Annals of Finance (2009) 5:15–48
DOI 10.1007/s10436-007-0090-2

RESEARCH ARTICLE

Small caps in international equity portfolios: the effects of variance risk

Massimo Guidolin · Giovanna Nicodano

Received: 23 August 2006 / Accepted: 21 November 2007 / Published online: 3 January 2008

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.

Keywords Intertemporal portfolio choice return predictability · Co-skewness and co-kurtosis · International portfolio diversification

JEL Classification G11 · G15 · F30 · C32 · G0 · G1

M. Guidolin (✉)
Accounting and Finance Group,
Manchester Business School and Collegio Carlo Alberto (CeRP),
Booth Street East, Manchester M13 9PL, UK
e-mail: Massimo.Guidolin@mbs.ac.uk

G. Nicodano
Faculty of Economics,
University of Turin and Collegio Carlo Alberto (CeRP),
Corso Unione Sovietica, 218bis,
10134 Turin, Italy
e-mail: giovanna.nicodano@unito.it

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 et al. 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, even in multivariate asset menus (Campbell et al. 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 2 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, regime 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 et al. (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 et al. (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

[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.

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 generate 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 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 subsections, 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

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. In addition to giving some basic information on the structure of the models, 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 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 $1, 2, \ldots k$, 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, 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

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

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 Sect. 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, in the case $m = 4$, $p = 1$, and $k = 3$ (which will represent a reasonable specification in our application) (1) implies 96 free parameters.

Equation (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 et al. (2003).[2] However, when multiple regimes are allowed (1) generates various sources of predictability. When either $\boldsymbol{\mu}_{S_t}$ or $\mathbf{A}_{j,S_t}$ $(j = 1, \ldots, p)$ 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. 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)—of normal variates, which is generally not Gaussian.

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 1994):

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

[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 $p = 0$.

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 $\mathbf{f}(\mathbf{r}_{t+1}; \hat{\boldsymbol{\theta}}_t)$ is the density of returns conditional on the regime, on past returns and on estimated parameters.

Regime switching models are estimated by maximum likelihood. As shown by Hamilton (1994), 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. 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 (1994) has proven consistency and asymptotic normality of the ML estimator $\hat{\boldsymbol{\theta}}$. Additional details on estimation algorithms and the statistical properties of the ML estimator can be found in Guidolin and Nicodano (2006).

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 a simple 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, $\boldsymbol{\varepsilon}_t = \mathbf{H}_t^{-1/2}\mathbf{u}_t$. Notice that this model does contain a large number of parameters, 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 comovements 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

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

[3] 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 $\boldsymbol{\Omega} = (\mathbf{I} - \mathbf{A} - \mathbf{B})\mathbf{S}$ applies, where $\mathbf{S}$ is the sample covariance matrix.

it is clear that both cross-products of past return shocks and past variances and covariances will affect subsequent conditional variances and covariances of returns. On the opposite, by construction (4) prevents any potential *explicit* co-variation 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 set to match their unconditional counterparts, $\bar{\rho}_{ij}$. Of course, whilst the assumption of constant correlations is questionable, the main advantage of the constant correlation M-GARCH model is that the number of parameters drops, for instance to 34 only in the case of $m = 4$ and $p = 1$.

2.3 Dynamic conditional correlation

Engle (2002) has recently proposed the DCC class as a way to overcome the well-known over-parameterization (hence, estimation) problems that plague the M-GARCH class. A DCC model is based on the idea 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), $\mathbf{D}_t = \text{diag}\{\sqrt{h_{1t}}, \sqrt{h_{2t}}, \ldots, \sqrt{h_{mt}}\}$, and $\mathbf{R}_t$ is a time-varying matrix which collects ones on the main diagonal and time-varying correlations $\rho_{ij,t}$ off its main diagonal. In this paper we entertain three alternative DCC specifications. The first one is a simple DCC-EGARCH(1,1) specification:[4]

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

$$\mathbf{e}_i'\mathbf{D}_t^2\mathbf{e}_i = exp\left[\omega_i + \beta_i ln(h_{i,t-1}) + \alpha_i|\varepsilon_{i,t}| + \gamma_i\varepsilon_{i,t}\right],$$

[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.

$$q_{ij,t} = \delta_0 + \delta_1 \varepsilon_{i,t-1}\varepsilon_{j,t-1} + \delta_2 q_{ij,t-1} \quad \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} \cdot q_{jj,t}}} \quad i \neq j. \tag{6}$$

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

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

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

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

where $\mathbf{C}$ is a full matrix. In this case variance risk comes from the ARCH-in mean component represented by $\mathbf{C} \cdot \text{diag}(\mathbf{D}_t)$ (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

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

which means that 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 $\mathbf{C}$. Equations (6) and (7) considerably reduce the number of parameters to be estimated, e.g., to 44 parameters in the DCC-EGARCH case.

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

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

i.e., both conditional variances and conditional covariances follow driftless integrated moving average processes. 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.[5]

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:

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

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

where $\exp(\mathbf{R}_{b+1}) \equiv \left[\exp\left(R_{1,b+1}\right) \exp\left(R_{2,b+1}\right) \ldots \exp\left(R_{m,b+1}\right)\right]'$ 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 special case is the buy-and-hold framework, $\varphi = T$. Under this assumption and similarly to Barberis (2000), the integral defining expected utility 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.[6] Guidolin and Nicodano (2006) provide details on the numerical techniques employed in the solutions.

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, $J\left(W_t, \mathbf{r}_t, \hat{\boldsymbol{\pi}}_t; \hat{\boldsymbol{\omega}}_t^R\right) \leq J(W_t, \mathbf{r}_t, \hat{\boldsymbol{\pi}}_t; \hat{\boldsymbol{\omega}}_t)$, i.e., the restrictions reduce the derived utility from wealth. The compensatory premium, λ_t^R, is computed as:

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

[6] 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.

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.

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.[7] 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. 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.

Table 1 reports 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% per year in the case of North American large caps to 13.6% 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.[8] 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

[7] While the MSCI Europe Benchmark index targets large capitalization firms, no equivalent for North America 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.

[8] 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.

Table 1 Summary statistics for international stock returns

Portfolio	Mean	Median	SD	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*

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 jth order Ljung-Box statistic

*5% significance

**Significance at 1%

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 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 possible conditional heteroskedastic patterns. Finally, we compute correlation coefficients. Pacific stock returns have lower correlations (0.45–0.55) with other portfolios than all other pairs. 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 comovements.

4.2 Model selection

We use two 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 model specification, we make an extensive effort. We estimate a large number of variants of (1) and of models in the multivariate ARCH family and use four in-sample criteria:

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) has derived an upper bound for the significance level of the LR test under nuisance parameters.

2–4. Three standard information criteria, i.e., the Akaike (AIC), Bayes–Schwartz (BIC), and Hannan–Quinn (H-Q) criteria. These statistics trade-off in-sample fit with prediction accuracy and rely on the principle that a well-specified model should not only provide an accurate in-sample fit, but also prove useful to forecast. 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.

Table 2 reports the outcomes of these model selection/specification tests.[9] 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. Therefore we proceed and inspect the additional columns of the table. When appropriate, the fourth column of Table 2 systematically tests the null of $k = 1$ against $k > 1$ and reports p-values calculated under Davies' upper bound. Obviously, even adjusting for the presence of nuisance parameters, the evidence against specifying single-state models is overwhelming: the smallest LR statistic takes a value of 162, which is clearly above any conceivable critical value. This gives a first, crucial implication: the data offer strong evidence of time-variation in 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 regime switching VARs, i.e., only the first two rows of Table 2. We therefore compare the performance of multivariate ARCH and regime switching VARs. On the one hand, within the regime 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 parsimonious and economically interesting. Columns 5–7 of Table 2 show that some tension exists

[9] In Table 2, 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.

Table 2 Model selection: information criteria

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**

Table 2 continued

Model	Number of parameters	Log-likelihood	LR test for linearity	AIC	BIC	Hannan–Quinn
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

The table reports model selection criteria for a range of multivariate regime switching and autoregressive conditional heteroskedasticity models. The regime switching models have structure: $r_t = \boldsymbol{\mu}_{s_t} + \sum_{j=1}^{p} A_{j,s_t} r_{t-1} + \boldsymbol{\varepsilon}_t$ 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) among different criteria. The BIC is minimized by a tightly parameterized three-state heteroskedastic model with $p = 0$ with 48 parameters. However, this is less than surprising as the BIC is generally known to select relatively small models in non-linear 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 Sect. 6 entertains the possibility that the data may be best described by a model in which $p = 0$ (see Guidolin and Nicodano 2006, for related evidence), in this section and the following we take this three-state time-invariant VAR(1) model as our baseline framework.

Rows 3–8 of Table 2 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). 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).

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 regime 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 regime switching framework implies a need to estimate 20 additional parameters. However, it is obvious that 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 mainly the ability of an econometric framework to produce accurate forecasts of the entire joint density of equity returns.[10] 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 (z_t). This is the probability of observing a value smaller than or equal to a given value $\tilde{\mathbf{r}}_{t+1}$ under the null that the model is correctly specified. If the model is correctly specified, z_t should be independently and identically distributed (IID) and uniform on the interval [0, 1]. 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: $z_t^* \equiv \Phi^{-1}(z_t)$. Provided that the model is correctly specified, z^* should be IID and normally distributed (IIN(0, 1)). In our tests, we use likelihood ratio tests that focus on a few salient moments of the return distribution. Guidolin and Nicodano (2006) provides details on this econometric tool.

[10] 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.

Table 3 Model selection: tests based on one-step ahead density forecasts

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*

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

*Significance at the 5% level

**Significance at the 1% level

Table 3 reports Berkowitz-style, transformed z-tests 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; the three-state heteroskedastic regime 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). 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 to our weekly stock return data would provide misleading inferences on optimal portfolio weights and therefore on the 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 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 lead. 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. All in all, we take these results 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.

4.3 Econometric estimates

Table 4, 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: e.g., a one standard deviation shock to European small caps forecasts an increasing 1-week ahead returns of 0.37% only.

Panel B of Table 4 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 regime-specific weekly mean returns in each of the regimes, using the formula $E[\mathbf{r}_t|S_t = i] = (\mathbf{I}_m - \mathbf{A})^{-1}\boldsymbol{\mu}_i$ (Table 5).

Regime 1 is clearly a crash state in which all international equity markets face large losses. However, notice that this characterization of regime 1 is not incompatible with an equilibrium interpretation (and common sense) because this state has low persistence, i.e., with probability in excess of 0.60 markets leave the crash state between t and $t+1$ to move to positive expected return states.[11] As a result, the

[11] We compute $E[\mathbf{r}_{t+4}|S_t = i, \mathbf{r}_t] = \sum_{i=1}^{3} \hat{\pi}^i_{t+4,t} \left(\boldsymbol{\mu}_i + \mathbf{A}E[\mathbf{r}_t|S_t = i]\right)$ to show that 1-month expected returns are positive:

	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

Table 4 Estimates of a three-state (time-invariant) VAR(1) regime switching model

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				
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***

Table 4 continued

Panel B—single state model	Pacific	Europe—small caps	Europe—large caps	North America large
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***	

The table shows estimation results for the regime switching model: $r_t = \boldsymbol{\mu}_{s_t} + Ar_{t-1} + \boldsymbol{\varepsilon}_t$ 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

*10% significance

**Significance at 5%

***Significance at 1%

Table 5 Regime-specific weekly mean returns in each of the three regimes

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

crash state has a negligible duration of less than 2 weeks, which fits the idea that extreme market crashes 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 are generated by this regime. Interestingly, when the world equity markets leave the crash state, 46% of the time this is to regime 3, where expected returns are high. In the crash regime, volatilities are high, especially in the case of European stocks, both small and large. Also correlations are high. Figure 1 gives a visual representation in terms of smoothed (full sample) probabilities which fits this interpretation, in the sense that this state appears relatively often but lasts at most for 3 weeks. Although “crashing weeks” seem to appear 2–3 times a year on average (e.g., the week of September 11, 2001), 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. Its average duration is 12 weeks, i.e., it fits periods in which the markets are hardly moving in any direction (and excess returns are low). This state is characterized by intermediate volatility, although the estimated values exceed their unconditional counterparts for three indices out of four. Consistently with the literature, 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 period 2001–2002 are captured by this state. 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.

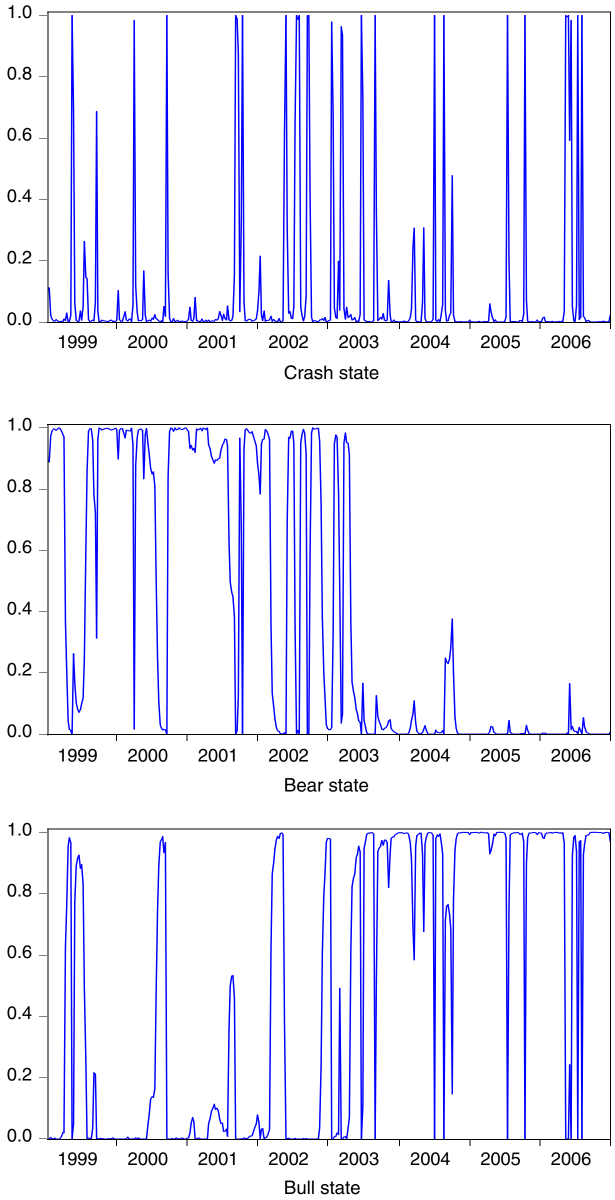

Fig. 1 Smoothed state probabilities from three-state VAR(1) regime switching model. The *graphs* plot the smoothed state probabilities for the multivariate three-state VAR(1) Switching model comprising weekly return series on four international equity portfolios (including European small cap firms).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

The bull state is characterized by low volatility and relatively reduced correlations. To complement to what could be seen for the bear regime, Fig. 1 shows that most of 2000 and then 2004–2006 are captured by this regime. Even though their persistence is similar, the 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.[12]

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. 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.[13] We then proceed in Sect. 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 Sect. 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 4. 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),

[12] We also perform a few diagnostic checks that exploit the fact that under the null of correct model specification, the one-step ahead forecast errors $\{\boldsymbol{\eta}_{t+1}\}$ should form a martingale difference sequence, $E[\boldsymbol{\eta}_{t+1}|\mathcal{F}_t] = 0$. We find very weak evidence of structure in the forecast errors (and their squares) generated by a three-state regime switching model. We also formally test a regime switching ARCH(1) specification in which $\boldsymbol{\Sigma}_t = \boldsymbol{\Omega}_{S_t} + \mathbf{A}_{S_t}\mathbf{u}_t\mathbf{u}_t'\mathbf{A}_{S_t}$. This specification implies 16 additional parameters, the elements of the matrix $\mathbf{A}_{1s_t}$. An LR test resoundingly rejects this model.

[13] We find substantial demand for North American large and Asian Pacific stocks (55 and 45% at a 2-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).

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 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.[14] Importantly, this last set of results corresponds to the case in which variance risk is disregarded altogether.

The demand for EUSC in Fig. 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 4 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%. 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%) and North American large stocks (60%).

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 1 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%).[15]

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

[14] 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.

[15] 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.

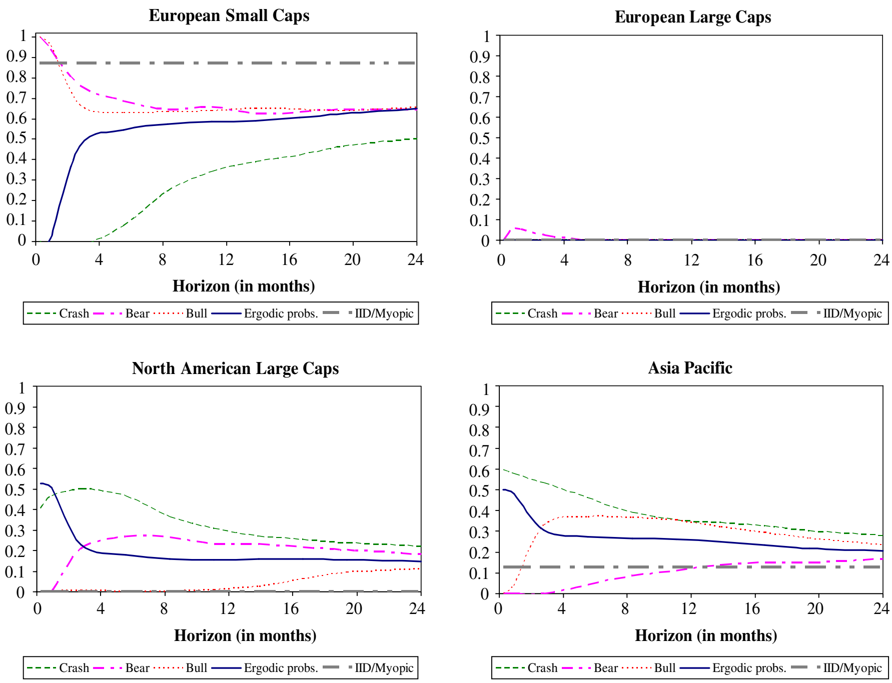

Fig. 2 Buy-and-hold optimal allocation. The *graphs* plot the optimal international equity portfolio weights when returns follow a three-state VAR(1) switching model as a function of the investment horizon. The vector autoregressive coefficients are constant over time. As a benchmark (*horizontal lines*) we also report the IID/Myopic allocation. Optimal portfolio weights are computed assuming constant relative risk aversion preferences with $\Lambda = 5$

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, even investors with relatively long horizons of 2 years express a moderate demand of EUSC, around 30–35% 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.

We also 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%).[16] Also in this case, ignoring variance risk would assign way too high a weight to EUSC, in

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

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 6 and 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 Sect. 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 $j = l$.[17] When $i = j = l$ we shall be looking at the standard own-skewness coefficient of some portfolio return. In Table 6, bold coefficients highlight point estimates' significance at standard levels (5%). There is a remarkable correspondence between signs and magnitudes of co-skewness coefficients in the data and the unconditional

[17] 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.

Table 6 Sample and implied co-skewness coefficients

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

The table reports the sample co-skewness coefficients,

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

and compares them with the coefficients implied by a three-state VAR(1) regime switching model. 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). Bold coefficients are significantly different from zero

estimates under our estimated regime switching model, in the sense that when the coefficient is statistically significant in the data, it is always so also under the regime 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.[18] Furthermore, notice that the co-skewness coefficients $S_{\text{EUSC,EUSC},j}$ 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., $S_{\text{EU\_ large,EU\_large},j}$ for varying js) are close to zero and sometimes positive. Worse, all of the $S_{\text{EUSC},j,j}$ 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

[18] 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.

Table 7 Co-skewness and co-kurtosis coefficients vs. an equally weighted international portfolio

	Co-skewness		Co-kurtosis		Co-skewness		Co-kurtosis	
	Sample	MS-ergodic	Sample	MS-ergodic	Sample	MS-ergodic	Sample	MS-ergodic
	European Small Caps				North American Large Caps			
$S_{EU\_small,EU\_small,EW\_ptf}$	**−1.266**	**−0.499**	–	–	0.087	−0.171	–	–
$S_{EU\_small,EW\_ptf,EW\_ptf}$	**−0.913**	**−0.628**	–	–	−0.297	−0.28	–	–
$S_{EU\_small,EU\_small,Pac,EW\_ptf}$	–	–	**4.174**	**3.298**	–	–	3.245	**3.844**
$S_{EU\_small,EU\_small,NA,EW\_ptf}$	–	–	3.039	2.944	–	–	1.914	2.046
$S_{EU\_small,EU\_small,EU\_large,EW\_ptf}$	–	–	**4.171**	2.172	–	–	2.456	**3.864**
$S_{EW\_ptf,EW\_ptf,EU\_small,Pac}$	–	–	**3.304**	**3.789**	–	–	2.13	2.045
$S_{EW\_ptf,EW\_ptf,EU\_small,NA}$	–	–	**2.977**	**3.666**	–	–	**3.273**	**4.233**
$S_{EW\_ptf,EW\_ptf,EU\_small,EU\_large}$	–	–	**5.945**	**4.645**	–	–	2.977	**3.666**
$S_{EW\_ptf,EW\_ptf,EU\_small,EU\_small}$	–	–	**5.697**	**4.675**	–	–	**3.479**	3.233
$S_{EW\_ptf,EW\_ptf,EU\_ptf,EU\_small}$	–	–	**8.234**	**5.71**	–	–	**3.857**	**4.844**
$S_{EU\_small,EU\_small,EU\_small,EU\_ptf}$	–	–	**4.937**	**4.157**	–	–	**3.682**	**3.49**

The table reports average sample co-skewness coefficients computed vs. an international portfolio and compares them with the coefficients implied by a three-state VAR(1) switching model. 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 become volatile. Therefore EUSC does display considerable variance risk. On the top of variance risk, from Tables 1 and 6 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 Fig. 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, $S_{\mathrm{i,EW\_ptf,EW\_ptf}}$ for the generic portfolio i has expression

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

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 $S_{\mathrm{EUSC,EUSC,EW\_ptf}} = -0.50$ and $S_{\mathrm{EUSC,EW\_ptf,EW\_ptf}} = -0.63$, i.e., 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. 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 $K_{\mathrm{EUSC,EUSC},j,j}$ 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

$K_{\text{EUSC,EUSC,EW\_ptf,EW\_pft}}$ is 4.7, which is one of the highest among these types of coefficients. $K_{\text{EUSC,EUSC,EW\_ptf,EW\_pft}}$ 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 Decomposing variance risk

The actual relative importance of the factors underlying variance risk for optimal asset allocation can be directly assessed through the computation of portfolio weights when all but one of the sources of variance risk are left in operation. We start with the case in which co-skewness effects are shut off so that only co-kurtosis may affect portfolio decisions. This can be obtained by calculating weights under a special regime switching model in which $\boldsymbol{\mu}$ is made independent of the state S_t, while $\boldsymbol{\Sigma}_{s_t}$ remains a function of the regime, i.e.,

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{A}\mathbf{r}_{t-1} + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{s_t}). \tag{13}$$

This model aims at capturing pure heteroskedasticity effects since expected stock returns are constant over time. Equation (13) implies that departures from multivariate IID normality come from *even* co-higher moments. We estimate (13) and calculate optimal weights for an investor with $\gamma = 5$. Insofar as volatilities and correlations are concerned (unreported) parameter estimates are close to those in Table 4. Detailed results are in Guidolin and Nicodano (2006), in the form of buy-and-hold optimal weights for alternative investment horizons. However, the qualitative thrust of the results is easily summarized, especially by difference to the results in Fig. 2, for the unrestricted model. The results suggest that odd high-order co-moments (co-skewness) have first-order effects in reducing the portfolio role of EUSC. In fact, under model (13) the ergodic optimal portfolio weights (89% in EUSC and the reminder in Pacific stocks) are essentially the same as those obtainable under the false assumption of a Gaussian IID model with no regimes. Otherwise, EUSC remain very important when the investor is given information on the current, initial state being of a bear or bull type, but in the crash state their weight differs dramatically from Fig. 2, being higher by a spread of 30–40%.

5.4 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 Sect. 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

Table 8 Utility cost of excluding small caps from the asset menu

	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

EUSC would be of a first-order magnitude for all investment horizons, as shown in Table 8.

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

6.1 The return generating process

Section 4.2 has left us with two potential questions. First, in Table 2 the information criteria gave conflicting indications as to whether a VAR(1) component was needed within a three-state switching framework. As a result, we have also estimated a simpler, three-state regime switching model with $p = 0$ (as in Guidolin and Nicodano 2006) and proceeded to compute optimal weights and welfare loss estimates. Optimal portfolio weights display the same patterns commented in Sect. 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 weights shows that the differences vs. the ones plotted in Fig. 2 are negligible. Therefore, it is hard to think that details of the regime switching process selected in this paper may entirely drive our results on the effects of variance risk on the demand of EUSC.

Second, even if Table 3 has provided evidence on the properties of the regime switching model, at least one of the multivariate ARCH—the DCC EGARCH(1,1)—has come close to provide low information criteria and good out-of-sample predictive accuracy. We have therefore computed afresh optimal weights when the joint process of returns is described by the DCC EGARCH(1,1) VAR(1). To save space, in Table 9 we report results for the optimal shares of EUSC when $\gamma = 5$ (for comparison, also results for the other models are shown):

Table 9 Optimal portfolio weights to European small caps under alternative econometric models

Weight to EUSC	Investment horizon (in weeks)			
	$T = 1$	$T = 12$	$T = 24$	$T = 104$
Three-state RS 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

Clearly, even if we had opted in favor of the DCC EGARCH model in Sect. 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 Fig. 2, i.e., at short horizons the most important portfolios are Asia Pacific and North America large, while at a 2-year horizon EUSC plays a major role.

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., $m = 5$. 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 4: 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% 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 indistinguishable from the one in Table 4, to the point that most estimates of the transition probabilities do fall in a 95% confidence interval around the estimates obtained in Table 4. 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 Fig. 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.

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.

Acknowledgments 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.

References

Acharya, V., Pedersen, L.: Asset pricing with liquidity risk. J Financ Econ **77**, 375–410 (2005)

Ang, A., Bekaert, G.: International asset allocation with regime shifts. Rev Financ Stud **15**, 1137–1187 (2002)

Ang, A., Chen, J.: Asymmetric correlations of equity portfolios. J Financ Econ **63**, 443–494 (2002)

Barberis, N.: Investing for the long run when returns are predictable. J Financ **55**, 225–264 (2000)

Barone-Adesi, G., Gagliardini, P., Urga, G.: Testing asset pricing anomalies with co-skewness. J Business Econ Stat **22**, 474–485 (2004)
Berkowitz, J.: Testing density forecasts with applications to risk management. J Bus Econ Stat **19**, 465–474 (2001)
Black, F.: Studies of stock market changes, proceedings of the american statistical association. In: Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section. Alexandria: American Statistical Association (1976)
Bollerslev, T.: Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. Rev Econ Stat **72**, 498–505 (1990)
Campbell, J., Chan, Y.L., Viceira, L.: A multivariate model of strategic asset allocation. J Financ Econ **67**, 41–80 (2003)
Cvitanić, J., Polimenis, V., Zapatero, F.: Optimal portfolio allocation with higher moments. Ann Financ (2007, forthcoming)
Das, S., Uppal, R.: Systemic risk and international portfolio choice. J Financ **59**, 2809–2834 (2004)
Davies, R.: Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **64**, 247–254 (1977)
De Santis, G., Gerard, B.: International asset pricing and portfolio diversification with time-varying risk. J Financ **52**, 1881–1912 (1997)
Diebold, F., Gunther, T., Tay, A.: Evaluating density forecasts. Int Econ Rev **39**, 863–883 (1998)
Dittmar, R.: Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. J Financ **57**, 369–403 (2002)
Engle, R.: Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J Business Econ Stat **20**, 339–350 (2002)
Engle, R., Kroner, K.: Multivariate simultaneous GARCH. Econom Theory **11**, 122–150 (1995)
Engle, R., Mezrich, J.: GARCH for groups. Risk **9**, 36–40 (1996)
Fenton, V., Gallant, R.: Qualitative and asymptotic performance of SNP density estimation. J Econom **74**, 77–118 (1996)
Garcia, R.: Asymptotic null distribution of the likelihood ratio test in Markov switching models. Int Econ Rev **39**, 763–788 (1998)
Guidolin, M., Nicodano, G.: Small caps in international equity portfolios: the effects of variance risk, Federal Reserve Bank of St. Louis Working Paper 2005-075A (2006)
Guidolin, M., Timmermann, A.: Size and value anomalies under regime switching. J Financ Econom (2008, forthcoming)
Guidolin, M., Timmermann, A.: International asset allocation under regime switching, skew and kurtosis preferences. Rev Financ Stud (2008, forthcoming)
Hamilton, J.: Time series analysis. Princeton University Press, Princeton (1994)
Harvey, C., Siddique, A.: Conditional skewness in asset pricing tests. J Financ **55**, 1263–1295 (2000)
Liu, J., Longstaff, F., Pan, J.: Dynamic asset allocation with event risk. J Financ **58**, 231–259 (2003)
Newey, W., McFadden, D.: Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (eds.) Handbook of Econometrics, vol. 4, pp. 2113–2245. New York: Elsevier (1994)
Pástor, L.: Portfolio selection and asset pricing models. J Finance **55**, 179–223 (2000)
Perez-Quiros, G., Timmermann, A.: Firm size and cyclical variations in stock returns. J Financ **55**, 1229–1262 (2000)