OSSIAM RESEARCH TEAM

Markov Switching Models and the Volatility Factor: A MCMC Approach.

October, 14, 2016

Markov Switching Models and the Volatility Factor: A MCMC Approach.

October, 14, 2016

Carmine De Franco, PhD
Quantitative analyst
carmine.de-franco@ossiam.com

Prof. Massimo Guidolin
Bocconi University
massimo.guidolin@unibocconi.it

Bruno Monnier, CFA
Quantitative analyst
bruno.monnier@ossiam.com

Abstract

This paper intends to explain the modeling strategy, the empirical Bayesian implementation (estimation), and to illustrate a simple but plausible application of the power of Markov switching models in classical performance and risk analysis. We apply such models for strategies based on US stocks and compare an extension of the standard four-factor model including a new volatility factor to a Markov-switching three-factor model.

Key words: Alternative beta strategies, CAPM, Fama-French three factor model, Gibbs estimation Markov-switching models, Markov-Chain MonteCarlo (MCMC) algorithm.

1 Introduction

Multi-factor models have come to represent the pillar of modern valuation techniques in asset management applications and beyond (see e.g., the growing popularity of three- and four-factor models in determining the cost of equity capital and in event studies). Yet, even since Fama and French (1993) (henceforth F-F) and then Carhart (1994) successfully maintained the necessity to overcome the rigidity of simple, single-factor models in the CAPM tradition, it has become evident that, as new investment strategies are developed and marketed in the asset management industry, new and increasingly complex multi-factor models had to be developed in order for academics, and practitioners alike, to make sense of the average realized excessive returns reported for these new strategies. In this paper, we tackle the issue of the evolving complexity of multi-factor models by examining the trade-off that may exist between the number of assumed factors and other features—in our case, the presence of time variation in the forms of persistent regime shifts in factor exposures—of a given and relatively parsimonious benchmark (such as Fama and French's three-factor model). In particular, our paper deals with the practically relevant case of the newly proposed volatility factor: the factor-mimicking portfolio that in general (i.e., "averaging" across a range of alternative definitions and applications) delivers a performance determined by the spread in returns between low and high-risk stocks, where "risk" may be defined in a variety of ways (total variance, residual variance from a market model, residual variance from some M-factor model, market beta, etc., see the review in Clarke et al. (2011)). Such volatility factors have been introduced to explain the realized, excess returns of a variety of trading strategies and financial instruments, for instance, hedge fund returns (see, e.g., Racicot and Théoret (2009)) and the minimum-variance portfolio puzzle. Indeed, the finding that long-only equity portfolios built to reduce the total portfolio volatility, either by a minimum variance optimization procedure (thus corresponding to the global minimum variance portfolio on the efficient frontier), or by ad-hoc selection of stocks with the lowest volatilities in the market, would systematically outperform classical market-capitalization benchmarks (see Blitz and Van Vliet (2007)). We investigate whether the realized excess performance of a range of popular strategies in the asset management industries (among others, minimum variance with different degrees of diversification, risk-efficient, maximum diversification, and risk parity, see Russo (2016)) does indeed require extending the classical three-factor F-F model to encompass an additional volatility factor in order for any (unlikely) statistical evidence of persistent, positive abnormal outperformance to disappear. As an alternative, we offer a simple three-factor extension that instead allows the factors themselves to exhibit time-varying, regime shifting properties (expected returns) and hence offer an explanation for the strategies' realized returns that does not require a volatility-driven additional factor. Importantly, and to avoid that our treatment may degenerate into a simple example, we avoid adopting one specific definition of the volatility factor encountered in the literature, but perform instead a comparison of several alternative definitions, because important properties such as the realized volatility pre-

mium and factor variability can vary significantly depending on the exact definition adopted. For the sake of brevity, we only report results relative to the more intuitive definition of the volatility factor, although the explanatory power of the models using different definitions of the volatility factor are very similar. Moreover, a wide range of ten different strategies are used in our application to represent alternative non-cap weighted approaches both known in the literature and available to investors under a variety of "wrappers" (among others, mutual and hedge funds, ETFs, customized mandates). The abnormal returns and variability of these portfolios are generally not fully explained by the standard CAPM one-factor linear model, and even by the F-F three-factor model which does a relatively poor job when it comes to explaining the variations of volatility based strategies. We first show that, as one would expect in the light of the existing literature (see e.g. the review in Ang (2014)), the power of standard, constant-exposure multi-factor models is enhanced by adding a new volatility factor. Yet this approach does not give a satisfactory explanation of the portfolios return dynamics because it essentially explains the abnormal performance of low volatility portfolios by adding those portfolios on the right hand side of the factor model. Furthermore, such volatility factors all prove to be very strongly negatively correlated with the initial three-factor model, making then the factor exposure artificially higher. [1]

The question then becomes whether and how it may be possible to do better—or at least, equally well—using an approach that explicitly recognizes the existence of dynamics in the time series properties of the standard, F-F factors to a point to make sense of the abnormal returns of many of the strategies entertained without moving one of their essential economic features—the exposure to low-variance assets—to the right-hand side of a multi-factor model to become one additional factor.

We do exactly this in the context of a rich, multivariate Markov switching (henceforth, MS) model for the vector of the three traditional Fama-French factors that extends the seminal work in Guidolin and Timmermann (2008) on time varying factor means and variances. To ease a number of numerical issues (poor identification, slow convergence, difficulty of imposing reasonable priors, and the possibility of many local maxima), we perform estimation adopting state-of-the-art Monte Carlo Markov Chain (henceforth, MCMC), simulation-based techniques. With reference to a December 1969-December 2014 sample of monthly data on excess market, size-sorted, and book-to-market (value)-sorted U.S. equity returns, we find that two Markov regimes are required to capture the dynamic properties of stock returns. This also holds when information criteria that penalize more richly parameterized models are applied. The two regimes isolated have a natural interpretation as bull and bear states and their definition appears to be decisively driven by factor volatility, especially for the market factor. When we proceed to regress—again allowing beta exposures to change over time in a regime switching fashion—the excess returns of the ten strategies on the time-varying risk premia of the three F-F factors implied by inferences from the two-state MS model, we find that for most strategies, any evidence of abnormal returns disappears from the first bear regime and appears to be limited in the bull regime.

Moreover, even though the number of factors is limited to three and the new volatility risk is not employed, all measures of goodness-of-fit are in line with the four-factor model. Two closely related papers are Racicot and Théoret (2009) and Russo (2016). With a similar motivation, Racicot and Théoret (2009) aim at expanding standard linear models in different ways to improve their explanatory power and reduce the alphas inferred from fund performances. Although obtained with different tools, their conclusions overlap with ours in the sense that part of the alphas can be attributed to unaccounted fat-tailed and skewed factors and dynamic exposures of strategies to such factors. Both fat tails and asymmetries are easily obtained within Markov-switching models for the distribution of the factors, as in our paper. Russo (2016) also bears a number of similarities with our paper. On the one hand, he investigates whether exposure to skewness may represent an explanation for the low-risk anomaly, in the sense of skewness being lower among low-beta or low-total risk stocks, these latter should carry a premium and eventually yield higher returns. Of course, we also account for distributional asymmetries but through the lenses of regime shifts in beta exposures to unstable factor returns. Russo also pursues a few multi-factor decomposition of smart beta, minimum variance strategies but in a constant exposure framework, differently from our approach. Although we are not aware of papers that have contrasted the option to increase the number of factors vs. modelling the time-varying properties of the traditional F-F factors, there is a growing literature that has studied the time series properties of the F-F factors. For instance, Black and McMillan (2004), Chincoli and Guidolin (2016), Guidolin and Timmermann (2008), Gulen et al. (2011) and Zhang et al. (2009) have examined regimes in Size, Value, and Momentum factors, also documenting that models with regime shifts outperform single-state benchmarks not only in terms of their in-sample fit, but also for their out-of-sample predictive accuracy. However, our contribution is distinctive because it emphasizes that a more careful approach to the dynamics of the classical F-F factors may represent a substitute for expanding the set of factors used in applied work.

The paper is organized as follows: Section 2 is devoted to the description of the factor model and the introduction of the volatility factor. Factor exposures of alternative beta strategies are studied in Section 3 when we consider static linear models. Section 4 introduces a simple MS model that captures dynamic changes in factor risk premia while Section 5 explains at length how such a flexible model may be estimated using Bayesian MCMC methods, and in particular relies on one implementation of the Gibbs sampling approach. Section 6 reproduces the results of alternative beta strategies factor exposure in a regime-dependent framework.

2 Data

Data used in the paper is taken from the Center for Research in Security Prices (CRSP) database. It contains monthly returns, prices and outstanding shares for US stock publicly traded on NYSE over the period December 1969 - December 2014. We exclude foreign equities, ADRs, REITs and Closed End Funds. Monthly returns include dividends.

Factor Model The baseline model for our analysis is a standard factor model à la Fama-French where the three pricing factors are the US equity market, the value and size effects. We collect the returns of US listed stocks from the CRSP database, and each month, we consider the capitalization-weighted portfolio composed of the 1,000 largest stocks as a proxy of the market factor. We also collect the monthly returns of the 3-month US T-bill as a proxy for the risk-free rate (henceforth RF). Unless expressly mentioned then, the market factor is the excess return over the risk-free rate (henceforth MC). We use standard size factor (henceforth SMB - small minus big) and value factor (henceforth HML - high minus low) from Kenneth French's database.[2]

Volatility factor In addition to the baseline model, we also present a pricing model that includes volatility. To compute this factor, we build a zero net investment portfolio representing the volatility characteristic, allegedly, a priced risk factor, responsible for a so-called low volatility anomaly, see, e.g., Ang et. al (2006).[3] This portfolio represents the performance difference between low volatility and high volatility stocks. There is no unique definition of volatility factor in the literature, and the volatility factors found in research papers and risk models differ in several important ways. We use total volatility as the risk measure, because its estimation is model-independent.[4] Other different sources of differences in the literature are percentile limits (low or high risk is defined using volatility deciles, quintiles, terciles). We use deciles and quintiles for defining what is meant by "low" or "high" volatility. We find that this choice has an important effect on the volatility factor premium, that is much higher for the volatility factor built on deciles than on quintiles.[5] Finally, we do not perform double sorting on size, value, nor optimization to make the factor orthogonal to the traditional Fama-French factors. Based on the possible variations discussed, we introduce the following version of the volatility factor:

- **SMV** - stable minus volatile. It is built as the difference between the monthly return of the lowest and highest volatility quantiles. Within the quantiles, stocks are value-weighted. Volatility quintiles are taken from Kenneth French website and are computed on all CRSP stocks, using NYSE breakpoints and 60-day estimation window for variance.

	MC	SMB	HML	SMV
Mean	0.0052*	0.0017	0.0039*	0.0035
Volatility	0.0452	0.0315	0.0298	0.0654
Median	0.0092	0.0002	0.0034	0.0043
Skewness	−0.4949	0.6103	−0.0194	−0.1691
Kurtosis	4.7649	9.2847	5.4826	5.7763
		Correlations		
SMB	0.2344*			
HML	−0.3167*	−0.2472*		
SMV	−0.6099*	−0.6623*	0.4529*	

Exhibit 1: Basic statistics and correlations for factors. The data concerns the sample from December 1969 to December 2014. The US risk free-rate averages 0.0041 per month (4.92% per annum) over the period with 0.0027 volatility. Stars refer to estimations of means significant at a size of 10% or less.

Exhibit 1 collects basic statistics of the factors' returns. The excess market returns' average is 0.0052 (or 6.24% per annum, which compares with a classical Mehra and Prescott (1985) estimate of the risk premium just above 6%); a much higher me-

dian at 0.0092 (11.04% per annum) and a standard deviation of 0.0452 (15.66% per annum). SMB and HML have average returns of 0.0017 and 0.0039 (approximately 2.04% and 4.68% per year) although not statistically significant for SMB. They also show more modest volatilities of 0.0315 and 0.0298 (10.91% and 10.32% per annum, respectively). Consistently with a large part of the literature, the low volatility portfolio SMV yields a positive average return of 0.0035 (4.20% per annum), although not statistically significant, because of the high volatility of its short leg which makes the estimation of the mean particularly noisy. The premium is yet smaller compared to MC, but in line with HML and higher than SMB, although with higher volatility. One could conclude that, in line with Clarke et al. (2010), the premium on volatility factor is questionable, yet it is a statistically significant risk factor. The volatility factor also has strong negative correlations with MC and SMB and positive correlation with HML. All portfolios (but SMB, which has features of a lottery bet, characterized by some positive skewness) imply negatively skewed returns and fat tails. In any event, even when portfolio returns exhibit approximately zero skewness, their excess kurtosis is always large enough to cause rejections of the null hypothesis of normally distributed returns in a Jarque-Bera test. This is consistent with the factor portfolio-mimicking returns being characterized by regime switching dynamics, consistently with the evidence in Guidolin and Timmermann (2008) for the market, SMB, and HML and in Bekaert et al. (2012) with reference to the volatility factor.

Alternative beta strategies We simulate ten different alternative beta strategies accordingly with the most popular methodologies available in the industry. Exhibit 2 lists the strategies and their main characteristics. The alternative beta strategies are rebalanced quarterly, at the end of March, June, September and December. All strategies consider the biggest 1,000 stocks of the investment universe at the rebalancing date. Although both the set of strategies and their methodologies are similar to those studied in De Franco et al. (2016), we applied different estimation procedures for dividends, volatilities and returns, to be consisted with the CRSP dataset. All strategies are long-only and fully invested and stock dividends are reinvested in the stocks themselves. For all strategies that use quantitative measures such as volatility ($\boldsymbol{\sigma}$) or expected return ($\boldsymbol{\mu}$), the window to estimate such quantities are set to be in line with the current industry practice. For the covariance matrix ($\boldsymbol{\Sigma}$), we take a 60 month rolling estimation period with optimal shrinkage (Ledoit and Wolf, 2003) to insure proper scaling.

Alternative Beta Strategies
Dividend (DIV35)
Diversity (DW07)
Equal Weight (EW)
Low Volatility (LV5Y)
Maximum Diversification (MD)
Minimum Variance concentrated (MV)
Minimum Variance sector constrained (MVs)
Minimum Variance diversified (MVsH)
Risk Efficient (RE)
Risk Parity (RP5Y)

Exhibit 2: Alternative beta strategies.

DW07: Diversity DW07 contains all stocks in MC whose weight is tilted with the power function $w \rightarrow w^p$, with $p = 0.7$.

Portfolios' weights are finally scaled to sum up to 1. We refer to Fernholz (1999) for detailed discussion on this strategy.

RP5Y: Risk Parity The strategy first removes all stocks from the MC portfolio that have a price history shorter than 60 months. Stocks are then weighted by the inverse of their 5-year volatility. Portfolios' weights are finally scaled to sum up to 1. For detailed discussion of such portfolio construction we refer to Demey et al. (2010) and Roncalli (2013)

LV5Y: Low Volatility The methodology is the same as RP5Y, except for the fact that we first exclude stocks with abnormally low volatility (essentially stocks that have shown flat prices for quite a long time), and we then select the remaining 100 stocks with the lowest volatility.

DIV35: Dividend The strategy first removes all stocks in the MC portfolio that have a price history shorter than 12 months. It ranks the remaining stocks by their rolling 1-year dividend yield and selects the top 35 percent (roughly 350 stocks). Stocks are finally equally weighted.

MD: Maximum Diversification The strategy first removes all stocks from the MC portfolio that have a price history shorter than 60 months. The final portfolio is then computed by maximizing the Diversification Ratio (Choueifaty and Coignard, 2008): $D(w) = (w'\boldsymbol{\sigma}) / (w'\boldsymbol{\Sigma}w)$. Volatility is estimated over the previous 36 months. Optimization is constrained to have each stock's weight bounded from above by the minimum between 1.5% and 20 times the weight in the MC portfolio.

RE: Risk Efficient The strategy first removes all stocks in the MC portfolio that have a price history shorter than 24 months. Portfolio's weights are defined as $w = \boldsymbol{\Sigma^{-1}} * \boldsymbol{\mu}$. Expected return vector $\boldsymbol{\mu}$ is set to be proportional to the semi-downside volatility. For detailed discussion on this allocation scheme and the relation between expected return and semi-downside deviation we refer to Amenc et al. (2011).

MV: Minimum Variance concentrated The strategy first removes all stocks in the MC portfolio that have a price history shorter than 60 months. Portfolio's weights are then set to minimize the total variance $w'\boldsymbol{\Sigma}w$, weights are constrained to be lower than 5%. For further details on this portfolio construction we refer to Clarke et al. (2006).

MVs: Minimum Variance sector constrained The weighting scheme is similar to MV, where we add a maximum sector constraint at 20% and maximum weight per stock at 3%. We use the sector classification established by Kenneth French, from the Standard Industrial Classification (SIC), to form 10 industrial sectors.

MVsH: Minimum Variance diversified The weighting scheme is the same as MVs, with the addition of a diversification constraint in the optimization problem. The diversification is measured by the Herfindahl-Hirschmann index: $H(w) = \sum_i w_i^2$ and the targer is set at 1/80. For further details on the Herfindahl-Hirschmann constraints and the impact of constraint on the Minimum Variance portfolio, we refer to Rulik (2013), De Franco and Monnier (2014) and references therein.

Strategy	Alpha	MC	R^2	ER	Std
DIV35	0.0022∗	0.7952∗	83.36%	0.63%	3.93%
DW07	0.0003	1.0317∗	98.86%	0.56%	4.69%
EW	0.0007	1.0863∗	93.00%	0.63%	5.09%
LV5Y	0.0034∗	0.5330∗	36.78%	0.61%	3.97%
MC	0	1.0000∗	100.00%	0.51%	4.52%
MD	0.0012∗	0.9202∗	85.49%	0.60%	4.50%
MV	0.0028∗	0.5179∗	48.89%	0.55%	3.35%
MVs	0.0027∗	0.6026∗	68.19%	0.58%	3.30%
MVsH	0.0027∗	0.6425∗	72.26%	0.60%	3.41%
RE	0.001	1.0313∗	88.93%	0.63%	4.94%
RP5Y	0.0018∗	0.9422∗	89.59%	0.66%	4.50%

Exhibit 3: CAPM regressions for alternative beta strategy excess returns, average excess returns (ER) and standard deviations (Std). The data concerns the sample from December 1969 to December 2014. Stars refer to estimations significant at a size of 10% or less.

3 Factor Exposures Of Alternative Beta Strategies: The Volatility Factor

When considering the classical CAPM equation for the excess return of the alternative beta strategy returns:

$$R_t^i - RF_t = \alpha_i + \beta^i MC_t + \epsilon_t, \qquad (3.1)$$

for $1 \leq i \leq nb$, where $nb = 10$ is the number of alternative beta strategies in our dataset, we find that the majority of these alternative beta strategies deliver significant positive alphas, as shown in Exhibit 3. Except for Diversity (DW07), Equal Weight (EW) and Risk Efficient (RE), all alternative beta strategies show significant alphas ranging from 0.0012 for Maximum Diversification (MD), roughly 1.44% per annum, to 0.0034 (4.08% per annum) for Low Volatility (LV5Y). It seems that there are better ways to achieve the long-term goals of equity investor rather than invest into the market portfolio. Although this fact may be challenged in several ways (MC is not exactly the market portfolio, the alternative beta strategies are rebalanced regularly so that they incur transaction costs, there are no liquidity considerations in the analysis, etc) it is a fact that the industry has seen a significant proliferation of investment products linked to such strategies. The analysis of variance shows that for at least Low Volatility (LV5Y), concentrated Minimum Variance (MV) and, to some extent, sector constrained and fully constrained Minimum Variance strategies (MVs and MVsH) the model does a poor job when it comes to explaining the variation of monthly returns. Indeed, the goodness-of-fit is only 36.78% for LV5Y and 48.89% for MV, which poorly compare with R^2 coefficients in excess of 80% for DIV35, DW07, EW, MD, RE and RP5Y. Generally speaking, when a strategy delivers non-zero alpha, it indicates that there are (risk-adjusted) excess returns that are either abnormal (similar to free lunches, although they exist only on average) or that

the model is misspecified. To see this, assume for example that the portfolio's excess return evolves following (3.2)

$$R_t^i - RF_t = \sum_{h=1}^{M} \beta_h^i F_{h,t} + \epsilon_t^i \qquad (3.2)$$

for $1 \leq i \leq nb$. If a factor is missed, (3.2) becomes

$$R_t^i - RF_t = \underbrace{\beta_1^i F_{1,t}}_{\alpha_t} + \sum_{h=2}^{M} \beta_h^i F_{h,t} + \epsilon_t,$$

where we assume that Factor F_1 is missed. In this case, part of the misspecified average exposure $\beta_1^i \mathbb{E}[F_1]$ may appear as alpha rather than factor exposure. When we expand the CAPM equation with the size and value factors (to obtain the Fama-French three factor model), we notice how the majority of alphas disappears. This fact, shown in Exhibit 4, is well known in the literature (see for example Arnott (2011)) and it implies that the majority of alternative beta strategies looks like a static combination of the market and (priced) risk factors. Interestingly, all alternative beta strategies have a pronounced value bias (positive and significant exposure of HML), which shows as significant positive premium over the period studied (see Exhibit 1). Moreover, we highlight how DW07, EW, MD, RE and RP5Y end up positively exposed to size, a direct consequence of the weighting schemes used. On the other hand, Dividend (DIV35) shows a strongly negative exposure to size, as small companies are naturally less prone to be stable dividend payers. The volatility based strategies have moderate negative exposure to size (even if not statistically significant for LV5Y, MVs and MVsH). Again, this is consistent with the naturally higher volatility associated with small companies. The alphas are now statistically indistinguishable from zero for all but MVs, MVsH (for which they are positive) and RE (for which it is negative). The diversified Minimum Variance strategies show significant alphas of 0.0013 (MVs, sector constrained, 1.56% per annum) and 0.0011 (MVsH, fully constrained, 1.32% per annum), while Risk Efficient (RE) has a small but significant alpha that equals -0.0008 (-0.96% per annum). Finally, under the three factor model, the goodness-of-fit is sufficiently high for all but LV5Y and MV, where the R^2 is only slightly better than the one implied by the CAPM equation. This first analysis suggests that the CAPM alpha shown in Exhibit 3 is indeed originated from factor exposure to size and value at least for DIV35, LV5Y, MD, RE and RP5Y. Both the presence of residual alphas and the variance analysis suggest that, at least for the volatility based strategies, equation 3.2 should also include a proxy for the volatility factor. Exhibit 5 collects the results of the regressions when we extend the Fama-French three factor model with the volatility factor SMV. In such a model (MC, SMB, HML and SMV), there is no significant (positive) alphas for alternative beta portfolios. We notice however that in this model, Risk Efficient (RE) shows a statistically significant -0.0010 alpha (-1.20% per annum). Market betas appear higher than the ones in Exhibit 4. Similarly, except for DIV35, all alternative beta strategies appear to be positively exposed to size. Both effects are originated in the strong negative correlation of the SMV factor to MC and SMB (Exhibit 1). All strategies remain characterized by a significant tilt to value and positively exposed to the volatility factor, with

Strategy	Alpha	MC	SMB	HML	R^2
DIV35	0.0005	0.8980*	−0.1485*	0.3760	93.17%
DW07	−0.0001	1.0238*	0.1166*	0.0535*	99.45%
EW	−0.0002	1.0552*	0.3589*	0.1318*	97.61%
LV5Y	0.001	0.6436*	−0.0539	0.4870*	49.51%
MC	0	1.0000*	0	0	100.00%
MD	−0.0003	0.9589*	0.1431*	0.2970*	89.24%
MV	0.0009	0.6096*	−0.0460*	0.4030*	61.19%
MVs	0.0013*	0.6665*	−0.0148	0.2942*	74.7%
MVsH	0.0011*	0.7105*	−0.0135	0.3148*	79.19%
RE	−0.0008*	1.0509*	0.2828*	0.3150*	94.06%
RP5Y	0.0001	0.9810*	0.1777*	0.3244*	94.28%

Exhibit 4: Fama-French 3 factor model regressions for alternative beta strategy excess returns. The data concerns the sample from December 1969 to December 2014. Stars refer to estimations significant at a size of 10% or less.

the exception of diversity (DW07), which by construction is very close to the market portfolio MC. Naturally, the magnitude of this exposure will be higher for volatility based strategies (LV5Y, MV, MVs, MVsH) and moderate for risk based ones (MD, RE and RP5Y). We also notice that Dividend (DIV35) is positively exposed to the volatility factor. Under the extended factor model, the goodness-of-fit is now satisfactory for MV, MVs and MVsH, while remaining low for LV5Y (the R^2 coefficient is equal to 55.86%). The increase with respect to the results in Exhibit 4 is significant for the MV, MVs and MVsH. The fact that for LV5Y and MV the R^2 is still low (compared to the rest of the alternative beta strategies) can be due to the lower diversification of these strategies, for which the idiosyncratic risk is a significant component of the total variance. This approach allows us to explain all the alphas generated by the CAPM in (3.1). Unfortunately the betas we find with the four factor model (Fama-French + SMV) in Exhibit 5 are affected by the strong (negative) correlation that the volatility factor has with both the MC and the SMB factor. As a consequence, the economic interpretation of factor exposure becomes less crisp. For example, we find it sensible that strategies based on volatility should generally be negatively exposed to SMB (because small companies tend to be more volatile) and the same should hold true for dividend based strategies. This interpretation does not hold any longer when we add the SMV factor. Indeed, under this model, low volatility strategies have positive exposure to SMB (for instance, in Exhibit 5, LV5Y with 0.2391 and MV with 0.2339). Also, the market exposure of such strategies appears artificially high (for example LV5Y with 0.8139 and MVsH with 0.86, while the industry reference usually ranges between 0.60 and 0.75).

Strategy	Alpha	MC	SMB	HML	SMV	R^2
DIV35	-0.0003	0.9688∗	-0.0267	0.3286∗	0.1140∗	94.28%
DW07	0.0000	1.0201∗	0.1102∗	0.0559∗	-0.006	99.45%
EW	-0.0001	1.0423∗	0.3366∗	0.1405∗	-0.0208∗	97.63%
LV5Y	-0.0008	0.8139∗	0.2391∗	0.3729∗	0.2741∗	55.86%
MC	0	1.0000∗	0	0	0	100.00%
MD	-0.0006	0.9810∗	0.1812∗	0.2822∗	0.0357∗	89.33%
MV	-0.0009	0.7723∗	0.2339∗	0.2939∗	0.2619∗	69.35%
MVs	-0.0005	0.8253∗	0.2584∗	0.1878∗	0.2556∗	82.71%
MVsH	-0.0006	0.8629∗	0.2488∗	0.2126∗	0.2454∗	86.07%
RE	-0.0010∗	1.0651∗	0.3073∗	0.3055∗	0.0230∗	94.09%
RP5Y	-0.0005	1.0341∗	0.2691∗	0.2888∗	0.0855∗	94.76%

Exhibit 5: Fama-French + SMV 4 factor model regressions for alternative beta strategies excess returns. The data concerns the sample from December 1969 to December 2014. Stars refer to estimations significant at a size of 10% or less.

4 Markov Switching Model

To avoid these strong collinearity issues, we replace the volatility factor with a more sophisticated Markov switching model to replace the plain-vanilla regression (3.2). More precisely, we assume that the generic model with M factors evolves as follows:

$$f_t^j = \mu_{S_t}^j + \sigma_{S_t}^j \epsilon_t^j \qquad \epsilon_t^j \sim N(0,1) \tag{4.1}$$

for $j = 1, 2, \ldots, M$, in which both means and variances switch according to the state variable S_t, where S_t takes a finite number of values, $S_t = 1, 2, ..., K$ and

$$Cov_t[\epsilon_t^h, \epsilon_t^k] = \rho_{S_t}^{hk} \qquad \text{for } h \neq k.$$

Note that the Markov state variable driving the parameters for all factors is only one and hence common across all factors in this specification.[6] S_t follows a K-state, first-order Markov chain with constant transition probabilities collected in the transition matrix $\mathbf{P}$:

$$\mathbf{P} \equiv \begin{bmatrix} p_{11} & \cdots & p_{1K} \\ \vdots & \ddots & \vdots \\ p_{K1} & \cdots & p_{KK} \end{bmatrix}, \tag{4.2}$$

where $p_{lq} \equiv \Pr(S_{t+1} = q | S_t = l)$ and by construction $\sum_{j=1}^{K} p_{lj} = 1$. In vector form, the model may be written as:

$$\mathbf{f}_t = \boldsymbol{\mu}_{S_t} + \mathbf{D}_{S_t} \epsilon_t \quad \epsilon_t \sim N(0, \boldsymbol{\Upsilon}_{S_t}), \tag{4.3}$$

where $\mathbf{D}_{S_t}$ is a diagonal matrix that collects regime-dependent standard deviations and

$$\boldsymbol{\Upsilon}_{S_t} \equiv \begin{bmatrix} 1 & \rho_{S_t}^{12} & \cdots & \rho_{S_t}^{1M} \\ \rho_{S_t}^{12} & 1 & \cdots & \rho_{S_t}^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{S_t}^{1M} & \rho_{S_t}^{2M} & \cdots & 1 \end{bmatrix}$$

is a regime-dependent correlation matrix, so that $Var(\mathbf{f}_t) = \mathbf{D}_{S_t} \boldsymbol{\Upsilon}_{S_t} \mathbf{D}_{S_t}$. Let us assume that the alternative beta strategy excess returns are now given by

$$R_t^i - RF_t = \sum_{h=1}^{M} \beta_h^i(S_t) F_{h,t} + \eta_t^i, \tag{4.4}$$

$$\eta \sim N(0, \Sigma), \Sigma_{i,j} = Cov\left[\eta_t^i, \eta_t^j\right] = \quad 0, i \neq j,$$

for $1 \leq i \leq nb$. To see why a misspecification of factor regimes may produce an artificial alpha, assume that we want to estimate the (misspecified) static linear model

in (3.2) in place of (4.3). From

$$\beta_h^i(S_t) = \tilde{\beta_h^i} + \left(\beta_h^i(S_t) - \tilde{\beta_h^i}\right),$$

where $\tilde{\beta_h^i}$ is the static beta we try to estimate from a linear model, we obtain

$$R_t^i - RF_t = \sum_{h=1}^{M} \beta_h^i(S_t) F_{h,t} + \eta_t^i$$

$$= \sum_{h=1}^{M} \tilde{\beta_h^i} F_{h,t} + \underbrace{\sum_{h=1}^{M} \left(\beta_h^i(S_t) - \tilde{\beta_h^i}\right) F_{h,t}}_{\alpha_t} + \eta_t^i.$$

It follows that

$$\mathbb{E}\left[\alpha_t\right] = \sum_{h=1}^{M} \mathbb{E}\left[\left(\beta_h^i(S_t) - \tilde{\beta_h^i}\right) F_{h,t}\right]$$

$$= \sum_{h=1}^{M} \sum_{k=1}^{K} p_t^k \mathbb{E}\left[\left(\beta_h^i(S_t) - \tilde{\beta_h^i}\right) F_{h,t} \mid S_t = k\right]$$

$$= \sum_{h=1}^{M} \sum_{k=1}^{K} p_t^k \left(\beta_h^i(k) - \tilde{\beta_h^i}\right) \mu_{h,k},$$

where p_t^k is the probability of being in the state k at time t. α is then a stochastic process that within a linear model will be improperly treated as an hidden factor. Should this α be a process with non-zero mean, we will end up with some spurious intercept in the linear regression in (3.2). In other words, ignoring possible switches in the factor dynamics and/or in the factor exposures can produce artificial an alpha when one uses simple linear models.

In Section 5 we will provide insights on the estimation of the model (4.3) while Section 6 will collect our findings regarding the estimation of regime-dependent alphas and betas in (4.4).

5 Estimation Results

We estimate four alternative specifications of the model in (4.3) in which the same Markov state drives both means and covariances of the factor returns. These correspond to selecting $K = 1, 2, 3, 4$. The maximum level of $K = 4$ is chosen in the light of the evidence in related empirical work by Guidolin and Timmermann (2008). The set of factor used in the model (4.3) is [MC, SMB, HML]. In a frequentist framework, inference on MS models consists of first estimating the model's unknown parameters by Maximum Likelihood, then making inferences on the unobserved MS variable, conditional on the parameter estimates (Guidolin and Timmermann, 2008), considered to be fixed in repeated random samples. Using an expectation-maximization-type algorithm, this process is then repeated until convergence in (point) parameter estimates is achieved. However, it is well-known that even conditioning on some assumed series for the state variable S_t, Maximum Likelihood estimation of the parameters may become rather complex for multivariate and richly parameterized models. Furthermore, it is not straightforward to prove that there is convergence to a global rather than a local maximum point parameter. In a Bayesian framework, both the parameters of the model and the MS state variable, S_t, $t = 1, 2, ..., T$, are treated as random. Thus, in contrast to the frequentist approach, inference is based on a joint distribution of data and parameters, not on the conditional distribution. In addition, by employing a Gibbs sampling methodology (following Albert and Chib (1993)), Bayesian analysis of Markov-Switching models is easy to implement. Both the parameters of the model and the unobserved Markov-Switching variables are treated as missing data and are generated from appropriate conditional distributions using Gibbs sampling (Gelfand et. al

(1990)). The Gibbs sampler is an iterative Monte Carlo Markov Chain (MCMC) technique that breaks down the problem of drawing successive samples from a multivariate density for the variables of interest and the Markov states into drawing successive samples from lower dimensional (ideally, univariate) densities. In practice, the Gibbs sampler builds a Markov chain of the variables of interest and the state variable such that the limiting distribution of the chain is the desired joint posterior density. A key to the Bayesian approach is that along with S_t, $t = 1, 2, ..., T$, the model's unknown parameters

$$\theta \equiv (\mu_1, \ldots, \mu_K, D_1, \ldots, D_K, \Upsilon_1, \ldots, \Upsilon_K)$$

are treated as random variables. For a given time series R, $\ddot{R}_T$ stands for the full sample trajectory R_t, $t = 1, 2, ..., T$. For Bayesian inference about these $T + KM + KM(M+1)/2 + K(K-1)$ variables we need to derive the joint posterior density:

$$\begin{aligned} & g\left(\ddot{S}_T, \theta, P \mid \ddot{\mathbf{f}}_{\mathbf{T}}\right) \\ & = g\left(\theta, P \mid \ddot{\mathbf{S}}_{\mathbf{T}}, \ddot{\mathbf{f}}_{\mathbf{T}}\right) \times g\left(\ddot{S}_T \mid \ddot{\mathbf{f}}_{\mathbf{T}}\right) \qquad (5.1) \\ & = g\left(\theta \mid \ddot{\mathbf{S}}_{\mathbf{T}}, \ddot{\mathbf{f}}_{\mathbf{T}}\right) \times g\left(P \mid \ddot{\mathbf{S}}_{\mathbf{T}}\right) \times g\left(\ddot{S}_T \mid \ddot{\mathbf{f}}_{\mathbf{T}}\right), \end{aligned}$$

assuming that, conditional on $\ddot{\mathbf{S}}_{\mathbf{T}}$, the transition probabilities are independent of both the other parameters of the model and the data, $\ddot{\mathbf{f}}_{\mathbf{T}}$. Here $\ddot{\mathbf{f}}_{\mathbf{T}}$ denotes the history of the multi-factor model.

Conditional on $\ddot{\mathbf{S}}_{\mathbf{T}}$, equation (4.3) simply represents a vector regression model with a known dummy variable, S_t. To implement Gibbs-sampling, we need to derive the distributions of the blocks of each of the above $T + KM + M(M+1)/2 + K(K-1)$ variables conditional on all the other blocks of variates. Thus using arbitrary starting values for the parameters of the model, the following three steps can be repeated until convergence occurs:

Step 1. Generate each S_t from $g\left(S_t \mid \ddot{\mathbf{S}}_{\neq \mathbf{t}}, \boldsymbol{\theta}, \ddot{\mathbf{f}}_{\mathbf{T}}\right)$, $t = 1, \ldots, T$ where $\ddot{\mathbf{S}}_{\neq t}$ refers to a $(T-1) \times 1$ vector of state variables that excludes S_t.[7]

Step 2. Generate the transition probabilities from $g\left(P \mid \ddot{\mathbf{S}}_{\mathbf{T}}\right)$.

Step 3. Generate $\theta \equiv (\mu_1, \ldots, \mu_K, D_1, \ldots, D_K, \Upsilon_1, \ldots, \Upsilon_K)$ from $g\left(\theta \mid \ddot{\mathbf{S}}_{\mathbf{T}}, \ddot{\mathbf{f}}_{\mathbf{T}}\right)$.

Following a cyclical iterative pattern that follows steps 1-3, the Gibbs sampler generates the desired joint posterior distribution $g\left(\ddot{S}_T, \theta, P \mid \ddot{\mathbf{f}}_{\mathbf{T}}\right)$. Tierney (1994) has proven the convergence of the Gibbs sampler under appropriate regularity conditions. In fact, the sampler produces a series of $n = 1, 2, ..., Q, ..., Q+N$ dependent drawings by cycling through steps 1-3. To avoid an effect of the starting values on the desired joint densities and to ensure convergence of the chain, the first Q draws are normally discarded and only the simulated values from the last N iterations are retained. The simulated values indexed by $n = Q+1$, $Q+2$, ..., $Q+N$ are then regarded as an approximate sample from $g\left(\ddot{S}_T, \theta, P \mid \ddot{\mathbf{f}}_{\mathbf{T}}\right)$. To compute the posterior densities of parameters and the state variables, averages of the simulations retained are then employed.

We measure the fit provided by a given model $\mathcal{M}$ and perform model selection by using Bayesian information criteria BIC (Schwarz, 1978). The model with the lowest BIC is preferred as it is based, in

K	No. of free parameter	BIC
1	9	-6320
2	20	-6596
3	33	-6543
4	48	-6492

Exhibit 6: Free Parameters and BICs for model (4.3) with different number of regimes K. The data concerns the sample from December 1969 to December 2014. Estimation performed with 20,000 drawn from the MCMC algorithm.

part, on the predictive density of the data, $g\left(\ddot{\mathbf{f}}_{\mathbf{T}} \mid \ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}; \theta^{\mathcal{M}}\right)$, interpreted as the likelihood (where $\theta^{\mathcal{M}}$ is the set of all parameters involved in model $\mathcal{M}$ and $\ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}$ denotes the history of the state process). $\ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}$ and $\theta^{\mathcal{M}}$ depend on the choice of model $\mathcal{M}$ because the latter defines the number of regimes, and hence the sequence of simulated states as well as the structure of the vector of parameters to be generated. The BIC[8] is defined as

$$\begin{aligned} BIC\left(\mathcal{M}\right) \equiv & -2L\left(\ddot{\mathbf{f}}_{\mathbf{T}}^{\mathcal{M}} \mid \ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}; \theta^{\mathcal{M}}\right) \\ & + \dim\left(\theta^{\mathcal{M}}\right) \ln T, \end{aligned}$$

where $L\left(\ddot{\mathbf{f}}_{\mathbf{T}}^{\mathcal{M}} \mid \ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}; \theta^{\mathcal{M}}\right)$ is the likelihood of the available data on the factors computed in correspondence to the mean of the posteriors of $\ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}$ and $\theta^{\mathcal{M}}$, T is the sample size, and $\dim\left(\theta^{\mathcal{M}}\right)$ is the size of the vector of parameters $\theta^{\mathcal{M}}$.

Exhibit 6 shows that there is strong evidence of the need of multiple regimes, because the BIC declines from -6320 under $K = 1$ to -6596 when $K = 2$. In particular, a rather parsimonious two-state Markov-Switching model characterized by 20 parameters minimizes the BIC.

Exhibit 6 reveals significant evidence against a single-state model but also some strong (never very strong) evidence in favor of $K = 3$ or $K = 4$ compared to $K = 1$. Supported by the indications of the BIC, we have selected a two-state model. The improvement of both the maximized log-likelihood function and the BIC associated with a move from a one to a two-state model is substantial.

In order to select an optimal number of draws that the MCMC algorithm needs in order to provide stable estimates of the 20 free parameters (6 means, 6 variances, 6 correlations and 2 transition probabilities), we look at the convergence of such estimates as a function of the number of draws. We take as a reference point the vector of parameters estimated with 200,000 draws, and we calculate the l^1-norm of the difference between this vector and the vector of parameters estimated with a lower number of draws (the l^1-norm of a vector is the sum of absolute values of its entries). Exhibit 7 shows our results when we select the 2-regimes model. We observe that with a number of draws higher than 20,000 we already obtain stable parameters. Of further interest, the biggest contribution to the difference with respect to the reference point is due to the estimations of the 2 free transition probability parameters. Nevertheless, the magnitude of this error is reasonably low:

$$\left| p_{i,j}^{n} - p_{i,j}^{200{,}000} \right| \leq 0.35\%,$$

where $p_{i,j}^{n}$ is the transition probability in (4.2) estimated with n draws. The second largest error contribution comes from the estimates of the 6 mean parameters, whose l^1-norm fluctuates in a 0.10% bound before converging to the reference value with only 7,500 draws. Variance and correlation parameters appear very stable even with a very limited number of draws. Consider-

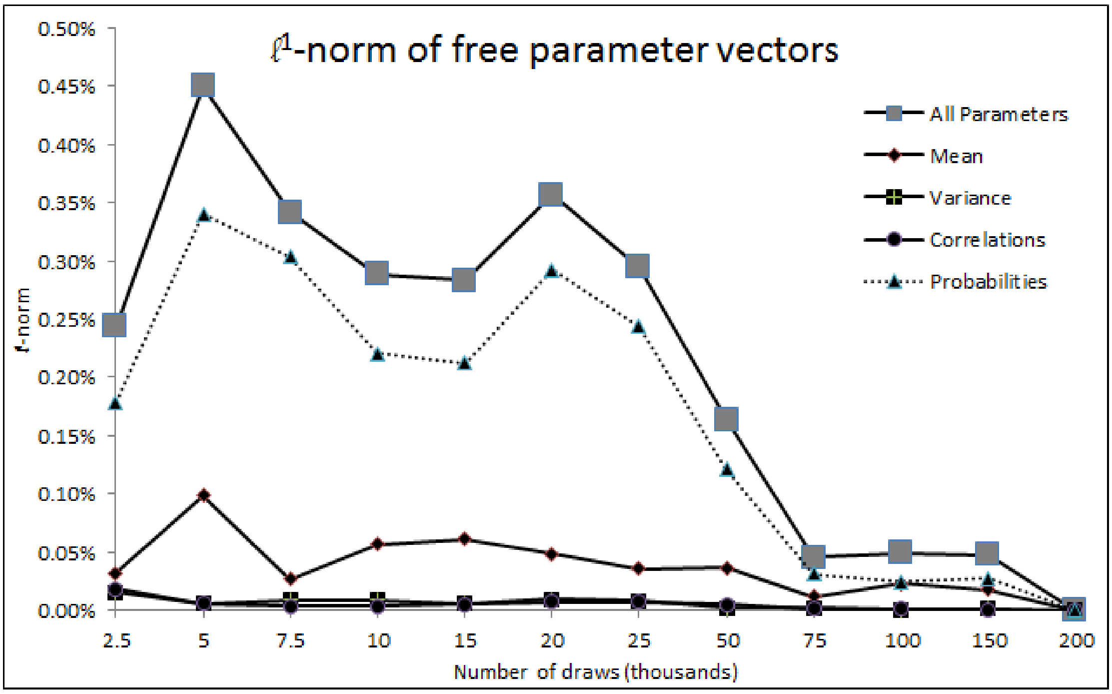


Exhibit 7: l^1-norm of the difference between the vector of all parameters estimated with 200,000 draws and the vector of all parameters estimated with a lower number of draws (All Parameters). The other lines represent l^1-norm of mean parameters vector only (Mean), variance parameters (Variance), correlation parameters (Correlation) and probability parameters (Probability).

Probabilities	Regime 1	Regime 2
Regime 1	0.7290*	0.2710*
Regime 2	0.0186*	0.9814*
Exp. Duration	3.69	53.66

Exhibit 8: Estimation of transition probabilities in (4.2). Stars refer to estimates significant at a size of 10% or less.

ing the convergence results in Exhibit 7, we chose 25,000 as the number of draws the MCMC algorithm will perform to estimate the model in (4.3) with K = 2. The results of the estimation are shown in Exhibits 8–9. The two regimes isolated have a natural interpretation as bull and bear states. Both the means (here, state-contingent risk premia) and the covariances contribute to the identification of the states. In particular, Regime 2 is a persistent state (its average duration is 53.66 months, slightly more than 4.5 years) in which the market risk premium is 0.0062 per month (roughly 7.4% per annum) and statistically significant, as indicated in Exhibit 9. Volatility is generally moderate, slightly below historical unconditional levels: 0.0452 monthly unconditionally standard deviation (Exhibit 1) versus 0.0421 in Regime 2, almost 1% lower in annualized terms). Correlations are mildly Markov switching, but the only radically different coefficient is the one estimating the correlation between shocks to excess market returns and value, that switches

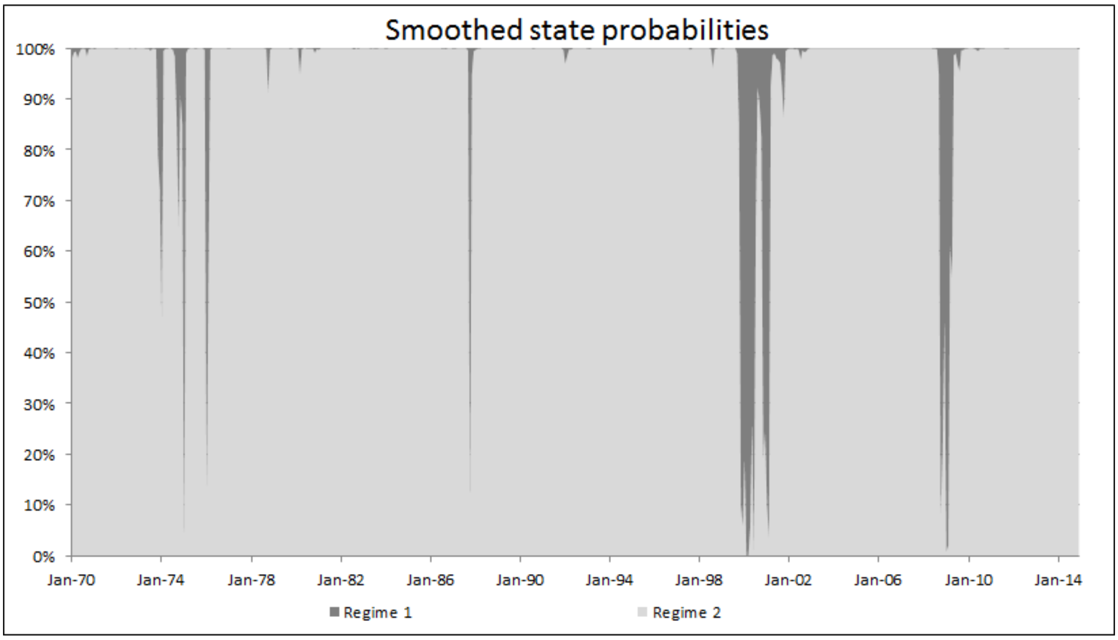


Exhibit 10: Smoothed state probabilities. The data concerns the sample from December 1969 to December 2014.

from -0.3848 in the bull state to -0.0996 in the bear one. Regime 1 is a less persistent state (its average duration is 3.69 months) and the market risk premium is negative (-0.0177, or -21.24% per annum) but not precisely estimated while market risk, measured by monthly standard deviation is twice higher than in Regime 2 (0.0926 against 0.0421). Exhibit 10 shows the unconditional state probabilities, over time, estimated with the specifications discussed above. The Markov switching nature of means and covariances makes the state identification in Exhibit 10 sharp, in the sense that the smoothed probabilities are always either close to zero or one, for both regimes. Exhibit 10 provides additional support to this interpretation of the two states. In particular, the plot of the ex-post, full-sample (mean posterior) probability of bear regime 1 spikes up in between the first and the second oil shock (1973 to

Means

	MC	SMB	HML
Regime 1	−0.0177	0.0164	0.0042
Regime 2	0.0062∗	0.0011	0.0038∗

Volatilities/Correlations

Regime 1	MC	SMB	HML
MC	0.0926∗		
SMB	0.4088∗	0.0864∗	
HML	−0.0996	−0.4765∗	0.0815∗

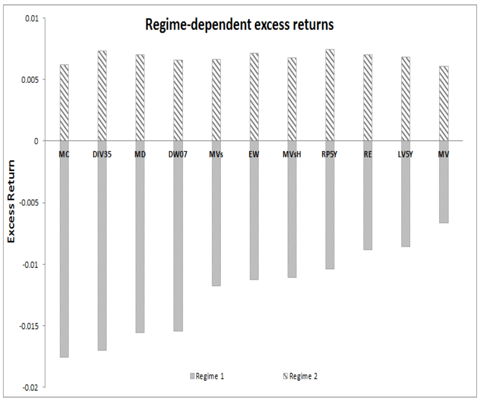
Regime 2	MC	SMB	HML
MC	0.0421∗		
SMB	0.2029∗	0.0271∗	
HML	−0.3848∗	−0.1577∗	0.0257∗

Exhibit 9: Estimation of parameters in model (4.3) with K = 2. Stars refer to estimates significant at a size of 10% or less.

1978), during the fall of 1987 bear market, the Russian crisis and subsequent Asian contagion in the late 1990, the dot-com stock market bubble bursting in 2000-2001, and more recently, the subprime-induced crisis of 2008-early 2009. Fama-French factors size and value show positive premia in both regimes, with a statistically significant monthly 0.0038 for value (roughly 4.5% per annum), while the premium for the size factor, although positive, is less significant. Premia are very high in Regime 1, when market has negative returns, even if the precision of the estimates is not very satisfactory. As usual, determining significant estimates for means is a difficult exercise, especially in a very volatile regime, where high standard deviations make estimations very noisy (see Ang and Bekaert (2002)). It is note worthy that the risk premia for the size factor is also not significant in the static case (K=1) as shown in Exhibit 1. The size factor remains positively correlated to the market in both regimes, even if the magnitude is bear Regime 1 is higher (0.4088 versus 0.2029, Exhibit 9) with both estimates statistically significant. On the other side, the value factor tends to be uncorrelated to the market in Regime 1 (correlation is -0.0996). Both size and value are negatively correlated in Regime 1 and 2, with a significant decrease in correlation in Regime 2.

6 Factor Exposures of Alternative Beta Strategies

We now use the results of the estimation model in Section 5 to estimate regime-dependent factor exposures in (4.4) and derive conditional excess returns and volatilities for each alternative beta strategy (Exhibit 11) together with the set of betas within each regime (Exhibit 12). It is noticeable that the difference between the highest and the lowest excess returns in Exhibit 3, roughly 0.0015 (or 1.81% per annum) is comparable to the same difference in Regime 2 (which turns out to be 0.0014). This is essentially due to the fact that Regime 2 is highly persistent (see Exhibit 8). The same holds true for the volatility (with a maximum spread of roughly 0.017, or 5.88% per annum). These spreads widen when we move into Regime 1, as it is graphically shown in Exhibit 11. The difference between the highest and the lowest excess return is now 0.011 (roughly 1.81% per annum) and the volatility spread goes up to 0.0342 (or 11.84% per annum). One can conclude then that the major differences between alternative beta strategies arise in market downturns. Exhibit 12 collects factor exposures, alphas and goodness-of-fit measures for regime-switching regressions of alternative beta strategies excess returns. We do not notice a significant increase in the R^2. In this respect, the key point of Markov switching techniques does not consist in improving the in-sample fit of the model, but in providing a more powerful framework to conduct inferences on average abnormal excess returns and factor exposures. However, an interesting path emerges from the analysis of factor exposures and alphas within regimes. For Dividend (DIV35), Low Volatility (LV5Y), concentrated and constrained Minimum Variances (MV, MVs and MVsH) and Risk Parity (RP5Y) there is a positive and statistically significant alpha in bull Regime 2 (ranging between 0.0003 (0.36% per annum) and 0.0017 (2.04% per annum). Indeed, except for RP5Y which anyway has a very small positive alpha, these strategies show a low market beta, in a regime



	Regime 1		Regime 2	
	ER	Std	ER	Std
DIV35	-0.0169	0.0841	0.0073	0.036
DW07	-0.0154	0.0958	0.0065	0.0434
EW	-0.0113	0.1077	0.0071	0.0469
LV5Y	-0.0086	0.0844	0.0067	0.0366
MC	-0.0175	0.0908	0.0061	0.0419
MD	-0.0155	0.0844	0.0069	0.0422
MV	-0.0066	0.0736	0.0060	0.0306
MVs	-0.0117	0.0739	0.0066	0.0299
MVsH	-0.0111	0.0782	0.0067	0.0308
RE	-0.0088	0.1075	0.0069	0.0452
RP5Y	-0.0103	0.0965	0.0074	0.0413

Exhibit 11: Average excess returns (ER) with relative standard deviations (Std) within regimes for alternative beta strategy excess returns. The data concerns the sample from December 1969 to December 2014.

where the average market excess return is significantly high at 0.0062 (see Exhibit 9). Nevertheless, these strategies shows average returns, in this regime, at least equal or higher than the market (see Exhibit 11), which is then translated into positive alpha. Similar to the static case, all alternative beta strategies (with lower magnitude for Equal Weight (EW) and Diversity (DW07)) have a significant value tilt, that increases when we switch to bear Regime 1. For all strategies, the exposure to SMB decreases from Regime 2 to Regime 1.

It is reasonable to think that during turbulent market conditions (Regime 1) smaller companies experience higher risk (so that exposure of volatility/risk based strategies diminishes). It can be noticed though that the average return of such factor is very high in Regime 1 (at 0.0164 as can be read in Exhibit 9) even if it does not appear statistically significant, among other things, because of the large dispersion in the returns of such smaller companies, which translates into very noisy estimates. Volatility and risk based strategies are able to show positive alphas in Regime 1 too, although for some of them the statistical precision is not sufficiently high (all Minimum Variances). Moreover, Risk Parity (RP5Y) delivers a significant 0.0057 alpha while Low Volatility (LV5Y, which shares the same weighting scheme as RP5Y but only selects the lowest 100 stocks by volatility) has no significant alpha. The highest diversification for RP5Y is definitely at work here. Market betas increase in Regime 1 for all but DIV35, LV5Y and Maximum Diversification (MD). In bear markets, volatility and correlations increase, so that the major factor becomes the market itself, which explains this consistent increase in betas. For those strategies for which this does not apply, we can suppose that

- Stocks with positive stable dividend records and recent price declines may appear interesting from dividend yield point of view, and become suitable candidates for DIV35, which in turn increases its idiosyncratic risk, and

Strategy	Regime	Alpha	MC	SMB	HML	R^2
DIV35	Regime 1	−0.0021	0.8355*	−0.1379*	0.4982*	93.65%
	Regime 2	0.0007*	0.8968*	−0.1242*	0.3292*	
DW07	Regime 1	0.001	1.0234*	0.0792*	0.0558*	99.48%
	Regime 2	−0.0001*	1.0235*	0.1312*	0.0468*	
EW	Regime 1	0.0023	1.0614*	0.2732*	0.1506*	97.76%
	Regime 2	−0.0002*	1.0517*	0.3932*	0.1093*	
LV5Y	Regime 1	0.0003	0.5733*	−0.0934*	0.6620*	50.77%
	Regime 2	0.0013*	0.6383*	−0.0012	0.4143*	
MD	Regime 1	−0.0019	0.8662*	−0.0111	0.3663*	90.55%
	Regime 2	−0.0002*	0.9658*	0.2310*	0.2553*	
MV	Regime 1	0.0037	0.6068*	−0.0926*	0.4934*	61.88%
	Regime 2	0.0009*	0.6026*	−0.0145	0.3615*	
MVs	Regime 1	0.0011	0.7096*	−0.1084*	0.3797*	75.72%
	Regime 2	0.0017*	0.6459*	0.0323*	0.2393*	
MVsH	Regime 1	0.0027	0.7543*	−0.1323*	0.4222*	80.56%
	Regime 2	0.0015*	0.6881*	0.0451*	0.2465*	
RE	Regime 1	0.0081*	1.1055*	0.0555*	0.4345*	95.54%
	Regime 2	−0.0006*	1.0275*	0.3768*	0.2205*	
RP5Y	Regime 1	0.0057*	0.9933*	−0.025	0.4537*	95.98%
	Regime 2	0.0003*	0.9638*	0.2719*	0.2339*	

Exhibit 12: Markov-switching Fama-French 3 factor model regressions for alternative beta strategy excess returns. The data concerns the sample from December 1969 to December 2014.

lowers the market exposure.

- A poorly diversified portfolio as LV5Y (with only 100 stocks) clearly suffers from lack of diversification, and then more exposed to extreme idiosyncratic risks.

With respect to static linear models (as in(3.2)), a Markov-switching regime framework provides a richer and more comprehensive performance and risk analysis. It enables us to infer the different factor exposures and risk-adjusted performances in different market conditions (regimes). Conversely, adding the set of explanatory factors (for example with volatility) certainly produces a better data fit for alternative beta strategies, as shown in Exhibit 5, but makes the interpretation more difficult because of the strong collinearities that such factor has with respect to the market and Fama-French size and value factor.

7 Conclusions

This paper compares factor exposures of alternative beta strategies in a Fama-French linear model extended with a volatility factor vs. time-varying beta exposures from a Markov-switching model. Alternative beta strategies have gained popularity in the last ten years because of their ability to generate positive alpha in a standard CAPM framework and also, at least for the Mini-

mum Variance strategies, when considering the standard three-factor model given by the market, size and value factors. These findings have been the empirical base for the introduction of a volatility factor. Indeed, within a three-factor model expanded to include volatility, the performance of alternative beta strategies are sufficiently well explained, except for highly concentrated LV5Y and MV. Nevertheless, increasing the number of priced factors in the model compromises its interpretability, since any natural proxy for the volatility factor is, by construction, highly negatively exposed to the market and the size factor. Another possibility to improve the explanatory power of the standard three-factor model is to consider the model within a Markov-switching framework. In this model, it turns out that empirically, volatility usually plays the role of the trigger that induces the switches between regimes. Each regime is characterized by a specific factor return distribution, which in turn yields regime-dependent risk premia. We apply Gibbs-sampling techniques to estimate a two-state Markov-switching model for the Fama-French factors. More precisely, we show that for the data studied in the paper, a two-state model is clearly preferred to the one-state model because it provides a better likelihood while being sufficiently parsimonious in the number of parameters to be estimated. Regime 1 turns out to be a bear market, characterized by negative market returns and high volatility, while Regime 2 is associated with bull market conditions. Regressions of alternative beta strategies return over the time-varying risk premia implied by this regime-switching factor model yields a few interesting results. The 10 non-market cap based strategies that we examine in this paper turn out not to display static exposure to factors. It appears indeed that the exposure to size and value factors strongly depends on the regime. Furthermore, the alphas in Regime 1 are rarely significant, although they are positive for low risk strategies (Minimum Variance, Low Volatility, and Dividend to some extent). In a bull regime instead, these strategies show positive and significant alphas, while for the others alphas are either negative but very small or not statistically significant.

Notes

[1]Of course, this issue could be mitigated by improving the orthogonality of the volatility factor using artificial adjustments, for instance by orthogonalization (see Guidolin and Pedio (2016) for details) but this will may often disrupt any links between the natural definition of a volatility factor and its economic rationale.

[2]Details on the definition of factor portfolios can be found in the data repository made available by Kenneth French at http://mba.tuck.dartmouth.edu/ pages/faculty/ken.french/data_library.html.

[3]A significant stream of empirical literature has studied the historical higher-than-expected performances of low volatile stocks compared to high volatile ones, which jeopardizes the CAPM conclusions for which expected returns are proportional to market beta so that higher risk should be rewarded. We refer to Haugen and Heins (1975), Haugen and Baker (1991), Blitz and Van Vliet (2007) and Haugen and Baker (2010) for further discussions.

[4]Other possible definitions would consider CAPM betas (see, e.g., Haugen and Heins (1975) or Blitz and Van Vliet (2007)) or idiosyncratic volatility (see Ang et. al (2006) Boyer et al. (2010) and Clarke et al. (2010)). In the case of idiosyncratic volatility, the minimum variance puzzle is not confirmed in Brockman et al. (2009) and Fu (2009), where they use of forecasting models for idiosyncratic volatility (EGARCH) instead of simple estimation on past data. Further, Cao and Xu (2010) report that decomposing idiosyncratic volatility in long-term and short-term components, long-run idiosyncratic volatility is positively priced (supports Fu (2009)) and short-term idiosyncratic volatility is negatively priced (supports Ang et. al (2006))

[5]Clarke et al. (2010) employ the definition of their volatility factor VMS (volatile-minus-stable) using volatility terciles, that are also double sorted on size. Although this definition is the one that is the most similar to the standard Fama-French factor construction, we believe that the explanatory power of the VMS factor will be limited, because it is ”too diversified”, that is contains too many stocks in its long and short legs. It is thus not focused enough to fully capture the behavior of the typical low volatility investments.

[6]Even when $\rho_{S_t}^{hk} = \rho^{hk}$ are constant over time, this does not imply that the correlations between factors will be constant over time, as these will also depend on the existence of common switching patterns across regime-specific intercepts, $\mu_{S_t}^{i}$ and $\mu_{S_t}^{k}$ for $i \neq k$, as well as variances.

[7]When possible, it is more convenient to generate the whole block of $g\left(\ddot{S}_T \mid \boldsymbol{\theta}, \ddot{\mathbf{f}}_{\mathbf{T}}\right)$.

[8] The BIC is based on an asymptotic result derived under the assumptions that the data distribution belongs to the exponential family, i.e., that the integral of the likelihood function times the prior probability distribution over the parameters of the model $\mathcal{M}$ is approximated as $-2L\left(\ddot{\mathbf{f}}_{\mathbf{T}} \mid \ddot{\mathbf{S}}_{\mathbf{T}}^{\mathcal{M}}; \theta^{\mathcal{M}}\right) + \dim\left(\theta^{\mathcal{M}}\right)(\ln T - \ln 2 - \ln \pi)$. As $T \to \infty$, this can be approximated by the BIC formula.

References

Albert, J. H. and Chib, S. (1993) Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance. *Journal of Business and Economic Statistics, Vol. 11, 1-15*.

Amenc, N. and Goltz, F. and Martellini, L. and Retkowsky, P. (2011). Efficient indexation: An alternative to cap-weighted indices. *Journal of Investment Management, Vol 9, 1-23*.

Ang, A. and Bekaert, G. (2002) International asset allocation with regime shifts. *Review of Financial Studies, Vol. 15, 1137-1187*.

Ang, A. and Hodrick, R.J. and Xing, Y. and Zhang, X. (2006) The cross-section of volatility and expected returns. *Journal of Finance, Vol. 61, 259-299*.

Ang, A. (2014) *Asset Management: A Systematic Approach to Factor Investing.* Oxford University Press.

Arnott, R. D. (2011) Better beta explained: demystifying alternative equity index strategies. *Journal of Index Investing, Vol. 2, 51-58.*

Bekaert, G. and Hodrick, R. J. and Zhang, X. (2012) Aggregate idiosyncratic volatility. *Journal of Financial and Quantitative Analysis, Vol. 47, 1155–1185*.

Black, A. J., and McMillan, D. G. (2004) Non-linear predictability of value and growth stocks and economic activity. *Journal of Business Finance and Accounting, Vol. 31, 439–474*.

Blitz, David and Van Vliet, Pim (2007) The volatility effect: Lower risk without lower return. *Journal of Portfolio Management, Fall Issue 2007, 102–113*.

Boyer, B. and Mitton, T. and Vorkink, K. (2007) Expected idiosyncratic skewness. *Review of Financial Studies, Vol. 23, 169–202*.

Brockman, P. and Schutte, M.G. and Yu, W. (2009) Is idiosyncratic risk priced? The international evidence. *Working Paper, University of Missouri - Columbia*.

Cao, X. and Xu, Y. (2010) Long-run idiosyncratic volatilities and cross-sectional stock returns. *Working Paper, University of Texas - Dallas*.

Carter, C. K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika, Vol. 81, 541–553*.

Carhart, M.M. (1997) On persistence in mutual fund performance. *Journal of Finance, Vol. 52, 57–82*.

Chincoli, F., and Guidolin, M. (2016) Linear and non-linear predictability in investment style factors: multivariate evidence. *Working Paper, Bocconi University - Milan.*

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *Journal of Portfolio Management, Vol. 35, 40–51*.

Clarke, R. and De Silva, H. and Thorley, S. (2006) Minimum-variance portfolios in the US equity market. *Journal of Portfolio Management, Vol. 33, 10–24*.

Clarke, R. and De Silva, H. and Thorley, S. (2010) Know your VMS exposure. *Journal of Portfolio Management, Vol. 36, 52–59*.

Clarke, R. and De Silva, H. and Thorley, S. (2011) Minimum-variance portfolio composition. *Journal of Portfolio Management, Vol. 37, 31–45*.

De Franco, C. and Monnier, B. (2014). The diversification constraint for minimum variance portfolios. *Research Paper, Ossiam*.

De Franco, C. and Monnier, B. and Nicolle, J. and Rulik, K. (2016) How different are alternative beta strategies? *Journal of Index Investing, Vol. 7, 57-77*.

Demey, P. and Maillard, S. and Roncalli, T. (2010). Risk-based indexation. *Research Paper, Lyxor*.

Fama, E. and French, K. (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, Vol. 33, 3-56*.

Fernholz, R. (1999) On the diversity of equity market. *Journal of Mathematical Economics, Vol 31, 3-56*.

Fu F. (2009) Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics, Vol 91, 24-37*.

Gelfand, A. E. and Hills, S. E. and Racine-Poon, A. and Smith, A. F. (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association, Vol. 85, 972-985*.

Guidolin, M., and Pedio, M. (2016) *Essentials of Applied Portfolio Management.* Bocconi University Press, Milan.

Guidolin, M. and Timmermann, A. (2008) Size and value anomalies under regime shifts. *Journal of Financial Econometrics, Vol. 6, 1-48*.

Gulen, H., Xing, Y., and Zhang, L. (2011) Value versus growth: time-varying expected stock returns. *Financial Management, Vol. 40, 381-407*.

Hamilton, J. D. (1994) *Time series analysis.* Princeton University Press, Princeton.

Haugen, R. A. and N. L. Baker (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management, Vol. 17, 35–40*.

Haugen, R. A. and N. L. Baker (2010). Case closed. *The Handbook of Portfolio Construction: Contemporary Applications of Markowitz techniques, Vol. 17, 601-619*.

Haugen, R. A. and A. J. Heins (1975). Risk and the rate of return on financial assets: Some old wine in new bottles. *Journal of Financial and Quantitative Analysis, Vol. 10, 775-784*.

Kim, C. J. and Nelson, C. R. (1998) Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics, Vol. 80, 188-201*.

Ledoit, O. and Wolf, M. (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, Vol. 10, 603-621*.

Mehra, R. and Prescott, E. C. (1985) The equity premium: A puzzle. *Journal of Monetary Economics, Vol 15, 145-161*.

Racicot, F.-E., and Théoret, R (2009) Integrating volatility factors in the analysis of the hedge fund alpha puzzle. *Journal of Asset Management, Vol 10, 37-62*.

Roncalli, T. (2013) *Introduction to Risk Parity and Budgeting.* CRC Press.

Rulik, K. (2013) Minimum variance portfolio: the art of constraints. *Research Paper, Ossiam*.

Russo, A. (2016) Low-risk equity investment - From theory to practice. *Journal of Asset Management, Vol. 17, 264-279*.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics, Vol 6, 461-464*.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *The Annals of Statistics, Vol. 22, 1701-1728*.

Zhang, Q. J., Hopkins, P., Satchell, S., and Schwob, R. (2009) The link between macroeconomic factors and style returns. *Journal of Asset Management, Vol. 10, 338-355*.

About Ossiam

Ossiam is a research-driven French asset management firm (authorized by the Autorité des Marchés Financiers) and specializes in delivering smart beta* solutions. Efficient indexing is at the core of Ossiam's business model. The firm was founded in response to a post-subprime crisis demand from investors for simplicity, liquidity and transparency. Given the environment, there was a growing need among investors for enhanced beta exposure and risk hedging. Ossiam is focused on the development of innovative investment solutions for investors via a new generation of indices.

*'Smart beta' refers to systematically managed, non-market-cap-weighted strategies covering any asset class.

80, avenue de la Grande Armée
75017 Paris France
info@ossiam.com
+ 33 1 78 40 56 90