Regime Shifts in Mean-Variance Efficient Frontiers: Some International Evidence

Authors	Massimo Guidolin, and Federica Ria
Working Paper Number	2010-040A
Creation Date	November 2010
Citable Link	https://doi.org/10.20955/wp.2010.040
Suggested Citation	Guidolin, M., Ria, F., 2010; Regime Shifts in Mean-Variance Efficient Frontiers: Some International Evidence, Federal Reserve Bank of St. Louis Working Paper 2010-040. URL https://doi.org/10.20955/wp.2010.040

Regime Shifts in Mean-Variance Efficient Frontiers: Some International Evidence

Massimo Guidolin*
Federal Reserve Bank of St. Louis and Manchester Business School

Federica Ria
Manchester Business School and Katarsis Capital Advisors S.a.

October 31, 2010

Abstract

Regime switching models have been assuming a central role in financial applications because of their well-known ability to capture the presence of rich non-linear patterns in the joint distribution of asset returns. This paper examines how the presence of regimes in means, variances, and correlations of asset returns translates into explicit dynamics of the Markowitz mean-variance frontier. In particular, the paper shows both theoretically and through an application to international equity portfolio diversification that substantial differences exist between bull and bear regime-specific frontiers, both in statistical and in economic terms. Using Morgan Stanley Capital International (MSCI) investable indices for five countries/macro-regions, it is possible to characterize the mean-variance frontiers and optimal portfolio strategies in bull periods, in bear periods, and in periods where high uncertainty exists on the nature of the current regime. A recursive back-testing exercise shows that between 1998 and 2010, adopting a switching mean-variance strategy may have yielded considerable risk-adjusted payoffs, which are the largest in correspondence to the 2007-2009 financial crisis.

JEL codes: C53, G12, C32..

Keywords: Multivariate Markov Switching, Mean-Variance Optimization, Asset Allocation, International Portfolio Diversification.

1. **Introduction**

Do investors rationally perceive a different trade-off between risk and (expected) returns during periods of financial crisis? If so, how do these perceptions affect optimal diversification even in times in which market conditions do not warrant the suspicion of an incipient crisis, especially when investors are characterized by long investment horizons? Can we find any models in the standard toolbox of financial econometricians that would be able to capture these recurrent switches between good (bull) and bear times and allow us to estimate the probabilities of such regime shifts occurring? One class of models that has gained growing attention in the financial econometrics and asset pricing literatures relies on multivariate extensions of the

*Correspondence to: Massimo Guidolin, Federal Reserve Bank of St. Louis, Research Division – St. Louis, MO 63166, USA. Phone: 314-444-8550, Fax: 314-444-8731. E-mail: Massimo.Guidolin@stls.frb.org; Massimo.Guidolin@mbs.ac.uk.

seminal work by Hamilton (1989) in macroeconomics and by Turner et al. (1989) in financial economics on the presence of Markov regimes in many important time series, including asset returns. This paper provides a short primer to the structure, estimation issues, and potential applications of multivariate Markov switching models. An illustrative application to an international equity portfolio diversification problem is provided with reference to standard Morgan Stanley Capital International (MSCI) indices.

Although our brief review of methods for and technical issues with multivariate Markov switching models competes with a number of alternative reference articles and books, our paper also addresses one question that has been receiving growing attention in the empirical finance literature:[1] How important and economically valuable can it be for an investor to use information on the current and predicted "state" of the capital markets when planning of her optimal diversification strategies? Clearly, if international equity returns are disconnected from the underlying economic regimes and as such tend to be display statistical properties (such as their moments, such as expected returns, variances, and covariances) that are not predictable over time, then the answer is trivial: because such states either fail to exist or they are irrelevant to the risk-reward opportunities offered by the equity markets, a rational investor may safely ignore the issue. This means that the now classical, Markowitz-style mean-variance recipes offered in most textbooks (such as Fabozzi, Focardi, and Kolm, 2006) would be correct and the investor ought to chose optimal portfolio weights on the basis of simple, naive historical estimates of means, variance, and covariances. If on the contrary, such regimes exist and may be identified, estimated, and predicted, then it is an open question whether an investor should take notice of them, and go through the relatively sophisticated econometric techniques – the subject of Section 2 in this paper – required by her acknowledging this state-dependence.

On the one hand, it is now well known that applications of classical mean-variance frontier (MVF) technology to dynamic asset allocation problems in which the MVF is allowed to depend on one more variables capturing the state of market investment opportunities, suffer from a number of issues (e.g., see Schöttle and Werner, 2006). For instance, the shape of the MVF together with the location of the efficient portfolios has been observed to change drastically as market data are progressively updated and expanded. Moreover, it is typical to observe that MVFs often occupy rather unrealistic regions of the mean-standard deviation space as a result of optimization based on error-prone estimates, resulting in large deviations between the ex-ante, in-sample and the ex-post, out-of-sample Sharpe ratios. For these reasons, a literature has been developing on the robustification of the MVF concept. Among many others, resampling approaches have been proposed that essentially bootstraps from the data the location and properties of the MVF (see e.g., Scherer, 2002). A number of papers have also stressed the advantages of using robust estimation and/or programming methods when computing the MVF (see e.g., Tütüncü and Koenig, 2004). One can see our paper as an attempt to produce more robust estimates of the MVF and hence—after appropriate mean-variance preferences have been assumed—more robust optimal portfolios not by changing methods of estimation or by resampling the data, but instead by exploring the implications of a simple and yet powerful parametric approach that explicitly tracks the time variation in the features

[1] For instance, a Reader is invited to consult Fruhwirth-Schnatter (2006), Hamilton (1994), Kim and Nelson (1999), or Krolzig (1997).

of the investment opportunity sets (means, variances, and correlations) as depending from a latent Markov state variable.

On the other hand, the issue of whether non-linear predictability patterns may be exploitable for dynamic asset allocation purposes is an important and increasingly researched one. There is an early applied portfolio management literature that has noticed that making portfolio choices state-dependent may deliver important *ex ante* performance improvements. For instance, Clarke and de Silva (1998) note that no static mix to be applied to standard mean-variance portfolios can be used to achieve a point along a state-dependent efficient frontier. The more efficient and desirable risk-reward combinations on the state-dependent frontier may be achieved only by systematically altering portfolio allocations in response to changes in the investment opportunities as the economy switches back and forth among different states.[2] Chow, Jacquier, Kritzman, and Lowry (1999) implicitly introduce the idea that even scarcely persistent, infrequent regimes may play an important role in portfolio choice, especially when it comes to estimate risk. They think about the issue by distinguishing between *time-measured* observations from *event-measured* observations. They notice that in some periods, the absence of any significant events will bring stock returns to reflect noise only, while in other periods there may be a number of discrete events that impress predictable patterns to realized stock returns. They propose to estimate risk parameters not from time-measured, but instead from event-measured data and conjecture that such an alternative approach may provide a better representation of the density from which portfolio returns may be drawn during turbulent market periods. In practice, their approach boils down to allow the investor to estimate and use a regime switching model with two separate covariance matrices, one for the quiet regime and the other for the turbulent one, although the probabilities are then not inferred from the data (as we do in this paper) but on the opposite simply "assigned". Another paper that carries out an interesting international portfolio choice application in some respects similar to ours, and that also resorts to some notion of "regime" (although only in an informal sense) is Butler and Joaquin (2002) who characterize the consequences of asymmetric correlations in international bear and bull markets to show that risk averse investor may want to tilt portfolio weights away from stock markets characterized by the highest correlations during downturns.

More recently, the literature has shifted towards writing and solving portfolio problems under the assumption of statistical frameworks in which asset returns follow a switching dynamics. Ramchand and Susmel (1998) examine the relationship between correlation and variance in a regime-switching ARCH model estimated on weekly stock returns data for the US and a few other major markets. They find that correlations between US and other world markers are 2 to 3.5 times higher when the US market is in a high variance state. They also calculate mean-variance portfolios and find that their switching framework leads to high Sharpe ratios. Ang and Bekaert (2002a) consider bivariate and tri-variate regime switching models that capture asymmetric correlations in volatile and stable markets and characterize a US investor's optimal asset allocation under constant relative risk aversion. Das and Uppal (2004) study

[2] The reason is that in the presence of state-dependence (say, when two states with probabilities $\bar{\pi}$ and $1-\bar{\pi}$ are possible), a mixture of Gaussian (more generally, elliptical) densities is never the same as a Gaussian density that has means and variances which are probability-weighted (with weights $\bar{\pi}$ and $1-\bar{\pi}$) averages of the state-dependent means and variances.

the effects of infrequent price changes on international equity portfolios. Equity returns are generated by a multivariate jump diffusion process where jumps are simultaneous and perfectly correlated across assets. Guidolin and Timmermann (2008b) use an international portfolio diversification application to propose a new tractable approach to solving asset allocation problems under Markov switching with a large number of assets. Investor preferences are assumed to be defined over moments of the wealth distribution such as its mean, variance, skew and kurtosis. They develop analytical methods that only require solving a small set of difference equations and can be applied even in the presence of large numbers of risky assets. Guidolin and Hyde (2008) apply these methods to investigate the time-varying linkages among the Irish stock market, one of the top world performers of the 1990s, and the US and UK stock markets. They also find that two regimes, characterized as bear and bull states, are required to characterize the dynamics of excess equity returns both at the univariate and multivariate level. This implies that the regimes driving the small open economy stock market are largely synchronous with those typical of the major markets. However, despite the existence of a persistent bull state in which the correlations among Irish and UK and US excess returns are low, they find that state co-movements involving the three markets are so relevant to reduce the optimal mean-variance weight carried by Irish stocks to at most one-quarter of the overall equity portfolio. Guidolin and Nicodano (2009) is a related application that shares with the current paper similar data and objectives, although their focus is mostly on the effects of higher-order moments (co-skewness and co-kurtosis with returns on the world market portfolio) on optimal international diversification decisions.

The rest of the paper is organized in the following way. Section 2 provides a brief introduction to the econometrics of multivariate Markov regime switching models. Section 3 describes our application to international portfolio diversification and reports a number of empirical results that we take – in the light of the literature cited above – to be representative of the typical results one may find when approaching and solving asset allocation problems with equity return data. Section 4 concludes.

2. Multivariate Markov Switching Models

Suppose that the $n \times 1$ random vector $\mathbf{y}_t$ follows a k-regime Markov switching (MS) $VAR(p)$ process with heteroskedastic component, compactly $MSIAH(k,p)$:

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t}\mathbf{y}_{t-j} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t \tag{1}$$

with $\boldsymbol{\epsilon}_t \sim NID(\mathbf{0}, \mathbf{I}_n)$. S_t is a latent state variable driving all the matrices of parameters appearing in (1). $\boldsymbol{\mu}_{S_t}$ is a $n \times 1$ vector that collects the n regime-dependent intercepts, while the $n \times n$ matrix $\boldsymbol{\Sigma}_{S_t}$ represents the factor applicable to state S_t in a state-dependent Choleski factorization of the variance covariance matrix of the variables of interest, $\boldsymbol{\Omega}_{S_t}$. Obviously, a non-diagonal $\boldsymbol{\Sigma}_{S_t}$ makes the n variables simultaneously cross-correlated. For instance, in Guidolin and Ono (2006) and Guidolin and Timmermann (2007) n is broken down in n_1 asset returns and n_2 macroeconomic predictors, with $n_1 + n_2 = n$. Then a non-diagonal $\boldsymbol{\Sigma}_{S_t}$ captures *simultaneous* co-movements between asset returns and macro factors, while *dynamic* (lagged) linkages across both different asset markets and between financial markets and macroeconomic influences are captured by the VAR(p) component. We assume the absence of roots outside the unit circle, thus making the process stationary.[3] In fact, conditionally on the unobservable state S_t, (1) defines a standard Gaussian reduced form VAR(p) model. On the other hand, when $k > 1$, alternative hidden states are possible and they will influence both the conditional mean and the volatility/correlation structures characterizing the multivariate process in (1), $S_t = 1, 2, ..., k$ $\forall t$. These unobservable states are generate by a discrete-state, homogeneous, irreducible and ergodic first-order Markov chain:[4]

$$\Pr(S_t = j | \{S_j\}_{j=1}^{t-1}, \{\mathbf{y}_j\}_{j=1}^{t-1}) = \Pr(S_t = j | S_{t-1} = i) = p_{ij}, \tag{2}$$

where p_{ij} is the generic $[i, j]$ element of the $k \times k$ transition matrix $\mathbf{P}$. Ergodicity implies the existence of a stationary vector of probabilities $\bar{\boldsymbol{\xi}}$ satisfying $\bar{\boldsymbol{\xi}} = \mathbf{P}'\bar{\boldsymbol{\xi}}$. Irreducibility implies that $\bar{\boldsymbol{\xi}} > \mathbf{0}$, meaning that all unobservable states are possible. In practice, $\mathbf{P}$ is unknown and hence $\bar{\boldsymbol{\xi}}$ can be at most estimated given knowledge on $\mathbf{P}$ extracted from the information set $\Im_t = \{\mathbf{y}_j\}_{j=1}^{t}$. For simplicity, we will also denote as $\bar{\boldsymbol{\xi}}$ such an estimated vector of ergodic (unconditional) state probabilities.

When n is large, model (1) implies the estimation of a large number of parameters, $k[n + pn^2 + n(n+1)/2 + (k-1)]$. For instance, for $k = 2$, $n = 8$, and $p = 1$ (the parameters characterizing the application in Guidolin and Ono, 2006), this implies the estimation of $2 \times [8 + 8^2 + 4 \times 9 + 1] = 218$ parameters![5] (1) nests a number of simpler models in which either some of the parameter matrices are not needed or some of these matrices are independent of the regime. These simpler models may greatly reduce the number of parameters to be estimated. Among them, the financial econometrics literature (see e.g., Ang and Bekaert, 2002a, and Guidolin and Nicodano, 2009) has devoted special attention to $MSIH(k)$ models,

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t,$$

in which $p = 0$, to $MSIA(k, p)$ homoskedastic models,

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t}\mathbf{y}_{t-j} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t,$$

in which the covariance matrix is constant over time, and to $MSIH(k, 0)$-$VAR(p)$ models (see Guidolin and Ono, 2006),

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_j\mathbf{y}_{t-j} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t, \tag{3}$$

which are a special case of (1) in which while intercepts and covariance matrices are regime-dependent, the VAR(p) coefficients are not.[6] For instance, model (3) implies the estimation of 'only' $k[n + n(n+1)/2 +$

[3] Ang and Bekaert (2002) have shown that formally, it is just sufficient for such a condition to be verified in at least one of the k alternative regimes, for covariance stationarity to obtain. NID stands for 'normal and identically distributed'.

[4] The assumption of a first-order Markov process is not restrictive, since a higher order Markov chain can always be reparameterized as a higher dimensional first-order Markov chain, i.e., substitutability exists between the order of the Markov chain driving S_t and the number of regimes k. Diebold et al. (1994) have generalized Markov switching models to the case of time-varying transition probabilities. Foley (2001) is an application to financial data.

[5] This is the sense in which Marron and Wand (1992) conclude that mixtures of normal distributions provide a flexible family that can be used to approximate many distributions. Mixtures of normals can also be viewed as a nonparametric approach to modeling the return distribution if the number of states, k, is allowed to grow with the sample size.

[6] Of course, also $MSI(k)$ models in which $p = 0$ and the covariance matrix is constant over time – in short, where only the intercepts depend on the hidden state S_t and $\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t$ – are estimable, although they appear less frequently in the finance literature, given the ample evidence of conditional heteroskedastic asset returns.

$(k-1)]+pn^2$ parameters. For the same configuration mentioned above, this means $2\times[8+4\times 9+1]+8^2 = 154 < 218$. Of course, a limit case of (1) is obtained when $k=1$:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}_t. \tag{4}$$

This is a standard multivariate Gaussian VAR(p) model, a benchmark in a large portion of the existing empirical macroeconomics and finance literature.

Certain applications in the literature (e.g., the seminal paper by Hamilton, 1989) have also entertain the following variation on (1),

$$(\mathbf{y}_t - \boldsymbol{\nu}_{S_t}) = \sum_{j=1}^{p} \mathbf{A}_{j,S_t}(\mathbf{y}_{t-j} - \boldsymbol{\nu}_{S_{t-j}}) + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t. \tag{5}$$

Krolzig (1997) shows that the dynamic implications of (1) and (5) are markedly different. First of all, notice that the definition of (5) directly implies that the conditional mean function is now governed by a $(p+1)$-th order Markov chain as the terms $(\mathbf{y}_{t-j} - \boldsymbol{\nu}_{S_{t-j}})$ make the entire sequence $\{S_{t-p}$, S_{t-p+1}, ..., S_{t-1}, $S_t\}$ relevant.[7]

2.1. *Estimation and Inference*

The first step towards estimation and prediction of a MSIAH model is to put the model in state-space form. Collect the information on the time t realization of the Markov chain in a random vector

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

where $I(S_t = i)$ is a standard indicator variable. In practice the sample realizations of $\boldsymbol{\xi}_t$ will always consist of unit "versors" $\mathbf{e}_i$ characterized by a 1 in the i-th position and by zeros everywhere else. Another important property is that $E[\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}] = \mathbf{P}'\boldsymbol{\xi}_{t-1}$. The state-space form is composed of two equations:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{X}_t\boldsymbol{\Psi}\left(\boldsymbol{\xi}_t \otimes \boldsymbol{\iota}_n\right) + \boldsymbol{\Sigma}^*\left(\boldsymbol{\xi}_t \otimes \mathbf{I}_n\right)\boldsymbol{\epsilon}_t \qquad &&\text{(measurement equation)} \\ \boldsymbol{\xi}_{t+1} &= \mathbf{F}\boldsymbol{\xi}_t + \mathbf{u}_{t+1} &&\text{(transition equation)} \end{aligned} \tag{6}$$

where $\mathbf{X}_t$ is a $n\times(np+1)$ vector of predetermined variables with structure $[1\ \mathbf{y}'_{t-1}...\mathbf{y}'_{t-p}]\otimes\boldsymbol{\iota}_n$, $\boldsymbol{\Psi}$ is a $(np+1)\times nk$ matrix collecting the VAR parameters, both matrices of means and autoregressive coefficients,

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\mu}_1' & \cdots & \boldsymbol{\mu}_k' \\ \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{p1} & \cdots & \mathbf{A}_{pk} \end{bmatrix},$$

$\boldsymbol{\Sigma}^*$ is a $n\times nk$ matrix collecting all the possible k "square root" (Choleski decomposition) factors $[\boldsymbol{\Sigma}_1\ \boldsymbol{\Sigma}_2\ \ldots\ \boldsymbol{\Sigma}_k]$ such that $\forall t$, $\boldsymbol{\Sigma}^*\left(\boldsymbol{\xi}_t\otimes\mathbf{I}_n\right)\left(\boldsymbol{\xi}_t\otimes\mathbf{I}_n\right)'\left(\boldsymbol{\Sigma}^*\right)' = \boldsymbol{\Omega}_{S_t}$, the S_t-regime covariance matrix of the asset

[7] Since a (5) specification implies an actual of number of regimes much higher than k, (5) is advised only when there are solid theoretical reasons for such a model to be investigated.

return innovations $\boldsymbol{\epsilon}_t$. Moreover, $\boldsymbol{\epsilon}_t \sim NID(\mathbf{0}, \mathbf{I}_n)$, and in the transition equation $\mathbf{u}_{t+1}$ is a zero-mean discrete random vector that can be shown to be a martingale difference sequence. Also, the elements of $\mathbf{u}_{t+1}$ are uncorrelated with $\boldsymbol{\epsilon}_{t+1}$ as well as $\boldsymbol{\xi}_{t-j}$, $\boldsymbol{\epsilon}_{t-j}$, $\mathbf{y}_{t-j}$, and $\mathbf{X}_{t-j}$ $\forall j \geq 0$. To operationalize the dynamic state-space system (6), assume that the multivariate process (1) started with a random draw from the unconditional probability distribution defined by the vector of state probabilities $\bar{\boldsymbol{\xi}}$. Finally, from the definition of transition probability matrix (2) it follows that since $E[\mathbf{u}_{t+1}|\boldsymbol{\xi}_t] = \mathbf{0}$ by assumption, then

$$E[\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t] = \mathbf{F}\boldsymbol{\xi}_t$$

implies that $\mathbf{F}$ corresponds to the transposed transition probability matrix $\mathbf{P}'$.[8]

The state-space representation of (5) is quite different. As already observed, the conditional mean is now governed by a $(p+1)$-th order Markov chain, so that it is now useful to collect the information on the realization of the Markov chain in a $k^{p+1} \times 1$ random vector

$$\boldsymbol{\xi}_t^{(p+1)} = \boldsymbol{\xi}_t \otimes \boldsymbol{\xi}_{t-1} \otimes ... \otimes \boldsymbol{\xi}_{t-p} = \begin{bmatrix} I(S_t = 1, S_{t-1} = 1, ..., S_{t-p} = 1) \\ I(S_t = 1, S_{t-1} = 1, ..., S_{t-p} = 2) \\ \vdots \\ I(S_t = 1, S_{t-1} = k, ..., S_{t-p} = k) \\ I(S_t = k, S_{t-1} = k, ..., S_{t-p} = k) \end{bmatrix}$$

so that $E[\boldsymbol{\xi}_t^{(p+1)}|\boldsymbol{\xi}_{t-1}^{(p+1)}] = \mathbf{P}'_{(p+1)}\boldsymbol{\xi}_{t-1}^{(p+1)}$ where $\mathbf{P}_{(p+1)} = P \otimes P \otimes ... \otimes P$ is the $k^{p+1} \times k^{p+1}$ transition matrix for the transformed set of regimes.[9] Therefore the transition equation will be now characterized by an $\mathbf{F}$ matrix that corresponds to $\mathbf{P}'_{(p+1)}$

$$\begin{aligned} \boldsymbol{\xi}_{t+1}^{(p+1)} &= \mathbf{F}\boldsymbol{\xi}_t^{(p+1)} + \boldsymbol{\varepsilon}_{t+1} \text{ or} \\ \boldsymbol{\xi}_{t+1}^{(p+1)} - \bar{\boldsymbol{\xi}} &= \mathbf{F}(\boldsymbol{\xi}_t^{(p+1)} - \bar{\boldsymbol{\xi}}) + \boldsymbol{\varepsilon}_{t+1} \end{aligned}$$

from the ergodic property that $\mathbf{F}\bar{\boldsymbol{\xi}} = \bar{\boldsymbol{\xi}}$, while the measurement equation becomes:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{X}_t\mathbf{B}\boldsymbol{\xi}_t^{(p+1)} + \boldsymbol{\Sigma}^* \left(\boldsymbol{\xi}_t^{(1)} \otimes \mathbf{I}_n\right) \mathbf{u}_t \\ &= \mathbf{X}_t\mathbf{B}\boldsymbol{\xi}_t^{(p+1)} + \boldsymbol{\Sigma}^* \left(((\mathbf{I}_k \otimes \boldsymbol{\iota}'_{k^p})\boldsymbol{\xi}_t^{(p+1)}) \otimes \mathbf{I}_n\right) \mathbf{u}_t, \end{aligned}$$

where $\boldsymbol{\xi}_t^{(1)}$ is the standard $k \times 1$ vector collecting state information for period t such that $\boldsymbol{\xi}_t^{(1)} = (\mathbf{I}_k \otimes \boldsymbol{\iota}'_{k^p})\boldsymbol{\xi}_t^{(p+1)}$, and the $(np+1) \times nk^{p+1}$ coefficient matrix $\mathbf{B}$ has structure:

$$\mathbf{B} = \boldsymbol{\Psi}(\mathbf{I}_k \otimes \boldsymbol{\iota}'_{k^p} \otimes \boldsymbol{\iota}_n) - \begin{bmatrix} \sum_{j=1}^p \mathbf{A}_{j,s_t^{(1)}}\boldsymbol{\nu}'_{s_{t-j}^{(1)}} & \cdots & \sum_{j=1}^p \mathbf{A}'_{j,S_t^{(k^{p+1}-1)}}\boldsymbol{\nu}'_{S_{t-j}^{(k^{p+1}-1)}} & \sum_{j=1}^p \mathbf{A}'_{j,S_t^{(k^{p+1})}}\boldsymbol{\nu}'_{S_{t-j}^{(k^{p+1})}} \\ & & \mathbf{O}_{(n(p-1)+1)\times nk^{p+1}} & \end{bmatrix}.$$

[8]In general, the dynamic state-space model in (6) is neither linear (as the state vector $\boldsymbol{\xi}_t$ also influences the covariance matrix of the process) nor Gaussian, as the innovations driving the transition equation are non-Gaussian random variables.

[9]Krolzig (1997, pp. 38-39) shows that $P_{(p+1)}$ has structure:

$$\mathbf{P}_{(p+1)} = \left(\mathbf{P}' \otimes \boldsymbol{\iota}_{k^p-1}\boldsymbol{\iota}'_{k^{p+1}}\right) \odot \left(\boldsymbol{\iota}_k \otimes \mathbf{I}_{k^p} \otimes \boldsymbol{\iota}'_k\right)$$

Multivariate Markov switching models are estimated by maximum likelihood. In particular, estimation and inferences are based on the EM (Expectation-Maximization) algorithm proposed by Dempster et al. (1977) and Hamilton (1989), a filter that allows the iterative calculation of the one-step ahead forecast of the state vector $\boldsymbol{\xi}_{t+1|t}$ given the information set $\Im_t$ and the consequent construction of the log-likelihood function of the data.[10] Appendix A gives a few additional details on the EM algorithm. Maximization of the log-likelihood function within the M-step is made faster by the fact that the first-order conditions defining the MLE may often be written down in closed form. Appendix B details the general form of such conditions. In particular, notice that the first-order conditions (20)-(21) in Appendix B (exact definitions of all symbols are introduced in the Appendix),

$$\begin{aligned} \sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_{t|T}(\boldsymbol{\theta},\boldsymbol{\rho}) \frac{\partial \ln \boldsymbol{\eta}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} &= \mathbf{0}' \\ \left(\sum_{t=1}^{T}\left(\hat{\boldsymbol{\xi}}_{t|T}^{(2)}\right)\right) \oslash \left(\boldsymbol{\iota}_k \otimes \left(\sum_{t=1}^{T} \hat{\boldsymbol{\xi}}_{t|T}\right)\right) &= \boldsymbol{\rho} \end{aligned}$$

all depend on the smoothed probabilities $\hat{\boldsymbol{\xi}}_{t|T} \equiv \Pr\left(\boldsymbol{\xi}_t|\Im_T;\boldsymbol{\theta},\boldsymbol{\rho}\right)$ (i.e., the state probabilities estimated on the basis of the full sample of data) and therefore they all present a high degree of non-linearity in the parameters $\boldsymbol{\gamma} \equiv [\boldsymbol{\theta}\ \ \boldsymbol{\rho}]'$. As a result, these first-order conditions have to be solved numerically, although convenient iterative methods exist. In fact, the expectation and maximization steps can be used in iterative fashion. Starting with arbitrary initial values $\tilde{\boldsymbol{\theta}}^0$ and $\tilde{\boldsymbol{\rho}}^0$, the expectation step is applied first, thus obtaining a sequence of smoothed probability distributions $\{\hat{\boldsymbol{\xi}}_{t|T}^1\}_{t=1}^T$. Given these smoothed probabilities, (21) is then used to calculate $\hat{\boldsymbol{\rho}}^1$, and (20) to derive $\tilde{\boldsymbol{\theta}}^1$. Based on $\tilde{\boldsymbol{\theta}}^1$ and $\tilde{\boldsymbol{\rho}}^1$, the expectation step can be applied again to find a new sequence of smoothed probability distributions $\{\hat{\boldsymbol{\xi}}_{t|T}^2\}_{t=1}^T$. This starts the second iteration of the algorithm. The algorithm keeps being iterated until convergence, i.e. until $[\tilde{\boldsymbol{\theta}}^l\ \ \tilde{\boldsymbol{\rho}}^l]' \simeq [\tilde{\boldsymbol{\theta}}^{l-1}\ \ \tilde{\boldsymbol{\rho}}^{l-1}]'$. Importantly, the likelihood function increases at each step and reaches an approximate maximum in correspondence to convergence (see Baum et al., 1970).

As for the properties of the resulting ML estimators, under standard regularity conditions (such as identifiability, stability and the fact that the true parameter vector does not fall on the boundaries) Hamilton (1989, 1993) and Leroux (1993) have proven consistency and asymptotic normality of the ML estimator $\tilde{\boldsymbol{\gamma}} = [\tilde{\boldsymbol{\theta}}\ \tilde{\rho}]'$:

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

where $\mathcal{I}_a(\boldsymbol{\gamma})$ is the asymptotic information matrix,

$$\mathcal{I}_a(\boldsymbol{\gamma}) \equiv \lim_{T\to\infty} -T^{-1} E\left[\frac{\partial^2 \ln \prod_{t=1}^{T} p(\mathbf{y}_t|\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right].$$

Although other choices exist – i.e., either to use the conditional scores or a numerical evaluation of the second partial derivative of the log-likelihood function with respect to $\tilde{\boldsymbol{\gamma}}$ – in applications it has become

[10] Some assumptions have to be imposed to guarantee at least the local identifiability of the parameters under estimation. One possibility relies on the results in Leroux (1992) to show that under the assumption of multivariate Gaussian shocks to the measurement equation, MSIAH models are identifiable up to any arbitrary re-labeling of unobservable states.

typical to employ a White-style 'sandwich' sample estimator of $\mathcal{I}_a(\boldsymbol{\gamma})$ which yields the estimate

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

where

$$\mathcal{I}_1(\tilde{\boldsymbol{\gamma}}) = T^{-1}\sum_{t=1}^{T}[\mathbf{h}_t(\tilde{\boldsymbol{\gamma}})]\,[\mathbf{h}_t(\tilde{\boldsymbol{\gamma}})]' \qquad \mathbf{h}_t(\tilde{\boldsymbol{\gamma}}) = \frac{\partial \ln p(\mathbf{y}_t|\Im_{t-1};\tilde{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}}$$

and

$$\mathcal{I}_2(\tilde{\boldsymbol{\gamma}}) = -T^{-1}\sum_{t=1}^{T}\left[\frac{\partial^2 \ln p(\mathbf{y}_t|\Im_{t-1};\tilde{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'}\right].$$

As a consequence, and with one important exception, standard inferential procedures are available to test statistical hypothesis. In particular, call $\phi : \mathcal{R}^q \to \mathcal{R}^r$ a (smooth) function that imposes $q-r$ restrictions on the q-dimensional parameter vector $\boldsymbol{\theta}$. We want to test $H_0 : \phi(\boldsymbol{\gamma}) = \mathbf{0}$ vs. $H_1 : \phi(\boldsymbol{\gamma}) \neq \mathbf{0}$ *under the assumption that under both hypothesis the number of regimes k is identical.*[11] Define $\tilde{\boldsymbol{\theta}}_r$ as the restricted estimator, obtained under the null hypothesis. Lagrange Multiplier (LM) tests are undoubtedly the preferred tests as they only require the estimation of the restricted model. While the scores of an unrestricted model,

$$\mathbf{s}_t(\tilde{\boldsymbol{\gamma}}) \equiv \sum_{\tau=1}^{t}\mathbf{h}_\tau(\tilde{\boldsymbol{\gamma}}) = \sum_{\tau=1}^{t}\left[\left.\frac{\partial \text{diag}(\boldsymbol{\eta}_\tau((\boldsymbol{\gamma}))F_\tau(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'}\right|_{\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}\right]'\hat{\boldsymbol{\xi}}_{\tau|t},$$

have zero mean vector by construction,[12] the scores of the restricted model obtained by MLE and imposing $\phi(\boldsymbol{\theta}) = \mathbf{0}$ can be used to obtain the standard test statistic:

$$LM_T \equiv \mathbf{s}_T(\tilde{\boldsymbol{\theta}}_r)'\left[\widetilde{Var}(\tilde{\boldsymbol{\theta}}_r)\right]^{-1}\mathbf{s}_T(\tilde{\boldsymbol{\theta}}_r) \xrightarrow{d} \chi^2_r$$

where $r = \text{rank}\left(\frac{\partial \phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)$ and $\tilde{\boldsymbol{\theta}}_r$ denotes the restricted estimator. For instance, a test of the hypothesis of homoskedasticity ($H_0 : vech(\Sigma_i) = vech(\Sigma_M)\ i = 1, 2, ..., M$) implies $r = (k-1)\frac{n(n+1)}{2}$ restrictions and can be formulated as a linear restriction on the matrix Σ^*. As an alternative, the Likelihood Ratio (LR) test might be employed:

$$LR \equiv 2\left[\ln L(\tilde{\boldsymbol{\theta}}) - \ln L(\tilde{\boldsymbol{\theta}}_r)\right] \xrightarrow{d} \chi^2_r.$$

Although very simple, this test requires the estimation of both the restricted and the unrestricted models, which for n high enough may be quite cumbersome and require a host of diagnostic checks on the performance of the EM algorithm in locating a truly global maximum of the likelihood function. The empirical application of this paper in Section 3 shows how the LR test is employed in practice.

[11] Notice though that hypothesis involving elements of $\boldsymbol{\rho}$ (the vec of the estimable elements of the transition matrix $\mathbf{P}$) set to equal zero cannot be entartained as they fall on the boundaries of the parameter space and imply a change in the number of actual regimes. However other hypothesis involving $\boldsymbol{\rho}$ can be tested without restrictions, for instance the important statistical hypothesis of independent regime switching (i.e. $\mathbf{P}$ has rank one).

[12] This is because $\mathbf{s}_T(\tilde{\boldsymbol{\theta}}) \equiv \sum_{t=1}^{T}\mathbf{h}_t(\tilde{\boldsymbol{\theta}}) = \sum_{t=1}^{T}\left[\left.\frac{\partial \text{diag}(\boldsymbol{\eta}_\tau((\boldsymbol{\theta}))F_\tau(\boldsymbol{\gamma})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}\right]'\hat{\boldsymbol{\xi}}_{t|T} = \mathbf{0}'$, from the MLE first-order conditions.

Finally, standard t and F statistics can be calculated in the form of a Wald test. Under asymptotic normality of the unrestricted ML estimator $\tilde{\boldsymbol{\theta}}$, it follows that

$$\sqrt{T}\left[\phi(\tilde{\boldsymbol{\theta}}) - \phi(\boldsymbol{\theta})\right] \xrightarrow{d} N\left(\mathbf{0}, \left.\frac{\partial\phi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \widetilde{Var}(\tilde{\boldsymbol{\theta}}) \left.\frac{\partial\phi'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}\right) \text{ and}$$

$$Wald \equiv T\phi'(\tilde{\boldsymbol{\theta}}) \left[\left.\frac{\partial\phi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \widetilde{Var}(\tilde{\boldsymbol{\theta}}) \left.\frac{\partial\phi'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}\right]^{-1} \phi(\tilde{\boldsymbol{\theta}}) \xrightarrow{D} \chi^2_r.$$

For instance, the hypothesis that in (1) the matrices of autoregressive coefficients are regime independent can be written as:

$$\begin{bmatrix} \mathbf{O}_n & \mathbf{O}_n & \cdots & \mathbf{I}_n & -\mathbf{I}_n & \mathbf{O}_n & \cdots & \mathbf{O}_n \\ \mathbf{O}_n & \mathbf{O}_n & \cdots & \mathbf{O}_n & \mathbf{I}_n & -\mathbf{I}_n & \cdots & \mathbf{O}_n \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_n & \mathbf{O}_n & \cdots & \mathbf{O}_n & \mathbf{O}_n & \mathbf{O}_n & \cdots & \mathbf{O}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \mathbf{A}'_{11}\mathbf{e}_1 \\ \vdots \\ \mathbf{A}'_{1k}\mathbf{e}_n \\ \vdots \\ \mathbf{A}'_{pk}\mathbf{e}_n \end{bmatrix} = \mathbf{R}vec(\mathbf{A}) = \mathbf{0}$$

and implies the (F) test statistic:

$$T\tilde{\boldsymbol{\theta}}'\mathbf{R}'\left[\mathbf{R}\widetilde{Var}(\tilde{\boldsymbol{\theta}})\mathbf{R}'\right]^{-1}\mathbf{R}\tilde{\boldsymbol{\theta}}.$$

The exception to standard inferential procedures mentioned above concerns the *number of non-zero rows* of the transition matrix $\mathbf{P}$, i.e. the number of regimes k. In this case, even under the assumption of asymptotic normality of the estimator $\tilde{\boldsymbol{\gamma}}$, standard testing procedures suffer from non-standard asymptotic distributions of the likelihood ratio test statistic due to the existence of *nuisance* parameters under the null hypothesis. We specifically discuss this problem in Section 2.3.

2.2. *Forecasting*

Under a mean squared prediction error (MSFE) criterion, the algorithms required to implement standard forecasting are relatively simple in spite of the nonlinearity of the MSIAH class and naturally derive from (6). Ignoring for the time being the issue of parameter uncertainty, i.e. the fact that the parameters of the multivariate Markov switching process are unknown and must therefore be estimated, the function minimizing the MSFE is the standard conditional expectation function. For instance, for a one-step ahead forecast we have:

$$E[\mathbf{y}_{t+1}|\Im_t] = \mathbf{X}_{t+1}\Psi\left(\hat{\boldsymbol{\xi}}_{t+1|t} \otimes \boldsymbol{\iota}_n\right)$$

where $\mathbf{X}_{t+1} = [1\ \mathbf{y}'_t ... \mathbf{y}'_{t-p+1}] \otimes \boldsymbol{\iota}_n$, Ψ collects the estimated conditional mean parameters, and $\hat{\boldsymbol{\xi}}_{t+1|t}$ is the one-step ahead, predicted latent state vector to be filtered out of the available information set $\Im_t$ according to transition equation

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P}'\hat{\boldsymbol{\xi}}_{t|t},$$

where also the transition matrix $\mathbf{P}$ will have to be estimated. It follows that

$$E[\mathbf{y}_{t+1}|\Im_t] = \mathbf{X}_{t+1}\Psi\left(\mathbf{P}'\hat{\boldsymbol{\xi}}_{t|t}\otimes\boldsymbol{\iota}_{n+m}\right). \tag{7}$$

For $h>1$-step ahead forecasts the task is much more challenging as: (1) $\mathbf{X}_{t+h}$ is unknown and must be predicted itself; (2) $E[\mathbf{X}_{t+T}|\Im_t]$ involves sequences of predictions $\{E[\mathbf{y}_{t+1}|\Im_t],...,E[\mathbf{y}_{t+T-1}|\Im_{t+T-2}]\}$ and as such $\{\hat{\boldsymbol{\xi}}_{t+1|t},...,\hat{\boldsymbol{\xi}}_{t+T-1|t}\}$ which are likely to impress patterns of cross-correlation to the unconditional values of the parameters to be used, because of the presence of regime switching. For instance, for $T=2$, $p=1$, and ignoring the presence of an intercept term, we have

$$\begin{aligned}
E[\mathbf{y}_{t+2}|\Im_t] &= E\left[\left(\mathbf{y}'_{t+1}\otimes\boldsymbol{\iota}_n\right)\Psi\left(\boldsymbol{\xi}_{t+2}\otimes\boldsymbol{\iota}_n\right)|\Im_t\right]\\
&= E\left[\left(\left(\mathbf{y}'_t\otimes\boldsymbol{\iota}_n\right)\Psi\left(\boldsymbol{\xi}_{t+1}\otimes\boldsymbol{\iota}_n\right)\otimes\boldsymbol{\iota}'_n+\boldsymbol{\Sigma}^*\left(\boldsymbol{\xi}_{t+1}\otimes\mathbf{I}_n\right)\mathbf{u}_t\otimes\boldsymbol{\iota}'_n\right)\Psi\left(\boldsymbol{\xi}_{t+2}\otimes\boldsymbol{\iota}_n\right)|\Im_t\right]\\
&= E\left[\left(\left(\mathbf{y}'_t\otimes\boldsymbol{\iota}_n\right)\Psi\left(\boldsymbol{\xi}_{t+1}\otimes\boldsymbol{\iota}_n\right)\otimes\boldsymbol{\iota}'_n\right)\Psi\left(\boldsymbol{\xi}_{t+2}\otimes\boldsymbol{\iota}_n\right)|\Im_t\right]
\end{aligned}$$

which is not the product of the conditional expectations of $\left((\mathbf{y}'_t\otimes\boldsymbol{\iota}_n)\Psi\left(\boldsymbol{\xi}_{t+1}\otimes\boldsymbol{\iota}_n\right)\otimes\boldsymbol{\iota}'_n\right)$ and $\Psi\left(\boldsymbol{\xi}_{t+2}\otimes\boldsymbol{\iota}_n\right)$ as the future state vectors $\boldsymbol{\xi}_{t+1}$ and $\boldsymbol{\xi}_{t+2}$ are correlated, from $\boldsymbol{\xi}_{t+2}=\mathbf{F}\boldsymbol{\xi}_{t+1}+\boldsymbol{\varepsilon}_{t+2}$. However, in applied work it is customary to follow the suggestion of Doan et al. (1984) and to substitute the sequence of predicted values of $\{\mathbf{y}_{t+1}, \mathbf{y}_{t+2}, ..., \mathbf{y}_{t+T-1}\}$ (as of time t) – i.e., $\{\hat{E}[\mathbf{y}_{t+1}|\Im_t], \hat{E}[\mathbf{y}_{t+2}|\Im_t], ..., \hat{E}[\mathbf{y}_{t+T-1}|\Im_t]\}$ for $\{E[\mathbf{y}_{t+1}|\Im_t], E[\mathbf{y}_{t+2}|\Im_{t+1}], ..., E[\mathbf{y}_{t+T-1}|\Im_{t+T-2}]\}$. In this case (7) generalizes to generic $T>2$-step ahead predictions:

$$E[\mathbf{y}_{t+T}|\Im_t] = E[\mathbf{X}_{t+T}|\Im_t]\Psi\left[\left(\mathbf{P}'\right)^T\hat{\boldsymbol{\xi}}_{t|t}\otimes\boldsymbol{\iota}_n\right],$$

which in practice gives a recursive formula since $E[\mathbf{X}_{t+T}|\Im_t]$ forces one to forecast a sequence of future $\mathbf{y}_{t+i}$ values, $i=1,2,...,T-1$. Similar problems apply to multi-step forecasts from the MSMVARH model (5). Using Doan et al.'s suggestion to recover linearity of the predictor in the last p observations and the regime inference, we obtain

$$E[\mathbf{y}_{t+T}|\Im_t] = E[\mathbf{X}_{t+T}|\Im_t]\mathbf{B}\bar{\boldsymbol{\xi}}+E[\mathbf{X}_{t+T}|\Im_t]\mathbf{B}\left[\left(\mathbf{P}'_{(p+1)}\right)^T(\hat{\boldsymbol{\xi}}_{t|t}^{(p+1)}-\bar{\boldsymbol{\xi}})\right].$$

2.3. *Model Selection and Diagnostic Checks*

In the absence of Markov switching dynamics in the matrices of autoregressive coefficients and in the covariance matrix of the vector process – i.e., for simple MSI$(k,0)$ and MSI(k)-VAR(p) processes, it is possible to show that general multivariate Markov switching models possess a standard VARMA representation that helps define a somewhat precise mapping between nonlinear Markov switching processes and their linear counterparts. In particular, under a few regularity conditions, (1) possesses a VARMA$(k+p-1,k-1)$ representation, where $k+p-1$ is the autoregressive order and $k-1$ is the moving average order. On the other hand, the MSMVAR(p) process (5) has a VARMA$(k+np-1,k+np-2)$ representation. In both cases, notice that the $VARMA(a,b)$ representation implies $a\geq b$. These results give a useful starting point in a simple-to-general specification approach:

1. A researcher may start out by conducting a standard Box-Jenkins' style model selection procedure applied to the class of VARMA models. The reason is that given the existence of $VARMA(a,b)$

representations for MSM processes, it is then possible to solve a simple bivariate system of linear equations to recover k and p from the selected values for a and b. Because in multivariate contexts, VARMA-style model selection is anyway quite difficult, noting that $k+p-1 \geq p$ and $k+np-1 \geq p$ suggests that the autoregressive order in the VARMA is never lower than the autoregressive order in the MSM model. Thus a standard VAR lag-selection procedure anyway provides a sensible upper bound to the correct value of p to employed in the MSM specification.

2. Given such a p^*, the focus shifts on the number of regimes k. Krolzig (1997) has suggested the analysis of each component of the vector $\mathbf{y}_t$ in isolation to detect the appropriate number of regimes for each of them, say k_i for y_{it}, $i = 1, 2, ..., n$. In this case the (V)ARMA equivalence results can be fully exploited. For each time series, the best fitting ARMA model could be selected using Box-Jenkins or any other ARMA specification criteria. Taking into account that the AR order p^* has been pre-selected, the optimal number of regimes k_i^* will simply correspond to the MA order plus one (plus two minus p^* in the (5) case). Call $\{k_i^*\}_{i=1}^n$ the sequence of resulting number of states for each univariate variable under study.

3. Given $\{k_i^*\}_{i=1}^n$, the total number of regimes characterizing the multivariate process might be in principle as high as $\prod_{i=1}^n k_i^*$ if the regimes are not simultaneously perfectly correlated with each other, i.e. if it does not occur that at least a subset of variables are governed by the same hidden Markov chain. This latter hypothesis is usually testable using standard inferential procedures.

4. Once the number of MSIAH (MSMAH) regimes k^* (M^*) has been selected, it is useful to test for the presence of regime-dependent heteroskedasticity and/or for the presence of regimes in the autoregressive component of the Markov switching model. For instance, an LM test might be employed. Or the MSM model might be estimated with and without heteroskedastic component and the LR test used to improve the specification.

As illustrated in our application in Section 3 as well as in a number of papers in the recent literature (see e.g., Guidolin and Timmermann, 2007), an alternative set of methods to perform data-driven model selection relies on information criteria, such the Schwartz, Hannan-Quinn, and Akaike criteria (see e.g., Sin and White, 1996, for evidence on information criteria performance in non-linear models). Interestingly, very few papers have addressed the issue of the small-sample and asymptotic performance of these information criteria specifically for the case of Markov switching models.

Once a restricted set of MSM models has been estimated, the need of further improvements could arise as the result of diagnostic checks.[13] Although the EM algorithm naturally delivers estimates of the parameters $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\xi}}_{1|0}^1$, besides the smoothed sequence of probability distributions $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^T$ and would therefore lead to define the (smoothed) residuals as

$$\tilde{\mathbf{u}}_t \equiv \mathbf{y}_t - \mathbf{X}_t \hat{\mathbf{B}} \hat{\boldsymbol{\xi}}_{t|T},$$

[13] In what follows we focus for simplicity on MSMAH models because they are logically and computationally more complicated than MSIAH models. However, all of our remarks apply once one replaces $\tilde{\mathbf{u}}_t \equiv \mathbf{y}_t - \mathbf{X}_t \hat{\mathbf{B}} \hat{\boldsymbol{\xi}}_{t|T}$ with $\tilde{\mathbf{u}}_t \equiv \mathbf{y}_t - \mathbf{X}_t \hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\xi}}_{t|T}$. and the number of regimes M^* with k^*.

these are not well suited to the use in diagnostic checks as they are full-sample random statistics and hence they structurally overestimate the explanatory power of the MSM model. On the contrary the one-step predictions errors

$$\tilde{\mathbf{e}}_{t|t-1} \equiv \mathbf{y}_t - \mathbf{X}_t\hat{\mathbf{B}}\mathbf{F}\hat{\boldsymbol{\xi}}_{t-1|t-1}$$

are limited information statistics (being based on filtered probabilities) and uncorrelated with the information set $\Im_{t-1}$ because $E[\mathbf{y}_t|\Im_{t-1}] = \mathbf{X}_t\hat{\mathbf{B}}\mathbf{F}\hat{\boldsymbol{\xi}}_{t-1|t-1}$ and therefore form a martingale difference sequence $E[\tilde{\mathbf{e}}_{t|t-1}|\Im_{t-1}] = \mathbf{0}$. Therefore standard tests of this hypothesis (such as Portmanteau tests of no serial correlation) could be used.[14] In the presence of Markov switching heteroskedastic components (i.e., the covariance matrices of shocks fail to depend on regimes), researchers in empirical finance (e.g., Kim and Nelson, 1999) have also suggested to check whether the smoothed, standardized residuals contain any residual ARCH effects. Standard LM-type as well as Ljiung-Box tests can be applied. This is a way to check whether Markov switching variances and covariances may be sufficient to capture most of the dynamics in volatility, else explicit ARCH-type modeling (even of a Markov switching nature, as in Hamilton and Susmel, 1994, or Guidolin, 2009) may be required.[15]

Another important type of diagnostic check concerns the number of regimes M. The problem is that under any number of regimes smaller than M there are a few structural parameters of the unrestricted model — the elements of the transition probability matrix associated with the rows that correspond to "disappearing states" — that can take any values without influencing the resulting likelihood function. We say that these parameters become a *nuisance* to the estimation. The result is that the presence of these nuisance parameters gives the likelihood surface so many degrees of freedom that computationally one can never reject the null that the nonnegative values of those parameters were purely due to sampling variation.[16] Different alternative ways have been proposed to develop sound inferential procedures concerning the number of regimes in multivariate Markov switching models. Hansen (1992) proposes to see the likelihood as a function of the unknown and non estimable nuisance parameters so that the asymptotic distribution is generated in each case numerically from a grid of transition and regime-dependent nuisance parameters. The test statistic becomes then

$$LW_T \leq \sup_{\boldsymbol{\rho}} LW_T(\boldsymbol{\rho})$$

where the right hand side converges in distribution to a function of a Brownian bridge. In most of the cases a closed form expression cannot be found and the bound must be calculated by simulation and becomes data-dependent. Also Davies (1977) bounds the LR test but avoids the problem of estimating the nuisance

[14] With the caveat that that the one-step prediction errors do not have a Gaussian density and hence the approximate validity of standard tests can only be guessed. For instance, Turner et al. (1989) devise similar tests in which the filtered probabilities are used as predictors of future variance and test the absence of serial correlation in the resulting regression residuals.

[15] Under an incorrect null of no Markov regimes, it is easy to show that asset returns may easily turn out to display non-linear stochastic structures that may show up as significant ARCH-type effects even in the absence of true ARCH in a correctly specified Markov switching generating process.

[16] Mathematically, the presence of unidentified nuisance parameters implies that the scores become identically zero and that the covariance matrix is singular.

parameters and derives instead an upper bound for the significance level of the LR test under nuisance parameters:

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

The bound holds if the likelihood has a single peak. A related test is proposed by Wolfe (1971) and applied by Turner et al. (1989). The modified LR test is:

$$LR^{Wolfe} = -\frac{2}{T}(T-3)\left[\ln L(\tilde{\boldsymbol{\gamma}}) - \ln L(\tilde{\boldsymbol{\gamma}}_r)\right] \xrightarrow{d} \chi_r^2$$

where $\tilde{\boldsymbol{\gamma}}_r$ is obtained under the null of simple multivariate normality and $r = M(M-1)$ because in the absence of regime switching there are $M(M-1)$ which cannot be estimated. Davidson and MacKinnon's (1981) J test for non-nested models can be also applied, because Markov switching models with M and $M-1$ regimes are logically nested but cannot be treated as such on a statistical basis. To implement a J test one has to estimate the model with M and $M-1$ states and calculate their full information fitted values, $\tilde{\mathbf{y}}_t^{(j)} = \mathbf{X}_t\hat{\mathbf{B}}^{(j)}\hat{\boldsymbol{\xi}}_{t|T}^{(j)}$; then estimate the (multivariate) regression

$$\mathbf{y}_t = (\mathbf{I}_n - \boldsymbol{\Delta})\mathbf{X}_t\hat{\mathbf{B}}\hat{\boldsymbol{\xi}}_t^{(M-1)} + \boldsymbol{\Delta}\tilde{\mathbf{y}}_t^{(M)} + \boldsymbol{\varepsilon}_t.$$

The p-value of an F-test for the matrix of coefficients $\boldsymbol{\Delta}$ gives the p-value for the null of M regimes.

Finally, common sense suggests that correct specification of a Markov switching model should give smoothed probability distributions $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$ that consistently signal switching among states with only limited periods in which the associated distribution is flatly spread out over the entire support and uncertainty dominates. Regime Classification Measures have been popularized as a way to assess whether the number of regimes M is adequate. In simple two-regime frameworks, the early work by Hamilton (1988) offered a rather intuitive regime classification measure:

$$RCM_1 = 100\frac{M^2}{T}\sum_{t=1}^{T}\prod_{m=1}^{M}\Pr\left(S_t = m|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}\right),$$

i.e., the sample average of the products of the smoothed state probabilities. Clearly, when a Markov switching model offers precise indications on the nature of the regime at each time t, the implication is that for at least one value of $m = 1, ..., M$, $\Pr\left(S_t = m|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}\right) \simeq 1$ so that $\sum_{m=1}^{M}\Pr(S_t = m|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}) \simeq 0$ because most other smoothed probabilities are zero. Therefore a good switching model will imply $RCM_1 \simeq 0$.[17] However, when applied to models such that $M > 2$, RCM_1 has one obvious disadvantage: a model can imply an enormous degree of uncertainty on the current regime, but still have $\sum_{m=1}^{M}\Pr\left(S_t = m|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}\right) \simeq 0$ for most values of t. For instance, when $M = 3$, it is easy to see that if $\Pr\left(S_t = 1|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}\right) = 1/2$, $\Pr\left(S_t = 2|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}\right) = 1/2$, and $\Pr\left(S_t = 3|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}}\right) = 0$ $\forall t$, then $RCM_1 = 0$ even though this remains a rather uninformative switching model to use in practice. As a result, it is rather common to witness that as M exceeds 2, almost all switching models (good and

[17] On the opposite, the worst possible Markov switching model implies $\Pr\left(S_t = 1|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\gamma}\right) = ... = \Pr\left(S_t = M|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\gamma}\right) = 1/M$ so that $\sum_{m=1}^{M}\Pr\left(S_t = m|\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\gamma}\right) = 1/K^2$ and $RCM_1 = 100$. Therefore $RCM_1 \in [0, 100]$ and lower values are to be preferred to higher ones.

bad) will automatically imply values of RCM_1 that decline towards 0. Guidolin (2009) proposes a number of alternative measures that may shield against this type of problems, for instance using:

$$RCM_2 = 100 \left\{ 1 - \frac{M^{2M}}{(M-1)^2} \frac{1}{T} \sum_{t=1}^{T} \prod_{m=1}^{M} \left[\Pr\left(S_t = k | \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_T; \tilde{\boldsymbol{\gamma}} \right) - \frac{1}{M} \right]^2 \right\}.$$

3. One Application: Regimes in International Stock Returns

In this Section we report one illustrative example of how multivariate Markov switching models may be used to capture the key dynamic features of large-scale, complicated financial phenomena and how they can be used to support sophisticated financial decision making. For reasons of space, we illustrate only the key points of our model specification search and of its asset management implications. Although the data are particular to this paper, methods and qualitative results may be considered a special case (extension) of results in related papers by Guidolin and Na (2009), Guidolin and Nicodano (2009), and Guidolin and Timmermann (2007, 2008b) to which a Reader is referred for additional applications and details. Our empirical analysis is also related to Ang and Bekaert (2004), who solve a large-scale international portfolio choice problem in which a version of the zero-beta CAPM is assumed ex-ante, i.e., $x_t^j = \beta_j \mu_{S_t}^W + \beta_j \sigma_{S_t}^W \epsilon_t^W + \eta_j \epsilon_t^j$, and the world market portfolio (indexed as W) follows a two-state Markov switching model in its mean, $\mu_{S_t}^W$.[18] Similarly to our paper, they consider a simple mean-variance portfolio framework. Using monthly MSCI net returns (expressed in US dollars) data for the period 1975:02-2000:12 and 20 national stock markets organized in 6 macro-regions (North America, UK, Japan, Europe ex-UK large countries, Europe ex-UK small countries, and Pacific ex-Japan), they document that Markov switching efficient frontiers and portfolio choices may be considerably different from classical, mean-variance asset allocation results.

3.1. *Data*

We examine MSCI international monthly equity index data for the sample 1988:01-2010:09, for a total of 273 observations. In this application we investigate the regime switching properties of the 5 major developed country/area value-weighted indices published by MSCI, i.e. (in order of declining capitalization), North America (US and Canada), Japan, Europe ex-UK, Pacific ex-Japan, and United Kingdom. The five indices are all expressed in US dollars. This means that we adopt the point of view of a US investor that is considering un-hedged international pure equity portfolio diversification decisions. The 1988-2010 sample period is a plausible estimation interval of time for many investors that could be interesting in performing econometric analysis on relatively recent equity data, while the fact that the sample extends well into 2008-2009 allows us to reach conclusions that are not only robust, but in fact fully affected by the recent turmoil in international equity markets.

Table 1 reports standard descriptive statistics. The table gives only one surprising indication: over a 23-year time span, the Japanese stock market has yielded on average a mean return that is close to

[18] Here x stands for excess return, ϵ_t^j and ϵ_t^W are both IID $N(0,1)$, and $\eta_j$is the idiosyncratic variance component. j indices the different markets under investigation.

zero (0.2% in annualized term) and therefore a negative Sharpe ratio. Otherwise, means, volatilities and realized Sharpe ratios (computed using as a reference the US 1-month T-bill rate) are all within expected ranges, i.e., annualized means are between 9.2 and 12.5% per year, volatilities are between 15 and 22% per year, and the monthly Sharpe ratios are between 0.09 and 0.12. Pacific, continental European, UK, and North American stocks all displayed significant deviations from a single-state IID Gaussian benchmark, as evidenced by the statistically significant Jarque-Bera statistics. All the four indices are characterized by negative skewness and kurtosis in excess of the Gaussian benchmark (three). However, only the continental European and North American skewness coefficients are significantly negative, while only the Pacific ex-Japan index returns have an excess kurtosis that is significantly positive. Yet, in overall terms, the Jarque-Bera test rejects normality for the four index return series. Interestingly, this occurs also in the case of Japanese stock returns, even though these display very modest excess kurtosis and positive skewness; it is therefore the sum of these modest deviations to cause a rejection of the null hypothesis of normality. Although none of the indices appears to be predictable based on its (linear) correlation structure, all the five indices present evidence of conditional heteroskedasticity, as shown by the Ljiung-Box portmanteau tests applied on squared returns.

3.2. *Model Selection*

Table 2 shows the results of a few model selection criteria applied to our 5×1 vector of international stock returns. Clearly, we have estimated a range of alternative MSIAH models, including simple single-state (i.e., Gaussian IID and Gaussian VAR) ones.[19] Significantly, with a total of 1,365 observations, one encounters difficulties at obtaining reliable estimates of richly parameterized models in which the number of parameters largely exceeds 100 so that the saturation ratio (i.e., the number of observations available to estimate each parameter, on average) is below 10. This is the case of the MSIAH(4,1) model, that would imply estimating almost 200 parameters.

The table shows that the Davies' approximate p-value for a test of the null of $k = 1$ regimes vs. the alternative hypothesis of $k \geq 2$ (the specific value for k depends then on each of the models considered) is always essentially zero for all Markov switching frameworks considered, independently on their specific structure in terms of choice of p and of whether variance and covariances ought to be allowed to be a function of the Markov state. The associated LR statistics are in fact so large that the same conclusion is likely to emerge regardless of the nuisance parameter correction applied to compute the p-values in the table. For instance, Wolfe's test statistic delivers p-values which are also essentially zero. Therefore, it is clear that the data seem to require the specification of Markov switching dynamics, which is consistent with earlier findings by Ramchand and Susmel (1998), Ang and Bekaert (2002a), and Guidolin and Timmermann (2008b).

Next, we employ three standard information criteria to select among multi-state regime switching models. For each of three criteria (Bayesian-Schwartz, Hannan-Quinn, and Akaike), Table 2 boldfaces the

[19] In financial applications it is atypical to find an interest in the fit of MSMAH models, given their complexity and the general finding of low-order VAR structures. Accordinly, in what follows we use k to denote the number of regimes.

three best models. Here, one has to recall that by construction (as they reward fit represented by the negative of the maximized average log-likelihood function and penalize over-parameterization by adding a positive term), information criteria illustrate an increasingly good trade-off between fit and parsimony as their values *decline*. As one would expect in the light of their relative penalties for model size, the BIC tends to favor small models, in this case with 2 or at most 3 regimes, while the single-state Gaussian IID model gives scores close to those of the three best models. On the contrary, AIC shows a bias in favor of relatively large, possibly over-parameterized models like the MSIH(4,0) model which enters the AIC's best-three set in spite of its modest saturation ratio of 14.8.[20] Hannan-Quinn is usually in an intermediate position when compared to BIC and AIC, although in our application it yields indications which are identical to BIC. However, in spite of these differences, Table 2 also stresses the existence of one and only model that receives "good scores" from all the information criteria deployed, and that is a relatively simply and parsimonious (42 parameters for a saturation ratio of 32.5) MSIH(2,0), i.e., a model with two regimes, regime-dependent covariance matrices, and no vector autoregressive component.[21]

In fact, the very last column of Table 2 also proceeds to test – using standard likelihood ratio tests – whether any expansion over the MSIH(2,0) may be required by the data. In particular, the null of a MSIH(2,0) against the alternative of a richer MSIAH(2,1) with $p = 1$ can be rejected with a p-value close to zero. At the same time, the null of a simpler MSI(2,0) model vs. the MSIH(2,0) (i.e., no regime-switching heteroskedasticity) can rejected with a p-value of zero; the null of a MSH(2,0) model vs. the MSIH(2,0) (i.e., no regime switching in conditional means) is also rejected with a p-value of zero. It is only the information criteria that have advised us to select MSIH(2,0) over a more complex and richly parameterized MSIAH(2,1). We have also tried to use Krolzig's (1997) VARMA-MSIA mapping method: because we find that a VARMA(2,1) seems to be required by the 5×1 vector at hand, we obtain that $k + p - 1 = 2$ and $k - 1 = 1$. Solving for k and p, this gives $k^* = 2$ and $p^* = 1$. However, both LR tests and information criteria advise us to select instead $k = 2$ and $p = 0$, augmented by a regime-dependent heteroskedastic component.

3.3. *A Two-State Model*

Table 3 shows the ML estimates for the two-state MSIH(2,0) model:

$$\mathbf{y}_t = I_{S_t=1}\boldsymbol{\mu}_1 + (1 - I_{S_t=1})\boldsymbol{\mu}_2 + [I_{S_t=1}\boldsymbol{\Sigma}_1 + (1 - I_{S_t=1})\boldsymbol{\Sigma}_2]\boldsymbol{\epsilon}_t, \tag{8}$$

where $\mathbf{y}_t$ denotes a 5×1 vector of (US-dollar denominated) returns and $I_{S_t=1}$ is a standard indicator variable that takes unit value when the system in the first regime. Table 3 reports estimates of the two-state model and, as a benchmark, of a matching single-state model – in this case a simple Gaussian IID model that implies constant means, variance, and covariances (means and variances are the same as in Table 1). Starting with the single-state model, it is clear that all pairs of stock indices are characterized by positive and highly statistically significant correlation coefficients, ranging between 0.48 and 0.82.

[20] In the non-linear literature, all models with saturation ratios below 20 are normally regarded with suspicion.

[21] The MSIH(2,0) model also yields a good RCM_1 of 15.1 and a RCM_2 of 16.9 which are relatively low. For instance, a MSIH(3,0) model returns a RCM_2 of 93.2, which is largely disappointing.

Although these correlations are high, from standard mean-variance diversification theory we know that not even 0.82 would be able to deprive a mean-variance investor from considerable gains from portfolio diversification.

The second panel of Table 3 reports instead the estimates of the Markov switching model. With one exception, for four of the five indices, the first regime is a bear state with low (not statistically significant in all 4 cases) mean returns and high (above-average, represented by the single-state panel of the table) correlations. In this state, expected returns fail to be statistically different from zero in the case of Japan, continental Europe, the UK, and the US. The only exception, mentioned above, occurs in the case of Pacific ex-Japan USD returns, which appear to have a high and mildly statistically significant expected return in the first regime, of 1.15% per month. Moreover, all pair-wise correlations are structurally higher in this first regime than they are in the second. For instance, the average correlation in regime 1 is 0.77 vs. an average of 0.63 in the single state model. For instance, during this bear regime, UK and continental European returns show a correlation of 0.94, i.e., from an investor's viewpoint there is only one single European stock market. We call this regime a bear state of high correlations, in which (notice, only within the regime itself), the value of international diversification may be attenuated by the strong tendency of international stock markets to linearly co-move in highly synchronous ways. Interestingly, and differently from what found by other papers with reference to shorter (or alternative) sample periods (e.g., Guidolin and Timmermann, 2008b), this bear state of high correlations fails to be also characterized by high volatilities: the estimated within-state standard deviations are higher than in the second state in the case of continental Europe and North America, but they are lower for the remaining 3 indices. The bear state has exceptionally high persistence— once in the bear regime, markets tend to stay in this state for 36 months on average— and as a result this regime characterizes approximately 56% of the data in the long run (equivalently, 0.56 is the ergodic probability of the bear regime).

Figure 1 plots the full-sample smoothed probabilities from the two-state model and shows that the most prolonged periods of modest mean stock returns but high correlations may be identified with 1996-1997 (the Asian flu), the dot-com market crash of 2000-2001, and more recently a long span extending itself between 2002 and mid-2008. This characterization of the smoothed probabilities and of the first regime should not surprise us because the model estimated in Table 2 seems to be primarily driven by the time-variation in pair-wise correlations, similarly to Ang and Bekaert (2002a). It then corresponds to common wisdom the fact that since 2000 and until the trough of the 2008-2009 financial crisis, developed financial markets have become increasingly correlated— some commentators have linked this observation to the alleged occurrence of two financial bubbles between the late 1990s and then 2004-2008. Moreover, both during the market downturn of 2000-2001 and the recent financial crisis, it has been obvious that some economies (such as those in the Far East) less dependent on the process of securitization and still relying on a solid web of export industries, have been affected by the crisis waves to a less extent than Europe or the US (not to mention Japan). This may justify why the first regime is essentially a bear state for four out of five of our indices, but not for the Pacific ex-Japan index, which groups a number of developed Asian economies, Australia, and New Zealand.

Again with the exception of Pacific ex-Japan, the second state is a bull regime characterized by positive (and statistically significant, for three indices out of four) mean returns and by moderate (below-average) correlations, in some cases not significantly different from zero when tests of size 1% are applied.[22] For instance, the average pair-wise correlation in this second regime is 0.51 vs. 0.77 in the first regime and 0.63 in the single-state model. Obviously, the diversification opportunities are maximum in this second state, even though our goal is to actually characterize diversification opportunities *across regimes*, and when investors are uncertain on the nature of the current regime, as it will be often the case in a latent regime switching environment. Also this regime displays considerable persistence— once in this regime, international stock markets tend to display this dynamics for 29 months on average; as a result, 44% of any long sample ought to be generated by this bull state of low correlations. Figure 1 shows that long stretches of time – well in excess of 29 months, in fact – such as 1988-1992, 1997-2000, and more recently early 2009 are characterized as draws from the bull regime, when markets do not strongly co-move.

We perform diagnostic checks based on the one-step predictions errors, similarly to Guidolin and Ono (2006). A number of alternative Portmanteau statistics and tests all indicate that the one-step errors are approximately martingale difference sequences. In particular, there is only weak evidence of residual ARCH effects in the prediction errors, which seems to be a rather common finding in similar applications (see Ang and Bekaert, 2002a, and Guidolin and Timmermann, 2008b).

3.4. *Time-Varying Efficient Frontiers*

One of the goals of this Section is to illustrate how multivariate Markov switching models may be put at work. We start doing this by computing regime-specific Markov switching mean-variance frontiers (MSMVF, for short). MSMVFs are a simple generalizations of standard textbook, Markowitz-style efficient frontiers to the case in which the (predicted) moments—in particular, means, variances, and covariances—of returns of assets in the choice menu are time-varying and driven by the realization of a Markov chain within a MSIAH (or MSMAH) process. Besides the econometric estimates of a Markov switching model—in our case to be identified with those presented and commented in Section 3.3—two basic ingredients inform the construction of a MSMVF: how to go from MSIAH parameter estimates to predictions of means, variances, and covariances of portfolio returns; the investment horizon for which the MSMVF is to be built.[23] Before presenting some illustrative results based on the MSIH(2,0) estimates, we describe how forecasts of future moments may be computed in a Markov switching framework.[24]

To get some intuition on the factors that determine the predictions of means and variances of asset

[22]It turns out that Japanese expected returns are not strongly affected by regimes and are always close to zero. This finding corresponds to what Guidolin and Timmermann (2008b) have reported within a different type of MS models and to the common perception that Japanese markets may have fallen in a "slump" since the early 1990s, i.e., over most of our historical sample. Moreover, Pacific ex-Japan stocks yield a positive but insignificant expected return, that is lower than the estimate obtained for the first regime.

[23]Notice that in standard unconditonal (implicitly, single-state Gaussian IID analysis) mean-variance analysis, the investment horizon makes no difference because the forecast of future means, variances, and covariances are the currently estimable means, variances, and covariances. Of course, the horizon will matter in the presence of predictability, as captured (both linearly and non-linearly) by a MSIAH model.

[24]To save space, we only deal with the simple case of a MSIH(2,0) model. A more complete treatment of how model estimates map into moment forecasts may be found in Guidolin and Timmermann (2009) or Ria (2008).

returns under a MSIH process for returns, consider first the case of a single risky asset ($n = 1$), $y_t = \mu_{S_t} + \sigma^2_{S_t}\varepsilon_t$, where $\varepsilon_t \sim N(0,1)$. Call $\hat{\boldsymbol{\xi}}_{t|t}$ the 2×1 vector of (filtered) probabilities of being in each of the two alternative regimes, based only upon the information available up to time t, $\hat{\boldsymbol{\xi}}_{t|t} \equiv [\Pr(S_t = 1|\mathcal{F}_t)\ \Pr(S_t = 2|\mathcal{F}_t)]'$. Notice that the (row) vector of time $t+1$ predicted probabilities of the two states can be computed as:

$$\hat{\boldsymbol{\xi}}_{t+1|t} = [\hat{\xi}_{1,t|t}\hat{p}_{11} + \hat{\xi}_{2,t|t}(1-\hat{p}_{22}) \quad \hat{\xi}_{1,t|t}(1-\hat{p}_{11}) + \hat{\xi}_{2,t|t}\hat{p}_{22}] = \hat{\boldsymbol{\xi}}'_{t|t}\hat{\mathbf{P}},$$

where $\hat{p}_{ij}$ is the row-i, column-j element of the transition probability estimate, $\hat{\mathbf{P}}$.[25] The predicted mean for period $t+1$ is then

$$\begin{aligned} E_t[y_{t+1}] &= \sum_{S_{t+1}=1}^{2} E_t[y_{t+1}|S_{t+1}]\Pr(S_{t+1}|\mathcal{F}_t) = \sum_{S_{t+1}=1}^{2} E_t[y_{t+1}|S_{t+1}]\Pr(S_{t+1}|\mathcal{F}_t, S_t)\Pr(S_t|\mathcal{F}_t) \\ &= \hat{\xi}_{1,t|t}\hat{p}_{11}\hat{\mu}_1 + \hat{\xi}_{1,t|t}(1-\hat{p}_{11})\hat{\mu}_2 + (1-\hat{\xi}_{1,t|t})(1-\hat{p}_{22})\hat{\mu}_1 + (1-\hat{\xi}_{1,t|t})\hat{p}_{22}\hat{\mu}_2. \end{aligned} \tag{9}$$

In general, and extending this result to the case of a T-step ahead forecast, we have that:

$$E_t[r_{t+T}] = \hat{\boldsymbol{\xi}}'_t\hat{\mathbf{P}}^T\hat{\boldsymbol{\mu}}^*, \tag{10}$$

where $\hat{\boldsymbol{\mu}}^*$ is a matrix that stacks in each row the regime-dependent mean return estimates for each asset (here $\hat{\boldsymbol{\mu}}^*$ is a 2×1 vector because there is only one asset). In the case of a generic number n of assets, this expression easily generalizes to: $E_t[\mathbf{y}_{t+T}] = \hat{\boldsymbol{\xi}}'_t\hat{\mathbf{P}}^T\hat{\boldsymbol{\mu}}^*$ where now $\hat{\boldsymbol{\mu}}^*$ is $2 \times n$ and $E_t[\mathbf{y}_{t+T}]$ yields a $1 \times n$ vector of predicted means. Next, consider the prediction of the variance of one asset return in period $t+1$ conditional on the information set $\mathcal{F}_t$:

$$\begin{aligned} Var_t[y_{t+1}] &= \sum_{S_{t+1}=1}^{2} E_t[(y_{t+1} - E_t[y_{t+1}])^2|S_{t+1}]\Pr(S_{t+1}|\mathcal{F}_t, S_t)\Pr(S_t|\mathcal{F}_t) \\ &= \hat{\xi}_{1,t|t}\hat{p}_{11}E[(\hat{\mu}_1 - E_t[y_{t+1}] + \hat{\sigma}_1\varepsilon_{t+1})^2] + \hat{\xi}_{1,t|t}(1-\hat{p}_{11})E[(\hat{\mu}_2 - E_t[y_{t+1}] + \hat{\sigma}_2\varepsilon_{t+1})^2] + \end{aligned}$$

$$\begin{aligned} &+ (1-\hat{\xi}_{1,t|t})(1-\hat{p}_{22})E[(\hat{\mu}_1 - E_t[y_{t+1}]\hat{\sigma}_1\varepsilon_{t+1})^2] + (1-\hat{\xi}_{1,t|t})\hat{p}_{22}E[(\hat{\mu}_2 - E_t[y_{t+1}] + \hat{\sigma}_2\varepsilon_{t+1})^2] \\ &= \hat{\boldsymbol{\xi}}'_{t|t}\hat{\mathbf{P}}\begin{bmatrix} (\hat{\mu}_1 - E_t[y_{t+1}])^2 + \hat{\sigma}_1^2 \\ (\hat{\mu}_2 - E_t[y_{t+1}])^2 + \hat{\sigma}_2^2 \end{bmatrix}. \end{aligned}$$

Once again, this is easily extended to the conditional variance of returns in period $t+T$:

$$Var_t[y_{t+T}] = \hat{\boldsymbol{\xi}}'_{t|t}\hat{\mathbf{P}}^T\begin{bmatrix} (\hat{\mu}_1 - E_t[y_{t+T}])^2 + \hat{\sigma}_1^2 \\ (\hat{\mu}_2 - E_t[y_{t+T}])^2 + \hat{\sigma}_2^2 \end{bmatrix}.$$

The implication is that unless $\hat{\mu}_1 = \hat{\mu}_2 = E_t[y_{t+T}]$ (i.e., unless there is no regime switching in the conditional mean function),

$$Var_t[y_{t+T}] > \hat{\boldsymbol{\xi}}'_{t|t}\hat{\mathbf{P}}^T\begin{bmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{bmatrix} = \hat{\xi}_{1,t+T|t}\hat{\sigma}_1^2 + \hat{\xi}_{2,t+T|t}\hat{\sigma}_2^2,$$

[25] In a similar fashion (see Timmermann, 2000), one can show that the vector of time $t+T$ predicted probabilities of the two states is given by $\hat{\boldsymbol{\xi}}'_{t+1|t} = \hat{\boldsymbol{\xi}}'_{t|t}\mathbf{P}^T$, where $\mathbf{P}^T \equiv \prod_{j=1}^{T}\mathbf{P}$ and $\mathbf{P}^0 = \mathbf{I}_2$.

where the right-hand side expression is simply a predicted state-probability weighted combination of the two-regime specific variance estimates. This means that Markov switching in the conditional mean always increases – as $(\hat{\mu}_k - E_t[y_{t+T}])^2 > 0$ for $k = 1, 2$ – the conditional variance of the asset return.

These results are easy to generalize to the case of multiple assets, $n \geq 2$, which is obviously the relevant case in an asset allocation perspective, see Guidolin and Timmermann (2009) and Ria (2008). Consider a portfolio of assets summarized by the $n \times 1$ vector of percentage portfolio weights at time t, $\boldsymbol{\omega}_t$. For simplicity, we still refer to the MSIH(2,0) model that well characterizes the MSCI international index data analyzed earlier on. Under a simple MSIH(2,0) switching model, the return on the portfolio, r^p_{t+1}, is:

$$r^p_{t+1} = \boldsymbol{\omega}_t'\mathbf{y}_{t+1} = \boldsymbol{\omega}_t'\boldsymbol{\mu}_{S_{t+1}} + \boldsymbol{\omega}_t'\boldsymbol{\Sigma}_{S_t}\boldsymbol{\varepsilon}_{t+1}.$$

The expected portfolio return next period is then simply

$$E_t[r^p_{t+1}] = \boldsymbol{\omega}_t'E_t[\mathbf{y}_{t+T}] = \hat{\boldsymbol{\xi}}_t'\hat{\mathbf{P}}^T\hat{\boldsymbol{\mu}}^*\boldsymbol{\omega}_t,$$

The variance of portfolio returns can be written in the following form

$$\begin{aligned} E_t\left[(r^p_{t+T} - \bar{\mu}^p_{t+T})^2\right] = \boldsymbol{\omega}_t'(&\hat{\xi}_{1,t+T|t}E[(\mathbf{y}_{t+T} - E_t[\mathbf{y}_{t+T}])(\mathbf{y}_{t+T} - E_t[\mathbf{y}_{t+T}])'|S_{t+T} = 1] + \\ &+ \hat{\xi}_{2,t+T|t}E[(\mathbf{y}_{t+T} - E_t[\mathbf{y}_{t+T}])(\mathbf{y}_{t+T} - E_t[\mathbf{y}_{t+T}])'|S_{t+T} = 2])\boldsymbol{\omega}_t, \end{aligned}$$

where the $n \times n$ matrix of squared return deviations from the mean in state S_{t+T} is given by

$$E[(\mathbf{y}_{t+T} - E_t[\mathbf{y}_{t+T}])(\mathbf{y}_{t+T} - E_t[\mathbf{y}_{t+T}])'|S_{t+T}] = \boldsymbol{\Omega}_{S_{t+T}} +$$

$$+ \begin{bmatrix} \left(\mu^1_{S_{t+T}} - E_t[y^1_{t+T}]\right)^2 & \left(\mu^1_{S_{t+T}} - E_t[y^1_{t+T}]\right)\left(\mu^2_{S_{t+T}} - E_t[y^2_{t+T}]\right) & \cdots & \left(\mu^1_{S_{t+T}} - E_t[y^1_{t+T}]\right)\left(\mu^n_{S_{t+T}} - E_t[y^n_{t+T}]\right) \\ \left(\mu^2_{S_{t+T}} - E_t[y^2_{t+T}]\right)\left(\mu^1_{S_{t+T}} - E_t[y^1_{t+T}]\right) & \left(\mu^2_{S_{t+T}} - E_t[y^2_{t+T}]\right)^2 & \cdots & \left(\mu^2_{S_{t+T}} - E_t[y^2_{t+T}]\right)\left(\mu^n_{S_{t+T}} - E_t[y^n_{t+T}]\right) \\ \vdots & \vdots & \ddots & \vdots \\ \left(\mu^n_{S_{t+T}} - E_t[y^n_{t+T}]\right)\left(\mu^1_{S_{t+T}} - E_t[y^1_{t+T}]\right) & \left(\mu^n_{S_{t+T}} - E_t[y^n_{t+T}]\right)\left(\mu^2_{S_{t+T}} - E_t[y^2_{t+T}]\right) & \cdots & \left(\mu^n_{S_{t+T}} - E_t[y^n_{t+T}]\right)^2 \end{bmatrix}$$

The second term, which shows the deviations of the state-specific conditional means from its overall expectation in each state and for each asset, does not arise in single-state models. This term could be potentially important for portfolio allocation purposes. The first term is the standard, regime-specific variance covariance matrix $\boldsymbol{\Omega}_{S_{t+1}}$. Again, the implication is that unless $\mu^i_1 = \mu^i_2$ $i = 1, 2, ..., n$ (no regime switching in the conditional mean, for none of the assets or portfolios under investigation), the conditional variance of the T-period ahead portfolio returns is not simply $\boldsymbol{\omega}_t'(\hat{\xi}_{1,t+T|t}\hat{\boldsymbol{\Omega}}_1 + \hat{\xi}_{2,t+T|t}\hat{\boldsymbol{\Omega}}_2)\boldsymbol{\omega}_t$, involving instead a complex matrix reflecting cross products of deviations of the conditional means from the unconditional means for the assets, taken in pairs.

A MSMVF reflects the basic intuition of Markowitz (1952): investors should decide optimal portfolio weights on the basis of the trade-off between portfolio risks and expected returns; risk should be measured by the variance of portfolio returns. Moreover, for any given level of expected return, a rational, risk-averse investor will choose the portfolio with minimum variance from amongst the set of all possible portfolios. In particular, suppose that the goal of our investor is to select a portfolio composed of n risky assets, in the form of a vector of portfolio weights $\boldsymbol{\omega}_t^T \equiv [\omega_{1t}^T \; \omega_{2t}^T \; ... \; \omega_{nt}^T]'$, such that $\sum_{i=1}^n \omega_{it}^T = (\boldsymbol{\omega}_t^T)' \boldsymbol{\iota}_n = 1$. Unless we shall state otherwise, notice that $\omega_{it}^T < 0$ (hence, $\omega_{jt}^T > 1$, $i \neq j$) is admissible, i.e. short sales are possible. Also notice that the notation $\boldsymbol{\omega}_t^T \equiv \boldsymbol{\omega}^T(\mathcal{F}_t)$ stresses that the weights are selected at time t and only conditional upon information available at time t. The investor has a risk-return trade-off goal over T periods; therefore she cares for "optimizing" the trade-off between the portfolio expected return and variance over the interval $[t, t+T]$. The investor's problem may be formulated as a simple constrained minimization:

$$\min_{\boldsymbol{\omega}_t^T} Var_t[r_{t+T}^p; \boldsymbol{\omega}_t^T] \qquad s.t. \text{ (i) } \breve{\mu} = E_t[r_{t+T}^p; \boldsymbol{\omega}_t^T]; \quad \text{(ii) } (\boldsymbol{\omega}_t^T)' \boldsymbol{\iota}_n = 1, \tag{11}$$

where r_{t+T}^p is the (continuously compounded) portfolio return T-period forward, $\breve{\mu}$ is the desired mean portfolio goal, and the notations $E_t[r_{t+T}^p; \boldsymbol{\omega}_t^T]$ and $Var_t[r_{t+T}^p; \boldsymbol{\omega}_t^T]$ want to stress that the (predicted) time t expectation and variance of a portfolio depend on the selected portfolio weights $\boldsymbol{\omega}_t^T$ in the ways discussed early on. As discussed in Fabozzi, Focardi, and Kolm (2006), this is a simple quadratic optimization problem that can be solved with the method of Lagrange multipliers. The resulting $\hat{\boldsymbol{\omega}}_t^T(\breve{\mu})$ will be a straightforward (yet, highly non-linear) function of the basic Markov switching parameter matrices, $\hat{\boldsymbol{\mu}}^*$, $\hat{\boldsymbol{\Sigma}}^*$, $\hat{\mathbf{P}}$, as well as the (filtered) state probabilities collected in the vector $\hat{\boldsymbol{\xi}}_t$. Since $\hat{\boldsymbol{\xi}}_t \in \mathcal{F}_t$ and the perceived state probabilities change over time, the result is that $\forall \breve{\mu}$, the variance-minimizing weights become themselves a function of either the current state S_t—if known—or at least of the vector of state probabilities $\hat{\boldsymbol{\xi}}_t$.

Consider now solving the program in (11) for all possible, different choices of $\breve{\mu} \in (-1, +\infty)$. This delivers a set of variance minimizing weights $\{\hat{\boldsymbol{\omega}}_t^T(\breve{\mu}); \breve{\mu} \in (-1, +\infty)\}$. In correspondence to each vector $\hat{\boldsymbol{\omega}}_t^T(\breve{\mu})$ it is then possible to computed the associated portfolio expected return and risk (where in fact $E_t[r_{t+T}^p; \hat{\boldsymbol{\omega}}_t^T(\mu^*)] = \breve{\mu}$ by construction). The set of all possible mean-variance combinations induced by $\{\hat{\boldsymbol{\omega}}_t^T(\breve{\mu}); \breve{\mu} \in (-1, +\infty)\}$ is the mean-variance frontier. Since in the presence of Markov switching dynamics all optimal variance-minimizing portfolios in $\{\hat{\boldsymbol{\omega}}_t^T(\breve{\mu}); \breve{\mu} \in (-1, +\infty)\}$ will generally depend on the state (or the perception of the state, as captured by the vector $\hat{\boldsymbol{\xi}}_t$) the resulting MSMVF will be state-dependent. Figure 3 shows three sets of MSMVFs computed in this way, each corresponding to a different choice of the horizon T, i.e., 1-, 6-, and 36-month ahead. Within each set, four different mean-variance frontiers are plotted. Three of them are MSMVFs and correspond to three alternative and key configurations of the (filtered) state vector $\hat{\boldsymbol{\xi}}_t$, i.e., when an investor has knowledge of the current regime being bear/high correlation, being bull/low correlation, or when an investor ignores the nature of the current regime and simply assigned to each of the two states a probability equal to their long-run, ergodic probabilities. The latter case corresponds to a plausible situation of ignorance on the nature of the current state. A fourth frontier is provided as a benchmark and simply corresponds to the (time invariant) frontier an investor would derive from a standard, Gaussian IID model as in top panel of Table 3.

Clearly, modelling Markov regimes has massive effects when the investment horizon is short: the bull and bear MSMVFs are substantially different, with the bull MSMVF implying a substantially better risk-return trade-off for low-to-intermediate levels of volatility but a worse trade-off for higher volatility levels. For instance, if an investor were to expect a portfolio return of $\breve{\mu} = 0.015$ (i.e., 18% per year), in the bull regime she would be able to reach this goal bearing a rather moderate risk of 16.5% per year; however, the same expected return target in a bear regime would force the investor to accept a much higher risk of 22.9% per year. However, if the target expected returns were to be $\breve{\mu} = 0.04$ (i.e., a rather aggressive 48% per year), in the bull regime she would be able to reach this goal bearing a higher risk (62% per year) than under the bear regime (54.8% per year). This means that while the higher regime-specific expected returns and lower correlations do help an investor to achieve a good risk-reward trade-off in the bull state, this works only for moderate levels of $\breve{\mu}$; for levels in excess of 3% per month, the fact that the two regimes do not possess a clear ranking across volatilities leads to the existence of a crossing point between the bull and bear MSMVF.

Interestingly, already for $T = 1$ month, the ergodic and single-state (IID) MSMVFs are very hard to tell apart from the Figure. Importantly, this does not have to happen as a result of any statistical property: Markov switching models produce ergodic joint densities for the variables that are neither Gaussian nor even approximately similar to single-state models (see Guidolin and Timmermann, 2007, for related comments). However, in this illustration, this turns out to be the case: the ergodic density implied by our MSIH(2,0) originates a MSMVF which is very close to the classical MVF. The bottom panel of Figure 2 shows that provided an investor has a sufficiently long horizon, the bull, bear, and ergodic MSMVFs all come to coincide. This is to be expected as the longer the horizon, the higher is the chance that the predicted state probabilities used to compute predicted means, variances, and covariances will come to essentially coincide with the model-implied ergodic state probabilities.[26] Since the ergodic MSMVF clearly cannot depend on T, for a sufficiently long horizon it also happens that bull, bear, and ergodic MSMVFs all converge to the single-state MSMVF. Finally, the intermediate panel of the Figure shows the case of $T = 6$ months. Clearly, this plot falls in-between the top and bottom panels, even though $T = 6$ has been selected to show that Markov switching effects are not entirely short-lived and will potentially affect optimal portfolio choices for horizons that are plausible in practice.

3.5. *Portfolio Implications*

As a last illustrative step, we have also computed portfolio weights using simple mean-variance preferences. One can interpret such an exercise as equivalent to computing the MSMVFs in Section 3.4 and then proceeding to select an optimal vector of weights $\tilde{\boldsymbol{\omega}}_t^T$ after super-imposing on the plots in Figure 2 some standard sets of mean-variance indifference curves that trade-off mean and variance to hold the investor indifferent across alternative portfolios. Similarly to Guidolin and Na (2009), consider an investor with

[26] The definition of ergodic state probabilities implies that $\lim_{T\to\infty} \boldsymbol{\xi}_{t+T|t} = \bar{\boldsymbol{\xi}}$.

the simple objective:[27]

$$\max_{\boldsymbol{\omega}_t^T}\left\{V_{t:t+T}(\boldsymbol{\omega}_t^T)\equiv E_t[r_{t:t+T}^p;\boldsymbol{\omega}_t^T]-\frac{\lambda}{2}Var_t[r_{t:t+T}^p;\boldsymbol{\omega}_t^T]\right\}\quad s.t.\ (\boldsymbol{\omega}_t^T)'\boldsymbol{\iota}_n=1, \tag{12}$$

where $r_{t:t+T}^p\equiv\sum_{\tau=1}^T r_{t+\tau}^p$ is the total, cumulative portfolio return between time t and time $t+T$, $\lambda>0$ is the coefficient of risk aversion characterizing the investor's preferences. As discussed in Fabozzi, Focardi, and Kolm (2006), such an objective may be derived from an expected utility maximization problem when the investor has preferences described by a quadratic utility function over final wealth at a certain future date $t+T$. At time t the investor maximizes the expected utility objective (12) by implementing a simple buy-and-hold strategy (no dynamic rebalancing) in which $\boldsymbol{\omega}_t^T$ is selected at time t and held up to time T.

In particular, we have proceeded to compute optimal international diversification weights (among developed markets) on the basis of the MSIH(2,0) estimates of Section 3.3 and setting λ to (locally) match the behavior of an investor with constant coefficient of relative risk aversion of 5.[28] We have performed this exercise recursively between 1998:01 and 2010:09. Table 4 reports a number of summary statistics for these recursive sets of (153) portfolio weights for the cases $T=1$, 12, and 120 months and for the MSIH(2,0) and the single-state models. The exercise considers both the case in which the constraints $(\boldsymbol{\omega}_t^T)'\mathbf{e}_j\in[0,1]$ $(j=1,...,n)$ are imposed (and this implies that the optimization in (12) has to be solved numerically), and the unconstrained case. The table reports means, medians, standard deviations and the 10% empirical confidence bands for recursive portfolio weights.[29] The upper panel of Table 4 considers the case in which short-sale constraints are imposed, while the lower panel deals with the unconstrained case. The table shows important differences between single- and Markov switching recursive portfolio weights. For instance, focussing on the case in which short sales are admitted, while a single-state model implies that Pacific ex-Japan for $T=1$ should receive a modest average weight of 0.6% (basically nothing when the median is used), the two-state MSIH model yields an average weight of 23% with rather large spikes in correspondence to periods of crisis in financial markets, as revealed by the fact that the median weight is 4% only; similarly, the weight to North American stocks is an overwhelming 87% (the median is 94%) under a no-predictability Gaussian IID model vs. 37% (35% using the median) under a regime switching MVF. The fact that Pacific ex-Japan, Japanese, and (to a smaller extent) UK stock weights are inflated by Markov switching dynamics is compensated by the lower weights assigned to North American stocks.

[27] At times we will impose the absence of short sales, $\boldsymbol{\omega}_t^T\geq\mathbf{0}$.

[28] Suppose $W_t=1$ and take a second order Taylor series expansion of a power utility function $W_{t+T}^{1-\gamma}/(1-\gamma)$ $(\gamma>0)$ around $\nu=\exp(r^fT)$:

$$u(W_{t+T})\simeq\frac{\nu^{1-\gamma}}{1-\gamma}+\nu^{-\gamma}(W_{t+T}-\nu)-\frac{1}{2}\gamma\nu^{-(\gamma+1)}(W_{t+T}-\nu)^2,$$

where we used that $u'(\nu)=\nu^{-\gamma}$, and $u''(\nu)=-\gamma\nu^{-(\gamma+1)}$. Expanding the powers of $(W_{t+T}-\nu)$ and taking the expectation conditional on information up to time t, one can show that

$$E_t[u(W_{t+T})]\simeq\kappa_0(\gamma)+\kappa_1'(\gamma)E_t[W_{t+T}]+\kappa_2(\gamma)Var_t[W_{t+T}]\propto E_t[W_{t+T}]-\lambda Var_t[W_{t+T}]$$

where $\kappa_2(\gamma)\equiv-\frac{1}{4}\gamma\nu^{-(1+\gamma)}\left[2+2(\gamma+1)+(\gamma+1)(\gamma+2)\right]<0$, i.e., λ may be interpreted as a complicated non-linear function of γ.

[29] In practice, the 10% lower bound is the 5th percentile and the 10% upper bound is the 95th percentile of the empirical distribution of optimal portfolio weights. In the table, we have boldfaced 10% confidence bands that fail to include zero, i.e., indices for which the exercise gives a clear indication as to the sign of the average committment to the portfolio.

Interestingly, when the constraints are not imposed in the lower panel of Table 4, some of the qualitative implications of the upper bound hold (basically, MS implies larger weights to Japanese and Pacific stocks and lower weights to continental European stocks), but others fail to hold: for instance, the average UK weights are higher under Markov switching in the case of no short sales, but are lower when no constraints are imposed.

A few other facts emerge from Table 4. As one would expect, MSIH weights are much more volatile than single-state weights are, and this difference is particularly strong in the $T = 1$ case and when no constraints are imposed. This should be expected because an investor that uses the Markov switching framework will actively try to time the international markets' bull and bear regimes and to tailor her optimal risk-return trade-off on the basis of the underlying dynamics. Interestingly, the differences between single- and two-state mean-variance weights *increases* as the investment horizon T grows larger. This is in no way a contradiction of our earlier remark that the MSMVFs converge to the single-state, classical MVF as T grows: notice in fact that while the MSMVFs simply reflect the trade-off between predicted means and variances for all possible portfolios T-step ahead, the portfolio problem in (12) concerns T-horizon, cumulative portfolio weights defined as $r^p_{t:t+T} \equiv \sum_{\tau=1}^{T} r^p_{t+\tau}$, where $r^p_{t+\tau} \equiv \boldsymbol{\omega}_t' \mathbf{y}_t$ and $\mathbf{y}_t$ is in our application a $n \times 1$ vector of portfolio weights. This means that the 120-month portfolio weights reported in Table 4 will reflect *cumulative* deviations of MSMVFs from the single-state, no predictability MVF, so that even though the MSMVFs do converge to the single-state MVF, the portfolio weights do not have to.

These differences are visualized by Figure 3, when a no short-sale constraint is imposed on the portfolio exercise. The most striking difference is the existence of periods (such as 2003-2004, i.e., during a span of time best characterized as a bear, high-correlation market) in which under Markov switching the demand for Japanese stocks should be positive and non-negligible, which is never the case under a classical single-state model. From Table 3, this is easily justified by the fact that—even though they give very small positive expected returns—Japanese stocks imply modest risk (their variance is among the lowest in the first regime) and are useful hedging tools, since their pair-wise correlations with returns on other equity indices are among the smallest (e.g., always below 0.7) in the bear regime. Equally visible is that fact that—especially in correspondence to high correlation periods (e.g., 2002-2007), while the Markov switching model suggests relative large investments in Pacific stocks, this is not the case under a single-state model. This is easily explained with the fact that Pacific stocks give high expected returns exactly in the bear, high-correlation regime and this provides an important hedging opportunity even to investors with medium- and long-investment horizons, given the high persistence of the first regime. Finally, it is clear that while a classical, Gaussian IID mean-variance strategy would suggest large and persistent weights invested in North American equities (never below 50% in our recursive exercise), this is not the case under a MSMVF case, when for instance the weight to North American stocks drops to zero between mid-2008 and early 2009, in correspondence to the recent financial crisis. The intuition is that during high correlation period, North American stocks yield low (Japanese-style) expected returns and have poor diversification properties, in the sense that their correlations with other indices (with the only exception of the Japanese) are generally close to 0.9.[30]

[30]During the period August 2008 - March 2009, a MS investor would have invested relatively large weights (in excess of

4. **Recursive Out-of-Sample Performance: Should We Be Tracking Regime Shifts in Efficient Frontiers?**

The empirical results in Section 3.5 cannot of course establish in any way that taking into account the presence of regime shifts in MVFs is important in practice, i.e., that it allows the creation of economic value to a portfolio manager (or its customers). In fact, so far we have purely documented a number of in-sample properties of the multivariate time series of MSCI international equity returns, such as the presence of two distinct regimes for expected returns and pair-wise correlations. However, economic value will be produced only if by taking into account regime shifts in MVFs, an investor could increase her realized, out-of-sample portfolio returns per unit of risk. In this Section we therefore proceed to set up a recursive experiment to assess whether—at least in the case of our application to international portfolio diversification—such improvement in risk-adjusted realized performance may be attained. Section 4.1 briefly describes the nature of the recursive experiment. Section 4.2 reports the key results. Section 4.3 briefly examines how the strategies would have performed during the recent financial crisis.

4.1. *Design of the Recursive Exercise*

We perform a standard exercise in real time asset allocation based on a fully recursive scheme of model estimation and portfolio optimization. In particular, we initialize our experiment using data from January 1988 up to December 1997 to estimate the parameters of our two competing models, a single-state Gaussian IID model and the MSIH(2,0) model that emerges as the best trade-off between fit and parsimony in Table 2. Based on the parameter (and state probability) estimates obtained over the sample 1988:01-1997:12 (for a total of 600 observations), we proceed to forecast multi-period means, variances, and covariances of returns on all equity portfolios, which allow us to determine mean-variance portfolio weights, taking into account of the nature of (inference on) the current state in the case of MSIH, as well as the likelihood of subsequent regime switches. This is done imputing to our "hypothetical" investor a range of alternative, potential investment horizons parameterized by T; in fact we use 3 alternative horizons, of 1, 12, and 120 months.[31] These recursive estimation and portfolio choice exercises are repeated on the following month, using data from January 1988 and up to January 1998 (for a total of 605 observations) to compute afresh forecasts of moments and to select optimal portfolio weights. Iterating this recursive scheme until September 2010 (when all the available 1,365 observations are employed) generates a sequence of 153 sets of optimal portfolio shares – importantly, one for each possible investment horizon – as well as realized portfolio returns from such ex-ante optimal choices, from which ex-post performance measures for these alternative portfolio strategies and horizons may be computed.

50% for T in excess of 6 months) in Japanese stocks. Similar plots are available upon request from the Authors for the case of no short sale constraints. Their qualitative implications are similar to those in Figure 3, although there is naturally much more variability in optimal weights.

[31] We perform these calculations for a range of alternative coefficients of risk aversion ($\gamma = 2$, 5, 10, and 20, when expressed in terms of the underlying coefficient of relative risk aversion). Although many qualitative conclusions are not sensitive to the specific risk aversion coefficient imputed, to save space we report only on the case of $\gamma = 5$. Further, detailed results are available upon request. Fugazza et al. (2008) report further comments on the logics and limitations of recursive back-testing exercises in asset allocation applications.

Tables 5 and 6 report average performances using a number of alternative metrics. Table 5 reports the annualized mean, the annualized standard deviation, and the corresponding Sharpe ratio for each of four strategies. For all these statistics, we have block-bootstrapped their 90% confidence interval using the realized performance measures, which means a total of $153 - T$ observations, for $T = 1$, 12, and 120.[32] Besides the single- and two-state Markov switching mean-variance strategies, we also report realized performance statistics for two additional benchmarks that have played an important role in the recent literature: an equally-weighted portfolio (also called "1/N" after De Miguel et al., 2009) in which 20% of the portfolio is invested in each of the five equity portfolios at all points in time; a value-weighted portfolio in which at each point in time the percentage invested equals the relative importance of each index on the sum of the aggregate market capitalizations of the five indices, which is the basic prescription of the (international) CAPM. In Table 5, we also proceed to present 90% confidence intervals for Sharpe ratios following Opdyke (2007). Following and generalizing seminal work by Jobson and Korkie (1981), Opdyke reminds us that in finite samples, computing the Sharpe ratio for a given strategy as a simple ratio between the sample mean excess return and sample standard deviation ($SR_T = \bar{r}_T/s_T$) leads to a biased estimation of the Sharpe ratio; for instance, using a Taylor approximation to the order $1/(153-T)^2$ (i.e., of order $O(1/(153-T)^2)$), they show that

$$E[SR_T] \simeq \frac{\mu}{\sigma}\left[1 + \frac{1}{4(153-T)}\left(\frac{E[(r-\mu)^4]}{\sigma^4} - 1\right)\right],$$

where μ is the mean excess return, σ is the standard deviation of excess returns, and $E[(r-\mu)^4]/\sigma^4$ is simply the kurtosis of portfolio returns. The asymptotic distribution of the statistic SR_T is obtained by an application of the delta method which holds under rather general assumptions on the multivariate distribution of portfolio returns (stationarity and ergodicity, not requiring normality):

$$SR_T = \frac{\bar{r}_T}{s_T} \overset{asy}{\sim} N\left(\frac{\mu}{\sigma}, \frac{1}{153-T}\left[1 + \frac{\mu^2}{4\sigma^2}\left(\frac{E[(r-\mu)^4]}{\sigma^4} - 1\right) - \frac{\mu}{\sigma}\frac{E[(r-\mu)^3]}{\sigma^3}\right]\right),$$

where the variance is derived from the Taylor series approximation. In Table 5 we present 90% (asymptotic) confidence intervals for the bias-corrected Sharpe ratios:

$$SR_T = \frac{\bar{r}_T/s_T}{\left[1 + \frac{1}{4(153-T)}\left(\left(T^{-1}\sum_{t=1}^{T}(r_t - \bar{r}_T)^4\right)/s_T^4 - 1\right)\right]}$$

from portfolio strategies with and without real estate, both in the classical and in the Bayesian framework.

Table 6 supplements the key information in Table 5 by also computing the Jensen's alpha of each strategy (along with asymptotic 90% confidence intervals), and the skewness and kurtosis of realized portfolio returns. Finally, as recently discussed in the literature (see e.g., Adcock, 2007) is that Sharpe ratios are highly sensitive to non-normally distributed returns. We find strong evidence of non-normal portfolio returns in our sample (see Table 6), especially because returns appear to be skewed. One common remedy consists of supplementing the presentation of Sharpe ratios with related reward-to-risk measures

[32] The application of a block bootstrap is particularly important because our recursive exercise generates realized performances for *overlapping* horizons, which is likely to cause complex serial correlation patterns in performance measures, especially in the case of $T = 12$ and 120.

that divide the numerator (mean excess return) of the Sharpe ratio by portfolio downside risk (semi-standard deviation). This ratio commonly called the Sortino ratio (also see Fishburn, 1977).[33]

4.2. *Performance Results*

Table 5 shows the key results of this paper. In qualitative terms, it is clear that taking into account the presence of regime shifts in the MVF tends to increase—relative to the single-state case, when regime shifts are ignored—realized mean returns but also to increase the realized volatility of portfolio returns. For most combinations of horizons and constraints (i.e., whether short sales are allowed or not), the first effect prevails and—both using standard and corrected Sharpe ratios—the two-state model outperforms the single-state model: this is the case of $T = 1$ month and no short sales and of $T = 12$ and 120 independently of constraints. In this sense, tracking regime shifts in MVFs certainly has a positive economic value. For instance, for the case of $T = 12$ and with no short-sale constraints, the mean-variance regime switching weights yield a (corrected) Sharpe ratio of 0.31 vs. 0.10 for the single-state model, with a remarkable increase of 0.21. This derives from a mean performance (10.6% per year) that is almost double the mean single-state performance (6%), and from an annualized volatility (23%) which is only slightly higher than the 21% obtained in the single-state case. However, a number of issues and caveats remain. First, the two-state strategy does outperform the two additional benchmarks that we have tracked only in one cases: $T = 12$ when short-sales are not admitted, when the (corrected) Sharpe ratio of 0.31 also exceeds the 0.24 ratio that the equally weighted portfolio may yield. Otherwise, as recently stressed in a number of papers, it remains the case that the equally weighted portfolio may often outperform both the two- and single-state mean-variance strategies. Second, even when the two-state strategy outperforms the single-state one, the corresponding 90% confidence interval remain relatively wide and often overlapping. For instance, even in the case of $T = 12$ with no short sales, the 90% CI for the two-state strategy ([0.07, 0.57]) does overlap with the CI for the single-state strategy ([-0.07, 0.27]). Longer data sets that would allow more extensive recursive out-of-sample exercises would be required to shrink these confidence intervals to allow sharper inferences.[34]

Table 6 extends these comparisons in a number of directions. Interestingly, the highest (and sometimes statistically significant) Jensen's alphas are yielded by the single-state model. However, it remains unclear why—in the presence of pervasive nonlinearities and non-normalities generated by the presence of Markov switching dynamics (see Guidolin and Timmermann, 2008b)—should the CAPM be able to exactly fit the realized portfolio returns generated by any of the strategies examined. This is in fact witnessed by the statistics on skewness and kurtosis of realized returns in Table 6: under no short sales and short horizons, all strategies yield negative skewness; while the mean-variance strategies tend to produce excess kurtosis (thick tails of realized performance), the opposite occurs for the equally- and value-weighted benchmarks. This justifies reporting in Table 6 the Sortino ratios, for which 95% confidence intervals have also been

[33]When returns are normally distributed, total variance and semi-variance (which conditions on returns being below their mean) are identical. Deviations from normality imply that total and downside variance differ.

[34]We limit our comments concerning the recursive results for the case $T = 120$ because with our sample this could generate only 33 overlapping realized performances.

block-bootstrapped using the empirical distribution of realized portfolio performances. Under the Sortino metrics, the findings are starkly in favor of the two-state strategy, especially for $T = 12$ and to some extent also $T = 120$ (although caution is suggested by the short out-of-sample period). For instance, when $T = 12$ and with no short sales, the two-state Sortino is 0.46 vs. 0.17 for the single-state model; the bootstrapped CIs in this case barely overlap, [0.11, 0.83] vs. [-0.11, 0.41], although they remain rather wide. However, it remains true that also in a Sortino dimension, it remains hard (although not impossible) to outperform the equally- and value-weighted benchmarks.

4.3. *Could This Have Mattered During the Financial Crisis?*

Last but not least, we here tackle the question that probably harbors in the mind of many Readers: would an effort at tracking the dynamics in MSMVFs have produced any significant payoff during the recent financial crisis? We base our attempt at addressing this question on a few dating efforts that have recently appeared in the literature (see e.g., Guidolin and Tam, 2010) and that have dated the "Great Financial Crisis" to span the period that goes between August 2007 and the late Spring of 2009. For symmetry, we will take the crisis period to consist of 24 months, and to correspond to the sub-sample 2007:08-2009:07.

Some promising, preliminary evidence comes already from a careful observation of Figure 1: there seems to be very interesting changes in the current regime classification occurring between 2008 and 2009. This gives hopes that our two-state strategy may have turned out to be sensitive to the occurrence of the crisis. This is confirmed by recursive calculations of realized out-of-sample performance for 1- and 12-month investment horizons, but limited to the crisis period. For instance, this means that between 2007:08 and 2009:07 we compute 24 sets of 1-month mean-variance weights under the four strategies/models covered by Tables 5 and 6 and then compute their realized performances (between 2007:09 and 2009:08), obtaining a vector of 24 realized returns. Our hopes are confirmed in the case of the two-state strategy when no short-sales constraints are imposed: for $T = 1$, we record an annualized mean of 12.6% with a standard deviation of 46.8%, for an overall Sharpe ratio of 0.067; for $T = 12$, we record an annualized mean of 14.4% with a standard deviation of 37.2%, for an excellent Sharpe ratio of 0.38. These performances have to be contrasted by the negative Sharpe ratios that would have been delivered by a classical, single-state mean-variance strategy, -0.38 for $T = 1$ and -0.61 for $T = 12$. In fact, during the crisis, a standard mean-variance strategy heavily tilted towards US, UK, and continental European stocks would have been disastrous for rather obvious reasons, while the two-state tilt towards Japanese and especially Pacific stocks would have paid off handsomely. Interestingly, even the equally-weighted strategy would have been losing with reference to this particular sub-sample, with implied Sharpe ratios of -0.18 and -0.07 for $T = 1$ and 12 months, respectively.[35]

[35] For completeness, we also report the corresponding statistics for the case in which short sales are not allowed. For the two-state model and $T = 1$, we record an annualized mean of -7.2% with a standard deviation of 28%, for an overall Sharpe ratio of -0.09; for $T = 12$, we record an annualized mean of -3.5% with a standard deviation of 34%, for an excellent Sharpe ratio of -0.11. Under the single-state model, the Sharpe ratios are instead -0.21 and -0.15, i.e., they remain substantially lower than in the regime switching case. The intuition is that the impossibility, under the MSMVF, to sell short European (including UK) stocks does hurt performance.

5. **Conclusion**

This paper has illustrated the potential effects of basing a standard but key financial decision making exercise—the construction and use of the mean-variance frontier in a simple international asset allocation framework—on models from the Markov switching class that imply that expected returns, their volatility, and their correlations may depend on a discrete but unobservable Markov chain that illustrates transitions from "bad" to "good" investment opportunities, and viceversa. Of course, a number of important aspects of the application of Markov switching methods to strategic asset allocation, international diversification, and risk management have failed to play a role in our paper, such as the interaction between regime shifts and hedging demands when an investor frequently (continuously) rebalances her portfolio (see Guidolin and Timmermann, 2007, 2008a), the role of preferences in which either predictions of conditional moments higher than mean and variance (see Guidolin and Timmermann, 2008b) or the entire predictive density for all portfolios/assets enter the portfolio problem, such as in the power utility (constant relative risk aversion) case (see Guidolin and Timmermann, 2007), the role of transaction costs in potentially reducing the out-of-sample, realized payoffs from portfolio strategies that exploit the presence of regime shifts (see Guidolin and Na, 2009). Another interesting issue concerns the existence of estimation errors even within a Markov switching framework, because we have been computing time-varying MSMVFs conditioning on both the current regime and MLE parameter estimates whose distribution is ignored. Of course, it would be possible or even advisable to combine the ideas on regime switching MVFs in this paper with recent advances in the portfolio management literature concerning the value of resampling and robust estimation, for instance along the Bayesian lines recently followed by Tu (2010).

A number of other extensions concern instead the structure of the econometric framework to be taken to the data, in order to exploit its ability to fit and – especially – forecast. For instance, it would seem interesting to write and estimate multivariate models in which two different (potentially uncorrelated or even negatively correlated) Markov chains drive the process of stock returns, one state variable to fit the dynamics of conditional means and another state variable to capture any dynamics in higher order moments. However, such models will probably imply the need to specify a high number of regimes, for instance regimes even to consider the simplest possible case. In a similar framework, Dueker and Sola (2008) have recently explored the possibility that variables (in our case, stock indices) with different size or economic importance be given a different weight in influencing the inference on the state of the overall system. Especially in macroeconomic applications (where different countries or states may carry a rather different importance) this extension may be crucial. A related point (see e.g., Guidolin, 2009, Henry, 2009) is that some data sets seem to need to more complex Markov chain models in which also ARCH-type effects follow a regime switching process or in which conditional means are modeled as vector autoregressive processes that are themselves subject to Markov switching effects (see e.g., Bohl et al., 2009, Guidolin and Ono, 2006). Finally, a third line of work that has received some attention (see e.g., Ang and Bekaert, 2002a, 2004, Diebold et al., 1994) has made the transition matrix governing the Markov chain a function of exogenous variables that would therefore influence the persistence and predicted duration of different regimes. However, how that may differ (both technically, as Markov switching models with time-varying transition probabilities are notoriously hard to estimate, and in terms of their forecasting performance, see Foley, 2001) from including the same variables as endogenous in the system by expanding the multivariate dimension of the model (see e.g., Guidolin and Ono, 2006) remains unclear and worthwhile of future research.

References

[1] Adcock, C., (2007) "Measuring Portfolio Performance Using a Modified Measure of Risk." *Journal of Asset Management*, 7(6): 388-403.

[2] Ang, A. and Bekaert, G. (2002a) "International Asset Allocation with Regime Shifts." *Review of Financial Studies*, 15(4): 1137-1187.

[3] Ang, A. and Bekaert, G. (2002b) "Short Rate Nonlinearities and Regime Switches." *Journal of Economic Dynamics & Control*, 26(7-8) 1244-1275.

[4] Ang, A. and Bekaert, G. (2004) "How Regimes Affect Asset Allocation." *Financial Analysts Journal*, 60(2): 86-99.

[5] Baum, L., T. Petrie, G. Soules, and N. Weiss. (1970) "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *Annals of Mathematical Statistics*, 41, 164-171.

[6] Bohl, M., J. Brzeszczyński, and B. Wilfling. (2009) "Institutional Investors and Stock Returns Volatility: Empirical Evidence from a Natural Experiment." *Journal of Financial Stability*, 5(2): 170-182.

[7] Butler, K. C. and Joaquin, D. C. (2002) "Are the Gains from International Portfolio Diversification Exaggerated? The Influence of Downside Risk in Bear Markets." *Journal of International Money & Finance*, 21(7): 931-1011.

[8] Chow, G., E. Jacquier, M. Kritzman, and Lowry, K. (1999) "Optimal Portfolios in Good Times and Bad." *Financial Analysts Journal*, 55(3): 65-73.

[9] Clarke, R. and de Silva, H. (1998) "State-Dependent Asset Allocation." *Journal of Portfolio Management*, 24(2): 57-64.

[10] Das, S. R. and Uppal, R. (2004) "Systemic Risk and International Portfolio Choice." *Journal of Finance*, 59(6): 2809-2834.

[11] Davidson, R. and MacKinnon, J. (1981) "Several Tests for Model Specification in the Presence of Alternative Hypothesis." *Econometrica*, 49(3): 781-793.

[12] Davies, R. (1977) "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative." *Biometrika*, 64(2): 247-254.

[13] DeMiguel, V., L. Garlappi, and R., Uppal, 2009, "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?," *Review of Financial Studies*, 22, 5, 1915-1953.

[14] Diebold, F., J., Lee, and Weinbach, G. (1994) "Regime Switching with Time-Varying Transition Probabilities." in Hargreaves (ed.) *Non-Stationary Time Series Analysis and Co-Integration*, Oxford University Press, Oxford, 283-302.

[15] Doan, T., R. Littermann, and C. Sims. (1984) "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*, 3(1): 1-14.

[16] Dempster, A. P., M. Schatzoff, and N. Wermuth. (1977) "A Simulation Study of Alternatives to Ordinary Least Squares." *Journal of the American Statistical Association*, 72(357): 77-91.

[17] Dueker, M. J. and Sola, M. (2008) "Multivariate Markov Switching with Weighted Regime Determination: Giving France More Weight than Finland." Working Papers: 2008-001, Federal Reserve Bank of St. Louis.

[18] Fabozzi, F., S. Focardi and P. Kolm. (2006) *Financial Modeling of Equity Market: From CAPM to Cointegration*. Hoboken, NJ: John Wiley & Sons, Inc.

[19] Fishburn, P., 1977, "Mean-Risk Analysis with Risk Associated with Below Market Returns", *American Economic Review*, 67(2), 116-126.

[20] Foley, A. (2001) "Regime Switching in Currency Markets and Portfolio Flows." *Journal of Asset Management*, 2(4), 353-367.

[21] Fruhwirth-Schnatter, S. (2006) *Finite Mixture and Markov Switching Models*. NY: Springer.

[22] Fugazza, C., M., Guidolin, and G., Nicodano (2008) "Diversifying in Public Real Estate: The Ex-Post Performance." *Journal of Asset Management*, 8(6), 361-373.

[23] Guidolin, M. (2009) "Detecting and Exploiting Regime Switching ARCH Dynamics in US Stock and Bond Returns." In: G. Gregoriou (Ed.), *Stock Market Volatility*. Chapman Hall: London.

[24] Guidolin, M. and Hyde, S. (2008) "Equity Portfolio Diversification under Time-varying Predictability: Evidence from Ireland, the US, and the UK." *Journal of Multinational Financial Management*, 18(4), 293-312.

[25] Guidolin, M. and Na, C. F. (2009) The Economic and Statistical Value of Forecast Combinations Under Regime Switching: An Application to Predictable US Returns. In: M. Wohar and D. Rapach (Eds.), Frontiers of Economics and Globalization, Volume 3, *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Publishing Ltd. & Elsevier Press: NY.

[26] Guidolin, M. and Nicodano, G. (2009) "Small Caps in International Equity Portfolios: The Effects of Variance Risk." *Annals of Finance*, 5(1): 15-48.

[27] Guidolin, M. and Ono, S. (2006) "Are the Dynamic Linkages Between the Macroeconomy and Asset Prices Time-Varying?" *Journal of Economics and Business*, 58(5-6): 480-518.

[28] Guidolin, M., and Y.-M., Tam (2010) "A Yield Spread Perspective on the Great Financial Crisis: Break-Point Test Evidence", Federal Reserve Bank of St. Louis working paper No. 2010-026.

[29] Guidolin, M. and Timmermann, A. (2007) "Asset Allocation under Multivariate Regime Switching." *Journal of Economic Dynamics & Control*, 31(11): 3503-3544.

[30] Guidolin, M. and Timmermann, A. (2008a) "Size and Value Anomalies under Regime Shifts." *Journal of Financial Econometrics*, 6(1): 1-48.

[31] Guidolin, M. and Timmermann, A. (2008b) "International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences." *Review of Financial Studies*, 21(2): 889-935.

[32] Guidolin, M. and Timmermann, A. (2009) "Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach." *Journal of Econometrics*, 150(2): 297-311.

[33] Hamilton, J. (1988) "Rational Expectations Econometric Analysis of Changes in Regime." *Journal of Economic Dynamics & Control*, 12(2-3): 385-413.

[34] Hamilton, J. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, 57(2): 357-384.

[35] Hamilton, J. (1993) Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime. In: G. Maddala, C. Rao, and H. Vinod, *Handbook of Statistics*, vol. 11. Amsterdam: North Holland.

[36] Hamilton, J. (1994) Time Series Analysis. Princeton, NJ: Princeton University Press.

[37] Hamilton, J. and Susmel, R. (1994) "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics*, 64(1-2): 307-333.

[38] Hansen, B. (1992) "The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Switching Model of GNP." *Journal of Applied Econometrics*, 7(S): S61-S82.

[39] Henry, O. (2009) "Regime Switching in the Relationship Between Equity Returns and Short-Term Interest Rates in the UK." *Journal of Banking and Finance*, 33, 405-414.

[40] Jobson, J., and B., Korkie (1981) "Performance Hypothesis Testing with the Sharpe and Treynor Measures." *Journal of Finance*, 36(4), 889-908.

[41] Kim, C. J. (1994) "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics*, 60(1-2): 1-22.

[42] Kim, C. J. and Nelson, C. R. (1999) *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA: MIT Press.

[43] Krolzig, H. M. (1997) *Markov-Switching Vector Autoregressions: Modeling, Statistical Inference, and Application to Business Cycle Analysis*. Berlin: Springer-Verlag.

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

[45] Markowitz, H. (1952) "Portfolio Selection." *Journal of Finance*, 7(1): 77-91.

[46] Marron, J. S. and Wand, M. P. (1992) "Exact Mean Integrated Squared Error." *Annals of Statistics*, 20(2): 712-736.

[47] Opdyke, J., (2007) "Comparing Sharpe Ratios: So Where Are the p-Values?." *Journal of Asset Management*, 8(5), 308-336.

[48] Ramchand, L. and Susmel, R. (1998) "Volatility and Cross Correlation across Major Stock Markets." *Journal of Empirical Finance*, 5(4): 397-416.

[49] Ria, F. (2008) *Scelte Ottime di Portafoglio in Presenza di Regimi Markoviani nella Frontiera Efficiente*. Unpublished M.Phil. thesis, University of Insubria.

[50] Scherer, B. (2002) "Portfolio Resampling: Review and Critique." *Financial Analysts Journal*, 58(6), 98-109.

[51] Schöttle, K., and Werner, R. (2006) "Towards Reliable Efficient Frontiers." *Journal of Asset Management*, 7(2), 128-141.

[52] Sin, C. Y. and White, H. (1996) "Information Criteria for Selecting Possibly Misspecified Parametric Models." *Journal of Econometrics*, 71(1-2): 207-225.

[53] Timmermann, A. (2000) "Moments of Markov Switching Models." *Journal of Econometrics*, 96(1): 75-111.

[54] Tu, J. (2010) "Is Regime Switching in Stock Returns Important in Portfolio Decisions?" *Management Science*, 56(7): 1198-1215.

[55] Turner, C., Startz R. and C. Nelson. (1989) "A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market." *Journal of Financial Economics*, 25(1): 3-22.

[56] Tütüncü, R. and Koenig, M. (2004) "Robust Asset Allocation", *Annals of Operations Research*, 132(1-4), 157-187.

Appendix A - The EM Algorithm

The algorithm is dividend in two logical steps, the Expectation and the Maximization steps. Start from the model written in state-space form (1). For the sake of argument, assume that all the parameters of the model in Ψ and $\mathbf{\Sigma}_M$ are known. We separately describe the expectation and maximization steps and then bring them together. To simplify notations and save space, in what follows we focus only on MSIAH models, although their extension to the MSMAH class is straightforward, once the expansion of the number of regimes induced by the "interaction" between k and p is accounted for.

The Expectation step. It is the product of a few smart applications of Bayes' law that allow to recursively derive a sequence of filtered probability distributions and then (going backwards) a sequence of smoothed probability distributions. Starting from a prior

$$\Pr(\boldsymbol{\xi}_t|\Im_{t-1}) = \sum_{\boldsymbol{\xi}_{t-1}} \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}\right) \Pr\left(\boldsymbol{\xi}_{t-1}|\Im_{t-1}\right)$$

the posterior distribution of $\boldsymbol{\xi}_t$ given $\Im_t = \{\Im_{t-1}, \mathbf{y}_t\}$ $\Pr(\boldsymbol{\xi}_t|\Im_t)$ is given by

$$\Pr(\boldsymbol{\xi}_t|\Im_t) = \frac{\Pr(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}) \Pr(\boldsymbol{\xi}_t|\Im_{t-1})}{\Pr(\mathbf{y}_t|\Im_{t-1})},$$

where $\Pr(\mathbf{y}_t|\Im_{t-1}) = \sum_{\boldsymbol{\xi}_t} \Pr(\mathbf{y}_t, \boldsymbol{\xi}_t|\Im_{t-1}) = \sum_{\boldsymbol{\xi}_t} \Pr(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}) \Pr(\boldsymbol{\xi}_t|\Im_{t-1})$ is the unconditional likelihood of the current observation given its past. For compactness it can also be expressed as

$$\boldsymbol{\eta}_t' \hat{\boldsymbol{\xi}}_{t|t-1} = \boldsymbol{\iota}_k' \left(\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1}\right)$$

where $\odot$ denotes the element by element (Hadamard) product and the $k \times 1$ vector $\boldsymbol{\eta}_t$ collects the possible log-likelihood values as a function of the realized state:

$$\boldsymbol{\eta}_t \equiv \begin{bmatrix} p(\mathbf{y}_t|\boldsymbol{\xi}_t = \mathbf{e}_1, \Im_{t-1}) \\ p(\mathbf{y}_t|\boldsymbol{\xi}_t = \mathbf{e}_2, \Im_{t-1}) \\ \vdots \\ p(\mathbf{y}_t|\boldsymbol{\xi}_t = \mathbf{e}_k, \Im_{t-1}) \end{bmatrix} = \begin{bmatrix} (2\pi)^{-1/2}|\mathbf{\Sigma}_1|^{-1/2} \exp\left[(\mathbf{y}_t - \mathbf{X}_t\mathbf{\Psi}\mathbf{e}_1)\mathbf{\Sigma}_1^{-1}(\mathbf{y}_t - \mathbf{X}_t\mathbf{\Psi}\mathbf{e}_1)\right] \\ (2\pi)^{-1/2}|\mathbf{\Sigma}_2|^{-1/2} \exp\left[(\mathbf{y}_t - \mathbf{X}_t\mathbf{\Psi}\mathbf{e}_2)\mathbf{\Sigma}_2^{-1}(\mathbf{y}_t - \mathbf{X}_t\mathbf{\Psi}\mathbf{e}_2)\right] \\ \vdots \\ (2\pi)^{-1/2}|\mathbf{\Sigma}_k|^{-1/2} \exp\left[(\mathbf{y}_t - \mathbf{X}_t\mathbf{\Psi}\mathbf{e}_k)\mathbf{\Sigma}_1^{-1}(\mathbf{y}_t - \mathbf{X}_t\mathbf{\Psi}\mathbf{e}_k)\right] \end{bmatrix}.$$

Since the *filtered* vector $\hat{\boldsymbol{\xi}}_{t|t}$ also corresponds to the discrete probability distribution of the possible states perceived on the basis of the information set $\Im_t$, we can re-write

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1}}{\boldsymbol{\iota}_k' \left(\boldsymbol{\eta}_t \odot \hat{\boldsymbol{\xi}}_{t|t-1} \right)}. \tag{13}$$

The algorithm is completed by the transition equation that implies that

$$E_t[\boldsymbol{\xi}_{t+1}] = \hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t}. \tag{14}$$

Assuming that the initial state probability vector $\hat{\boldsymbol{\xi}}_{1|0}$ somehow known, (13)-(14) define an iterative algorithm that allows one to generate a sequence of filtered state probability vectors $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$.[36] Notice that the filtered probabilities are the product of a limited information technique, since despite the availability of a sample of size T, each $\hat{\boldsymbol{\xi}}_{t|t}$ is filtered out of the information set $\Im_t$ only, ignoring $\{\mathbf{y}_\tau\}_{\tau=t+1}^{T}$. However, once $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$ has been calculated, Kim's (1994) smoothing algorithm is then easily implemented to recover the sequence of *smoothed* probability distributions $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$ by iterating the following algorithm backwards, starting from the filtered (and smoothed) probability distribution $\hat{\boldsymbol{\xi}}_{T|T}$ produced by (13)-(14). Observe that

$$\hat{\boldsymbol{\xi}}_{t|T} = \Pr\left(\boldsymbol{\xi}_t|\Im_T\right) = \sum_{\boldsymbol{\xi}_{t+1}} \Pr\left(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t+1}|\Im_T\right)$$

$$\begin{aligned}
&= \sum_{\boldsymbol{\xi}_{t+1}} \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \Im_T\right) \Pr\left(\boldsymbol{\xi}_{t+1}|\Im_T\right) \\
&= \sum_{\boldsymbol{\xi}_{t+1}} \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \Im_t, \{\mathbf{y}_\tau\}_{\tau=t+1}^{T}\right) \Pr\left(\boldsymbol{\xi}_{t+1}|\Im_T\right) \\
&= \sum_{\boldsymbol{\xi}_{t+1}} \frac{\Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t+1}, \Im_t\right) \Pr\left(\{\mathbf{y}_\tau\}_{\tau=t+1}^{T}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t+1}, \Im_t\right)}{\Pr\left(\{\mathbf{y}_\tau\}_{\tau=t+1}^{T}|\boldsymbol{\xi}_{t+1}, \Im_t\right)} \Pr\left(\boldsymbol{\xi}_{t+1}|\Im_T\right) \\
&= \sum_{\boldsymbol{\xi}_{t+1}} \frac{\Pr\left(\boldsymbol{\xi}_t|\Im_t\right) \Pr\left(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t, \Im_t\right)}{\Pr\left(\boldsymbol{\xi}_{t+1}|\Im_t\right)} \Pr\left(\boldsymbol{\xi}_{t+1}|\Im_T\right)
\end{aligned}$$

since the Markovian structure implies that $\Pr\left(\{\mathbf{y}_\tau\}_{\tau=t+1}^{T}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t+1}, \Im_t\right) = \Pr\left(\{\mathbf{y}_\tau\}_{\tau=t+1}^{T}|\boldsymbol{\xi}_{t+1}, \Im_t\right)$. Hence $\hat{\boldsymbol{\xi}}_{t|T}$ can be re-written as

$$\hat{\boldsymbol{\xi}}_{t|T} = \left(\mathbf{F}'\left(\hat{\boldsymbol{\xi}}_{t+1|T} \oslash \hat{\boldsymbol{\xi}}_{t+1|t}\right)\right) \odot \hat{\boldsymbol{\xi}}_{t|t}, \tag{15}$$

where $\oslash$ denotes element-by-element division and $\Pr\left(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t, \Im_t\right)$ equals by construction the transition matrix driving the first order Markov chain and therefore $\mathbf{F}'$ in the transition equation. (15) is initialized by setting $t = T-1$ thus obtaining

$$\hat{\boldsymbol{\xi}}_{T-1|T} = \left(F'\left(\hat{\boldsymbol{\xi}}_{T|T} \oslash \hat{\boldsymbol{\xi}}_{T|T-1}\right)\right) \odot \hat{\boldsymbol{\xi}}_{T-1|T-1}$$

and so forth, proceeding backwards until $t = 1$.[37]

[36] Alternatively, $\hat{\boldsymbol{\xi}}_{1|0}$ might be assumed to correspond to stationary unconditional probability distribution such that $\bar{\boldsymbol{\xi}} = \mathbf{P}\bar{\boldsymbol{\xi}}$.

[37] Notice that while $\hat{\boldsymbol{\xi}}_{T|T}$ and $\hat{\boldsymbol{\xi}}_{T-1|T-1}$ will be known from the application of Hamilton's smoothing algorithm, $\hat{\boldsymbol{\xi}}_{T|T-1} = \mathbf{F}\hat{\boldsymbol{\xi}}_{T-1|T-1}$.

The Maximization step. Call $\boldsymbol{\theta}$ the vector collecting the parameters appearing in the measurement equation and $\boldsymbol{\rho}$ the vector collecting the transition probabilities in $\mathbf{P}$, i.e. $\boldsymbol{\theta} \equiv [vec\,(\boldsymbol{\Psi})\,|vec(\boldsymbol{\Sigma}_M)]$ and $\boldsymbol{\rho} \equiv vec(\mathbf{P})$. Write the likelihood function as

$$L\left(\{\mathbf{y}_t\}_{t=1}^T|\{\boldsymbol{\xi}_t\}_{t=1}^T, \boldsymbol{\theta}\right) = \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right) \tag{16}$$

where $\Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right) = \sum_{s_0=1}^M \xi_{s_0} \prod_{t=1}^T p_{s_{t-1},s_t}$ and the first summation spans the space defined by

$$\boldsymbol{\xi}_1 \otimes \boldsymbol{\xi}_2 \otimes ... \otimes \boldsymbol{\xi}_T$$

for a total of k^T possible combinations. Then the parameters $[\boldsymbol{\theta}'\ \boldsymbol{\rho}']'$ can be derived by maximization of (16) subject to the natural constraints:

$$\mathbf{P}\boldsymbol{\iota}_k = \boldsymbol{\iota}_k \quad \boldsymbol{\xi}_0'\boldsymbol{\iota}_k = 1 \tag{17}$$

$$\boldsymbol{\rho} \geq \mathbf{0},\ \boldsymbol{\xi}_0 \geq \mathbf{0}, \text{ and } \boldsymbol{\Sigma}_M\mathbf{e}_j \text{ is positive definite } \forall j = 1, 2, ..., M. \tag{18}$$

At this point it is common place to assume the "nonnegativity" constraints in (18) are satisfied and to take the first-order conditions of a Lagrangian that explicitly enforces the adding-up constraints:

$$L^*\left(\{\mathbf{y}_t\}_{t=1}^T|\{\boldsymbol{\xi}_t\}_{t=1}^T, \boldsymbol{\theta}\right) = \ln\left[\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right)\right] - \boldsymbol{\lambda}_1'\left(\mathbf{P}\boldsymbol{\iota}_M - \boldsymbol{\iota}_k\right) - \lambda_2\left(\boldsymbol{\xi}_0'\boldsymbol{\iota}_k - 1\right). \tag{19}$$

Appendix B - First-order conditions useful in the M-step

Derivation of the logarithm of (19) with respect to $\boldsymbol{\theta}$ gives the score function:

$$\begin{aligned}
\frac{\partial L^*}{\partial \boldsymbol{\theta}'} &= \frac{1}{L}\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial \prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right) \\
&= \frac{1}{L}\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial \ln\left[\prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}'} \prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right) \\
&= \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \sum_{t=1}^T \frac{\partial \ln p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Pr\left(\boldsymbol{\xi}_t|\Im_T; \boldsymbol{\theta}, \boldsymbol{\rho}\right)
\end{aligned}$$

since from the definition of conditional probability

$$\frac{\prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right)}{\sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right)} = \frac{\prod_{t=1}^T p(\mathbf{y}_t|\boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr\left(\boldsymbol{\xi}_t|\boldsymbol{\xi}_0; \boldsymbol{\rho}\right)}{L\left(\{\mathbf{y}_t\}_{t=1}^T|\{\boldsymbol{\xi}_t\}_{t=1}^T, \boldsymbol{\theta}\right)} = \Pr\left(\boldsymbol{\xi}_t|\Im_T; \boldsymbol{\theta}, \boldsymbol{\rho}\right).$$

Therefore

$$\sum_{t=1}^T \hat{\boldsymbol{\xi}}_{t|T}(\boldsymbol{\theta}, \boldsymbol{\rho}) \frac{\partial \ln \boldsymbol{\eta}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \mathbf{0}' \tag{20}$$

provides the first set of FOCs w.r.t. $\boldsymbol{\theta}$. Notice that these conditions involves the smoothed probabilities of the state vector, $\{\hat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^T$. Furthermore, these are simply smoothed probability-weighted standard FOCs (score conditions) in a general MLE problem. For VAR(p) models, GLS-type, closed form solutions are available (see Hamilton, 1994).

The FOCs w.r.t. the transition probabilities are determined as follows. Since

$$\begin{aligned}
\frac{\partial \ln L}{\partial \boldsymbol{\rho}'} &= \frac{1}{L} \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial \Pr(\boldsymbol{\xi}_t | \boldsymbol{\xi}_0; \boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} \prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \\
&= \frac{1}{L} \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \frac{\partial \ln \Pr(\boldsymbol{\xi}_t | \boldsymbol{\xi}_0; \boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} \prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\xi}_t, \Im_{t-1}; \boldsymbol{\theta}) \Pr(\boldsymbol{\xi}_t | \boldsymbol{\xi}_0; \boldsymbol{\rho}) \\
&= \sum_{\{\boldsymbol{\xi}_t\}_{t=1}^T} \sum_{t=1}^T \frac{\partial \ln \Pr(\boldsymbol{\xi}_t | \boldsymbol{\xi}_0; \boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} \Pr(\boldsymbol{\xi}_t | \Im_T; \boldsymbol{\theta}, \boldsymbol{\rho}),
\end{aligned}$$

for each component p_{ij} of $\boldsymbol{\rho}$ this implies:

$$\begin{aligned}
\frac{\partial \ln L}{\partial p_{ij}} &= \sum_{t=1}^T \sum_{\boldsymbol{\xi}_{t-1}=\mathbf{e}_i} \sum_{\boldsymbol{\xi}_t=\mathbf{e}_j} \frac{\partial \ln \Pr\left(\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}; \boldsymbol{\rho}\right)}{\partial p_{ij}} \Pr\left(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1} | \Im_T; \boldsymbol{\theta}, \boldsymbol{\rho}\right) \\
&= \sum_{t=1}^T \sum_{\boldsymbol{\xi}_{t-1}=\mathbf{e}_i} \sum_{\boldsymbol{\xi}_t=\mathbf{e}_j} \frac{1}{p_{ij}} I(\boldsymbol{\xi}_{t-1} = \mathbf{e}_i, \boldsymbol{\xi}_t = \mathbf{e}_j) \Pr\left(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1} | \Im_T; \boldsymbol{\theta}, \boldsymbol{\rho}\right) \\
&= \sum_{t=1}^T \sum_{\boldsymbol{\xi}_{t-1}=\mathbf{e}_i} \sum_{\boldsymbol{\xi}_t=\mathbf{e}_j} \frac{\Pr(\boldsymbol{\xi}_{t-1} = \mathbf{e}_i, \boldsymbol{\xi}_t = \mathbf{e}_j | \Im_T; \boldsymbol{\theta}, \boldsymbol{\rho})}{p_{ij}},
\end{aligned}$$

which originates the vector expression

$$\frac{\partial \ln L}{\partial \boldsymbol{\rho}'} = \left(\sum_{t=1}^T \left(\hat{\boldsymbol{\xi}}_{t|T}^{(2)} \right)' \right) \oslash \boldsymbol{\rho}'$$

where $\hat{\boldsymbol{\xi}}_{t|T}^{(2)}$ is a k^2 vector of (smoothed) probabilities related concerning the state $\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{\xi}_t$. Since the k adding-up restrictions in $\mathbf{P}\boldsymbol{\iota}_M = \boldsymbol{\iota}_M$ can equivalently be written as $(\boldsymbol{\iota}_k' \otimes \mathbf{I}_k)\boldsymbol{\rho} = \boldsymbol{\iota}_k$, it follows that the FOCs can be written as

$$\frac{\partial L^*}{\partial \boldsymbol{\rho}'} = \left(\sum_{t=1}^T \left(\hat{\boldsymbol{\xi}}_{t|T}^{(2)} \right)' \right) \oslash \boldsymbol{\rho}' - \boldsymbol{\lambda}_1'(\boldsymbol{\iota}_k' \otimes \mathbf{I}_k) = \mathbf{0}'.$$

In other words,

$$\boldsymbol{\rho} = \left(\sum_{t=1}^T \left(\hat{\boldsymbol{\xi}}_{t|T}^{(2)} \right) \right) \oslash (\boldsymbol{\iota}_k \otimes \boldsymbol{\lambda}_1),$$

implying

$$(\boldsymbol{\iota}_k' \otimes \mathbf{I}_k) \left(\sum_{t=1}^T \left(\hat{\boldsymbol{\xi}}_{t|T}^{(2)} \right) \right) \oslash (\boldsymbol{\iota}_k \otimes \boldsymbol{\lambda}_1) = \left(\sum_{t=1}^T \left(\hat{\boldsymbol{\xi}}_{t|T} \right) \right) \oslash \boldsymbol{\lambda}_1 = \boldsymbol{\iota}_k$$

so that $\boldsymbol{\lambda}_1 = \left(\sum_{t=1}^T \hat{\boldsymbol{\xi}}_{t|T} \right)$ results. Finally, we have

$$\boldsymbol{\rho} = \left(\sum_{t=1}^T \left(\hat{\boldsymbol{\xi}}_{t|T}^{(2)} \right) \right) \oslash \left(\boldsymbol{\iota}_k \otimes \left(\sum_{t=1}^T \hat{\boldsymbol{\xi}}_{t|T} \right) \right), \tag{21}$$

which is a highly nonlinear function of smoothed regime probabilities, but that can also be easily evaluated.

Table 1

Summary Statistics for International Stock Returns

	Mean	St. Dev.	Sharpe ratio	Median	Min.	Max.	Skewness	Kurtosis	Jarrque-Bera	LB(12)	LB(12)-squares
Pacific ex-Japan	1.045**	6.137	0.117	1.055**	-25.0	20.8	-0.326	4.955*	48.30**	14.29	35.63**
Japan	0.171	6.355	-0.025	-0.208	-19.4	24.3	0.272	3.757	9.88**	11.65	42.67**
Europe ex-UK	0.933**	5.444	0.111	1.293**	-22.3	15.7	-0.653*	4.641	48.40**	13.12	65.20**
United Kingdom	0.763*	4.903	0.089	0.670**	-19.0	15.1	-0.102	4.048	12.97**	8.08	45.51**
North America	0.860**	4.344	0.122	1.300**	-17.9	10.6	-0.645*	4.233	36.21**	10.05	49.33**

** = significant at 1%; * = signifcant at 5%

Table 2

Model Selection Statistics

Model (k,p)	Log-likelihood	LR Statistic	Davies' approx. p-value	BIC	HQ	AIC	Number of parameters	Number of obs.	Saturation ratio	Tests
	Single-state models									
MSIA(1,0)	-3789.41	__	__	28.172	28.014	27.908	20	1365	68.3	
MSIA(1,1)	-3759.23	__	__	28.569	28.212	27.972	45	1360	30.2	
	Two-state models									
MSI(2,0)	-3766.82	45.195	0.000	28.151	27.937	27.794	27	1365	50.6	
MSIH(2,0)	-3702.17	174.495	0.000	**27.985**	**27.653**	**27.430**	42	1365	32.5	H: 129.30 (0.000)
MSH(2,0)	-3690.65	177.537	0.000	**27.971**	**27.678**	27.482	37	1365	36.9	I: 23.04 (0.000)
MSIA(2,1)	-3722.71	73.034	0.001	28.960	28.349	27.939	77	1360	17.7	VAR: 88.21 (0.001)
MSIAH(2,1)	-3655.23	208.011	0.000	28.773	28.043	27.553	92	1360	14.8	H: 134.98 (0.000)
	Three-state models									VAR: 93.88 (0.000)
MSI(3,0)	-3756.87	65.097	0.000	28.263	27.978	27.787	36	1365	37.9	
MSIH(3,0)	-3650.64	277.544	0.000	**28.101**	**27.578**	**27.228**	66	1365	20.7	H: 212.45 (0.000)
MSIA(3,1)	-3679.42	159.626	0.000	29.342	28.462	27.871	111	1360	12.3	VAR: 154.90 (0.000)
MSIAH(3,1)	-3601.06	316.346	0.000	29.384	28.266	27.515	141	1360	9.6	H: 156.72 (0.000)
	Four-state models									VAR: 99.17 (0.032)
MSI(4,0)	-3741.76	95.301	0.000	28.378	28.006	27.757	47	1365	29.0	
MSIH(4,0)	-3637.06	304.704	0.000	28.536	27.807	**27.319**	92	1365	14.8	H: 209.40 (0.000)
MSIA(4.1)	-3615.01	288.436	0.000	29.611	28.444	27.662	147	1360	9.3	VAR: 253.50 (0.000)
MSIAH(4,1)		No converge achieved (too many parameters)					192	1365	7.1	

Table 3

Estimates of Two-State Markov Switching Model

Panel A - SINGLE STATE MODEL

	Pacific EX JP	Japan	Europe EX UK	UK	North America
1. Mean returns	1.045**	0.171	0.933**	0.763*	0.860**
2. Correlations/Volatilities					
Pacific EX JP	6.137				
JP	0.480**	6.335			
Europe EX UK	0.690**	0.512**	5.444		
UK	0.695**	0.522**	0.816**	4.903	
North America	0.685**	0.429**	0.759**	0.729**	4.344

Panel B - TWO-STATE MODEL

	Pacific EX JP	JP	Europe EX UK	UK	North America
1. Mean returns					
Bear/High Correlation State	1.145*	0.104	0.606	0.550	0.419
Bull/Low Correlation State	0.943	0.240	1.265**	0.979*	1.308**
2. Correlations/Volatilities					
Bear/High Correlation State					
Pacific EX JP	5.882				
JP	0.633**	4.904			
Europe EX UK	0.865**	0.607**	5.950		
UK	0.868**	0.610**	0.938**	4.834	
North America	0.838**	0.573**	0.898**	0.853**	4.528
Bull/Low Correlation State					
Pacific EX JP	6.362				
JP	0.399*	7.524			
Europe EX UK	0.508**	0.483**	4.832		
UK	0.540**	0.482**	0.677**	4.944	
North America	0.542**	0.351*	0.566**	0.595**	4.084

3. Transition probabilities	Bear/High Correlation State	Bull/Low Correlation State
Bear/High Correlation State	0.973**	0.027
Bull/Low Correlation State	0.035	0.965**

Panel C - MARKOV CHAIN PROPERTIES, TWO-STATE MODEL

	Bear	Bull		Bear	Bull
Ergodic Probs	0.557	0.443	**Avg. dur.**	36.34	28.91

** = significant at 1% size or lower; * = significant at 5% size.

Table 4

Summary Statistics for Recursive, Optimal Mean-Variance Portfolio Weights (Sample Period: 1998:01 – 2010:09)

	T=1 month					T=12 months					T=120 months				
	Mean	Median	Std. Dev.	10% Confidence bands		Mean	Median	Std. Dev.	10% Confidence bands		Mean	Median	Std. Dev.	10% Confidence bands	
				5% Lower	5% Upper				5% Lower	5% Upper				5% Lower	5% Upper
No short sales															
Single-State Model, Mean-Variance Preferences															
Pacific ex-Japan	0.006	0.000	0.023	0.000	0.002	0.006	0.000	0.021	0.000	0.000	0.008	0.000	0.031	0.000	0.000
Japan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Europe ex-UK	0.064	0.000	0.105	0.000	0.221	0.066	0.000	0.104	0.000	0.242	0.072	0.000	0.126	0.000	0.292
United Kingdom	0.055	0.000	0.089	0.000	0.203	0.049	0.000	0.080	0.000	0.185	0.029	0.000	0.063	0.000	0.132
North America	0.874	0.942	0.150	**0.621**	**1.000**	0.878	0.959	0.151	**0.628**	**1.000**	0.891	1.000	0.167	**0.602**	**1.000**
Two-State Model, Mean-Variance Preferences															
Pacific ex-Japan	0.233	0.040	0.311	0.000	0.662	0.329	0.315	0.309	0.000	0.786	0.381	0.304	0.376	0.000	0.957
Japan	0.181	0.000	0.298	0.000	0.730	0.167	0.000	0.254	0.000	0.556	0.206	0.000	0.324	0.000	0.812
Europe ex-UK	0.036	0.000	0.168	0.000	0.000	0.013	0.000	0.099	0.000	0.000	0.012	0.000	0.102	0.000	0.000
United Kingdom	0.180	0.000	0.311	0.000	0.759	0.141	0.000	0.283	0.000	0.649	0.130	0.000	0.290	0.000	0.687
North America	0.370	0.352	0.343	0.000	0.897	0.350	0.334	0.306	0.000	0.739	0.271	0.225	0.312	0.000	0.761
Short sales admitted															
Single-State Model, Mean-Variance Preferences															
Pacific ex-Japan	-0.362	-0.451	0.406	-0.832	0.200	-0.376	-0.495	0.422	-0.868	0.217	-0.534	-0.717	0.591	-1.229	0.289
Japan	-1.089	-0.990	0.253	**-1.475**	**-0.857**	-1.130	-1.026	0.261	**-1.530**	**-0.895**	-1.651	-1.503	0.370	**-2.250**	**-1.319**
Europe ex-UK	0.474	0.375	0.406	**0.042**	**1.175**	0.503	0.418	0.424	**0.043**	**1.207**	0.692	0.580	0.590	**0.036**	**1.669**
United Kingdom	0.602	0.646	0.262	**0.303**	**0.891**	0.605	0.633	0.258	**0.271**	**0.853**	0.782	0.823	0.353	**0.312**	**1.106**
North America	1.374	1.443	0.424	**0.735**	**1.869**	1.398	1.481	0.440	**0.722**	**1.918**	1.712	1.842	0.614	**0.759**	**2.436**
Two-State Model, Mean-Variance Preferences															
Pacific ex-Japan	0.733	0.794	1.817	-1.334	3.165	1.040	1.157	1.683	-1.345	3.281	1.245	1.698	1.963	-2.379	3.352
Japan	0.476	0.168	1.373	-0.531	1.891	0.604	0.521	0.690	-0.079	1.296	0.786	0.557	1.010	-0.268	1.865
Europe ex-UK	-0.292	-0.369	1.972	-2.736	1.831	-2.628	-2.693	1.908	**-4.983**	**-0.031**	-3.281	-4.352	2.141	-4.983	0.080
United Kingdom	-0.910	-1.848	2.547	-3.683	2.728	-0.292	-0.887	2.561	-3.249	3.847	-1.086	-1.053	3.066	-4.863	2.918
North America	0.994	1.469	2.165	-2.353	3.057	2.277	2.489	1.643	-0.160	4.484	3.336	3.629	2.274	**0.127**	**5.854**

Table 5

Realized, Recursive Out-of-Sample Performance of Alternative Portfolio Strategies

	Mean	Mean		Volatility	Volatility		Sharpe ratio	Sharpe ratio		Corrected Sharpe r.	Corrected Sharpe r.	
		5% Lower	5% Upper		5% Lower	5% Upper		5% Lower	5% Upper		5% Lower	5% Upper
1-month Horizon (152 obs.)												
No Short Sales												
Single-State Model (No Predictability)	6.91	-2.27	15.92	**16.61**	14.44	18.69	0.052	-0.105	0.221	0.051	-0.104	0.220
Two-State MVF Model	8.03	-2.79	18.82	17.81	15.40	20.22	0.065	-0.126	0.271	0.066	-0.126	0.270
Equally-Weighted (1/N)	**8.26**	-1.23	17.44	16.99	14.62	19.29	**0.073**	-0.085	0.243	**0.073**	-0.085	0.242
Value-Weighted (International CAPM)	7.54	-1.61	16.61	16.62	14.31	18.80	0.063	-0.093	0.233	0.062	-0.093	0.232
Short Sales Admitted												
Single-State Model (No Predictability)	16.10	-4.38	36.98	37.92	31.73	43.73	**0.093**	-0.068	0.242	0.092	-0.068	0.241
Two-State MVF Model	**16.25**	-13.98	47.79	55.58	43.58	66.82	0.064	-0.105	0.210	0.064	-0.105	0.209
Equally-Weighted (1/N)	8.26	-1.23	17.44	16.99	14.62	19.29	0.073	-0.085	0.243	**0.073**	-0.085	0.242
Value-Weighted (International CAPM)	7.54	-1.61	16.61	**16.62**	14.31	18.80	0.063	-0.093	0.233	0.062	-0.093	0.232
12-month Horizon (141 obs.)												
No Short Sales												
Single-State Model (No Predictability)	5.99	*2.56*	*9.33*	**20.68**	18.23	22.86	0.099	-0.065	0.275	0.099	-0.065	0.274
Two-State MVF Model	**10.57**	*6.07*	*15.02*	22.60	20.08	24.90	**0.291**	*0.07*	*0.569*	**0.307**	*0.070*	*0.566*
Equally-Weighted (1/N)	9.45	*5.73*	*13.23*	22.91	20.34	25.24	0.240	*0.077*	*0.424*	0.239	*0.077*	*0.422*
Value-Weighted (International CAPM)	8.52	*4.82*	*12.13*	22.33	19.92	24.43	0.205	*0.038*	*0.384*	0.204	*0.038*	*0.382*
Short Sales Admitted												
Single-State Model (No Predictability)	-11.94	*-17.77*	*-6.03*	35.48	31.76	38.84	-0.448	*-0.648*	*-0.273*	-0.445	*-0.645*	*-0.271*
Two-State MVF Model	3.19	-4.27	11.22	46.75	37.06	56.37	-0.016	-0.201	0.141	-0.016	-0.200	0.141
Equally-Weighted (1/N)	**9.45**	*5.73*	*13.23*	22.91	20.34	25.24	**0.240**	*0.077*	*0.424*	**0.239**	*0.077*	*0.422*
Value-Weighted (International CAPM)	8.52	*4.82*	*12.13*	**22.33**	19.92	24.43	0.205	*0.038*	*0.384*	0.204	*0.038*	*0.382*
120-month Horizon (33 obs.)												
No Short Sales												
Single-State Model (No Predictability)	3.45	*2.04*	*4.94*	**13.47**	10.56	15.23	-0.117	-0.547	0.221	-0.114	-0.535	0.216
Two-State MVF Model	6.04	*3.98*	*8.33*	18.80	12.12	23.99	0.376	-0.110	0.777	0.369	-0.107	0.758
Equally-Weighted (1/N)	**7.92**	*6.33*	*9.64*	15.49	11.92	17.91	**0.812**	*0.580*	*1.105*	**0.793**	*0.567*	*1.080*
Value-Weighted (International CAPM)	6.34	*4.85*	*7.94*	14.58	11.17	16.83	0.520	*0.244*	*0.812*	0.508	*0.238*	*0.793*
Short Sales Admitted												
Single-State Model (No Predictability)	-9.53	*-9.92*	*-8.92*	**4.94**	0.51	7.96	-8.62	-85.83	-5.11	-8.42	*-83.86*	*-4.99*
Two-State MVF Model	3.60	-9.71	27.62	69.94	14.23	84.12	-0.018	-0.607	0.214	-0.017	-0.548	0.209
Equally-Weighted (1/N)	**7.92**	*6.33*	*9.64*	15.49	11.92	17.91	**0.812**	*0.580*	*1.105*	**0.793**	*0.567*	*1.080*
Value-Weighted (International CAPM)	6.34	*4.85*	*7.94*	14.58	11.17	16.83	0.520	*0.244*	*0.812*	0.508	*0.238*	*0.793*

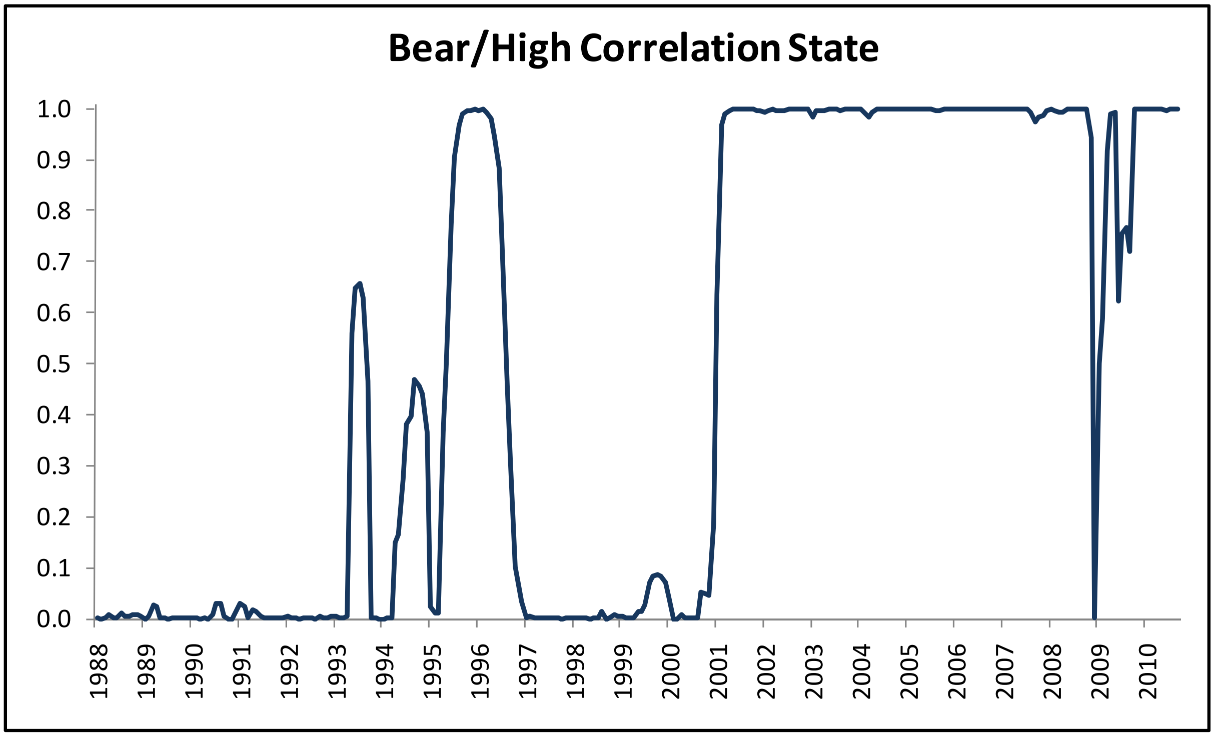
Table 6

Realized, Recursive Out-of-Sample Performance of Alternative Portfolio Strategies: Other Performance Measures

	Jensen's alpha	Jensen's alpha 5% Lower	Jensen's alpha 5% Upper	Skewness	Kurtosis	Sortino ratio	Sortino ratio 5% Lower	Sortino ratio 5% Upper
				1-month Horizon (152 obs.)				
Single-State Model (No Predictability)	**9.55**	3.55	15.33	-0.57	3.31	0.069	-0.143	0.305
Two-State MVF Model	3.16	-2.75	9.00	-0.42	4.00	0.089	-0.165	0.385
Equally-Weighted (1/N)	-0.14	-0.29	-0.02	-0.23	3.89	**0.100**	-0.116	0.356
Value-Weighted (International CAPM)	2.52	0.74	4.18	**-0.14**	**2.77**	0.085	-0.128	0.336
Single-State Model (No Predictability)	**35.64**	0.25	70.85	0.84	4.49	**0.164**	-0.110	0.478
Two-State MVF Model	-9.15	-67.47	43.19	**1.13**	7.83	0.125	-0.177	0.486
Equally-Weighted (1/N)	-0.14	-0.29	-0.02	-0.23	3.89	0.100	-0.116	0.356
Value-Weighted (International CAPM)	2.52	0.74	4.18	-0.14	**2.77**	0.085	-0.128	0.336
				12-month Horizon (141 obs.)				
Single-State Model (No Predictability)	**10.72**	7.57	14.33	-0.37	2.45	0.167	-0.113	0.410
Two-State MVF Model	6.18	1.48	11.17	-0.46	2.76	**0.464**	*0.107*	*0.833*
Equally-Weighted (1/N)	0.09	-0.01	0.20	**-0.01**	2.85	0.419	*0.140*	*0.675*
Value-Weighted (International CAPM)	2.25	0.91	3.41	-0.15	**2.28**	0.373	*0.074*	*0.628*
Single-State Model (No Predictability)	**44.71**	23.25	67.79	0.54	2.75	-1.011	-1.510	-0.597
Two-State MVF Model	0.291	-29.35	25.74	**1.28**	5.89	0.033	-0.268	0.357
Equally-Weighted (1/N)	0.09	-0.01	0.20	-0.01	2.85	**0.419**	*0.140*	*0.675*
Value-Weighted (International CAPM)	2.25	0.91	3.41	-0.15	**2.28**	0.373	*0.074*	*0.628*
				120-month Horizon (33 obs.)				
Single-State Model (No Predictability)	-24.46	-32.75	-17.26	0.84	2.10	-0.404	-1.805	0.874
Two-State MVF Model	-22.54	-61.41	21.88	**1.80**	6.72	0.884	-0.196	1.949
Equally-Weighted (1/N)	0.08	-0.39	0.62	0.54	**1.68**	**2.925**	*1.746*	*5.384*
Value-Weighted (International CAPM)	-7.24	-10.80	-4.09	1.07	2.78	2.198	*0.857*	*4.538*
Single-State Model (No Predictability)	-48.77	-69.63	-37.79	3.51	13.68	-16.41	-120.78	-6.30
Two-State MVF Model	-36.21	-110.35	-16.09	**5.12**	28.38	-0.337	-3.438	0.501
Equally-Weighted (1/N)	0.08	-0.39	0.62	0.54	**1.68**	**2.925**	*1.746*	*5.384*
Value-Weighted (International CAPM)	-7.24	-10.80	-4.09	1.07	2.78	2.198	*0.857*	*4.538*

Figure 1

Smoothed (Full-Sample) Probabilities from Two-State Markov Switching Model



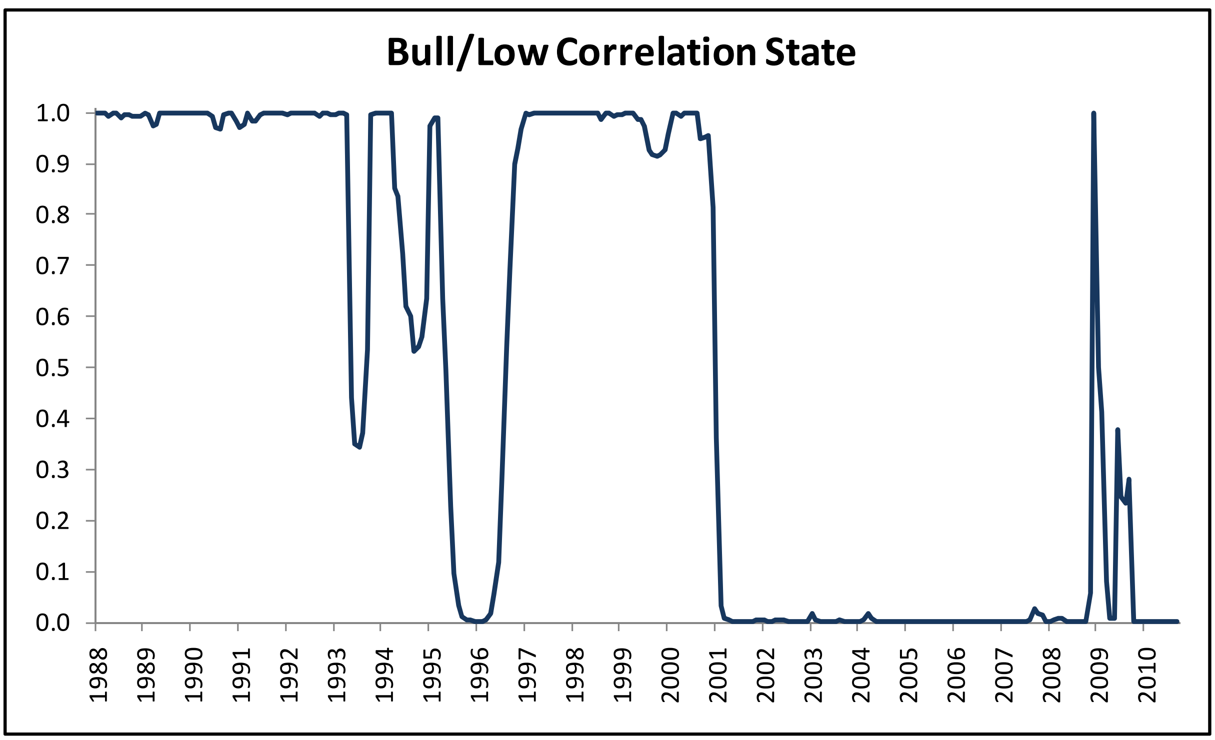


Figure 2

Single-State vs. Markov Switching Mean-Variance Frontiers

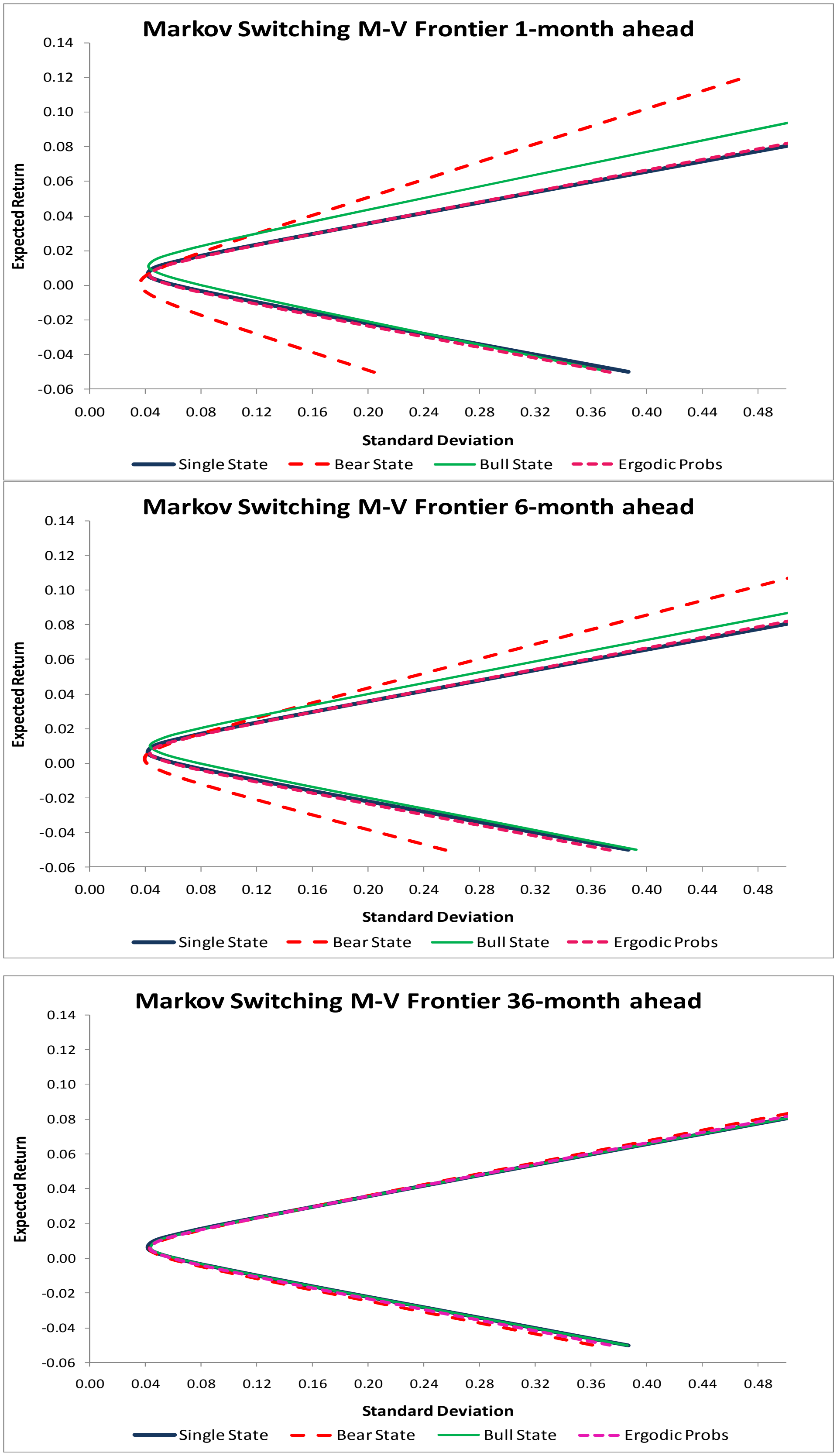


Figure 3

Recursive Optimal Mean-Variance Optimal Portfolio Weights under Markov Switching vs. Single-State Model

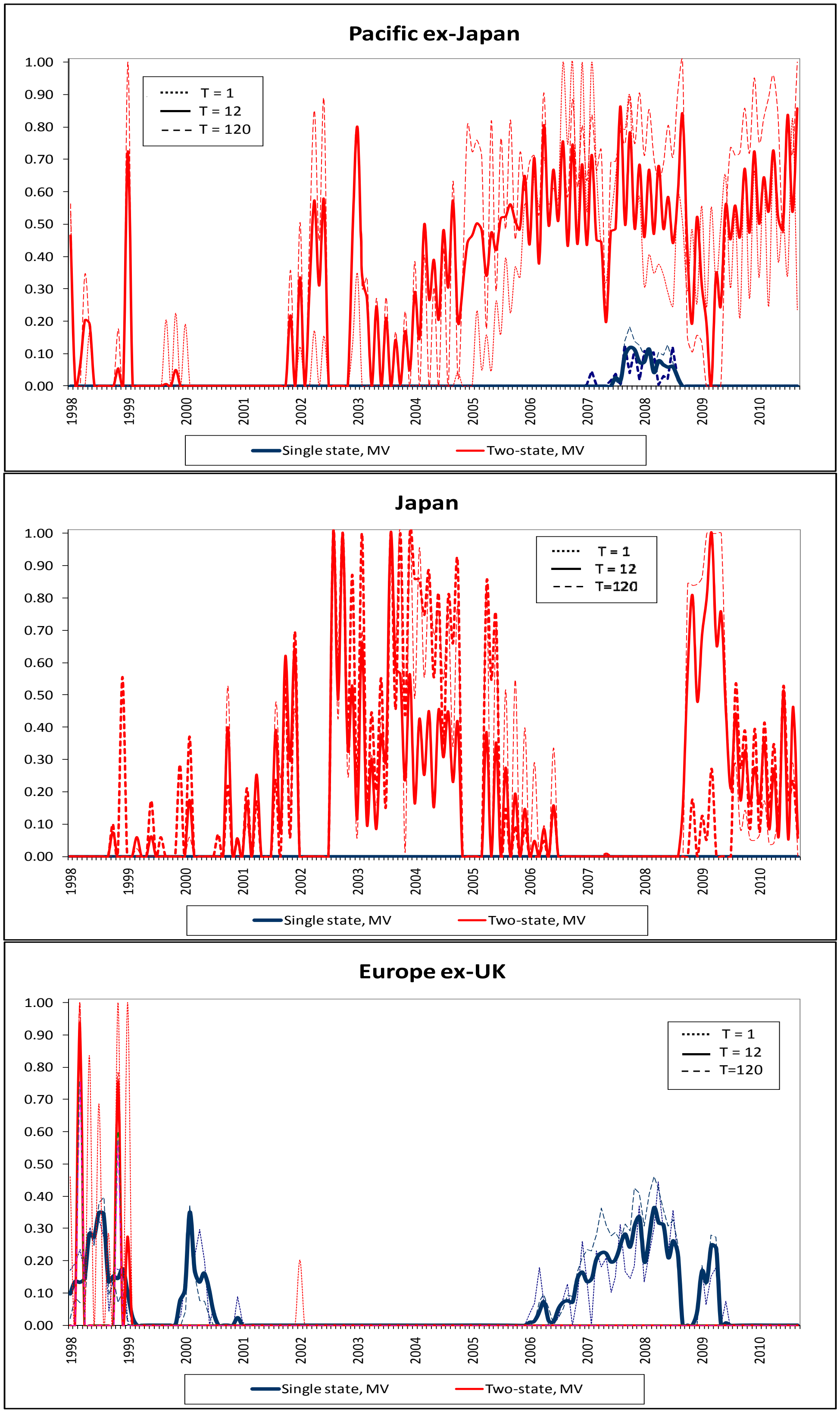


Figure 3 [continued]
Recursive Optimal Mean-Variance Optimal Portfolio Weights under Markov Switching vs. Single-State Model