocation model or the asset menu. Section 4 describes the data, documents the outcomes of a model specification

[1] Das and Uppal (2004) study the effects of jumps on international equity portfolios when jumps are simultaneous and perfectly correlated across assets. We also assume that regimes are perfectly correlated across stock portfolio returns, but allow for persistence of regimes. While this prevents us from obtaining Das and Uppal's simple analytic results, it allows to compute portfolio allocations conditional on a given regime when the investor anticipates the probability of a regime shift next period.

strategy based on the in- and out-of-sample performance, and gives econometric estimates. Section 5 reports our findings on international portfolio diversification. This section contains the core results of the paper and is organized around three sub-sections, each describing alternative sets of experiments useful to document the effects of variance risk. Section 6 extends the asset menu to include North American small, besides European small stocks and investigates the robustness of our results to the choice of the econometric framework. Section 7 concludes.

2. **Econometric Models of Variance Risk**

In general terms, any multivariate econometric model implying a non-zero correlation between levels vs. squares and squares vs. squares of individual as well as aggregate (market) returns can be used to capture and forecast what we have defined as variance risk. In this Section we provide an introduction to a variety of parametric frameworks that display such properties, ranging from classical multivariate ARCH models, to Dynamic Conditional Correlation (DCC) models, to models with regimes driven by Markov switching processes. Clearly, all these models have a non-linear nature, in the sense that either their second moments are predictable (and often all these moments are tied together, like in GARCH-in-mean models) or at least their first moments are subject to discrete shifts driven by some switching mechanism. In addition to giving some basic information on the structure of the models in each class and on related estimation issues, in this Section we describe which specific components in each model are responsible for generating variance risk.

2.1. *Markov Switching VARs*

The popular press often acknowledges the existence of stock market states by referring to them as "bull" and "bear" markets. Here we consider that the distribution of a set of international equity indices may depend on states characterizing international equity markets. We write the joint distribution of a vector of m returns, conditional on an *unobservable* state variable S_t, as:

$$\mathbf{r}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t} \mathbf{r}_{t-j} + \mathbf{u}_t \quad \mathbf{u}_t | \mathcal{F}_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{S_t}), \tag{1}$$

where $\mathbf{r}_t$ is the $m \times 1$ vector collecting stock returns, $\boldsymbol{\mu}_{S_t}$ is a vector of intercepts (these correspond to expected returns when either $p = 0$ or $\mathbf{r}_{t-j} = \mathbf{0}$ for $j = 1, ..., p$) in state S_t, $\mathbf{A}_{j,S_t}$ is the matrix of autoregressive coefficients at lag j in state S_t, and $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{S_t})$ is the vector of return innovations which are assumed to be jointly normally distributed with zero mean and state-specific covariance matrix $\boldsymbol{\Sigma}_{S_t}$. S_t is an indicator variable taking values $1, 2, ...k$, where k is the number of states. The presence of heteroskedasticity is allowed in the form of regime-specific covariance matrices.

Crucially, S_t is never observed and the nature of the state at time t may at most be inferred (filtered) by the econometrician using the history of asset returns. Similarly to most of the literature on regime switching models (see e.g. Ang and Bekaert, 2002), we assume that S_t follows a first-order Markov chain. Moves between states are assumed to be governed by a constant transition probability matrix, $\mathbf{P}$, with generic element p_{ij} defined as

$$\Pr(s_t = j | s_{t-1} = i) = p_{ij}, \quad i, j = 1, .., k, \tag{2}$$

i.e. the probability of switching to state i between t and $t+1$ given that at time t the market is in state j. While we allow for the presence of regimes, we do not exogenously impose or characterize them, consistently with the true unobservable nature of the state of markets in real life. On the contrary, in Section 4 we will conduct a thorough specification search letting the data endogenously determine the number of regimes k (as well as the VAR order, p) required to provide an accurate fit to the data and/or to correctly predict their distribution one-step ahead. Although highly flexible, Markov switching VARs may imply a need to estimate a relatively large number of parameters. For instance, (1) implies $km[1+pm+(m+1)/2]+k(k-1)$ parameters, e.g. as many as 96 free parameters in the case $m=4$, $p=1$, and $k=3$, which will represent a reasonable specification in our application.

(1) nests several return processes as special cases. If there is a single market regime, we obtain the linear VAR model with predictable mean returns that is commonly used in the literature on strategic asset allocation, see e.g. Campbell and Viceira (1999).[2] However, when multiple regimes are allowed, (1) generates various sources of predictability. When either $\boldsymbol{\mu}_{S_t}$ or $\mathbf{A}_{j,S_t}$ $(j=1,...,p)$ depend on the latent state, then expected returns vary over time. Similarly, when the covariance matrices differ across states there will be predictability in higher order moments such as volatilities, correlations, skews and tail thickness, see Timmermann (2000). Predictability is therefore not confined to mean returns but carries over to the entire return distribution. Notice also that while current returns are normally distributed conditional on the state, the one- period ahead return distribution is *not* simply normal with regime dependent conditional mean and/or regime dependent conditional volatility. It is instead a mixture – i.e., a probability weighted combination, with time-varying weights (the regime probabilities) that are updated as new return data arrive – of normal variates, which is generally not Gaussian. Furthermore, because the T-period ahead distribution is a mixture of Gaussian densities, higher order moments generally become more relevant (i.e. departures from the baseline, conditional multivariate normal get stronger and stronger) as T grows.

Since we treat the state of the market as unobservable – which is consistent with the idea that investors cannot observe the true state but can use the time-series of returns to filter the state – we model the evolution of investors' beliefs using a standard Bayesian updating algorithm (see Hamilton, 1990):

$$\boldsymbol{\pi}_{t+1}(\hat{\boldsymbol{\theta}}_t) = \frac{\left(\boldsymbol{\pi}_t'(\hat{\boldsymbol{\theta}}_t)\hat{\mathbf{P}}_t\right)' \odot \mathbf{f}(\mathbf{r}_{t+1};\hat{\boldsymbol{\theta}}_t)}{[(\boldsymbol{\pi}_t'(\hat{\boldsymbol{\theta}}_t)\hat{\mathbf{P}}_t)' \odot \mathbf{f}(\mathbf{r}_{t+1};\hat{\boldsymbol{\theta}}_t)]'\boldsymbol{\iota}_k}. \tag{3}$$

Here $\boldsymbol{\pi}_t(\hat{\boldsymbol{\theta}}_t)$ collects the $k\times 1$ vector of state probabilities and $\hat{\boldsymbol{\theta}}_t$ all the estimated parameters characterizing (1) and estimated at time t; $\odot$ denotes the element-by-element product, $\hat{\mathbf{P}}_t$ is the Markov transition matrix, and $f(\cdot)$ is the density of returns conditional on the regime, on past returns and on estimated parameters

$$\mathbf{f}(\mathbf{r}_{t+1};\hat{\boldsymbol{\theta}}_t) = \begin{bmatrix} (2\pi)^{-\frac{N}{2}}|\hat{\boldsymbol{\Sigma}}_1^{-1}|^{\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\mathbf{r}_t-\hat{\boldsymbol{\mu}}_1-\sum_{j=0}^{p-1}\hat{\mathbf{A}}_{1j}\mathbf{r}_{t-j}\right)'\hat{\boldsymbol{\Sigma}}_1^{-1}\left(\mathbf{r}_t-\hat{\boldsymbol{\mu}}_1-\sum_{j=0}^{p-1}\hat{\mathbf{A}}_{1j}\mathbf{r}_{t-j}\right)\right] \\ \vdots \\ (2\pi)^{-\frac{N}{2}}|\hat{\boldsymbol{\Sigma}}_k^{-1}|^{\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\mathbf{r}_t-\hat{\boldsymbol{\mu}}_k-\sum_{j=0}^{p-1}\hat{\mathbf{A}}_{kj}\mathbf{r}_{t-j}\right)'\hat{\boldsymbol{\Sigma}}_1^{-1}\left(\mathbf{r}_t-\hat{\boldsymbol{\mu}}_k-\sum_{j=0}^{p-1}\hat{\mathbf{A}}_{kj}\mathbf{r}_{t-j}\right)\right] \end{bmatrix},$$

[2]The i.i.d. Gaussian model – also often adopted as a benchmark in the portfolio choice literature (see e.g. Barberis, 2000) – obtains instead assuming $k=1$ and $p=0$.

which exploits the fact that conditional on the state, stock returns have a Gaussian distribution. (3) implies that the probability of the states at time $t+1$ is a weighted average of the one-step ahead predicted probabilities ($\boldsymbol{\pi}_t'(\hat{\boldsymbol{\theta}}_t)\hat{\mathbf{P}}_t$), with weights provided by the likelihood of observing the realized returns $\mathbf{r}_{t+1}$ conditional on each of the possible states, as represented by scaled versions of $\mathbf{f}(\mathbf{r}_{t+1};\hat{\boldsymbol{\theta}}_t)$.

Regime switching models are estimated by maximum likelihood. As shown by Hamilton (1990), the relevant algorithms are considerably simplified if (1) is put in its state-space form. In particular, estimation and inferences are based on the EM (Expectation-Maximization) algorithm proposed by Dempster et al. (1977), a filter that allows the iterative calculation of the one-step ahead forecast of the state vector

$$\boldsymbol{\xi}_t = [I(S_t = 1)\, I(S_t = 2)\ I(S_t = k)]'$$

where $I(S_t = i)$ is a standard indicator variable. As for the properties of the resulting maximum likelihood (MLE) estimators, under standard regularity conditions (such as identifiability, stability and the fact that the true parameter vector does not fall on the boundaries) Hamilton (1989, 1990) and Leroux (1992) have proven consistency and asymptotic normality of the ML estimator $\hat{\boldsymbol{\theta}}$:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathcal{I}_a(\boldsymbol{\theta})^{-1}\right)$$

where $\mathcal{I}_a(\boldsymbol{\theta})$ is the asymptotic information matrix. In our empirical results we are going to provide standard results based on a 'sandwich' sample estimator of $\mathcal{I}_a(\boldsymbol{\theta})$ by which:

$$\widetilde{Var}(\hat{\boldsymbol{\theta}}) = T^{-1}\left[\mathcal{I}_2(\hat{\boldsymbol{\theta}})\left(\mathcal{I}_1(\hat{\boldsymbol{\theta}})\right)^{-1}\mathcal{I}_2(\hat{\boldsymbol{\theta}})\right],$$

where

$$\mathcal{I}_1(\hat{\boldsymbol{\theta}}) \equiv T^{-1}\sum_{t=1}^{T}\left[\mathbf{h}_t(\hat{\boldsymbol{\theta}})\right]\left[\mathbf{h}_t(\hat{\boldsymbol{\theta}})\right]' \quad \mathbf{h}_t(\hat{\boldsymbol{\theta}}) \equiv \frac{\partial \ln p(\mathbf{y}_t|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \quad \mathcal{I}_2(\hat{\boldsymbol{\theta}}) \equiv -T^{-1}\sum_{t=1}^{T}\left[\frac{\partial^2 \ln p(\mathbf{y}_t|\mathcal{F}_{t-1};\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right],$$

($p(\mathbf{y}_t|\Im_{t-1};\tilde{\boldsymbol{\gamma}})$ is the conditional density of the data).

2.2. *Multivariate GARCH*

Consider the single-state VAR(p) model

$$\mathbf{r}_t = \boldsymbol{\mu} + \sum_{j=1}^{p}\mathbf{A}_j\mathbf{r}_{t-j} + \mathbf{u}_t \quad \mathbf{u}_t|\mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t)$$

where $\mathbf{H}_t = E_{t-1}[\mathbf{u}_t\mathbf{u}_t']$ is the conditional variance covariance matrix of the $m \times 1$ vector of asset returns. Engle and Kroner (1995) propose the following model for the conditional covariance matrix that generalizes to the multivariate case Bollerslev's (1986) univariate GARCH(1,1):

$$\mathbf{H}_t = \boldsymbol{\Omega} + \mathbf{A}(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}')\mathbf{A}' + \mathbf{B}\mathbf{H}_{t-1}\mathbf{B}', \tag{4}$$

where $\boldsymbol{\Omega}$, $\mathbf{A}$, and $\mathbf{B}$ are $m \times m$ parameter matrices and $\boldsymbol{\varepsilon}_t$ is the standardized vector of residuals, $\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{-1/2}\mathbf{u}_t$. $\mathbf{H}_t$ is guaranteed to positive definite as long as $\boldsymbol{\Omega}$ is positive definite, which can be ensured by re-parameterizing $\boldsymbol{\Omega}$ as $\boldsymbol{\Omega} = \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}'$ with $\boldsymbol{\Upsilon}$ lower triangular. Notice that this model does contain a large number of parameters ($m(1+pm)$ from the conditional mean function and $2m^2+ m(m+1)/2$ from the conditional variance covariance one), as many as 62 in the case of $m = 4$ and $p = 1$ as in most of the empirical work that follows.[3]

Interestingly, the baseline M-GARCH model in (4) does capture only one possible source of variance risk, i.e., the existence of co-movements and predictability across conditional variances (and covariances) of different assets: because the $[i, j]$ element of $\mathbf{H}_t$ can be written as a function of

$$\sum_{l=1}^{m}\sum_{k=1}^{m}\alpha_{lk}^{ij}\varepsilon_{l,t}\varepsilon_{k,t}+\sum_{l=1}^{m}\sum_{k=1}^{m}\beta_{lk}^{ij}h_{lk,t-1},$$

it is clear that both cross-products of past return shocks and past variances and co-variances will affect subsequent conditional variances and covariances of returns. On the opposite, by construction (4) prevents any potential *explicit* covariation of expected returns and second moments across different assets and portfolios.

A simplified yet popular version of (4) is Bollerslev's (1990) constant correlation multivariate GARCH in which conditional correlations (ρ_{ij}) are assumed to be constant over time so that the conditional covariances in $\mathbf{H}_t$ are derived by the simple product

$$h_{ij,t}=\rho_{ij}\sqrt{h_{i,t}}\sqrt{h_{j,t}},$$

where the conditional variances $h_{i,t}$ and $h_{j,t}$ are estimated from plain vanilla, univariate GARCH(1,1) models (usually with long-run variance matching restrictions imposed),

$$h_{i,t}=(1-\alpha_i-\beta_i)\bar{\sigma}_i+\beta_i h_{i,t-1}+\alpha_i\varepsilon_{i,t}^2, \tag{5}$$

and the constant correlations ρ_{ij} are simply set to match their unconditional counterparts, $\bar{\rho}_{ij}$. Of course, whilst the assumption of constant correlations may be questionable, the main advantage of the constant correlation M-GARCH model is that the number of parameters drops to $m(1+pm)+ m[2+(m-1)/2]$, for instance to 34 only in the case of $m = 4$ and $p = 1$. Importantly, the constant correlation model is proposed purely as benchmark − since it has been widely used in the empirical finance literature − even though it fails to generate any variance risk: although variances and covariances (not correlations) become stochastic and therefore time-varying, expected returns fail to co-vary with variances or co-variances, while the narrow univariate GARCH(1,1) specification in (5) fails to capture any co-variation between conditional variances or covariances across assets.

2.3. *Dynamic Conditional Correlation*

Engle (2002) has recently proposed the DCC class as a way to overcome the well-known over-parameterization (and hence, estimation) problems that plague the M-GARCH class. A DCC model is based on the idea

[3] The considerable growth in the number of parameters to be estimated derives from the fact that (4) implies that both conditional variances ($h_{i,t}$, on the main diagonal of $\mathbf{H}_t$, $i = 1, ..., m$) and conditional covariances ($h_{ij,t}$, $i \neq j$) are allowed to depend on past conditional covariances as well as cross-products of return residuals, $\varepsilon_{i,t}\varepsilon_{j,t}$. In our application we have actually imposed Engle and Mezrich's (1996) restriction that the long run variance matrix corresponds to the sample covariance matrix, i.e. that in (4) the restriction $\mathbf{\Omega} = (\mathbf{I}-\mathbf{A}-\mathbf{B})\mathbf{S}$ applies, where $\mathbf{S}$ is the sample covariance matrix. The result of the estimation differs from the maximum likelihood estimator (MLE) only in finite samples but reduces the number of parameters and often gives improved performance in forecasting applications.

of a two-step approach: first estimate conditional variances at the univariate level, and second directly parameterize conditional covariances, using the first-stage estimates of conditional variances to go from the conditional covariances to conditional correlations. Formally, a DCC model writes the $m \times m$ covariance matrix $\mathbf{H}_t$ as

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t,$$

where $\mathbf{D}_t$ is a diagonal matrix which collects the time t volatilities obtained from any of the well-known GARCH model (or combination thereof), $\mathbf{D}_t = diag\{\sqrt{h_{1t}}, \sqrt{h_{2t}}, ..., \sqrt{h_{mt}}\}$, and $\mathbf{R}_t$ is a time-varying matrix which collects ones on the main diagonal, while its generic $[i, j]$ element is:

$$\rho_{ij,t} = \frac{E_{t-1}[\varepsilon_{i,t}\varepsilon_{j,t}]}{\sqrt{E_{t-1}[\varepsilon_{i,t}^2]E_{t-1}[\varepsilon_{j,t}^2]}} \quad i \neq j,$$

where $\varepsilon_{i,t}$ is the standardized residual from the conditional mean function $E_{t-1}[r_{i,t}]$,

$$\varepsilon_{i,t} = \frac{r_{i,t} - E_{t-1}[r_{i,t}]}{\sqrt{h_{i,t}}} \quad i = 1, 2, ..., m.$$

In this paper we entertain three alternative DCC specifications. The first one is a simple DCC-EGARCH(1,1) specification which can be written as[4]

$$\begin{aligned}
\mathbf{r}_t &= \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{r}_{t-j} + \mathbf{u}_t \quad \mathbf{u}_t | \mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t) \\
\mathbf{e}_i' \mathbf{D}_t^2 \mathbf{e}_i &= exp\left[\omega_i + \beta_i ln(h_{i,t-1}) + \alpha_i |\varepsilon_{i,t}| + \gamma_i \varepsilon_{i,t}\right] \\
q_{ij,t} &= \delta_0 + \delta_1 \varepsilon_{i,t-1}\varepsilon_{j,t-1} + \delta_2 q_{ij,t-1} \quad \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} \cdot q_{jj,t}}} \quad i \neq j.
\end{aligned} \tag{6}$$

The correlation estimator $\rho_{ij,t}$ is guaranteed to be positive definite as the matrix $\mathbf{Q}_t$ with generic element $q_{ij,t}$ can be shown to be a weighted-average of a positive definite and a positive semi-definite matrix, $\mathbf{Q}_t = \boldsymbol{\Delta}_0 + \boldsymbol{\Delta}_1(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}') + \boldsymbol{\Delta}_2 \mathbf{Q}_{t-1}$. Also notice that by construction, when $i = j$ then $\rho_{ii,t} = 1$, as it should be. When the restriction that the long run variance matrix corresponds to the sample covariance matrix is imposed, the process for $q_{ij,t}$ can be simply re-written as

$$\begin{aligned}
q_{ij,t} &= (1 - \delta_1 - \delta_2)\bar{v}_{ij} + \delta_1 \varepsilon_{i,t-1}\varepsilon_{j,t-1} + \delta_2 q_{ij,t-1} \\
&= \bar{v}_{ij} + \delta_1(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \bar{\rho}_{ij}) + \delta_2(q_{ij,t-1} - \bar{\rho}_{ij})
\end{aligned}$$

where $\bar{v}_{ij}$ is the unconditional covariance between the residuals from assets i and j. In this case, variance risk is generated by the asymmetric component (sometimes called *leverage*, from the fact that negative stock returns imply declining stock prices and therefore a reduction of the value of the equity relative to corporate debt, and thus an increase in corporate leverage) $\gamma_i \varepsilon_{i,t}$: assuming that $\alpha_i > 0$ and $\gamma_i < 0$, it is clear that

$$\left.\frac{\partial lnh_{i,t}}{\partial \varepsilon_{i,t}}\right|_{\varepsilon_{i,t} \geq 0} = \alpha_i + \gamma_i \leq \alpha_i - \gamma_i = \left.\frac{\partial lnh_{i,t}}{\partial \varepsilon_{i,t}}\right|_{\varepsilon_{i,t} < 0},$$

[4] In what follows the EGARCH label refers to the conditional variance component. As far as the conditional correlations are concerned, we simply use the GARCH(1,1)-type structure suggested in Engle (2002, p. 341). $\mathbf{e}_i' \mathbf{D}_t^2 \mathbf{e}_i$ selects the $i-$th (squared) element of the diagonal matrix $\mathbf{D}_t$, where $\mathbf{e}_i$ is an $m \times 1$ vector with a 1 in the $i-$th position and zeros elsewhere.

i.e., when $\varepsilon_{i,t} < 0$ there is a higher impact of a shock on the predicted variance than when $\varepsilon_{i,t} \geq 0$. The leverage effect creates a positive correlation between variance and expected returns. In this sense, assets or portfolios with higher (absolute) values of the coefficient γ_i will be characterized by higher variance risk.

The second DCC model used in this paper is a DCC-GARCH(1,1)-in-mean model:

$$\begin{aligned}
\mathbf{r}_t &= \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{r}_{t-j} + \mathbf{C} \cdot diag(\mathbf{D}_t) + \mathbf{u}_t \quad \mathbf{u}_t | \mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t) \\
\mathbf{e}_i' \mathbf{D}_t^2 \mathbf{e}_i &= (1 - \alpha_i - \beta_i)\bar{\sigma}_i + \beta_i h_{i,t-1} + \alpha_i \varepsilon_{i,t}^2 \\
q_{ij,t} &= \bar{v}_{ij} + \delta_{ij,1}(\varepsilon_{i,t-1}\varepsilon_{j,t-1} - \bar{v}_{ij}) + \delta_{ij,2}(q_{ij,t-1} - \bar{v}_{ij}) \quad \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} \cdot q_{jj,t}}} \quad i \neq j,
\end{aligned} \tag{7}$$

where $\mathbf{C}$ is a full matrix. In this case variance risk comes from the ARCH-in mean component represented by $\mathbf{C} \cdot diag(\mathbf{D}_t)$ (where $diag(\cdot)$ is the operator that stacks the elements on the main diagonal in $\mathbf{D}_t$ into an $m \times 1$ vector): when the $[i, j]$ element of $\mathbf{C}$ is non-zero, then

$$\left. \frac{\partial E_{t-1}[r_{i,t}]}{\partial (h_{j,t}^{1/2})} \right| = c_{ij}$$

which means than a change in the volatility of asset j will affect the expected return on asset i. Notice that while basic finance theory suggests that $c_{ii} > 0$ (i.e., higher own-volatility increases expected returns), no priors are commonly expressed with reference to the off-diagonal elements of $\mathbf{C}$.[5]

One final version of (7) that we estimate in this paper is the integrated DCC-GARCH(1,1) version, in which the restrictions $\alpha_i + \beta_i = 1$ and $\delta_1 + \delta_2 = 1$ are imposed so that

$$\begin{aligned}
\mathbf{r}_t &= \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{r}_{t-j} + \mathbf{C} \cdot diag(\mathbf{D}_t) + \mathbf{u}_t \quad \mathbf{u}_t | \mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t) \\
\mathbf{e}_i' \mathbf{D}_t^2 \mathbf{e}_i - \mathbf{e}_i' \mathbf{D}_{t-1}^2 \mathbf{e}_i &= \Delta h_{i,t-1} = \alpha_i (\varepsilon_{i,t}^2 - h_{i,t-1}) \\
\Delta q_{ij,t} &= \delta_1 (\varepsilon_{i,t-1}\varepsilon_{j,t-1} - q_{ij,t-1}) \quad \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} \cdot q_{jj,t}}} \quad i \neq j,
\end{aligned} \tag{8}$$

i.e. both conditional variances and conditional covariances follow driftless integrated moving average processes. The comments expressed above with regard to the sources of variance risk apply to this case as well, with the difference that the elements of $\mathbf{D}_t$ will obviously be considerably more persistent in determining expected returns. Of course, an advantage the integrated DCC is that $2m$ less parameters have to be estimated.

Thanks to the joint normality assumptions in (6) and (7) one can easily write down the likelihood function for a DCC model and proceed to numerical optimization to obtain MLE efficient (i.e., reaching the Cramer bound) estimates. Alternative sets of sufficient conditions (see e.g., Newey and McFadden, 1994) may be employed to yield consistency and asymptotic normality. However, Engle (2002) shows that a number of tricks may be used to obtain consistent but inefficient estimators that greatly simplify our task.[6] In particular, it is easy to show that the log-likelihood function may be written as the sum of a volatility part and of a correlation part.

[5] (6) and (7) considerably reduce the number of parameters to be estimated, to $m(1 + pm) + 6m$ in the DCC-EGARCH case (e.g., for $m = 4$ and $p = 1$ these are 44 parameters) and to $m(1 + pm) + 4m$ in the DCC-GARCH case in which long-run restrictions are imposed on both variances and correlations (e.g., for $m = 4$ and $p = 1$ this yields a tight set of 36 parameters only).

[6] The existence of these tricks and the notorious difficulties with the estimation of multivariate ARCH models with more

2.4. *The Economics of Variance Risk*

In this paper what we have defined as variance risk derives from the presence of either switches in the expected returns and/or variances and covariances of portfolio returns, or from the presence of peculiar features within the class of multivariate ARCH models, in the form of asymmetries (like in the EGARCH case), cross-asset dependence of second moments (e.g., when the volatility of asset i is persistently affected by the volatility of asset j), and/or the presence of time-varying second moments in the conditional mean function (ARCH-in-mean effects). Therefore variance risk is directly caused by regime switching and/or conditional autoregressive heteroskedasticity. As a result, discussing the economics of variance risk implies researching the origins of these important and widely documented statistical features.

As for regime switching, there are solid economic reasons why the equilibrium joint distribution of a number of stock portfolio returns may contain regimes. Suppose that investors have constant relative risk aversion and that asset returns are determined from the standard no-arbitrage, equilibrium relation $E_t[M_{t+1}(1 + r_{i,t})] = 1$, where M_{t+1} is the pricing kernel which is commonly restricted to be $M_{t+1} \equiv \beta(C_{t+1}/C_t)^{-\gamma}$ and $g_{t+1} \equiv C_{t+1}/C_t$ is real per-capita consumption growth. The risk premium on risky assets (over and above the conditionally risk-free rate, r_t^f) is then given by

$$E_t[r_{i,t+1} - r_t^f] = -\frac{Cov_t[M_{t+1}, (r_{i,t+1} - r_t^f)]}{E_t[M_{t+1}]}.$$

If the consumption growth rate follows a simple regime switching process, $g_{t+1} \sim N(\mu_{S_{t+1}}, \sigma^2_{S_{t+1}})$ ($S_{t+1} = 1, ..., k$), i.e. both the mean and the variance of the rate of growth of fundamentals may take a number of different values, according to the state of the economy (e.g., expansions and recessions). This implies that also the pricing kernel will follow a $k-$state process. It is then straightforward to show that

$$E_t[r_{i,t+1} - r_t^f] = -\frac{\sum_{s_{t+1|t}=1}^{k} \pi_{s_{t+1|t}} Cov[M_{t+1}, (r_{i,t+1} - r_t^f)|s_{t+1}]}{\sum_{s_{t+1|t}=1}^{k} \pi_{s_{t+1|t}} E[M_{t+1}|s_{t+1}]},$$

where $\pi_{s_{t+1|t}} = E[S_{t+1}|\mathcal{F}_t]$, a prediction of the probability of the future state, conditional on the information currently available. This simple model implies that returns on risky assets follow a regime switching process driven by the states in the underlying pricing kernel that reflect time-varying expected consumption growth and time-varying conditional covariances between asset returns and consumption growth.[7] Importantly, a number of papers have recently documented the presence of regimes in a number of fundamental series commonly employed in the equilibrium asset pricing literature, see e.g. Cecchetti, Lam, and Mark (2000). In this sense, variance risk may be linked to business cycle variations in economic growth (cash flows) associated with the economic cycle (e.g., Whitelaw, 2001), breaks in macroeconomic volatility (e.g., Lettau, Ludvigsson and Wachter, 2006), large macroeconomic shocks (e.g. oil prices) or institutional changes.

than two or three return series explains why we have framed the discussion and estimation of the relatively rich multivariate asymmetric (EGARCH), symmetric (GARCH) integrated, and GARCH-in mean models within the DCC framework. Of course, the parametric assumption of a conditional multivariate normal distribution for the shocks might be removed, although in this paper it is sensible to compare models that rely on a homogeneous assumption of conditional normality throughout.

[7] The extension of these claims to an international framework is straightforward when g_{t+1} is interepreted as real per-capita world consumption growth.

The literature on the economic origins of time-varying volatilities and covariances is older. Four key ideas seem to have emerged. First, conditional heteroskedasticity may simply result from the fact that in modern financial markets information flows in uneven ways and tends to cluster in short periods of time (see e.g., Tauchen and Pitts, 1983, and more recently Fong and Wong, 2005). However, this explanation preferentially applies to explain the existence of ARCH effects in asset returns at the daily or even infra-daily level, while it is considerably more difficult to use it to understand the presence of conditional heteroskedasticity at lower frequencies (such as weekly or monthly), without taking stance on difficult issues of statistical aggregation. Second, a literature has emerged (see e.g., DenHaan and Spear, 1998) that shows that frictions – such as borrowing constraints and transaction costs – may cause transactions and therefore price movements to lump in clusters that would cause ARCH in asset returns. Third, since the work by Timmermann (1996) (also see Veronesi, 2000, for recent results on time-varying intensity in ARCH effects, and Guidolin and Timmermann, 2007) it has become known that the learning of a representative investor concerning the process followed by the fundamentals priced in equilibrium may generate strong conditional heteroskedasticity. Finally, recent work by Kurz and Motolese (2001) and Kurz, Jin, and Motolese (2005) has shown that – even in the absence of market frictions and learning – it may be simply be that the interaction among investor with heterogeneous beliefs may produce realistic conditional heteroskedastic patterns. In particular, Kurz, Jin, and Motolese take interest in a few ARCH processes that may potentially generate variance risk. In this sense, it is microstructure effects (i.e., concerning the arrival process of information), market frictions, learning, as well as investor heterogeneity that may ultimately cause variance risk to be important in portfolio choices.

3. **The Asset Allocation Problem**

Consider an investor with power utility defined over terminal wealth, W_{t+T}, coefficient of relative risk aversion $\gamma > 0$, and horizon T:

$$u(W_{t+T}) = \frac{W_{t+T}^{1-\gamma}}{1-\gamma} \tag{9}$$

The investor maximizes expected utility by choosing a vector of portfolio shares at time t, that can be adjusted every $\varphi = \frac{T}{B}$ months at B equally spaced points. When $B = 1$ the investor simply implements a buy-and-hold strategy. Let $\boldsymbol{\omega}_b$ be the portfolio weights on $m \geq 1$ risky assets at these rebalancing times. Defining $W_B \equiv W_{t+T}$, and assuming for simplicity a unit initial wealth, the investor's optimization problem is:

$$\max_{\{\boldsymbol{\omega}_j\}_{j=0}^{B-1}} \quad E_t\left[\frac{W_B^{1-\gamma}}{1-\gamma}\right]$$
$$s.t.\ W_{b+1} = W_b \boldsymbol{\omega}_b' \exp\left(\mathbf{R}_{b+1}\right) \tag{10}$$

where $\exp\left(\mathbf{R}_{b+1}\right) \equiv [\exp\left(R_{1,b+1}\right)\ \exp\left(R_{2,b+1}\right)\ \ldots\ \exp\left(R_{m,b+1}\right)]'$ denotes an $m \times 1$ vector of cumulative, gross returns between two rebalancing points (under continuous compounding). The derived utility of wealth function can be simplified, for $\gamma \neq 1$, to:

$$J(W_b, \mathbf{r}_b, \boldsymbol{\theta}_b, \boldsymbol{\pi}_b, t_b) \equiv \max_{\{\boldsymbol{\omega}_j\}_{j=b}^{B-1}} E_b\left[\frac{W_B^{1-\gamma}}{1-\gamma}\right] = \frac{W_b^{1-\gamma}}{1-\gamma} Q(\mathbf{r}_b, \boldsymbol{\theta}_b, \boldsymbol{\pi}_b, t_b), \tag{11}$$

i.e. the optimal value function can be factored in such a way to be homogeneous in wealth ($\boldsymbol{\theta}_b$ and $\boldsymbol{\pi}_b$ are vectors that collect the parameters of the return generating process, conditional on information at time b).

One interesting special case is the buy-and-hold framework in which $\varphi = T$. Under this assumption, Appendix A shows that, similarly to Barberis (2000), the integral defining the expected utility functional can be approximated as follows:

$$\max_{\boldsymbol{\omega}_t} N^{-1} \sum_{n=1}^{N} \frac{\left[\boldsymbol{\omega}_t' \exp\left(\sum_{i=1}^{T} \mathbf{r}_{t+i,n}\right)\right]^{1-\gamma}}{1-\gamma},$$

where N is the number of simulations, and $\boldsymbol{\omega}_t' \exp\left(\sum_{i=1}^{T} \mathbf{r}_{t+i,n}\right)$ is the portfolio return in the n-th Monte Carlo simulation when the portfolio structure is given by $\boldsymbol{\omega}_t$. Each simulated path of portfolio returns is generated using draws from the assumed econometric model, for instance (1)-(2) that allows regimes to shift randomly as governed by the transition matrix, $\mathbf{P}$. We use $N = 30,000$ simulations.[8] Appendix A provides details on the numerical techniques employed in the solutions and extends these methods to the case of an investor who adjusts portfolio weights every $\varphi < T$ months. Here we only stress that because the backward solution of (10) implies the relationship

$$Q(\mathbf{r}_b, \boldsymbol{\pi}_b, t_b) = \max_{\boldsymbol{\omega}_b} E_{t_b}\left[\left(\frac{W_{b+1}}{W_b}\right)^{1-\gamma} Q\left(\mathbf{r}_{b+1}, \boldsymbol{\pi}_{b+1}, t_{b+1}\right)\right],$$

it is clear that portfolio choices will reflect not only hedging demands for assets due to stochastic shifts in investment opportunities but also a hedging motive caused by changes in investors' beliefs concerning future state probabilities, $\boldsymbol{\pi}_{b+1}$.

3.1. *Welfare Cost Measures*

To quantify the utility costs of restricting the investor's asset allocation problem, we follow Ang and Bekaert (2002) and Guidolin and Timmermann (2005). Call $\hat{\boldsymbol{\omega}}_t^R$ the vector of portfolio weights obtained by imposing restrictions on the portfolio problem, for instance, when the investor is forced to avoid small capitalization firms. We aim at comparing the investor's expected utility under the unrestricted model $-$ leading to some optimal set of controls $\hat{\boldsymbol{\omega}}_t$ $-$ to the utility derived assuming the investor is constrained. Since a restricted model is a special case of an unrestricted model, the following relationship between the value functions holds:

$$J(W_t, \mathbf{r}_t, \hat{\boldsymbol{\pi}}_t; \hat{\boldsymbol{\omega}}_t^R) \leq J(W_t, \mathbf{r}_t, \hat{\boldsymbol{\pi}}_t; \hat{\boldsymbol{\omega}}_t),$$

i.e. imposing restrictions reduces the derived utility from wealth. The compensatory premium, λ_t^R, is then computed as:

$$\lambda_t^R = \left[\frac{J(W_t, \mathbf{r}_t, \hat{\boldsymbol{\pi}}_t; \hat{\boldsymbol{\omega}}_t)}{J(W_t, \mathbf{r}_t, \hat{\boldsymbol{\pi}}_t; \hat{\boldsymbol{\omega}}_t^R)}\right]^{\frac{1}{1-\gamma}} - 1. \tag{12}$$

The interpretation is that an investor would be willing to pay λ_t^R in order to get rid of the restriction. In what follows we report annualized percentage measures of such a certainty equivalent loss λ_t^R, to be interpreted

[8] Of course, if a different econometric framework were to be postulated, this would simply change the baseline process from which return paths ($\{\mathbf{r}_{t+i,n}\}_{i=1}^{T}$) are simulated. Experiments indicated that for $m = 4$, a number of simulations N between 20,000 and 40,000 trials delivers satisfactory results in terms of accuracy and minimization of simulation errors vs. computational speed.

as the annual percentage fee that the investor stands ready to pay to purchase services that remove the constraint.

4. **Empirical Results**

4.1. *The Data*

We use weekly data from the MSCI total return indices data base for Pacific, North American Small, European Small Caps and European Large Caps (MSCI Europe Benchmark). Returns on North American Large Caps are computed as a weighted average of the MSCI U.S. Large Cap 300 Index and the D.R.I. Toronto Stock Exchange 300, using as weights the relative capitalizations of U.S. and Canada.[9] We use total return data series, inclusive of dividends, adjusted for stock splits, etc. Returns are expressed in the local currencies as provided by MSCI. This implies a rather common assumption − see e.g. De Santis and Gerard (1997) and Ang and Bekaert (2002) − that our investor is sophisticated enough to fully hedge her currency positions.

The sample period is January 1, 1999 - January 3, 2007. A Jan. 1, 1999 starting date for our study is justified by the evidence of substantial portfolio reallocations induced by the disappearing currency risk in the European Monetary Union (Galati and Tsatsaronis, 2001). We use data at a weekly frequency, which guarantees the availability of 417 observations for each of the series. Furthermore, notice that our sample straddles at least two complete stock market cycles, capturing both the last months of the stock market rally of 1998-1999, its fall in March 2000, the crash of September 11, 2001, and the subsequent recovery of 2003-2006.

Tables 1 and 2 report summary statistics for stock returns. Since we have a well-balanced sample in terms of sequence of bear and bull markets, average mean returns appear to take typical values for all portfolios under consideration, i.e., from a low of 3.8 percent per year in the case of North American large caps to 13.6 percent for European small cap stocks. However − as discussed in the Introduction − small caps represent an exception. In particular, European small caps are characterized by a high annualized 39% positive median return, followed by European large and North American small caps with 21 and 17% per year.[10] The resulting (median-based) Sharpe ratios for small capitalization firms make them highly appealing from a portfolio perspective: European American small caps display a stunning 0.27 weekly Sharpe ratio.

On the other hand, Table 1 questions the validity of an approach that relies only on the Sharpe ratio: the small cap skewness is negative and large, indicating that there are asymmetries in the marginal density that make negative returns more likely than positive ones; their kurtosis exceeds the Gaussian benchmark (three), indicating that extreme realizations are more likely than in a simple Gaussian i.i.d. framework. Second, opposite remarks apply to other stock indices, in particular North American large caps and Asian Pacific ones: their skewness is either positive or nil, which may be seen as an expected utility-enhancing feature by many investors; their kurtosis is moderate, close to what a Gaussian i.i.d. framework implies. These remarks

[9] While the MSCI Europe Benchmark index targets mainly large capitalization firms, no equivalent for North America (i.e. US and Canada) is available from MSCI. In practice, the U.S. large caps index receives a weight of 94.5% vs. a 5.5% for the Canadian index.

[10] In addition to the mean, we also use the median of returns as an estimator of location: for variables characterized by substantial asymmetries (negative skewness), the median is a more representative location parameter than the mean.

beg our core question: When and how much do higher order moments matter for asset allocation?

The last two columns reveal that while serial correlation in levels is limited to European and small caps portfolios, the evidence of volatility clustering − i.e. the tendency of squared returns to be serially correlated − is widespread, which points to the possible need to capture conditional heteroskedastic patterns.

Finally, Table 2 reports correlation coefficients. Pacific stock returns have lower correlations (around 0.45 - 0.55 only) with other portfolios than all other pairs in the table. This feature makes Pacific stocks an excellent hedging tool. All other pairs display correlations in the order of 0.55 - 0.8, which is fairly high but also expected in the light of the evidence in the literature that all major international stock markets are becoming increasingly prone to synchronous co-movements (e.g. Longin and Solnik, 2001).

4.2. *Model Selection*

We use two alternative sets criteria to select econometric frameworks able to effectively capture the properties of the data. Since our main hypothesis that small caps would be plagued by pervasive variance risk requires achieving an accurate specification of a sufficiently rich model from the set provided in Section 2, we make an extensive effort. We estimate a large number of variants of (1) and of models in the multivariate ARCH family and use five criteria to gauge their correct specification:

1. Davies (1977)-corrected likelihood ratio tests of the presence of multiple regimes, i.e. formal tests of the null hypothesis of $k = 1$ against the alternative of $k \geq 2$. As discussed in Garcia (1998), testing for the number of regimes may be tricky as under the null a few parameters of the unrestricted model − i.e. the elements of the transition probability matrix associated to the rows that correspond to "disappearing states" — can take any values without influencing the likelihood function; these parameters are said to become a nuisance to the estimation. In the presence of nuisance parameters, even asymptotically the LR statistic fails to have a standard chi-square distribution. Davies (1977) derives an upper bound for the significance level of the LR test under nuisance parameters:

$$\Pr\left(LR > x\right) \leq \Pr\left(\chi_1^2 > x\right) + \sqrt{2x}\exp\left(-\frac{x}{2}\right)\left[\Gamma\left(\frac{1}{2}\right)\right]^{-1},$$

   where $\Gamma(\cdot)$ is the standard gamma function.

2. - 4. Three standard information criteria, i.e. the Akaike (AIC), Bayes-Schwartz (BIC), and Hannan-Quinn (H-Q) criteria. These statistics are supposed to trade-off in-sample fit with prediction accuracy and rely on the principle that a correctly specified model should not only provide an accurate in-sample fit, but also prove useful to forecast out-of-sample. In practice, information criteria identify the ex-ante potential of good out-of-sample performance by penalizing models with a large number of parameters. A well-performing model ought to minimize each of the information criteria. Information criteria do not explicitly suffer from nuisance parameter issues and are therefore employed to compare models with different number of regimes, as well models in the same k-class but with different structure.[11]

[11] These criteria are now relatively well-established in the regime switching literature, see e.g. Sola and Driffill (1994). Roeder and Wasserman (1997) formally argue in favor of using information criteria in mixtures of normals.

Table 3 reports the outcomes of these model selection/specification tests.[12] Tests appear in the same order in which they have been listed above, and each of them corresponds to columns 4 through 7 of the table; columns 2 and 3 simply report the number of parameters implied by each model and the corresponding maximized log-likelihood function. Obviously, simply maximizing the log-likelihood function does not represent a useful model selection criterion, since the likelihood keeps increasing by simply adopting progressively more complicated models which end up requiring hundreds of parameters (and saturation ratios below a standard minimum value of 20 observations per estimated parameter). Therefore we proceed and inspect the additional columns of the table. When appropriate, the fourth column of Table 3 systematically tests the null of $k = 1$ against $k > 1$ (the exact number of regimes varies) and reports p-values calculated under Davies' upper bound. Obviously, even adjusting for the presence of nuisance parameters, the evidence against specifying traditional single-state models is overwhelming: the smallest LR statistic takes a value of 162, which is clearly above any conceivable critical value regardless of the number of restrictions imposed. This gives a first, crucial implication: the data offer strong evidence of time-variation in the coefficients of models capturing the dynamics of international stock returns.

Once we establish that $k \geq 2$ is appropriate, this only rules out single-state models which are nested by Markov switching VARs, i.e. only the first two rows of Table 3. We therefore compare the performance of multivariate ARCH and Markov switching VARs. On the one hand, within the Markov switching class, the range of models estimated is wide and spans models with $k = 2, 3, 4$, $p = 1, 2$, and with and without a regime-dependent covariance matrix. Also models in which the VAR coefficients are constant over time and fail to depend on the regime are estimated, since they are relatively parsimonious and economically interesting (see Guidolin and Ono, 2006). Columns 5-7 of Table 3 show that some tension exists among different criteria. The BIC is minimized by a tightly parameterized three-state heteroskedastic model with $p = 0$ in which 48 parameters have to be estimated. However, this is less than surprising as the BIC is generally known to select relatively small models in nonlinear frameworks (see e.g. Fenton and Gallant (1996)). Next, the AIC and H-Q criteria point towards a richer three-state heteroskedastic model with time-invariant VAR(1) matrix (i.e., a MSIH(3,0)-VAR(1) model with a saturation ratio of 26). Even though Section 6 also entertains the possibility that the data may be best described by a model in which $p = 0$ (see Guidolin and Nicodano, 2007, for related evidence), in this Section and the following we take this three-state time-invariant VAR(1) model as our baseline framework.

On the other hand, rows 3-8 of Table 3 estimate a range of multivariate ARCH models. Although the best fit, in terms of maximizing the log-likelihood function, is provided by a multivariate GARCH(1,1)-in-mean (including a VAR(1) component), once the log-likelihood is penalized by the relative large number of parameters (78), both BIC and H-Q indicate that the most promising fit is instead achieved by a more parsimonious (44 parameters) DCC EGARCH(1,1) VAR(1).[13] Therefore it seems natural to proceed and compare the statistical properties of the latter DCC EGARCH(1,1) model with a MSIH(3,0)-VAR(1).

[12] In the table, the switching models are classified as MSIAH(k, p), where I, A and H refer to state dependence in the intercept, vector autoregressive terms and heteroskedasticity. p is the autoregressive order. Models in the class MSIH$(k, 0)$-$VAR(p)$ have regime switching in the intercept but not in the VAR coefficients.

[13] The AIC selects instead the larger full multivariate GARCH(1,1)-in-mean. Notice that both in the regime switching and in the ARCH case, the AIC favors relatively large models which pose considerable risks of over-parametrization.

A first piece of evidence that favors the regime switching framework is that all the information criteria attain lower values (16.04 for the AIC, 16.76 for BIC, and 16.38 for H-Q) in the Markov switching case than in the ARCH case (16.55, 16.98, and 16.72, respectively). This indicates that the increase in the log-likelihood function caused by the adoption of multi-state models more than compensates the fact that the Markov switching framework implies a need to estimate 20 additional parameters. However, it is obvious that – even when penalized by a function that is monotone increasing in the number of estimated parameters, therefore discounting the potential for parameter uncertainty and/or instability – a model should be selected not for its in-sample fit, but for its potential of producing a useful out-of-sample performance. In particular, notice that in an asset allocation application, what matters is chiefly the ability of an econometric model to produce accurate forecasts of the entire joint density of equity returns.[14] The seminal work of Diebold et al. (1998) has spurred increasing interest in specification tests based on the h-step ahead accuracy of a model for the underlying density. These tests are based on the probability integral transform, or z-score. This is the probability of observing a value smaller than or equal to the realization $\tilde{\mathbf{r}}_{t+1}$ (assuming $h = 1$) under the null that the model is correctly specified. Appendix B provides further details on this econometric tool in general and on its specific application in our paper. The general idea is that the z-scores, being a function of the forecast errors, should obey a number of statistical restrictions under the null of correct model specification.

Table 4 reports Berkowitz-style, transformed z-tests (for pseudo-out of sample one week-ahead scores) for four models: a benchmark Gaussian IID model that implies the absence of predictability (i.e., constant expected returns, variances, and covariances); a Gaussian VAR(1) inspired by the literature on linear predictability in finance; the three-state heteroskedastic Markov switching VAR(1) model; the DCC EGARCH(1,1) VAR(1) model. Strikingly, the popular Gaussian IID and VAR(1) models are both resoundingly rejected by most tests and for three out of four international equity markets (the exception is Asian Pacific returns, which is relatively unsurprising in the light of Table 1). Rejections tend to be harsh, in the sense that even for the VAR(1) 13 out of 16 tests give p-values below 0.05 (11 are highly statistically significant): applying standard linear methods (see e.g. Barberis, 2000) to our weekly international stock return data would provide misleading inferences on optimal portfolio weights and therefore on the reasons for the negligible importance of small capitalization stocks in internationally diversified portfolios.

The picture drastically improves when either of the two non-linear frameworks are fitted to the data. For all international portfolios, the test statistics drastically drop sometimes declining by a factor of 600-700. This means that either regimes or ARCH effects are needed to correctly forecast the joint density of returns. Although for three out of four portfolios (the exception is European large firms) the differences in scores between ARCH and regime switching are not substantial, we generally notice that a three-state VAR(1) model seems to obtain a small hedge. For instance, out of 16 tests, 7 are significant in the ARCH case, vs. 3 only for the regime switching model. The z-score tests are in practice perfect – i.e., there is no sign of departures from the null of Gaussian IID scores – for European small caps and North American large caps, although some marginal problems are left for the other two portfolios.[15] All in all, we take these results

[14] Very simply put, notice that a risk-averse investor with concave utility function will attach weight not only to mean wealth, but to the entire density of wealth, which is obviously a function of the joint predictive density of future equity returns.

[15] We also produce and analyze (unreported) plots that display the empirical distributions of $\{z^{*}_{t+1}\}$ for each of the four return series and compare them with a normal variate with identical mean and variance. The models' faults are obvious for most series.

as evidence of a superior out-of-sample performance of the regime switching framework vs. the ARCH one. Therefore the Sections that follow treat the three-state VAR(1) model as our baseline case, and deal with the DCC EGARCH(1,1) for robustness check purposes.

4.3. *Econometric Estimates*

Table 5, panel B reports ML parameter estimates for the three-state, heteroskedastic VAR(1) model. As a benchmark, panel A shows estimates for a single-state VAR(1) model. Panel A shows typical linear results: there is very weak evidence of non-zero expected returns. Linear predictability is weak, in the sense that while European small cap returns are mildly persistent, European large caps are anti-persistent (i.e., a high return today forecasts a lower return tomorrow) and are weakly affected by past returns on North American large caps. However, the economic effects are almost negligible: even limiting our attention to the statistically significant coefficients, a one standard deviation shock to European small (large) caps forecasts an increase (decrease) in one-week ahead returns of 0.37% (-0.77%), while a one standard deviation shock to North American large caps forecasts an increase in one-week ahead European large returns of 0.25%.[16]

Panel B of Table 3 reports parameter estimates for the three-state process. While the state-independent VAR matrix remains similar to the one in panel A, the intercepts are significant only in the two "outside" regimes. The interpretation of such regimes is made possible by computing the (unconditional, long-run) regime-specific weekly mean returns in each of the regimes, using the formula $E[\mathbf{r}_t|S_t = i] = (\mathbf{I}_m - \mathbf{A})^{-1}\boldsymbol{\mu}_i$:

	Pacific	EU Small	EU Large	NA Large
$E[\mathbf{r}_t\|S_t = 1]$	-1.11	-3.23	-2.41	-1.09
$E[\mathbf{r}_t\|S_t = 2]$	-0.01	0.33	0.35	0.08
$E[\mathbf{r}_t\|S_t = 3]$	0.44	0.90	0.36	0.30

Regime 1 is clearly a crash state in which all international equity markets face large losses (these are weekly means, i.e. a -1 percent corresponds a whopping annualized -52%). However, notice that this characterization of regime 1 is not incompatible with an equilibrium interpretation (as well as common sense) because this state has very low persistence, i.e. with probability in excess of 0.60 the international economy leaves the crash state between t and $t+1$ to move to positive expected return states.[17] As a result, the crash state has a negligible duration of less than 2 weeks, which well-fits the idea that extreme market crashes mostly

For single-state models, the score distributions are either leptokurtic or even multi-modal. Further analysis, reveals that the forecast errors for Asian Pacific returns show signs of conditional heteroskedasticity (albeit weak) that cannot be captured by a regime switching structure. The forecast errors for European large caps show some evidence of a need for two regimes only.

[16] In table 5, the VAR panels have to be read in the following way: each coefficient illustrates the effect a shock to the variable in the column on the variable in the row.

[17] We explicitly compute $E[\mathbf{r}_{t+4}|S_t = i, \mathbf{r}_t] = \sum_{i=1}^{3} \hat{\pi}^i_{t+4,t} \left(\boldsymbol{\mu}_i + \mathbf{A}E[\mathbf{r}_t|S_t = i]\right)$, when in each state the system is inizialized at its unconditional mean, and obtain:

	Pacific	EU Small	EU Large	NA Large
$E[\mathbf{r}_t\|S_t = 1]$	0.08	0.22	0.07	0.06
$E[\mathbf{r}_t\|S_t = 2]$	0.01	0.20	0.18	0.05
$E[\mathbf{r}_t\|S_t = 3]$	0.21	0.34	0.01	0.11

showing that at a one-month horizon all expected returns become positive.

correspond to short-lived episodes. However, this is not an irrelevant state, because the overall structure of the estimated transition matrix implies that approximately 9.9% of the data will be generated by this regime. Interestingly, when the world equity markets leave the crash state, roughly 46% of the time this is to regime 3, when expected returns are high. In the crash regime, volatilities are relatively high, especially in the case of European stocks, both small and large caps. Also correlations are relatively high (with the exception of the coefficient concerning the Pacific-EU large pair). Figure 1 gives a visual representation in terms of smoothed (ex-post, full sample) probabilities which fits this interpretation, in the sense that this state appears relatively often but lasts at most for three weeks. Although 'crashing weeks' seem to appear 2-3 times a year on average (e.g. the week of September 11, 2001 is picked up by this state), these have become relatively more frequent and persistent during the crisis period of 2002-2003.

Regime 2 is a persistent bear state in which expected returns are negligible and statistically insignificant. The average duration of the state is 12 weeks, i.e. it fits periods in which the markets are hardly moving in any direction (and excess returns are low or even negative). This state is characterized by intermediate volatility, although the estimated values exceed their unconditional counterparts for three indices out of four. Consistently with Longin and Solnik (2001), in this bear state correlations are higher than their unconditional counterparts. Figure 1 shows that some portions of the turbulent 1999 and then most of the years 2001-2002 are captured by this bear state. Clearly, a sequence of visits to the bear state intertwined by sporadic hits of the crash regime, may impress a negative trend to equity prices. Regime 3 is instead a persistent bull state in which expected returns are positive and statistically significant. The average duration of this state is 11 weeks. The bull state is characterized by low volatility and relatively reduced correlations. To complement to what could be seen for the bear regime, Figure 1 shows that most of 2000 and then 2004-2006 are captured by this regime. Even though their persistence is similar, the peculiar structure of the transition matrix implies that while on average 52.4% of the data will come from the bear state, 37.7% will be generated by the bull regime.

4.4. *Diagnostic Tests*

Standard, residual-based diagnostic checks are made difficult within the multivariate Markov switching class by the fact that in (1) $\mathbf{u}_t \sim N(0, \boldsymbol{\Sigma}_{s_t})$ only conditional on a given regime. Since for most t, the vector of state probabilities $\hat{\boldsymbol{\pi}}_t$ will differ from $\mathbf{e}_i$ $(i = 1, ..., k)$, the generalized residuals,

$$\sum_{i=1}^{k}(\mathbf{e}_i'\hat{\boldsymbol{\pi}}_t)\left(\mathbf{r}_t - \hat{\boldsymbol{\mu}}_i - \hat{\mathbf{A}}\mathbf{r}_{t-1}\right),$$

will fail to be either i.i.d. or normally distributed. Therefore standard residual-based tests will fail if focussed on testing the i.i.d. properties of the residuals and will anyway run into difficulties when tests rely on normality. However, Krolzig (1997) shows that under the assumption of correct specification, one important property ought to pin down at least the one-step ahead forecast errors,

$$\boldsymbol{\eta}_{t+1} \equiv \mathbf{r}_{t+1} - \sum_{i=1}^{k}(\mathbf{e}_i'\hat{\mathbf{P}}\hat{\boldsymbol{\pi}}_t)\left(\hat{\boldsymbol{\mu}}_i + \hat{\mathbf{A}}\mathbf{r}_t\right)$$

(where $\hat{\boldsymbol{\pi}}_t$ is the vector of real-time, filtered state probabilities and $\hat{\boldsymbol{\pi}}_t'\hat{\mathbf{P}}\mathbf{e}_i$ is the one-step ahead prediction of the probability of state $i = 1, ..., k$): $\{\boldsymbol{\eta}_{t+1}\}$ should define a martingale difference sequence (MDS), i.e.

$$E[\boldsymbol{\eta}_{t+1}|\mathcal{F}_t] = 0.$$

This hypothesis is testable in standard ways, i.e. looking at the ability of elements of the information set at time t (e.g. current returns and forecast errors) to forecast both elements of $\boldsymbol{\eta}_{t+1}$ as well as their powers (since $E[\boldsymbol{\eta}_{t+1}|\mathcal{F}_t] = 0$ is more restrictive than $Cov[\boldsymbol{\eta}_{t+1}, Y_t] = 0$, where Y_t is any variable that belongs to $\mathcal{F}_t$).

We implement two types of residual-based tests. In each case, we make an effort to provide intuition for what a rejection of the null of the forecast errors being a martingale difference sequence would imply in economic terms. To gain additional insights, we generally apply tests to the each of the elements of $\{\boldsymbol{\eta}_{t+1}\}$ in isolation (i.e. to the univariate series of forecast errors concerning portfolio returns). We start by testing whether any lagged return predicts current and future forecast errors. Rejections of the null of zero predictive power, would point to misspecification in the conditional mean function implied by our MSIH(3,0)-VAR(1) model in particular (but not exclusively) in the VAR order (p). While for the Pacific, European small, and North American large past returns fail to be correlated with current forecast errors, for European large returns we find that at one lag such correlation is 0.24 and with a p-value below 0.05. This is hard to interpret because in Section 4.1 it became clear that either $p = 2$ or a fully-fledged Markov switching VAR(1) structure are not required by the data.

Obviously, similar restrictions apply to the ability of past forecast errors to predict future errors, i.e. on the implied serial correlation structure of the forecast errors themselves. If past forecast errors help predict future errors, clear improvements in the model are possible. Here we find once more that while all indices but European large cap errors have no appreciable serial correlation structure (e.g. their Ljung-Box order 12 p-values are 0.57 and 0.14, respectively), once more European large stocks are negatively serially correlated at lag one (-0.12), which is borderline significant. All in all, we interpret this evidence as roughly consistent with the absence of obvious misspecifications in our conditional mean functions.[18]

Next, we examine the ability of variables in the information set to predict squared forecast errors. In case of rejections of the no predictability restriction, this test can be interpreted as a test of omitted volatility clustering and ARCH effects in the model. There is borderline evidence of some positive and significant first-order serial correlation in squared forecast errors for Asian Pacific and North American large returns, while both past own and cross-returns fail to predict subsequent squared forecast errors. We notice that these two portfolios are the ones for which the GARCH(1,1) and EGARCH(1,1) (univariate) models estimated in Section 3.2 turned out to be the most persistent. Even though these diagnostic results may reveal a weak need for a more persistent conditional heteroskedastic process than what the combination of regime-dependent covariance matrices and persistent Markov states imply, we take the evidence of an overwhelming inability to reject the MDS null as a sign that there is no strong need to specify ARCH effects on the top of making $\boldsymbol{\Sigma}_{s_t}$ a function of the state.[19]

[18] We also examine the ability of lagged excess returns of market i to predict forecast errors of market j, $i \neq j$. We fail to find any appreciable linear (cross-) correlation structure in the forecast errors.

[19] We formally test a regime switching ARCH(1) specification in which

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Omega}_{S_t} + \mathbf{A}_{S_t}\mathbf{u}_t\mathbf{u}_t'\mathbf{A}_{S_t}$$

5. **International Portfolio Diversification**

In this section we present the core results of the paper. We start by computing optimal portfolio weights for an asset menu which allows for European small caps in addition to traditional stock portfolios, such as Asian Pacific, North American large, and European large caps portfolios ($m = 4$). We impose no-short sale restrictions and focus on the simpler buy-and-hold case.[20] We also solve a traditional portfolio problem in which the asset menu includes no small cap portfolios ($m = 3$). The purpose of this exercise is to enable us to compute the welfare gains obtainable by expanding the asset menu to include small caps. Since this portfolio problem is merely entertained as a benchmark, details are available upon request.[21] We then proceed in Section 5.2 to make sense of the results using the notion of variance risk and to link such concept with the notions of co-skewness and co-kurtosis that have played a central role in the recent work by Harvey and Siddique (2000) and Dittmar (2002). To stress the importance of variance risk, in Section 5.3 we provide a decomposition of our results to distinguish between the contribution of co-skewness and co-kurtosis.

5.1. *Implied portfolio weights*

We discuss two sets of portfolio weights. A first exercise computes optimal asset allocation at the beginning of 2007 for an investor who, using all past data for estimation purposes, has obtained the estimates in Table 5. This is a simulation exercise in which the unknown model parameters are calibrated to coincide with the full-sample estimates. In such a type of exercise the assessment of the role played by the different equity portfolios in international diversification may dramatically depend on the peculiar set of parameter estimates one obtains. As a result, we supplement this first exercise with calculations of real time optimal portfolio weights, each vector being based on a recursively updated set of parameter estimates.

The role of European small caps (henceforth EUSC) in portfolio choice may strongly depend on the regime: indeed they have the best and second-best Sharpe ratios in the bear and bull states (a non-negligible 0.17 and a stellar 0.62, respectively), and display the worst possible combination (negative mean and high variance) in the crash state. However, it is not clear how this contrasting information may influence the choice of investors who cannot observe the state. Furthermore, speculating on the Sharpe ratio to trace back portfolio implication may be incorrect when portfolios have higher-moment properties featuring high variance risk, see Table 1.

Figure 2 shows optimal portfolio shares as a function of the investment horizon (from 1 week to 2 years) for a buy-and hold investor who employs parameter estimates at the beginning of January 2007. Results are computed for a level of risk aversion $\gamma = 5$. Each plot concerns one of the available equity portfolios and reports five schedules: three of them condition on knowledge of the initial state of the markets (crash, bear or bull); one further schedule implies the existence of uncertainty on the state and assumes that the regime

This specification implies specifying 16 additional parameters, the elements of the matrix $\mathbf{A}_{1s_t}$. A LR test resoundingly rejects this specification.

[20] Guidolin and Nicodano (2007) show that these two restrictions hardly matter for the main result of this paper.

[21] We find substantial demand for North American large and Asian Pacific stocks (55 and 45% at a two-year horizon) and negligible weights for European large firms. The utility loss of ignoring regimes is rather small, less than 100 basis points for a long-horizon investor, which is consistent with the findings in Ang and Bekaert (2002).

probabilities are set to match the long run, ergodic frequencies (0.10, 0.52, 0.38, for crash, bear and bull); one last schedule depicts the optimal choice by a myopic investor who incorrectly believes that international stock returns are drawn by a multivariate IID Gaussian model.[22] Importantly, this last set of results corresponds to the case in which variance risk is disregarded altogether.

The demand for EUSC in Figure 2 is high but monotone decreasing in the investment horizon in the bear and bull state, and rather modest but increasing in the horizon in the crash state. However, since for long horizons and because of the ergodic nature of the Markov chain estimated in Table 5 the initial state tends to be of minor importance in determining the shape of the joint density of stock returns, the three schedules show rapid convergence for $T \geq 1$ year, all reaching weights between 45 and 65 percent. Importantly, the schedule for the crash state provides first evidence that using the Sharpe ratio may be misleading: in regime 1 and for horizons below 4 months, EUSC are never demanded as all the weight is given to Asian Pacific (40 percent) and North American large stocks (60 percent).

Even more interesting is the result concerning the 'steady-state' allocation to EUSC, when the investor assumes that all regimes are possible with a probability equal to the long-run measure. In this case − the most realistic situation since regimes are not observable − EUSC plays a limited role. Their weight is zero for short horizons ($T = 1, 2$) and grows to a reasonable 50% for intermediate horizons of about one year. Once more, at short horizons the steady-state portfolio puts almost identical weights on North American and Pacific equities. On the opposite, the IID myopic portfolio would be grossly incorrect, when compared to the steady-state regime switching weights, as it would place high weights on EUSC (87%) and Pacific stocks (13%).[23]

We repeat these calculations using different levels of risk aversion, γ. For instance, when $\gamma = 10$, the demand for EUSC becomes steeply decreasing in the bear and bull regimes, and flat (although always increasing) in the crash state. While in the crash state the weight assigned to EUSC remains nil for $T \leq$ 7-8 months, we notice that even investors with relatively long horizons of two years express a moderate demand of EUSC, around 30-35 percent at most. Interestingly, very risk averse investors shift their portfolios away from small caps and towards North American large firms, which is consistent with the notion that small firms are very risky.

Figure 3 shows our estimates of the (annualized) welfare costs of ignoring the existence of variance risks (regimes). Since Figure 2 stresses the existence of large differences between regime-switching and IID myopic weights, it is less than surprising to see that the utility loss from ignoring variance risk is of a first-order magnitude: for instance, a moderately risk-averse ($\gamma = 5$), long-horizon ($T = 2$ years) investor who assigns ergodic probabilities to the states would be indifferent to regimes if compensated by an annualized, riskless fee equal to 80 basis points. These sums are of course much larger should we endow the investor with precise information on the nature of the current state (especially when the information is profitable, as it is in the

[22] Similarly to Barberis (2000), the investment horizon is irrelevant for asset allocation purposes. We have also computed optimal portfolio weights under a Gaussian VAR(1) model and obtained similar results, since the linear predictability patterns are weak.

[23] There is no reason to think that the IID schedule ought to be an average of the regime-specific ones: the unconditional (long-run) joint distribution implied by a Gaussian IID and a multivariate regime switching model need not be the same; on the opposite, our specification tests offer evidence that the null of a Gaussian IID model is rejected, an indication that the unconditional density of the data differs from the one implied by a switching model.

crash regime when the horizon is short), as the welfare loss climbs to an annualized level in excess of 10%.

Next, we recursively estimate our three-state switching VAR(1) model and compute optimal portfolio weights with data covering the expanding samples Jan. 1999 - Dec. 2002, Jan. 1999 - first week of Jan. 2003, etc. up to the full sample Jan. 1999 - Jan. 2007. The previous results do not entirely depend on the point in time in which they have been performed. The average weight assigned to EUSC remains only approximately 44%, while European large caps acquire importance (13%), with North American large and Pacific stocks still playing the role of quality stocks useful for diversification at long horizons and in the crash regime (22 and 21%).[24] Also in this case, ignoring variance risk would assign way too high a weight to EUSC, in excess of 75% on average (the remaining goes to Pacific stocks). As a result, our recursive estimates of the welfare loss of ignoring regime switching (not reported) are extremely large over certain parts of the sample, exceeding annualized compensatory variation of 3-5% even under the most adverse parameters and investment horizons.

5.2. *Making sense of the results: variance risk*

Our simulations find that, under realistic assumptions concerning knowledge of the state, a rational investor should invest a limited proportion of her wealth in EUSC despite their high Sharpe ratio. Tables 5-7 report several findings that help us put this result into perspective. It is well known that investors with power utility functions are not only averse to variance and high correlations between pairs of asset returns − as normally recognized − but also averse to negative co-skewness and to high co-kurtosis, i.e. to properties of the higher order co-moments of the joint distribution of asset returns. For instance, investors dislike assets whose returns tend to become highly volatile at times in which the price of most of the other assets declines: in this situation, the expected utility of the investor is hurt by both the low expected mean portfolio returns as well as the high variance contributed by the asset. Similarly, investors ought to be wary of assets the price of which declines when the volatility of most other assets increases. Investors will also dislike assets whose volatility increases when most other assets are also volatile. We say that an asset that suffers from this bad higher co-moment properties is subject to high *variance risk.*

Tables 6 and 7 pin down these undesirable properties of EUSC. In Table 6 we calculate the co-skewness coefficients,

$$S_{i,j,l} \equiv \frac{E[(r_i - E[r_i])(r_j - E[r_j])(r_l - E[r_l])]}{\{E[(r_i - E[r_i])^2]E[(r_j - E[r_j])^2]E[(r_l - E[r_l])^2]\}^{1/2}},$$

between all possible triplets of portfolio returns i, j, l. We do that both with reference to the data as well for the three-state VAR model estimated in Section 4.2. In the latter case, since closed-form solutions for higher order moments are hard to come by, we employ simulations to produce estimates of the co-moments. Calculations are performed both unconditionally (i.e. averaging across regimes) and conditioning on knowledge of the initial regime. In the latter case, the conditional co-moments refer to the one-step ahead predictive joint density of asset returns. Based on our definition, variance risk relates to the cases in which the triplet boils down to a pair, i.e. either $i = j$, or $i = l$, or $j = l$.[25] When $i = j = l$ we shall be looking at

[24] These weights are obtained by averaging across investment horizons, although slopes tend to be moderate, consistently with the shapes reported in Figure 2. These results are for the $\gamma = 5$ case.

[25] Coefficient estimates for the cases in which $i \neq j \neq l$ are available but are hard to interpret. However our comments concerning the general agreements between sample and model-implied co-moment estimates also extend to the $i \neq j \neq l$ case.

the standard own-skewness coefficient of some portfolio return. In Table 6, bold coefficients highlight point estimates' significance at standard levels (5 percent). There is a remarkable correspondence between signs and magnitudes of co-skewness coefficients in the data and the unconditional estimates under our estimated Markov switching model, in the sense that when the coefficient is statistically significant in the data, it is always so also under the Markov switching model, with the correct sign and appropriate magnitude. Similarly to Das and Uppal (2004) we interpret this result as a sign of correct specification of the model, although we record some tendency to generate more negative co-skewness than one can actually find in the data.[26] Furthermore, notice that the co-skewness coefficients $S_{EUSC,EUSC,j}$ are all negative and large in absolute value: the volatility of EUSC is indeed higher when each of the other portfolios performs poorly. On the opposite, similar co-skewness coefficients for most other indices (e.g. $S_{EU\_\text{large},EU\_\text{large},j}$ for varying js) are close to zero and sometimes positive. Worse, all of the $S_{EUSC,j,j}$ coefficients are also large and negative (particularly, when j = Pacific and Europe large), an indication that EUSC may be losing ground exactly when some of the other assets become volatile. Therefore EUSC does display considerable variance risk. On the top of variance risk, from Tables 1 and 5 it emerges that EUSC also show high and negative own-skewness, another feature a risk-averse investor ought to dislike.

The results in the third column of Table 6 are relevant to interpret long-run portfolio choices, when the statistical properties of stock returns are well-approximated by their unconditional density. Table 6 also reports regime-specific, one-step ahead co-skewness coefficients, when the initial state is known. In the highly persistent bull and bear regime 2 and 3, departures from multivariate normality are minimal and in fact none of the co-skewness coefficients is significantly different from zero. Therefore, at least for short investment horizons of a few weeks at most, using the Sharpe ratio for portfolio allocation purposes may be justified and − consistently with the results in Figure 2 − EUSC ought to receive considerable weight. On the opposite, the crash state 1 implies some important departures of the joint predictive density of stock returns even over short investment horizons. In particular, EUSC have a tendency to decline when the volatility of Pacific and North American stocks is above average, while the volatility of EUSC tends to be high when each of the other markets is bear.

Of course, it may be hard to balance off co-skewness coefficients involving EUSC with different magnitudes or signs. Therefore it is helpful to calculate quantities similar to those in Table 6 for portfolio returns vs. some *aggregate* portfolio benchmark. For our purposes we use an equally weighted portfolio (EW_ptf, 25% in each stock index), although results proved fairly robust to other notions (e.g. value-weighted) of benchmark portfolio. For instance, S_{i,EW_ptf,EW_ptf} for the generic portfolio i has expression

$$S_{i,EW\_ptf,EW\_ptf} \equiv \frac{E[(r_i - E[r_i])(r_{EW\_ptf} - E[r_{EW\_ptf}])^2]}{\sqrt{Var[r_i]Var[r_{EW\_ptf}]}},$$

the notion of co-skewness between a security i and the market portfolio employed in Harvey and Siddique (2000). Once more the match between data- and model-implied coefficients is striking. In particular, in panel A of Table 7 we obtain model estimates $S_{EUSC,EUSC,EW\_ptf} = -0.50$ and $S_{EUSC,EW\_ptf,EW\_ptf} = -0.63$, i.e.

[26] In particular, there is some evidence of negative skewness and co-skewness affecting Asian Pacific stocks which is relatively weak in the data. This probably explains why in Table 4 the regime switching model seems to generate z-scores with some structure in squares, a sign that some conditional heteroskedastic patterns are incorrectly specified.

the variance of EUSC is high when equally weighted returns are below average, and EUSC returns are below average when the variance of the equally weighted portfolio is high. This is another powerful indication of the presence of variance risk plaguing EUSC. For comparison purposes, in panel B of Table 7 we repeat calculations for high-quality North American large stocks and obtain negligible (or even positive) coefficients, both from the model and in our sample of data.

We perform an operation similar to Table 6 with reference to the fourth co-moments of equity returns. In table 8, we find a striking correspondence between co-kurtosis coefficients measured in the data and unconditional coefficients implied by our regime switching model. Generally speaking, EUSC have dreadful co-kurtosis properties: for instance $K_{EUSC,EUSC,j,j}$ exceeds 2.2 for all js and tends to be higher than all other similar coefficients involving other portfolios, which means that the volatility of EUSC is high exactly when the volatility of all other portfolios is high. As already revealed by Table 1, also the own-kurtosis of EUSC substantially exceeds a Gaussian reference point of 3. These results confirm that also the model-implied $K_{EUSC,EUSC,EW\_ptf,EW\_pft}$ is 4.7, which is one of the highest among these types of coefficients. $K_{EUSC,EUSC,EW\_ptf,EW\_pft}$ is reminiscent of an indicator of covariance between EUSC illiquidity and market illiquidity. All in all, we have also some evidence that the extreme tails of the marginal density of EUSC tends to be fatter than for other portfolios and that their volatility might be dangerously co-moving with that of other assets.

5.3. *Welfare Costs of Ignoring European Small Caps*

Our evidence concerning the high variance risk of EUSC may in principle be able to explain their neglect as higher moments of their return distribution increase undesired skewness and kurtosis of wealth. However: Does this mean that there is no utility loss from restricting the available asset menu to exclude small caps? We provide an answer for the case of EUSC. We compute compensatory variations, using the approach illustrated in Section 3.We assume that the investor chooses the best specification for the return generating process for each asset menu. The conclusion that can be drawn is that – in spite of their limited optimal weight – the loss from disregarding EUSC would be of a first-order magnitude for all investment horizons:

	Investment Horizon (in weeks)			
	$T = 1$	$T = 12$	$T = 24$	$T = 104$
	Ergodic state probabilities			
$\gamma = 5$	3.37	3.35	3.59	5.81
$\gamma = 10$	9.55	7.83	8.67	18.94
	Recursive out-of-sample results (Means)			
$\gamma = 5$	10.19	5.58	5.88	4.25

When faced with compensatory variation in excess of 3% per year, it is difficult to think that small caps are not important for international diversification purposes. Although it is well-known that the effective costs paid when transacting on small caps tend to be high, it is unlikely that any sensible estimate of the costs implied by long-run buy-and-hold positions may systematically exceed the spectrum of welfare loss estimates we have found. So, modest optimal weights and high doses of variance risk are still compatible with a claim that small caps are key to expected utility enhancing international portfolio diversification.

6. **Further Discussion**

In this Section we briefly deal with two residual issues raised by the preceding analysis. In Section 6.1 we ask whether our portfolio results are specific to the Markov switching VAR framework assumed in this paper. In Section 6.2 we ask whether our results on European small caps may be generalized to small capitalization firms at large, by performing the exercise afresh when the asset menu includes North American small caps in addition to European ones.

6.1. *The Return Generating Process*

Section 4.2 has left us with two potential questions. First, in Table 3 the information criteria gave conflicting indications as to whether a VAR(1) component was needed within a three-state heteroskedastic switching framework. As a result, we have also estimated a simpler, three-state Markov switching model with $p = 0$ (as in Guidolin and Nicodano, 2007) and proceeded to compute optimal weights and welfare loss estimates. The interpretation and dynamic properties of the regimes are essentially unchanged vs. Section 4.3. Optimal portfolio weights display the same patterns commented in Section 5.1: when the investor ignores the nature of the current regime, the demand for EUSC is zero at short horizons and grows to approximately 55% at longer horizons, but in any event remains well below the share of 85-90% that a Gaussian IID model that ignores variance risk would imply. A careful examination of the resulting weights in fact shows that the differences vs. the ones plotted in Figure 2 are negligible.[27] Therefore, it is hard to think that details of the regime switching process selected in this paper (apart from the obvious, that the model should not be blatantly misspecified) may entirely drive our results on the effects of variance risk on the demand of EUSC.

More interestingly, even though Table 4 has provided reassuring evidence on the properties of the regime switching VAR model, at least one of the multivariate ARCH frameworks – in particular, the DCC EGARCH(1,1), in which variance risk is generated by leverage effects – did come close to provide relatively low information criteria and appreciable out-of-sample predictive accuracy. We have therefore computed afresh optimal portfolio weights when the joint process of returns is described by the DCC EGARCH(1,1) VAR(1). To save space, we simply report results for the optimal shares of EUSC when $\gamma = 5$ (for comparison, also results for the three-state Markov switching model, a Gaussian IID and a single-state VAR(1) model are presented):[28]

Weight to EUSC	Investment horizon (in weeks)			
	$T = 1$	$T = 12$	$T = 24$	$T = 104$
Three-state MS VAR(1)	0.00	0.46	0.58	0.65
DCC EGARCH(1,1) VAR(1)	0.18	0.31	0.36	0.76
Single-state VAR(1) (linear)	0.89	0.88	0.87	0.87
Gaussian IID (no predictability)	0.86	0.86	0.86	0.86

[27] Detailed results and parameter estimates are available upon request.

[28] Some more details are needed for replicability: for purposes of computation of predicted returns, the VAR(1) is initialized at the unconditional means for each of the return series (these unconditional means are regime-dependent in case of regime switching); in the DCC EGARCH variances and covariances are initialized at their unconditional levels (these are by construction equal to their sample estimates). For the regime switching model, calculations initialize state probabilities at their ergodic values.

Clearly, even if we had opted in favor of the DCC EGARCH model in Section 4.1, the qualitative results would have not changed: an appropriate multivariate ARCH framework is able to capture sufficient variance risk in the data to yield a EUSC weight schedule which is upward sloping and implies relatively low investments in EUSC for short horizons. Unreported results show that in practice the qualitative structure of the optimal DCC EGARCH portfolio is similar to the one shown in Figure 2, i.e., at short horizons the most important portfolios are Asia Pacific and North America large, while at a two-year horizon EUSC plays a major role. The only remarkable difference is that now for $T \geq 52$ weeks also European large caps receive a positive weight.

6.2. *Expanding the Asset Menu*

How general are our results for the role of small capitalization firms in internationally diversified equity portfolios? To answer this question, we proceed to generalize the problem to also include North American small caps (NASC), besides the North American large portfolio, i.e. $m = 5$. We repeat the analysis of Section 5 and therefore omit many details to save space.

Estimation of a MSIH(3,0)-VAR(1) model for the expanded asset menu leads to a characterization of the regimes which is very similar to one in Table 5: the second regime is a normal/bear highly persistent state (average duration is 11 weeks) in which expected returns (with the exception of NASC) are moderate and often not statistically positive, while volatilities and correlations are close to their unconditional values. In this state, only small caps (both European and North American) yield statistically significant, positive expected returns. The first regime is non-persistent a bear/crash state in which mean returns are significantly negative and large (between -2.5 and -2.7 percent per week for European large, EUSC, and NASC), volatilities are high (between 25 and 250% higher than in the normal state), and correlations high. The third regime is a persistent (average duration is 10 weeks) bull state implying high and significant means, high volatilities and modest correlations. Strikingly, the structure of the estimated transition matrix is virtually indistinguishable from the one in Table 5, to the point that most estimates of the transition probabilities do fall in a 95% confidence interval around the estimates obtained in Table 5. This is an important finding that corroborates the validity of our three-state regime switching model. The ergodic probabilities of the regimes are almost unchanged, 0.11, 0.50, and 0.39, respectively.

NASC have properties similar to those that characterize EUSC: for instance, the NASC weekly Sharpe ratio is 0.076 (vs. 0.095 for EUSC). Unreported plots of the optimal portfolio schedules similar to Figure 2 show that a myopic investor that ignores variance risk would invest most of her wealth (85%) in EUSC, and another important proportion in NASC (10%), and the remainder (5%) in Pacific stocks, essentially for hedging reasons. This portfolio recommendation would be once more incorrect: regime switching portfolio schedules contain dramatic departures from the IID myopic assumption: focussing on the case of $\gamma = 5$ and assuming the investor ignores the current regime, her commitment to EUSC would remain large (and increasing in T) but would be in the 45-55% range; once more, EUSC imply large amounts of variance risk and poor third- and fourth-order moment properties, which brings their weights down by a full 35% (i.e., 85% minus 50%). A similar argument applies to NASC, whose weight declines from 10% in the single-state case to essentially zero when variance risk is taken into account. Optimal allocations also turn out to be strongly regime-dependent: for instance, the crash state 1 is highly favorable to North American large cap and Pacific investments as these stocks have the highest Sharpe ratio in this regime, while Pacific stocks provide a relatively good hedge.[29]

7. Conclusion

We have measured three important components of the variance risk of an asset that adversely affect the skewness and the kurtosis of wealth. These are the negative covariance between its returns and the volatility of other assets, the negative covariance between its volatility and returns of other assets, and the covariance between volatilities, that are reminiscent of the priced factors in the cross section of returns reported by Harvey and Siddique (2000) and Dittmar (2002). In this metric, small caps have large variance risk. A powerful display of the effects of variance risk on portfolio choice is our result that the optimal portfolio share of European small caps under state-dependent returns – when the state of the stock market is unobservable – is always less than 50%, while their optimal weight in a myopic portfolio ought to be close to 90%. Interestingly, this finding does not depend on the details of the econometric model employed, such as the number of regimes, the presence of linear predictability patterns, or even the adoption of a multivariate, asymmetric ARCH framework to capture the presence of non-linear dynamics in returns and variance risk.

References

[1] Acharya V., and L., Pedersen, 2005, "Asset Pricing with Liquidity Risk", *Journal of Financial Economics,* 77, 375-410

[2] Ang A., and G., Bekaert, 2002, "International Asset Allocation with Regime Shifts", *Review of Financial Studies*, 15, 1137-1187.

[3] Ang A. and J., Chen, 2002, "Asymmetric Correlations of Equity Portfolios", *Journal of Financial Economics,* 63, 443-494.

[4] Barberis, N., 2000, "Investing for the Long Run When Returns Are Predictable", *Journal of Finance*, 55, 225-264.

[5] Barone-Adesi, G., Gagliardini P., and G. Urga, 2004, "Testing Asset Pricing Anomalies with Coskewness", *Journal of Business and Economic Statistics,* 22, 474-485.

[6] Berkowitz, J., 2001, "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, 465-474.

[29] We have also computed co-skewness and co-kurtosis coefficients vs. an equally weighted portfolio, both under the available data and under the three-state regime switching model. We find that for both small cap portfolios there is evidence that their variance increases when the variance of the market is high, that their variance is high when the market is bear, and that their returns are below average when the market is unstable. These properties (along with own kurtosis and skewness) explain why our portfolio results do not completely reflect simple Sharpe ratio-based arguments and why both portfolios receive a much higher weight under the myopic IID calculations than under regime switching.

[7] Black, F., 1976, "Studies of Stock Market Changes", in Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section (American Statistical Association, Alexandria, VA).

[8] Bollerslev, T., 1986, "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

[9] Bollerslev, T., 1990, "Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model", *Review of Economics and Statistics,* 72, 498-505.

[10] Campbell, J., Chan, and L., Viceira, 2003, "A Multivariate Model of Strategic Asset Allocation", *Journal of Financial Economics*, 67, 41-80.

[11] Campbell, J., and L., Viceira, 1999, "Consumption and Portfolio Decisions when Expected Returns Are Time Varying", *Quarterly Journal of Economics*, 114, 433-495.

[12] Cecchetti, S., P. Lam, and N., Mark, 2000, "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good To Be True?", *American Economic Review*, 90, 787-805.

[13] Cvitanić, J., V., Polimenis, and F., Zapatero, 2007, "Optimal Portfolio Allocation with Higher Moments", *Annals of Finance*, forthcoming.

[14] Das, S., and R., Uppal, 2004, "Systemic Risk and International Portfolio Choice", *Journal of Finance*, 59, 2809-2834.

[15] Davies, R., 1977, "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative", *Biometrika*, 64, 247-254.

[16] De Santis, G., and B., Gerard, 1997, "International Asset Pricing and Portfolio Diversification with Time-Varying Risk", *Journal of Finance,* 1997, 52, 1881-1912.

[17] Den Haan, W., and S., Spear, 1998, "Volatility Clustering in Real Interest Rates: Theory and. Evidence", *Journal of Monetary Economics*, 41, 431-453.

[18] Diebold, F., T., Gunther, and A., Tay, 1998, "Evaluating Density Forecasts", *International Economic Review*, 39, 863-883.

[19] Dittmar, R., 2002, "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns", *Journal of Finance*, 57, 369-403.

[20] Engle, R., 2002, "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business and Economic Statistics,* 20, 339-350.

[21] Engle, R., and K., Kroner, 1995, "Multivariate Simultaneous GARCH", *Econometric Theory,* 11, 122-150.

[22] Engle R., and J., Mezrich, 1996, "GARCH for Groups", *Risk*, 9, 36-40.

[23] Fenton, V., and R., Gallant, 1996, "Qualitative and Asymptotic Performance of SNP Density Estimation", *Journal of Econometrics*, 74, 77-118.

[24] Fong, W., and W., Wong, 2006, "The Modified Mixture of Distributions Model: A Revisit", *Annals of Finance*, 2, 167-178.

[25] Garcia, R., 1998, "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *International Economic Review*, 39, 763-788.

[26] Guidolin, M. and G., Nicodano, 2007, "Managing International Portfolios with Small Capitalization Stocks", Federal Reserve Bank of St. Louis Working Paper 2007-031A.

[27] Guidolin, M. and S., Ono, 2006, "Are the Dynamic Linkages Between the Macroeconomy and Asset Prices Time-Varying?", *Journal of Economics and Business*, 58, 480-510.

[28] Guidolin, M. and A., Timmermann, 2005, "Size and Value Anomalies under Regime Switching", Federal Reserve Bank of St. Louis Working Paper 2005-007A.

[29] Guidolin, M. and A., Timmermann, 2006, "International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences", Federal Reserve Bank of St. Louis Working Paper 2006-034B.

[30] Guidolin, M. and A., Timmermann, 2007, "Properties of Asset Prices under Alternative Learning Schemes", *Journal of Economic Dynamics and Control*, 31, 161-217.

[31] Hamilton, J., 1990, "Analysis of Time Series Subject to Changes in Regime", *Journal of Econometrics*, 45, 39-70.

[32] Hamilton, J., 1994, *Time Series Analysis*. Princeton University Press.

[33] Harvey, C., and A., Siddique, 2000, "Conditional Skewness in Asset Pricing Tests", *Journal of Finance*, 55, 1263-1295.

[34] Kim, T., and E., Omberg, 1996, "Dynamic Nonmyopic Portfolio Behavior", *Review of Financial Studies*, 9, 141-161.

[35] Krolzig, H.-M., 1997, *Markov-Switching Vector Autoregressions*, Berlin, Springer-Verlag.

[36] Kurz, M., H., Jin, and M., Motolese, 2005, "Determinants of Stock Market Volatility and Risk Premia", *Annals of Finance*, 1, 109-147.

[37] Kurz, M., and M., Motolese, 2001, "Endogenous Uncertainty and Market Volatility", *Economic Theory*, 17, 497-544.

[38] Lettau, M., S., Ludvigson, and J., Wachter, 2006, "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?", *Review of Financial Studies*, forthcoming.

[39] Leroux, B., 1992, "Maximum Likelihood Estimation for Hidden Markov Models", *Stochastic Processes and their Applications*, 40, 127-143.

[40] Liu, J., F., Longstaff, and J., Pan, 2003, "Dynamic Asset Allocation with Event Risk", *Journal of Finance*, 58, 231-259.

[41] Longin, F., and B., Solnik, 2001, "Correlation Structure of International Equity Markets During Extremely Volatile Periods", *Journal of Finance*, 56, 649-676.

[42] Newey W., and D., McFadden, 1994, "Large Sample Estimation and Hypothesis Testing", in *Handbook of Econometrics,* vol. 4, R. Engle and D. McFadden (eds.), New York: Elsevier, pp. 2113-2245.

[43] Pástor, L., 2000, "Portfolio Selection and Asset Pricing Models", *Journal of Finance*, 55, 179-223.

[44] Peres-Quiros G., and A., Timmermann, 2000, "Firm Size and Cyclical Variations in Stock Returns", *Journal of Finance,* 55, 1229-1262.

[45] Roeder, K., and L., Wasserman, 1997, "Practical Bayesian Density Estimation Using Mixtures of Normals", *Journal of the American Statistical Association*, 92, 367-377.

[46] Rosenblatt, M., 1952, "Remarks on a Multivariate Transformation", *Annals of Mathematical Statistics*, 23, 470-472.

[47] Sola, M., and J., Driffill, 1994, "Testing the Term Structure of Interest Rates Using a Stationary Vector Autoregression with Regime Switching", *Journal of Economic Dynamics and Control*, 18, 601-628.

[48] Tauchen, G., and M., Pitts, 1983, "The Price Variability-Volume Relationship on Speculative Markets", *Econometrica*, 51, 485-505.

[49] Timmermann, A., 1996, "Excess Volatility and Predictability of Stock Prices in Autoregressive Dividend Models with Learning", *Review of Economic Studies*, 63, 523-557.

[50] Timmermann, A., 2000, "Moments of Markov Switching Models", *Journal of Econometrics*, 96, 75-111.

[51] Veronesi, P., 2000, "How Does Information Quality Affect Stock Returns?", *Journal of Finance*, 55, 807-837.

[52] Whitelaw, R., 2001, "Stock Market Risk and Return: An Equilibrium Approach", *Review of Financial Studies*, 13, 521-548.

Appendix A – Solution Methods

A variety of solution methods have been applied in the literature on portfolio allocation under time-varying investment opportunities. Barberis (2000) employs simulation methods and studies a pure allocation problem without interim consumption. Ang and Bekaert (2002) solve for the optimal asset allocation using quadrature methods. Campbell and Viceira (1999, 2002) derive approximate analytical solutions for an infinitely lived investor when interim consumption is allowed and rebalancing is continuous. Finally, some papers have derived closed-form solutions by working in continuous-time, e.g. Kim and Omberg (1996) for the case without interim consumption.

In our paper we make two choices that simplify the computational task with respect to competing approaches. First, solving (10) by standard backward induction techniques is, unfortunately, a formidable task (see e.g.

the discussion in Barberis, 2000, pp. 256-260). Under standard discretization techniques the investor first needs to use a sufficiently dense grid of size $G, \{\boldsymbol{\theta}_b^j,\ \boldsymbol{\pi}_b^j\}_{j=1}^G$ to update both $\boldsymbol{\theta}_{b+1}$ and $\boldsymbol{\pi}_{b+1}$ from $\boldsymbol{\theta}_b$ and $\boldsymbol{\pi}_b$. In the presence of a high number of parameters implied by (1), standard numerical techniques are not feasible for this problem or would at best force us to use a very rough discretization grid, introducing large approximation errors. Therefore our approach simply assumes that investors condition on their current (as opposed to future ones, $\boldsymbol{\theta}_{b+1}$) parameter estimates, $\hat{\boldsymbol{\theta}}_t$. Under this assumption, since W_b is known at time t_b, $Q(.)$ simplifies to:

$$Q(\mathbf{r}_b, \boldsymbol{\pi}_b, t_b) = \max_{\boldsymbol{\omega}_b} E_b\left[\left(\frac{W_{b+1}}{W_b}\right)^{1-\gamma} Q\left(\mathbf{r}_{b+1}, \boldsymbol{\pi}_{b+1}, t_{b+1}\right)\right].$$

Second, we resort to simulation methods similarly to Barberis (2000). Ang and Bekaert (2002) were the first to study this problem under regime switching. They consider pairs of international stock market portfolios under regime switching with observable states, so the state variable simplifies to a set of dummy indicators. This setup allows them to apply quadrature methods based on a discretization grid (see also Guidolin and Timmermann, 2005). Our framework is quite different since we treat the state as unobservable and calculate asset allocations under optimal filtering (3).

To deal with the latent state we use Monte-Carlo methods for expected utility approximation. In the case in which dynamic rebalancing is admitted ($B \geq 2$), suppose that the optimization problem has been solved backwards at the rebalancing points t_{B-1}, ..., t_{b+1} so that $Q(\boldsymbol{\pi}_{b+1}^j, t_{b+1})$ is known for all values $j = 1, 2, ..., G$ on the discretization grid. For each $\boldsymbol{\pi}_b = \boldsymbol{\pi}_b^j$, it is then possible to find $Q(\boldsymbol{\pi}_b^j, t_b)$ at time t_b. For concreteness, consider the case of $p = 0$, i.e. the conditional mean function does not imply any autoregressive structure. Approximating the expectation in the objective function

$$E_{t_b}\left[\left\{\boldsymbol{\omega}_b' \exp\left(\mathbf{R}_{b+1}\right)\right\}^{1-\gamma} Q(\boldsymbol{\pi}_{b+1}^j, t_{b+1})\right]$$

by Monte Carlo methods requires drawing N samples of asset returns $\{\mathbf{R}_{b+1,n}(\boldsymbol{\pi}_b^j)\}_{n=1}^N$ from the $(b+1)\varphi$-step-ahead joint density of asset returns conditional on $\widehat{\boldsymbol{\theta}}_t$, assuming that $\boldsymbol{\pi}_b^j$ is optimally updated.

The algorithm consists of the following steps:

1. For a given $\boldsymbol{\pi}_b^j$ calculate the $(b+1)\varphi$-step ahead probability of being in each of the possible future regimes $s_{b+1} = j$ as $\boldsymbol{\pi}_{b+1|b} = (\boldsymbol{\pi}_b^j)'\hat{\mathbf{P}}_t^{\varphi}$, using that $\hat{\mathbf{P}}_t^{\varphi} \equiv \prod_{j=1}^{\varphi} \hat{\mathbf{P}}_t$ is the φ-step ahead transition matrix.

2. For each possible future regime, simulate N $\varphi-$period returns $\{\mathbf{R}_{b+1,s}(s_b)\}_{n=1}^N$ in calendar time from the switching model:

$$\mathbf{r}_{t_b+i,n}(s_b) = \hat{\boldsymbol{\mu}}_{s_{t_b+i}} + \boldsymbol{\varepsilon}_{t_b+i,n}.$$

   At all rebalancing points this simulation allows for stochastic regime switching as governed by the transition matrix $\hat{\mathbf{P}}_t$. For example, if we start in regime 1, between t_b+1 and t_b+2 there is a probability $\hat{p}_{12} \equiv \mathbf{e}_1'\hat{\mathbf{P}}_t\mathbf{e}_2$ of switching to regime 2, and $\hat{p}_{11} \equiv \mathbf{e}_1'\hat{\mathbf{P}}_t\mathbf{e}_1$ of staying in regime 1.

3. Combine the simulated $\varphi-$period asset returns $\{\mathbf{R}_{b+1,n}\}_{n=1}^N$ into a random sample of size N, using the probability weights contained in the vector $\boldsymbol{\pi}_b^j$:

$$\mathbf{R}_{b+1,n}(\boldsymbol{\pi}_b^j) = \sum_{i=1}^{k}(\boldsymbol{\pi}_b^j)'\mathbf{e}_i\mathbf{R}_{b+1,n}(s_b = i).$$

4. Update the future regime probabilities perceived by the investor using the formula:

$$\boldsymbol{\pi}_{b+1,n}(\boldsymbol{\pi}_b^j) = \frac{\left(\boldsymbol{\pi}_b'(\hat{\boldsymbol{\theta}}_b)\hat{\mathbf{P}}_b^{\varphi}\right)' \odot \boldsymbol{\eta}(\mathbf{r}_{b+1};\hat{\boldsymbol{\theta}}_b)}{[(\boldsymbol{\pi}_b'(\hat{\boldsymbol{\theta}}_b)\hat{\mathbf{P}}_b^{\varphi})' \odot \boldsymbol{\eta}(\mathbf{r}_{b+1};\hat{\boldsymbol{\theta}}_b))]'\boldsymbol{\iota}_k}$$

   obtaining an $N \times 4$ matrix $\{\boldsymbol{\pi}_{b+1,n}(\boldsymbol{\pi}_b^j)\}_{n=1}^N$, each row of which corresponds to a simulated row vector of perceived regime probabilities at time t_{b+1}.

5. For all $n = 1, 2, ..., N$, calculate the value $\tilde{\boldsymbol{\pi}}_{b+1,n}^j$ on the discretization grid ($j = 1,\ 2, ...,\ G$) that is closest to $\boldsymbol{\pi}_{b+1,n}(\boldsymbol{\pi}_b^j)$ according to the metric $\sum_{i=1}^3 |(\boldsymbol{\pi}_{b+1}^j)'\mathbf{e}_i - \boldsymbol{\pi}_{b+1,n}'\mathbf{e}_i|$, i.e.

$$\tilde{\boldsymbol{\pi}}_{b+1,n}^j(\boldsymbol{\pi}_b^j) \equiv \arg\min_{\mathbf{x}\in\boldsymbol{\pi}_{b+1}^j} \sum_{i=1}^{3} |\mathbf{x}'\mathbf{e}_i - \boldsymbol{\pi}_{b+1,n}'\mathbf{e}_i|.$$

   Knowledge of the vector $\{\tilde{\boldsymbol{\pi}}_{b+1,n}^j(\boldsymbol{\pi}_b^j)\}_{n=1}^N$ allows us to build $\{Q(\boldsymbol{\pi}_{b+1}^{(j,n)}, t_{b+1})\}_{n=1}^N$, where $\boldsymbol{\pi}_{b+1}^{(j,n)} \equiv \tilde{\boldsymbol{\pi}}_{b+1,n}^j(\boldsymbol{\pi}_b^j)$ is a function of the assumed vector of regime probabilities $\boldsymbol{\pi}_b^j$.

6. Solve the program

$$\max_{\boldsymbol{\omega}_b(\boldsymbol{\pi}_b^j)} N^{-1}\sum_{n=1}^{N}\left[\left\{\boldsymbol{\omega}_b' \exp\left(\mathbf{R}_{b+1,n}(\boldsymbol{\pi}_b^j)\right)\right\}^{1-\gamma} Q(\boldsymbol{\pi}_{b+1}^{(j,n)}, t_{b+1})\right],$$

   which for large values of N provides an arbitrarily precise Monte-Carlo approximation of the expectation $E\left[\left\{\boldsymbol{\omega}_b' \exp\left(\mathbf{R}_{b+1,n}(\boldsymbol{\pi}_b^j)\right)\right\}^{1-\gamma} Q(\boldsymbol{\pi}_{b+1}^j, t_{b+1})\right]$. The value function corresponding to the optimal portfolio weights $\hat{\boldsymbol{\omega}}_b(\boldsymbol{\pi}_b^j)$ defines $Q(\boldsymbol{\pi}_b^j, t_b)$ for the jth point on the initial grid.

This algorithm is applied to values $\boldsymbol{\pi}_b^j$ on the discretization grid until all values of $Q(\boldsymbol{\pi}_b^j, t_b)$ are obtained for $j = 1,\ 2, ...,\ G$. It is then iterated backwards until $t_{b+1} = t + \varphi$. At that stage the algorithm is applied one last time, taking $Q(\boldsymbol{\pi}_{t+\varphi}^j, t+\varphi)$ as given and using one row vector of perceived regime probabilities $\boldsymbol{\pi}_t$, the vector of smoothed probabilities estimated at time t. The resulting vector of optimal portfolio weights $\hat{\boldsymbol{\omega}}_t$ is the desired optimal portfolio allocation at time t, while $Q(\boldsymbol{\pi}_t, t)$ is the optimal value function.

Appendix B – Density Specification Tests

Under a k-regime mixture of normals, the z-score is given by

$$\Pr\left(\mathbf{r}_{t+1} \leq \tilde{\mathbf{r}}_{t+1}|\mathcal{F}_t\right) = \sum_{i=1}^{k}\Phi_m\left(\boldsymbol{\Sigma}_i^{-1}\left[\mathbf{r}_{t+1} - \boldsymbol{\mu}_i - \sum_{j=1}^{p}\mathbf{A}_{j,i}\mathbf{r}_{t+1-j}\right]\right)\Pr(S_{t+1}=i|\mathcal{F}_t) \equiv z_{t+1} \in \mathcal{R}, \qquad (13)$$

where $\Phi_m(\cdot)$ is the standard m-variate normal c.d.f. As stressed by Rosenblatt (1952), if the model is correctly specified, z_{t+1} should be independently and identically distributed (IID) and uniform on the interval $[0, 1]$. The uniform requirement relates to the fact that deviations between realized values and predicted ones should be conditionally normal and as such describe a uniform distribution once it is 'filtered through' an appropriate Gaussian cdf. The IID condition reflects the fact that if the model is correctly specified, forecast errors ought to be unpredictable and fail to show any detectable structure. Unfortunately, testing whether a distribution is uniform is not a simple task. Berkowitz (2001) has recently proposed a likelihood-ratio test that inverts Φ to get a transformed z-score,

$$z_{t+1}^* \equiv \Phi^{-1}(z_{t+1}),$$

which essentially turns the z-score back into a bell-shaped random variable. Provided that the model is correctly specified, z^* should be IID and normally distributed (IIN(0, 1)). We follow Berkowitz (2001) and use a likelihood ratio test that focuses on a few salient moments of the return distribution. Suppose the log-likelihood function is evaluated under the null that $z_{t+1}^* \sim IIN(0,1)$:

$$L_{IIN(0,1)} \equiv -\frac{T}{2}\ln(2\pi) - \sum_{t=1}^{T}\frac{(z_t^*)^2}{2}.$$

Under the alternative of misspecification, the likelihood incorporates deviations from the null, $z_{t+1}^* \tilde{} \; IIN(0,1)$:

$$z_{t+1}^* = \eta + \sum_{j=1}^{w}\sum_{i=1}^{r}\psi_{ji}(z_{t+1-i}^*)^j + \nu u_{t+1}, \tag{14}$$

where $u_{t+1} \sim IIN(0,1)$. The null of a correct model implies $w \times r + 2$ restrictions – i.e., $\eta = \psi_{ji} = 0$ ($j = 1, ..., w$ and $i = 1, ..., r$) and $\nu = 1$ – in (14). Let $L(\hat{\eta},\, \{\hat{\psi}_{ji}\}_{j=1\;i=1}^{w\;\;\;r},\, \hat{\nu})$ be the maximized log-likelihood obtained from (14). To test that a null model is correctly specified, we can then use the following test statistic:

$$LR_{wr+2} \equiv -2\left[L_{IIN(0,1)} - L(\hat{\eta}, \{\hat{\psi}_{ji}\}_{j=1\;i=1}^{w\;\;\;r}, \hat{\nu})\right] \xrightarrow{d} \chi^2_{wr+2}.$$

In addition to the standard Jarque-Bera test that considers skew and kurtosis in the z-scores to detect non-normalities in z_{t+1}^*, it is customary to present three likelihood ratio tests, namely a test of zero-mean and unit variance ($w = r = 0$), a test of lack of serial correlation in the z-scores ($w = 1$ and $r = 1$) and a test that further restricts their squared values to be serially uncorrelated in order to test for omitted volatility dynamics ($w = 2$ and $r = 2$). Notice that a rejection of the null of normal transformed z-scores has the same meaning as rejecting the null of a uniform distribution for the raw z-scores, i.e. the model fails to generate a density with the appropriate shape. A rejection of the zero-mean, unit variance restriction points to specific problems in the location and scale of the density underlying the model. A rejection of the restriction that $\{z_{t+1}^*\}$ is IID points to dynamic misspecifications (serial correlation or heteroskedasticity).

In our paper, density specification tests are applied not the vector $\mathbf{r}_{t+1}$, but to each of its components:

$$\Pr\left(r_{t+1}^j \leq \tilde{r}_{t+1}^j|\Im_t\right) = \sum_{i=1}^{k}\Phi\left(\sigma_{j,i}^{-1}\left[r_{t+1}^j - \mu_{j,i} - \mathbf{e}_j'\sum_{u=1}^{p}\mathbf{A}_{j,i}\mathbf{r}_{t+1-u}\right]\right)\Pr(S_{t+1} = i|\mathcal{F}_t) \equiv z_{t+1}^j \qquad j = 1, ..., m,$$

where $\sigma_{j,i}$ is the volatility of variable j in state i.

Table 1

Summary Statistics for International Stock Returns

The table reports basic moments for weekly percentage equity total return series (including dividends, adjusted for stock splits, etc.) for a few international portfolios and two sample periods. The sample period is January 1999 – January 2007. All returns are expressed in local currencies. Means, medians, and standard deviations are annualized by multiplying weekly moments by 52 and $\sqrt{52}$, respectively. LB(j) denotes the j-th order Ljung-Box statistic.

Portfolio	Mean	Median	St. Dev.	Skewness	Kurtosis	LB(4)	LB(4)-squares
	January 1999 – June 2003						
DJ Stoxx Europe – Large Caps	6.137	20.881	23.900	-0.288	6.520	31.817**	84.586**
MSCI Europe – Small Caps	13.575	38.844	19.838	-1.922	12.649	11.037*	14.708**
North America – Large Caps	3.812	6.874	16.525	0.065	4.965	12.588*	55.783*
MSCI North America – Small Caps	12.846	16.519	23.559	-0.435	4.936	18.395**	15.535**
MSCI Asia Pacific	5.942	17.058	16.772	-0.259	3.672	4.343	10.602*

* denotes 5% significance, ** significance at 1%.

Table 2

Correlation Matrix of International Stock Returns

The table reports linear correlation coefficients for weekly equity total return series (including dividends, adjusted for stock splits, etc.) for a few international portfolios. The sample period is January 1999 – June 2003. All returns are expressed in local currencies.

	EU – Large	EU – Small	North Am. – Large	North Am. – Small	Pacific
EU – Large Caps	1	0.669	0.735	0.662	0.545
EU – Small Caps		1	0.562	0.591	0.548
North Am. – Large Caps			1	0.811	0.491
North Am. – Small Caps				1	0.446
Pacific					1

Table 3

Model Selection: Information Criteria

The table reports model selection criteria for a range of multivariate regime switching and autoregressive conditional heteroskedasticity models. The regime switching models have structure:

$$r_t = \mu_{s_t} + \sum_{j=1}^{p} A_{j,s_t} r_{t-1} + \varepsilon_t$$

where $\mathbf{r}_t$ is a 4×1 vector collecting weekly total return series, $\boldsymbol{\mu}_{s_t}$ is the intercept vector in state s_t, A_{j,s_t} is the j-th order matrix of VAR coefficients that characterizes state s_t, and $\boldsymbol{\varepsilon}_t = [\varepsilon_{1t}\ \varepsilon_{2t}\ \varepsilon_{3t}\varepsilon_{4t}]' \sim N(\mathbf{0}, \Sigma_{s_t})$. The unobservable state s_t is governed by a first-order Markov chain that can assume three values. The multivariate ARCH models span full multivariate GARCH and DCC models with and without 'ARCH-in-mean' components (in this case the indication of the model is followed by '-M'). In one DCC case, an integrated version is estimated. The data are weekly. The sample period is January 1999 – January 2007 (for a total of 1,668 observations).

Model	Number of parameters	Log-likelihood	LR test for linearity	AIC	BIC	Hannan-Quinn
Baseline model: Single-state, Homoskedastic						
Gaussian IID	14	-3660.66	NA	17.6243	17.7597	17.6778
Gaussian VAR(1)	30	-3612.42	NA	17.5116	17.8023	17.6266
Baseline model: Single-state, Conditional Heteroskedastic VAR(1)						
MGARCH(1,1) VAR(1)	62	-3374.92	NA	16.4841	17.0837	16.7211
M-CCORR GARCH(1,1) VAR(1)	34	-3533.96	NA	17.1125	17.4413	17.2425
MGARCH(1,1)-M VAR(1)	78	-3345.64	NA	16.4203	17.1747	16.7186
DCC EGARCH(1,1) VAR(1)	44	-3406.75	NA	16.5504	16.9759	16.7185
DCC GARCH(1,1)-M VAR(1)	56	-3392.98	NA	16.5419	17.0835	16.7560
Integrated DCC GARCH(1,1)-M VAR(1)	48	-3458.23	NA	16.8165	17.2807	17.0000
Baseline model: Two-state, Regime Switching						
MMSI(2,0)	20	-3579.84	161.6538 (0.000)	17.2654	17.4588	17.3419
MMSIH(2,0)	30	-3398.37	524.5905 (0.000)	16.4430	16.7332	16.5577
MMSIAH(2,1)	62	-3345.62	533.5979 (0.001)	16.3828	16.9835	16.6203
MMSIAH(2,2)	94	-3304.88	568.5333 (0.001)	16.3802	17.2926	16.7410
Baseline model: Three-state, Regime Switching						
MMSI(3,0)	28	-3564.39	192.5370 (0.000)	17.2297	17.5005	17.3368
MMSIH(3,0)	48	-3339.88	641.5576 (0.000)	16.2488	**16.7131**	16.4324
MMSIAH(3,1)	96	-3266.88	691.0677 (0.000)	16.1677	17.0979	16.5355
MMSIH(3,0)-VAR(1)	64	-3292.25	640.3429 (0.000)	**16.0358**	16.7559	**16.3810**
Baseline model: Four-state, Regime Switching						
MMSI(4,0)	38	-3527.62	266.0879 (0.000)	17.1013	17.4688	17.2466
MMSIH(4,0)	68	-3315.97	689.3923 (0.000)	16.2301	16.8877	16.4901
MMSIH(4,0)-VAR(1)	84	**-3262.39**	700.0574 (0.000)	16.0884	16.9023	16.4102

Table 4

Model Selection: Tests Based on One-Step Ahead Density Forecasts

This table reports model specification tests based on the principle that under a correct specification, the properly transformed one-step-ahead standardized residuals should follow an independently and identically distributed normal distribution with zero mean and unit variance (see Berkowitz (2001)). Significant tests indicated by stars show that the model is misspecified. Jarque-Bera tests whether the normalized residuals have zero skew and excess kurtosis. LR_2 is a test for correct mean and variance (zero and one, respectively); LR_3 tests for first order serial correlation, while LR_6 tests for first and second order serial correlation in the normalized residuals and their squares. This gives the ability to detect the presence of residual ARCH effects.

Model	Number of parameters	Jarque-Bera test	LR_2	LR_3	LR_6
Asian Pacific Returns					
Single-state Gaussian IID	14	11.6962**	4.5063	8.0395*	23.1629**
Single-state VAR(1)	30	12.2945**	3.0559	9.1113*	24.4797**
DCC EGARCH (1,1) VAR(1)	44	1.7125	3.8076	6.0539	16.0505*
Three-state switching VAR(1)	64	0.2435	2.1002	4.5020	15.0220*
European Small Caps Returns					
Single-state Gaussian IID	14	1842.85**	9.0504*	23.1624**	29.0764**
Single-state VAR(1)	30	1713.56**	7.8042*	15.0302**	24.0624**
DCC EGARCH (1,1) VAR(1)	44	6.9604*	4.5064	7.0936	15.9039*
Three-state switching VAR(1)	64	3.8467	1.3234	5.5732	10.3962
European Large Caps Returns					
Single-state Gaussian IID	14	1373.65**	10.3795**	32.2275**	42.8775**
Single-state VAR(1)	30	1815.78**	9.9389**	17.9249**	24.9474**
DCC EGARCH (1,1) VAR(1)	44	14.7638**	4.9884	8.7494*	14.1148*
Three-state switching VAR(1)	64	7.5521*	1.8200	4.1316	6.2666
North American Large Caps Returns					
Single-state Gaussian IID	14	43.1248**	2.8740	13.3314**	24.5946**
Single-state VAR(1)	30	33.5435**	2.9638	6.6524	17.2950**
DCC EGARCH (1,1) VAR(1)	44	8.0634*	3.0495	7.4425	11.8231
Three-state switching VAR(1)	64	0.3220	1.3368	3.1414	13.4038*

*denotes significance at the 5% level, ** significance at the 1% level.

Table 5

Estimates of a Three-State (Time-Invariant) VAR(1) Regime Switching Model

The table shows estimation results for the regime switching model:

$$r_t = \mu_{s_t} + A r_{t-1} + \varepsilon_t$$

where $\mathbf{r}_t$ is a 4×1 vector collecting weekly total return series, $\boldsymbol{\mu}_{s_t}$ is the intercept vector in state s_t, and $\boldsymbol{\varepsilon}_t = [\varepsilon_{1t}\ \varepsilon_{2t}\ \varepsilon_{3t}\varepsilon_{4t}]' \sim N(\mathbf{0}, \Sigma_{s_t})$. The unobservable state s_t is governed by a first-order Markov chain that can assume three values. The first panel refers to the single-state case of a single-state Gaussian VAR(1). Asterisks attached to correlation coefficients refer to covariance estimates. Transition probabilities have to be read as the probability of switching from the state in the row to the state in the column.

Panel A – Single State Model				
	Pacific	Europe – Small caps	Europe – Large caps	North America Large
1. Intercept	0.1159	0.2393*	0.0820	0.0610
2. VAR Coefficients				
Pacific	-0.0510	-0.0124	0.0213	0.0879*
Europe – Small caps	-0.0501	0.1349**	-0.0199	0.0748
Europe – Large caps	-0.0381	0.0551	-0.2333***	0.1082***
North America - Large caps	-0.0859	0.1542**	-0.0908*	-0.0770
2. Correlations/Volatilities				
Pacific	2.3164***			
Europe – Small caps	0.5504***	2.7134***		
Europe – Large caps	0.5258***	0.6592***	3.4494***	
North America - Large caps	0.5117***	0.5746***	0.7012***	2.2471***

Panel B – Three State Model				
	Pacific	Europe – Small caps	Europe – Large caps	North America Large
1. Intercepts				
Crash State	-1.1424**	-3.0750***	-2.8111***	-1.1692***
Bear State	0.0080	0.3367	0.4240	0.1000
Bull State	0.4378***	0.8178***	0.3748**	0.2853***
2. VAR Coefficients				
Pacific	0.0204	-0.0404	-0.0323	0.1376*
Europe – Small caps	0.0289	0.0608	-0.1056**	0.1645***
Europe – Large caps	0.0199	0.0037	-0.2928***	0.2443***
North America - Large caps	-0.0356	0.0840*	-0.1294**	-0.0021
3. Correlations/Volatilities				
Crash state:				
Pacific	2.9458***			
Europe – Small caps	0.6584***	6.1870***		
Europe – Large caps	0.3047**	0.5268***	6.7203***	
North America - Large caps	0.4108***	0.6295***	0.7058***	2.8631***
Bear state:				
Pacific	2.7288***			
Europe – Small caps	0.5531***	1.9938***		
Europe – Large caps	0.6324***	0.7604***	3.5717***	
North America - Large caps	0.5385***	0.6546***	0.7455***	2.9000***
Bull state:				
Pacific	1.6663***			
Europe – Small caps	0.5816***	1.3236***		
Europe – Large caps	0.5548***	0.7627***	1.7849***	
North America - Large caps	0.4631**	0.6067***	0.6368***	1.2638***

3. Transition probabilities	Crash State	Bear State	Bull State
Crash State	0.3973**	0.3263**	0.2763*
Bear State	0.0436	0.9097***	0.0467
Bull State	0.0821*	0.0035	0.9145***

* denotes 10% significance, ** significance at 5%, *** significance at 1%.

Table 6

Sample and Implied Co-Skewness Coefficients

The table reports the sample co-skewness coefficients,

$$S_{i,j,l} \equiv \frac{E[(r_i - E[r_i])(r_j - E[r_j])(r_l - E[r_l])]}{\{E[(r_i - E[r_i])^2]E[(r_j - E[r_j])^2]E[(r_l - E[r_l])^2]\}^{1/2}}$$

(*i, j, l* = Europe large, North America large, Pacific, Europe small) and compares them with the co-skewness coefficients implied by a three-state VAR(1) regime switching model:

$$\mathrm{r}_t = \mu_{s_t} + \mathrm{A}\mathrm{r}_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim I.I.D.\ N(\mathbf{0}, \mathrm{I}_4)$ is an unpredictable return innovation. Coefficients under regime switching are calculated employing simulations (50,000 trials) and averaging across simulated samples of length equal to the available data (January 1999 – January 2007). In the table NA stands for 'North American small caps', and Pac for 'Pacific' portfolios. Bold coefficients are significantly different from zero.

Coeff.	Sample	MS – ergodic	Regime 1	Regime 2	Regime 3
$S_{EU\_large,EU\_large,NA}$	-0.181	-0.314	-0.318	-0.304	-0.227
$S_{EU\_large,EU\_large,Pac}$	-0.219	-0.157	-0.158	-0.151	-0.164
$S_{EU\_large,EU\_large,EU\_small}$	**-0.518**	**-0.341**	-0.142	-0.136	-0.148
$S_{NA,NA,Pac}$	-0.109	**-0.547**	**-0.552**	-0.233	-0.267
S_{NA,NA,EU_small}	-0.257	**-0.631**	**-0.536**	-0.315	**-0.352**
S_{NA,NA,EU_large}	0.006	-0.063	-0.063	-0.064	-0.065
S_{Pac,Pac,EU_small}	**-0.577**	**-0.376**	**-0.381**	**-0.261**	-0.295
S_{Pac,Pac,EU_large}	-0.274	-0.254	-0.256	-0.244	-0.265
$S_{Pac,Pac,NA}$	-0.215	**-0.730**	**-0.735**	**-0.312**	**-0.453**
$S_{EU\_small,EU\_small,EU\_large}$	**-0.791**	**-0.358**	-0.160	-0.152	-0.167
$S_{EU\_small,EU\_small,NA}$	**-0.603**	**-0.444**	**-0.348**	-0.230	-0.262
$S_{EU\_small,EU\_small,Pac}$	**-0.881**	**-0.446**	**-0.249**	-0.235	-0.261
$S_{EU\_large, EU\_large, EU\_large}$	-0.288	-0.233	-0.235	-0.225	-0.243
$S_{NA,NA,NA}$	0.065	-0.017	0.072	-0.017	-0.014
$S_{Pac,Pac,Pac}$	-0.259	**-0.693**	**-0.701**	-0.272	-0.079
$S_{EU\_small, EU\_small, EU\_small}$	**-1.922**	**-1.222**	-0.231	-0.217	-0.244

Table 7

Sample and Implied Co-Skewness and Co-Kurtosis Coefficients of European Small Caps vs. an Equally Weighted International Equity Portfolio

The table reports average sample co-skewness coefficients,

$$S_{i,j,l} \equiv \frac{E[(r_i - E[r_i])(r_j - E[r_j])(r_l - E[r_l])]}{\{E[(r_i - E[r_i])^2]E[(r_j - E[r_j])^2]E[(r_l - E[r_l])^2]\}^{1/2}}$$

$$K_{i,j,l,b} \equiv \frac{E[(r_i - E[r_i])(r_j - E[r_j])(r_l - E[r_l])(r_b - E[r_b])]}{\{E[(r_i - E[r_i])^2]E[(r_j - E[r_j])^2]E[(r_l - E[r_l])^2]E[(r_b - E[r_b])^2]\}^{1/2}}$$

(*i, j, l* = Europe large, North America large, Pacific, Europe small, Equally weighted portfolio) and compares them with the co-kurtosis coefficients implied by a three-state regime switching model with a time-invariant VAR(1) component. Coefficients under multivariate regime switching are calculated employing simulations. Bold co-skewness coefficients are significantly different from zero; bold co-kurtosis coefficients are significantly different from their Gaussian counterparts.

	Co-Skewness		Co-Kurtosis	
	Sample	MS - ergodic	Sample	MS - ergodic
	European Small Caps			
$S_{EU\_small,EU\_small,EW\_ptf}$	**-1.266**	**-0.499**	–	–
$S_{EU\_small,EW\_ptf,EW\_ptf}$	**-0.913**	**-0.628**	–	–
$S_{EU\_small,EU\_small,Pac,EW\_ptf}$	–	–	**4.174**	**3.298**
$S_{EU\_small,EU\_small,NA,EW\_ptf}$	–	–	3.039	2.944
$S_{EU\_small,EU\_small,EU\_large,EW\_ptf}$	–	–	**4.171**	2.172
$S_{EW\_ptf,EW\_ptf,EU\_small,Pac}$	–	–	**3.304**	**3.789**
$S_{EW\_ptf,EW\_ptf,EU\_small,NA}$	–	–	**2.977**	**3.666**
$S_{EW\_ptf,EW\_ptf,EU\_small,EU\_large}$	–	–	**5.945**	**4.645**
$S_{EW\_ptf,EW\_ptf,EU\_small,EU\_small}$	–	–	**5.697**	**4.675**
$S_{EW\_ptf,EW\_ptf,EU\_ptf,EU\_small}$	–	–	**8.234**	**5.710**
$S_{EU\_small,EU\_small,EU\_small,EU\_ptf}$	–	–	**4.937**	**4.157**
	North American Large Caps			
$S_{NA\_large,NA\_large,EW\_ptf}$	0.087	-0.171	–	–
$S_{NA\_large,EW\_ptf,EW\_ptf}$	-0.297	-0.280	–	–
$S_{NA\_large,NA\_large,EU\_large,EW\_ptf}$	–	–	3.245	**3.844**
$S_{NA\_large,NA\_large,Pac,EW\_ptf}$	–	–	1.914	2.046
$S_{NA\_large,NA\_large,EU\_small,EW\_ptf}$	–	–	2.456	**3.864**
$S_{EW\_ptf,EW\_ptf,NA\_large,Pac}$	–	–	2.130	2.045
$S_{EW\_ptf,EW\_ptf,NA\_large,EU\_large}$	–	–	**3.273**	**4.233**
$S_{EW\_ptf,EW\_ptf,NA\_large,EU\_small}$	–	–	2.977	**3.666**
$S_{EW\_ptf,EW\_ptf,NA\_large,NA\_large}$	–	–	**3.479**	3.233
$S_{EW\_ptf,EW\_ptf,EW\_ptf,NA\_large}$	–	–	**3.857**	**4.844**
$S_{NA\_large,NA\_large,NA\_large,EU\_ptf}$	–	–	**3.682**	**3.490**

Table 8

Sample and Implied Co-Kurtosis Coefficients

The table reports the sample co-kurtosis coefficients,

$$K_{i,j,l,b} \equiv \frac{E[(r_i - E[r_i])(r_j - E[r_j])(r_l - E[r_l])(r_b - E[r_b])]}{\{E[(r_i - E[r_i])^2]E[(r_j - E[r_j])^2]E[(r_l - E[r_l])^2]E[(r_b - E[r_b])^2\}^{1/2}}$$

(*i, j, l, b* = Europe large, North America large, Pacific, Europe small) and compares them with the co-kurtosis coefficients implied by a three-state VAR(1) regime switching model:

$$r_t = \mu_{s_t} + Ar_{t-1} + \varepsilon_t,$$

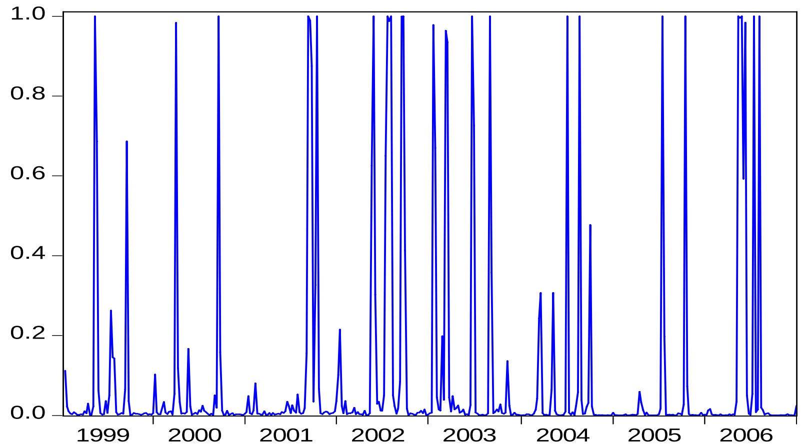
where $\varepsilon_t \sim I.I.D.\ N(\mathbf{0}, I_4)$ is an unpredictable return innovation. Coefficients under multivariate regime switching are calculated employing simulations (50,000 trials) and averaging across simulated samples. In the table NA stands for 'North American small caps', and Pac for 'Pacific' equity portfolios. Bold coefficients are significantly different from their Gaussian counterparts.

Coeff.	Sample	MS – erg.	Regime 1	Regime 2	Regime 3
$K_{EU\_large, EU\_large, NA, EU\_small}$	**2.477**	**2.534**	1.533	1.510	1.559
$K_{EU\_large, EU\_large, NA, Pac}$	1.448	1.745	1.741	1.325	1.272
$K_{EU\_large, EU\_large, Pac, EU\_small}$	1.704	1.644	1.640	1.629	1.458
$K_{NA,NA,EU\_large,Pac}$	1.524	1.416	1.421	1.386	1.365
$K_{NA,NA,EU\_large,EU\_small}$	**2.189**	**1.996**	**1.999**	1.858	1.401
K_{NA,NA,Pac,EU_small}	1.179	**2.180**	1.183	1.157	1.141
$K_{Pac,Pac,EU\_large,EU\_small}$	1.846	1.762	1.758	1.123	1.255
$K_{Pac,Pac,EU\_large,NA}$	1.462	**2.014**	**2.017**	1.743	1.585
$K_{Pac,Pac,EU\_large,NA}$	1.350	1.224	1.228	1.156	1.270
$K_{EU\_small,EU\_small,EU\_large,NA}$	**2.336**	**2.595**	1.594	1.294	1.626
$K_{EU\_small,EU\_small,EU\_large,Pac}$	**2.662**	1.694	1.687	1.675	1.711
$K_{EU\_small,EU\_small,,NA,Pac}$	1.769	**2.482**	**2.482**	1.430	1.540
$K_{EU\_large,EU\_large,NA,NA}$	**3.341**	**2.840**	2.246	1.794	1.896
$K_{EU\_large,EU\_large,Pac,Pac}$	1.657	**2.167**	1.765	1.545	1.793
$K_{EU\_large,EU\_large,EU\_small,EU\_small}$	**3.370**	**2.979**	1.975	**1.959**	**1.997**
$K_{NA,NA,Pac,Pac}$	**1.811**	**2.906**	**4.910**	1.843	1.800
$K_{NA,NA,EU\_small,EU\_small}$	**2.228**	**2.982**	**2.985**	**2.015**	2.058
$K_{Pac,Pac,EU\_small,EU\_small}$	**3.585**	**3.312**	2.314	1.847	1.389
$K_{EU\_large,EU\_large,EU\_large,NA}$	**4.375**	**3.299**	2.304	**2.264**	1.229
$K_{EU\_large,EU\_large,EU\_large,Pac}$	1.504	1.928	1.921	1.620	1.436
$K_{EU\_large,EU\_large,EU\_large,EU\_small}$	**3.918**	**2.860**	1.856	1.448	1.371
$K_{NA,NA,NA,Pac}$	**1.963**	**2.663**	**5.659**	2.612	1.757
K_{NA,NA,NA,EU_small}	**2.461**	**3.484**	**3.486**	1.809	1.580
$K_{Pac,EU\_small,EU\_small,EU\_small}$	**5.824**	**4.082**	**3.246**	**2.190**	1.554
K_{NA,NA,NA,EU_large}	**3.501**	**4.473**	**3.475**	**2.412**	**2.271**
K_{Pac,Pac,Pac,EU_large}	1.808	**2.360**	2.356	1.337	1.548
$K_{EU\_small,EU\_small,EU\_small,EU\_large}$	**5.209**	**3.064**	2.058	1.541	1.940
$K_{Pac,Pac,Pac,NA}$	1.604	1.594	1.601	1.524	1.336
$K_{EU\_small,EU\_small,EU\_small,NA}$	**3.656**	**2.628**	**2.495**	1.572	2.064
K_{Pac,Pac,Pac,EU_small}	**2.485**	**4.082**	**4.091**	1.999	1.596
$K_{EU\_large,EU\_large,EU\_large,EU\_large}$	**6.520**	3.673	3.474	3.164	3.450
$K_{NA,NA,NA,NA}$	**4.965**	**5.061**	**4.023**	3.648	3.049
$K_{Pac,Pac,Pac,Pac}$	3.672	**4.958**	3.981	3.286	3.394
$K_{EU\_small,EU\_small,EU\_small,EU\_small}$	**12.649**	**8.422**	**4.422**	3.354	3.249

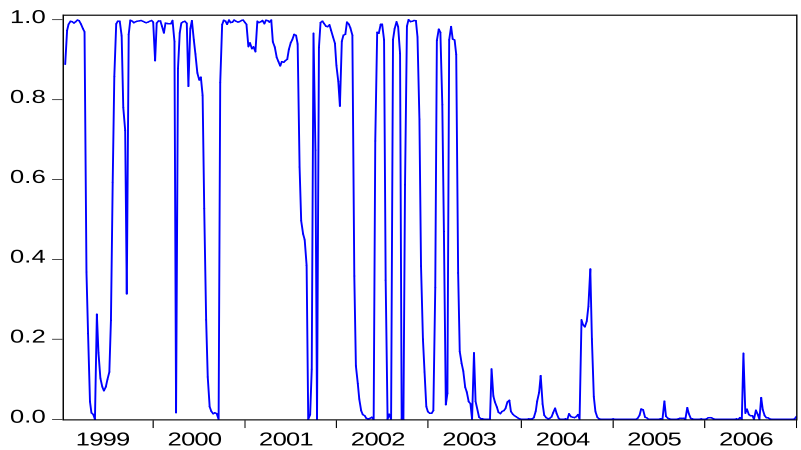
Figure 1

Smoothed State Probabilities from Three-State VAR(1) Regime Switching Model

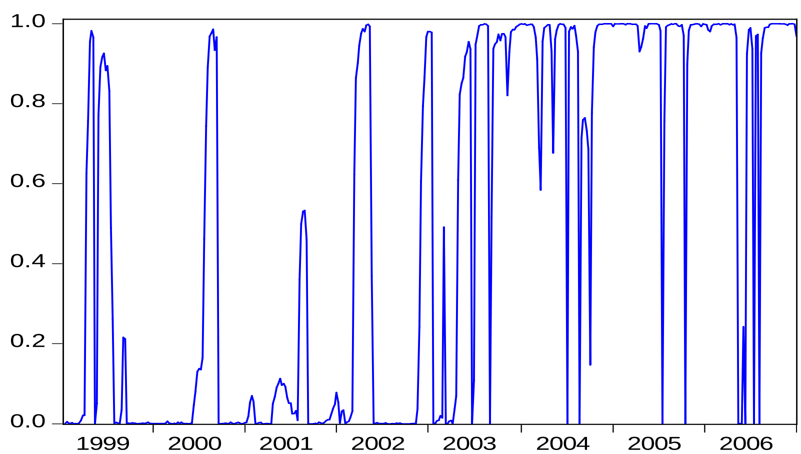
The graphs plot the smoothed state probabilities for the multivariate four-state VAR(1) Switching model comprising weekly return series on four international equity portfolios (including European small cap firms). The sample period is Jan. 1999 – Jan. 2007. Parameter estimates underlying these plots are reported in Table 4.



Crash state



Bear state



Bull state

Figure 2

Buy-and-Hold Optimal Allocation

The graphs plot the optimal international equity portfolio weights when returns follow a three-state VAR(1) switching model as a function of the investment horizon. The vector autoregressive coefficients are constant over time. As a benchmark (horizontal lines) we also report the IID/Myopic allocation. Optimal portfolio weights are computed assuming constant relative risk aversion preferences with $\gamma = 5$.

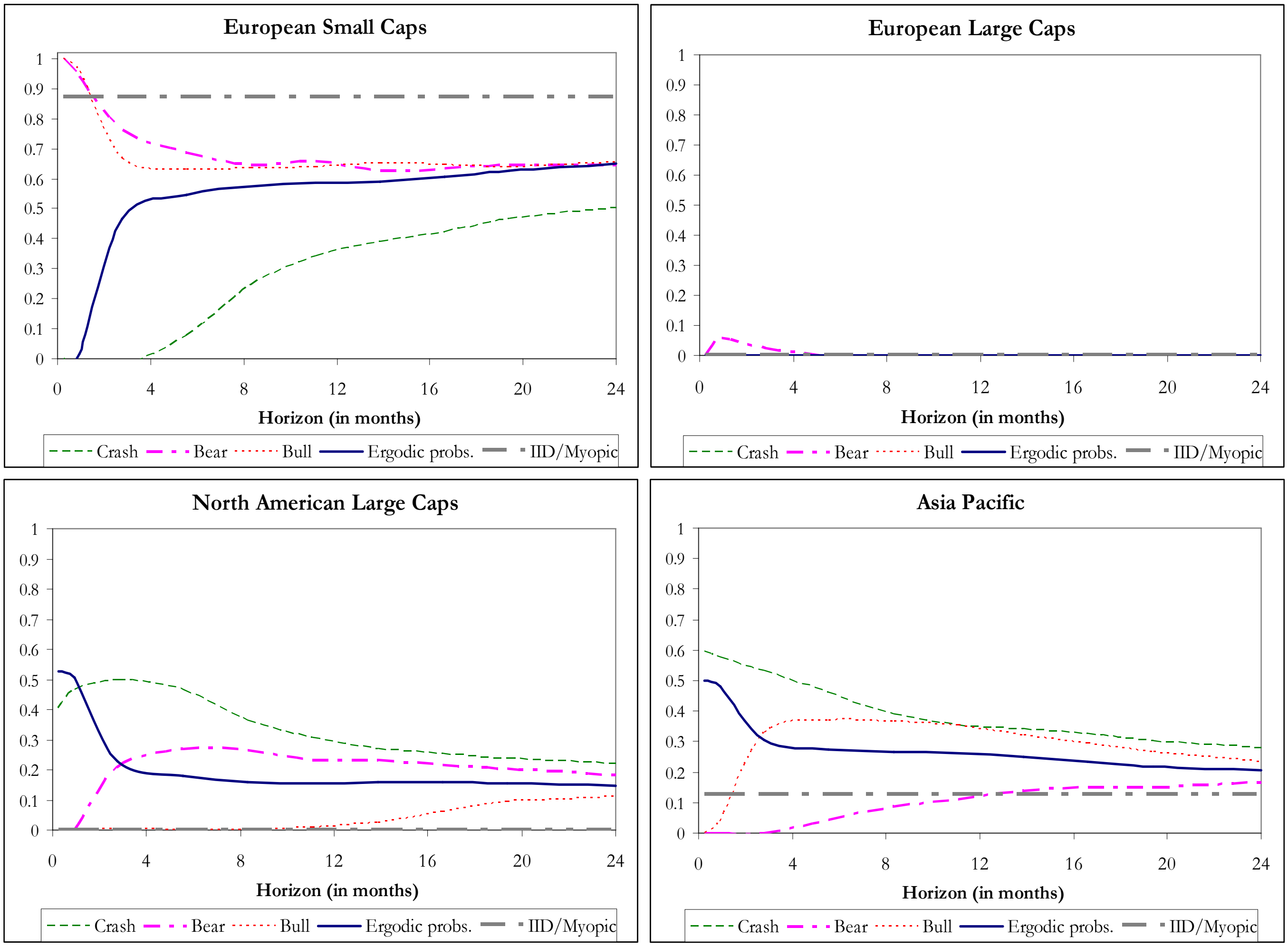
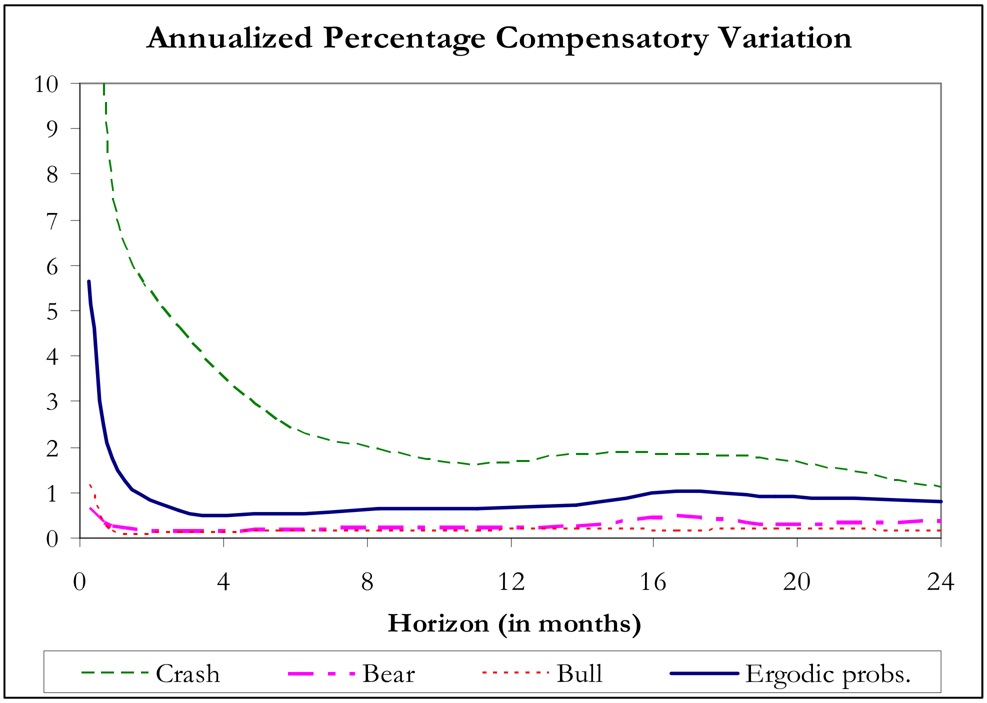


Figure 3

Welfare Costs of Ignoring Regime Switching

The graphs plot the (annualized) percentage compensatory variation from ignoring the presence of regime switches in the multivariate process of asset returns. The graphs plot the annualized welfare costs as a function of the investment horizon; calculations were performed for two alternative levels of the coefficient of relative risk aversion. The investor is assumed to have a simple buy-and-hold objective.

$\gamma = 5$



$\gamma = 10$