European Journal of Operational Research 265 (2018) 685–702

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Estimating stochastic discount factor models with hidden regimes: Applications to commodity pricing

Marta Giampietro[a], Massimo Guidolin[b,*], Manuela Pedio[a]

[a] Bocconi University, Italy
[b] IGIER, Bocconi University, Italy

ARTICLE INFO

Article history:
Received 30 November 2015
Accepted 13 July 2017
Available online 2 August 2017

Keywords:
Finance
Commodities
Stochastic discount factor
Hidden Markov model

ABSTRACT

We develop new likelihood-based methods to estimate factor-based Stochastic Discount Factors (SDF) that may accommodate Hidden Markov dynamics in the factor loadings. We use these methods to investigate whether it is possible to find a SDF that jointly prices the cross-section of eight U.S. portfolios of stocks, Treasuries, corporate bonds, and commodities. In particular, we test a range of possible different specification of the SDF, including single-state and Hidden Markov models and compare their statistical and pricing performances. In addition, we assess whether and to which extent a selection of these models replicates the observed moments of the return series, and especially correlations. We report that regime-switching models clearly outperform single-state ones both in term of statistical and pricing accuracy. However, while a four-state model is selected by the information criteria, a two-state three-factor full Vector Autoregression model outperforms the others as far as the pricing accuracy is concerned.

1. Introduction

The interest of institutional investors in commodities has significantly increased over the last two decades. Over the same period, due to the appearance of exchange traded funds, retail investors have developed a growing taste for this asset class. This is due to the fact that commodities can generate equity-like returns in the long-run, act as risk diversifiers both in the short- and in the long-run, and serve as an inflation hedge (see, e.g., Erb & Harvey, 2006; Gorton & Rouwenhorst, 2006). Despite their growing importance in institutional and retail portfolios, our understanding of this asset class remains unsatisfactory. While the literature is extensive as far as traditional asset classes, especially where equities are concerned, there is no consensus on whether there is an asset pricing model which can effectively explain both the cross-sectional and the time series variation in commodity returns (see, e.g., the discussion in Bakshi, Gao, & Rossi, 2014, henceforth BGR).

Part of the reason of this lack of in-depth understanding of commodities, is that they require the development of new methods and operational approaches. Our paper explores such combination of novel methods, i.e., Hidden Markov chain models (HMM), to describe time heterogeneity (see Dias, Vermunt, & Ramos, 2015) in the fundamental pricing measure governing the cross-section of asset returns. We also develop a methodology to operationalize the estimation of HMM stochastic discount factors (henceforth, HMM SDF) that relies on variations of standard Expectations-Maximization algorithms to maximize the log-likelihood function.[1]

When applied to commodities, the classical asset pricing research issues—for instance the nature of the factors driving risk premia and volatility—have been approached in a number of ways: some researchers have employed commodity-specific factors (see, e.g., Szymanowska, de Roon, Nijman, & Van den Goorbergh, 2012; Yang, 2013; BGR, 2014); others have used models designed to price all assets (see, e.g., Asness, Moskowitz, & Pedersen, 2013; Koijen, Moskowitz, Pedersen, & Vrugt, 2013). The former approach considers commodities as a separate asset class for which specific pricing factors are needed; the latter assumes that financial markets are perfectly integrated and hence that a unique measure able to price all assets can be found. Moreover, most studies have tested asset pricing models on commodity return data in a stand-alone fashion while few others have augmented the test asset menu with other asset classes (see e.g., Asness et al., 2013).

Amid this heterogeneity of approaches, a conclusion as to the best asset pricing model for commodities has not been found yet. Following the second strand of literature, we develop a range of different specifications of SDFs based on macroeconomic factors,

* Corresponding author.
E-mail address: massimo.guidolin@unibocconi.it (M. Guidolin).

[1] SDF models are based on the result that in the absence of arbitrage, the price of an asset at time t is the expected discounted value of the asset payoff in period $t+s$, based on information available at time t, when discounting and integration are performed under the risk neutral measure implied by the SDF.

http://dx.doi.org/10.1016/j.ejor.2017.07.045
0377-2217/

constructed by extracting three, five (and to some limited extent, ten) principal components from a broad set of variables concerning prices, production, and the labor and housing markets, as in Ludvingson and Ng (2009). Our objective is to assess whether there exist one or more specifications of the SDF that jointly prices the cross-section of stock, bond, and (spot) commodity returns and that replicates the empirically observed moments of returns, with a specific interest for the matrix of correlations among pairs of commodities as well as across asset classes. Our tested models include both standard linear projections of latent SDFs on the pricing factors and a set of different specification of HMM SDFs which incorporate latent regime shifts governed by one ergodic and irreducible Markov chain that drives shifts in the coefficients that map the K priced factors into the SDF.

In our application to commodity research, we use a cross section of eight portfolios of stocks, Treasuries, corporate bonds, and commodity indices (the S&P-GS Agriculture and livestock, precious metals, industrial metals, and energy indices) over the period January 1989–December 2011 as test assets. To test which, if any, model specification(s) is best able to price the cross-section of the returns of our test assets, we employ a number of different tests. First, we conduct a standard specification search by comparing the values of three different information criteria that provide a measure of the statistical accuracy of the different models. Second, we test the pricing accuracy of our alternative specifications. In particular, we assess whether the in-sample differences between the observed and predicted returns implied by each model are statistically significant using a standard Chi-squared test. In addition, we adopt Hansen–Jagannathan's (henceforth HJ) distance measure to quantify the distance between the candidate SDFs and what is empirically required to price our assets. Finally, we explore whether model-implied moments (mean, standard deviation, skewness, kurtosis, and, above all, correlations) are able to match empirical ones.

To the best of our knowledge, our contribution is original for (at least) two reasons. First, when we extend our model to include latent regime shifts governed by one ergodic and irreducible Markov chain, we test whether there is evidence of a need of additional parametric flexibility to price a range of asset classes that includes commodities. It is generally accepted that financial markets follow boom-and-bust cycles that involve both the mean and the volatility of asset returns (see Guidolin, 2011). Recently, the awareness that also correlations between returns on different asset classes would undergo massive changes has emerged (see Bae, Kim, & Mulvey, 2014). While the literature has mainly focused on stocks and bonds (see, e.g., Guidolin & Timmermann, 2006), the number of researchers who has applied Hidden Markov models (HMM) to commodities is limited (see, e.g., Alizadeh, Nomikos, & Pouliasis, 2008; Bae et al., 2014; Lee & Yoder, 2007*).* We contribute to this strand of research even though our focus is not simply on the modeling of regimes and time-varying moments in commodity returns, but the on-time variation that is induced by the presence of regime shifts in the relationships between the SDF and an interpretable set of priced risk factors.

Our second contribution is that we look for an SDF that not only prices the cross-section of stock, bond, and commodity returns, but also replicates their observed correlations. This is particularly useful in the light of the recent developments in the commodity markets. While there is some evidence that prior to the early 2000s, commodities shared co-movements with stocks (see Gorton & Rouwenhorst, 2006) or each other (see Erb & Harvey, 2006), it seems well established in the minds of investors and researchers that the commodity markets have recently undergone deep changes: commodity prices have experienced booms followed by significant busts and the correlation between stock and commodity returns has risen dramatically since mid-2008 (see, e.g., Büyüksahin & Robe, 2014; Tang & Xiong, 2012).[2]

Our empirical results can be summarized as follows. First, we find strong evidence that HMMs outperform their single-state counterparts both in terms of statistical and pricing accuracy. Indeed, not only all the information criteria select HMMs over their (more parsimonious) single state counterparts, but also the latter strongly underperforms the former with respect in the pricing performance space. Indeed, standard chi-square tests on the differences between model-implied and observed returns leads to rejection of the null of equal values for all the single-state SDF specification. On the contrary, for four different specifications of HMM SDF we fail to reject the null that the predicted returns are different from the observed ones. In particular, a two-state three-factor full MSVAR model seems to rank first in terms of pricing accuracy. In addition, although single-state and HMM models seem to deliver similar performances in terms of matching the empirically observed moments of the asset returns, the two-state HMM model more accurately matches the pair-wise correlations between commodities and traditional asset classes.

Second, despite some models (i.e., regime-switching ones) are closer to HJ's bound than others, when we formally test the null of a zero distance, all models—both single-state and HMM—are rejected. This result may be read in the light of the claim of many researches (see, e.g., Szymanowska et al., 2012; BGR, 2014) that commodities are segmented from other asset classes and needs to be priced using commodity specific factors. Therefore, we test the robustness of our top performing model (namely a two-state three factor full MSVAR) against an SDF-version (with obvious goals of comparability) of the linear factor model proposed by BGR based on three commodity-specific factors: Average, Carry, and Momentum. We find that the pricing performance of our model dominates the one of the benchmark in terms of HJ distance. This establishes that a large HJ distance does not exclusively plague our modeling efforts and appears to be pervasive even to important benchmarks in the literature.

The rest of the paper has the following structure. Section 2 introduces SDF models. Linear factor SDF models are presented in Section 2.1 and are generalized to the HMM case in Section 2.2. Section 2.3 presents the maximum likelihood estimation strategy and explains in detail how the Baum–Welch algorithm can be used to implement it. Section 3 introduces the data used in our application. Section 4 contains our application and key empirical findings. Section 5 concludes.

2. The methodology: factor-based SDF models

SDF models provide a general framework for pricing assets: many existing asset pricing methods, such as the capital asset pricing model, the general equilibrium, consumption-based, intertemporal capital asset pricing model, and also Black and Scholes' formula, can all be shown to be specializations of SDF models to reflect specific assumptions. One SDF is defined by a simple proposition: the price of an asset at time t is equal to the expected payoff of the asset in $t+1$, based on information available at time t:

$$p_{i,t} = E[M_{t+1}X_{i,t+1}|\Im_t] \qquad i = 1, \ldots, n, \tag{1}$$

where $p_{i,t}$ is the price of the ith asset at time t, $X_{i,t+1}$ is the asset payoff in $t+1$ (inclusive of future sale prices and any cash flows paid between t and $t+1$), M_{t+1} is the SDF (a random variable), and $E[\cdot|\Im_t]$ indicates the expectation based on the information available

[2] There is no consensus as to the explanations for this evidence but many commentators agree on the financialization of commodities as a cause (see, e.g., Basak and Pavlova, 2013; Büyükşahin and Robe, 2014).

at time t. The existence of an SDF is equivalent to the law of one price, its positivity is equivalent to the absence of arbitrage opportunities, and its uniqueness is equivalent to market completeness.

If we divide both sides of Eq. (1) by $p_{i,t}$, we obtain

$$1 = E\left[M_{t+1}\frac{X_{i,t+1}}{p_{i,t}}|\Im_t\right] = E[M_{t+1}(1+R_{i,t+1})|\Im_t], \tag{2}$$

where $1+R_{i,t+1} = X_{i,t+1}/p_{i,t}$ is the gross return of asset i. (2) is the standard Euler condition under a generic SDF, M_{t+1}, derived in a representative agent framework. We define m_{t+1} $\ln M_{t+1}$, $r_{i,t+1}$ $\ln(1+R_{i,t+1})$ and $u_{i,t+1}$ $m_{t+1}+r_{i,t+1}$. If the random vector $\boldsymbol{y}_{t+1}$ $[m_{t+1}, r_{1,t+1}, r_{2,t+1}, \dots r_{n,t+1}]'$, with $\Psi_t \equiv \{\boldsymbol{y}_{t-s}\}_{s\geq 0} \subset \Im_t$, has a stationary, homoskedastic multivariate Gaussian distribution, Appendix A shows that

$$\begin{aligned} E[r_{i,t+1}|\Psi_t] &= -E[m_{t+1}|\Psi_t] - \frac{1}{2}Var[m_{t+1}] \\ &\quad -\frac{1}{2}Var[r_{i,t+1}] - Cov[r_{i,t+1}, m_{t+1}], \end{aligned} \tag{3}$$

which establishes a functional link between conditional risk premia on any asset or portfolio, the corresponding first and second moments of the assumed SDF, and the covariance between such an operator and the net returns on the very asset. Moreover, this general expression can be specialized further when assumptions are imposed on the SDF functional form. In the following, we distinguish between two cases, the single- vs. the multi-state, HMM case. Moreover, we do not take a stance as to whether the SDF is unique or not, but we focus—under any assumed model structure—to the SDF that maximizes some statistical criterion and therefore emerges from the data. Therefore, in the case of market incompleteness (as in Marroquìn-Martinez & Moreno, 2013), we assume that the portfolios are replicable, so that their equilibrium return does not depend on the SDF selected.

2.1. Linear factor SDF models

Suppose the SDF has the log-linear structure,

$$M_{t+1} = exp\left(\gamma_0 + \sum_{j=1}^{K}\gamma_j f_{j,t+1}\right) > 0, \tag{4}$$

where $f_{1,t+1}, f_{2,t+1}, \dots, f_{K,t+1}$ are K systematic factors driving the way in which all assets are priced. In the following, also for identification purposes when the model is generalized to the HMM case, we assume that the K factors are observable. Positivity of M_{t+1} ensures the absence of arbitrage opportunities. Notice that the vector $\boldsymbol{\gamma} \equiv [\gamma_0, \gamma_1, \gamma_2, \dots \gamma_K]'$ restricts the relationship between the prices of all assets and the priced risk factors to be homogeneous across assets according to the fundamental Euler condition in (2). There is of course considerable latitude in the definition of what are the priced risk factors that enter M_{t+1}. Either expanding their number ($K \to \infty$) or by carefully selecting their identity, these will improve the fit of the model.[3] Although it would be interesting to derive (4) from a general equilibrium framework, in this paper we take such a structure of the SDF as a primitive but estimable functional form and discuss instead the identity, the number of the factors, and (in Section 2.2) the number of hidden states affecting the loadings.

[3] A few special cases may provide meaningful benchmarks. Clearly, when in (4) $K=1$ and $f_{1,t+1}$ is the gross return on the market portfolio, then the SDF becomes the standard CAPM. When $K=1$ and $f_{1,t+1}$ is the log-consumption growth rate, γ_0 is the subjective rate of discount, and γ_1 is the opposite of the constant coefficient of relative risk aversion, we obtain the classical consumption (C)CAPM. De Roon and Szymanowska (2010) test whether commodity futures returns vary cross-sectionally due to differences in consumption risk and they find that at quarterly horizon, the CCAPM explains about 50% of the cross-sectional variation in mean futures returns, while the conditional version explains up to 60%.

At this point, if we assume that the random vector $\boldsymbol{y}_{t+1}$ $[m_{t+1}, f_{1,t+1}, f_{2,t+1}, \dots f_{K,t+1}, r_{1,t+1}, r_{2,t+1}, \dots r_{n,t+1}]'$ expanded to include the factors driving the SDF, has a stationary multivariate Gaussian distribution, then Appendix A shows that the asset pricing model

$$E[r_{i,t+1}|\Psi_t] = -E[m_{t+1}|\Psi_t] - \frac{1}{2}\sigma_m^2 - \frac{1}{2}\sigma_i^2 - \sigma_{i,m} \tag{5}$$

obtains, where $\sigma_m^2 \equiv Var[m_{t+1}]$, $\sigma_i^2 \equiv \frac{1}{2}Var[r_{i,t+1}]$, and $\sigma_{i,m} \equiv Cov[r_{i,t+1}, m_{t+1}]$. Because it is well known that the time t price of a riskless, one-period zero coupon bond equals $1/\ E[M_{t+1}|\Psi_t]$, and that as a result, the continuously compounded riskless return is

$$r_t^f = -E[m_{t+1}|\Psi_t] - \frac{1}{2}\sigma_m^2, \tag{6}$$

the model is equivalently re-written as:

$$E[r_{i,t+1}|\Psi_t] + \frac{1}{2}\sigma_i^2 - r_t^f = -\sigma_{i,m} \tag{7}$$

where $0.5\sigma_i^2$ is a standard Jensen's inequality correction and the driving force behind the risk premium is therefore $-\sigma_{i,m}$, as one would expect: the higher the covariance between asset returns and the SDF (hence, the higher the covariance between asset returns and marginal utility that is high in bad states), the lower the risk premium on the asset.

Appendix A also shows that under a rather standard stationary VAR(P) representation for the process followed by the factors, the full joint model for the SDF and asset returns is:

$$\begin{aligned} m_{t+1} &= \gamma_0 + \sum_{j=1}^{K}\gamma_j f_{j,t+1} + w_{t+1} \qquad (w_{t+1} \equiv m_{t+1} - E[m_{t+1}|\Psi_t]) \\ f_{j,t+1} &= \varphi_{j,0} + \sum_{k=1}^{K}\sum_{p=1}^{P}\varphi_{j,k,p} f_{k,t+1-p} + \delta_{j,t+1} \qquad j = 1, \dots, K \\ r_{i,t+1} &= \underbrace{-\left(0.5\sigma_m^2 + 0.5\sigma_i^2 + \sigma_{i,m}\right)}_{\mu_i} - m_{t+1} + \nu_{i,t+1} \\ &= \mu_i - m_{t+1} + \nu_{i,t+1}, \quad i = 1, \dots, n \end{aligned} \tag{8}$$

which can also be represented as ($Ind_{\{\cdot\}}$ is a standard indicator function):

$$\begin{aligned} &\begin{bmatrix} 1 & \boldsymbol{0}'_K & \boldsymbol{0}'_n \\ \boldsymbol{0}_K & \boldsymbol{I}_K & \boldsymbol{O}_{K\times n} \\ \boldsymbol{0}_n & \boldsymbol{O}_{n\times K} & \boldsymbol{I}_n \end{bmatrix}\boldsymbol{y}_{t+1} \\ &= \boldsymbol{\mu} + \sum_{l=0}^{P-1}\begin{bmatrix} 0 & \ddot{\boldsymbol{\gamma}}_l' Ind_{\{l=0\}} & \boldsymbol{0}'_n \\ \boldsymbol{0}_K & \boldsymbol{\Phi}_l & \boldsymbol{O}_{K\times n} \\ -\boldsymbol{\iota}_n Ind_{\{l=0\}} & \boldsymbol{O}_{n\times K} & \boldsymbol{O}_{n\times n} \end{bmatrix}\boldsymbol{y}_{t+1-l} + \boldsymbol{\eta}_{t+1} \text{ or} \\ &\boldsymbol{A}_0\boldsymbol{y}_{t+1} = \boldsymbol{\mu} + \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l} + \boldsymbol{\eta}_{t+1} \ \ with\ \ \boldsymbol{\Phi}_0 = \ \boldsymbol{O}_{K\times K} \end{aligned} \tag{9}$$

where $\boldsymbol{\mu} \equiv [\gamma_0, \varphi_{1,0}, \varphi_{2,0}, \dots \varphi_{K,0} - (0.5\sigma_m^2(\boldsymbol{\gamma}) + 0.5\sigma_1^2 + \sigma_{1,m}), \dots - (0.5\sigma_m^2(\boldsymbol{\gamma}) + 0.5\sigma_n^2 + \sigma_{n,m})]'$, or

$$\boldsymbol{\mu} \equiv \begin{bmatrix} \gamma_0 \\ \boldsymbol{\varphi}_0 \\ -0.5\sigma_m^2(\boldsymbol{\gamma}) - 0.5diag_{2:n+1}(\boldsymbol{\Sigma}) - \boldsymbol{\Sigma}\boldsymbol{e}_1 \end{bmatrix} \quad \ddot{\boldsymbol{\gamma}}_l \equiv \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_K \end{bmatrix}$$

$$\boldsymbol{\Phi}_l \equiv \begin{bmatrix} \varphi_{l,1,1} & \varphi_{l,1,2} & \cdots & \varphi_{l,1,K} \\ \varphi_{l,2,1} & \varphi_{l,2,2} & \cdots & \varphi_{l,2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{l,K,1} & \varphi_{l,K,2} & \cdots & \varphi_{l,K,K} \end{bmatrix}$$

and $\boldsymbol{\eta}_{t+1} \equiv [w_{t+1}, \delta_{1,t+1}, \delta_{2,t+1}, \dots \delta_{K,t+1}, \nu_{1,t+1} - w_{t+1}, \dots \nu_{n,t+1} - w_{t+1}]' \sim IID\ N(0, \boldsymbol{\Sigma})$.

2.2. HMM factor SDF models

One of the explanations sometimes reported for the negative correlation between commodities and other asset classes is the different behavior of stocks, bonds and commodities over the business cycle (see, e.g., Jensen, Mercer, & Johnson, 2002). Indeed, Gorton and Rouwenhorst (2006) find that commodity futures perform well in the early stages of a recession, a time when stock returns generally disappoint; in later stages of recessions, commodity returns fall off, but this is generally a very good time for equities. This suggests that a well-specified model for the SDF ought to account for persistent, good and bad states. Moreover, BGR (2014) have noted that conditional pricing models that allow for state dependence in the sensitivity of the stochastic discount factor to the risk factors can often outperform their unconditional counterparts. Following their lead, suppose the SDF has instead a regime switching log-linear structure:

$$M_{t+1}(S_{t+1}) = exp\left(\gamma_{0,S_{t+1}} + \sum_{j=1}^{K} \gamma_{j,S_{t+1}} f_{j,t+1}\right) > 0, \tag{10}$$

where $f_{1,t+1}$, $f_{2,t+1}$, …, $f_{K,t+1}$ are again K observable, systematic factors driving the way in which all assets are priced and S_{t+1} follows a J-state Markov chain. The state-specific vector $\boldsymbol{\gamma}_{S_{t+1}} \equiv [\gamma_{0,S_{t+1}},\ \gamma_{1,S_{t+1}},\ \gamma_{2,S_{t+1}}, \dots\ \gamma_{K,S_{t+1}}]'$ restricts the within-regime relationship between the prices of all assets and the priced risk factors to be homogeneous across assets according to the standard Euler condition in (1). By construction the same regime shifts affect the mapping between all factors and the SDF and as such these are transmitted to all priced assets or portfolios. Because of the MS structure in (9), the empirical fit of the model to the cross-section of asset returns may be improved not only by carefully selecting/expanding the K factors, but also the number of regimes (say, J) or the features of the Markov chain driving the shifts in the coefficients loading the factors on the SDF.

Following steps analogous to those in Appendix A and conditioning on both the state S_{t+1}, and the past of observable returns and factors, if the random vector $\boldsymbol{y}_{t+1}$ $[m_{t+1}, f_{1,t+1}, f_{2,t+1}, \dots f_{K,t+1}, r_{1,t+1}, r_{2,t+1}, \dots r_{n,t+1}]'$, with $\Psi_t \equiv (\{\boldsymbol{y}_{t-s}\}_{s\geq 0}, S_t) \subset \Im_t$, has a conditional multivariate Gaussian distribution then the conditional asset pricing model,

$$\begin{aligned} &E[r_{i,t+1}|\Psi_t, S_t] \\ &= \underbrace{-\left(0.5\sigma_m^2(S_{t+1}) + 0.5\sigma_i^2 + \sigma_{i,m}(S_{t+1})\right)}_{\mu_i(S_{t+1})} - E[m_{t+1}(S_{t+1})|\Psi_t, S_t] \\ &= \mu_i(S_{t+1}) - E[m_{t+1}(S_{t+1})|\Psi_t, S_t] \quad i = 1, \dots, n \end{aligned} \tag{11}$$

is obtained. This is a model in which regimes in the SDF are reflected in expected asset returns both directly through the one-step ahead forecast $E[m_{t+1}(S_{t+1})|\Psi_t, S_t]$ and indirectly, via the asset-specific terms $\mu_i(S_{t+1}) \equiv -0.5\sigma_m^2(S_{t+1}) - 0.5\sigma_i^2 - \sigma_{i,m}(S_{t+1})$ that also reflect state-dependent covariances between asset returns and the SDF. Also in this case, even though the forcing, observable state variables, $f_{1,t+1}, f_{2,t+1}, \dots, f_{K,t+1}$, follow a linear process and may enter linearly the model, the latter becomes non-linear because of the role played by the latent Markov state, S_t, that governs the switches in the parameters appearing in the process of the factors. For instance, the full model may be written as:

$$m_{t+1} = \gamma_{0,S} + \sum_{j=1}^{K} \gamma_{j,S} f_{j,t+1} + w_{t+1} \quad (w_{t+1} \equiv m_{t+1} - E[m_{t+1}|\Psi_t, S_t])$$

$$f_{j,t+1} = \varphi_{j,0} + \sum_{k=1}^{K}\sum_{p=1}^{P} \varphi_{j,k,p} f_{k,t+1-p} + \delta_{j,t+1}, \quad j = 1, \dots, K$$

$$\begin{aligned} r_{i,t+1} &= \mu_i(S_{t+1}) - E[m_{t+1}(S_{t+1})|\Psi_t, S_t] + \upsilon_{i,t+1} \\ &= \mu_i + m_{t+1} + \upsilon_{i,t+1}, \quad i = 1, \dots, n \end{aligned} \tag{12}$$

where the vector of K factors follow a VAR (P) process and that can also be represented as:

$$\begin{bmatrix} 1 & \mathbf{0}'_K & \mathbf{0}'_n \\ \mathbf{0}_K & \boldsymbol{I}_K & \boldsymbol{O}_{K\times n} \\ \mathbf{0}_n & \boldsymbol{O}_{n\times K} & \boldsymbol{I}_n \end{bmatrix} \boldsymbol{y}_{t+1} = \boldsymbol{\mu}(S_{t+1}) + \sum_{l=0}^{P-1} \begin{bmatrix} 0 & \ddot{\boldsymbol{\gamma}}'_{l,S_{t+1}} Ind_{\{l=0\}} & \mathbf{0}'_n \\ \mathbf{0}_K & \boldsymbol{\Phi}_l & \boldsymbol{O}_{K\times n} \\ -\boldsymbol{\iota}_n Ind_{\{l=0\}} & \boldsymbol{O}_{n\times K} & \boldsymbol{O}_{n\times n} \end{bmatrix} \boldsymbol{y}_{t+1-l} + \boldsymbol{\eta}_{t+1} \tag{13}$$

where $\boldsymbol{\mu} \equiv [\gamma_0,\ \varphi_{1,0},\ \dots\ \varphi_{K,0},\ -(0.5\sigma_m^2(\boldsymbol{\gamma}_{S_{t+1}}) + 0.5\sigma_1^2 + \sigma_{1,m}), \dots - (0.5\sigma_m^2(\boldsymbol{\gamma}_{S_{t+1}}) + 0.5\sigma_n^2 + \sigma_{n,m})]'$, or

$$\boldsymbol{\mu}(S_{t+1}) \equiv \begin{bmatrix} \gamma_0 \\ \boldsymbol{\varphi}_0 \\ -0.5\sigma_m^2\left(\boldsymbol{\gamma}_{S_{t+1}}\right) - 0.5 diag_{2:n+1} - \boldsymbol{\Sigma}\left(\boldsymbol{\gamma}_{S_{t+1}}\right)\boldsymbol{e}_1 \end{bmatrix}$$

$$\ddot{\boldsymbol{\gamma}}_{l,S_{t+1}} \equiv \begin{bmatrix} \gamma_{1,S_{t+1}} \\ \gamma_{2,S_{t+1}} \\ \vdots \\ \gamma_{K,S_{t+1}} \end{bmatrix} \qquad \boldsymbol{\Phi}_l \equiv \begin{bmatrix} \varphi_{l,1,1} & \varphi_{l,1,2} & \cdots & \varphi_{l,1,K} \\ \varphi_{l,2,1} & \varphi_{l,2,2} & \cdots & \varphi_{l,2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{l,K,1} & \varphi_{l,K,2} & \cdots & \varphi_{l,K,K} \end{bmatrix}$$

and $\boldsymbol{\eta}_{t+1} \equiv [w_{t+1},\ \delta_{1,t+1},\ \delta_{2,t+1}, \dots\ \delta_{K,t+1},\ \upsilon_{1,t+1} - w_{t+1}, \dots\ \upsilon_{n,t+1} - w_{t+1}]' \sim IID\ N(0, \boldsymbol{\Sigma}_{S_{t+1}})$.[4] For simplicity, we assume a time homogeneous Markov chain, although a literature shows that generalizations may be fruitful (see e.g., Dias et al., 2015). However, in our case, an element of time heterogeneity is already impressed by the fact that the SDF depends on factors characterized by potentially rich dynamics, while the HMM modeling truly affects not directly the factors, but the loadings with which the factors appear in the SDF, in Eq. (10).

2.3. Maximum likelihood estimation strategy

Our estimation strategy is based on the principle of maximum likelihood under parametric assumptions concerning the joint normal distribution of random vector $\boldsymbol{y}_{t+1}$ $[m_{t+1}, f_{1,t+1}, f_{2,t+1}, \dots f_{K,t+1}, r_{1,t+1}, r_{2,t+1}, \dots r_{n,t+1}]'$ and, when appropriate, the Markov state variable S_t. In the single-state, covariance stationary case, because the inverse of the $(n+K+1) \times (n+K+1)$ matrix $\boldsymbol{A}_0$ obviously exists (its determinant is 1), and the corresponding Jacobian is unity, the log of the joint density function of the sample, conditioned on the initial values of the variables ($\boldsymbol{y}_0$), is given by[5]

$$\begin{aligned} &ln\ f\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ \dots, \boldsymbol{y}_T; \boldsymbol{\theta}\right) \\ &= -\frac{T}{2} \ln \det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{A}_0\boldsymbol{y}_{t+1} - \boldsymbol{\mu} - \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l}\right)' \\ &\times \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{A}_0\boldsymbol{y}_{t+1} - \boldsymbol{\mu} - \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l}\right) \end{aligned} \tag{14}$$

(up to an omitted constant term), when the matrices are replaced by the appropriate objects. Maximizing such a log joint density function with respect to the components of $\boldsymbol{\theta} \equiv [\boldsymbol{\gamma}\ \boldsymbol{\varphi}_0\ \{\boldsymbol{\Phi}_l\}_{l=0}^{P-1}\ \boldsymbol{\Omega}]'$

[4] HMM dynamics in the covariance matrix of $\boldsymbol{\eta}_{t+1}$ derives from regimes in the SDF loadings and therefore occurs residually. The assumption of normal shocks is typical of the literature, see Bae et al. (2014).

[5] In what follows $\ddot{\boldsymbol{\eta}}_{t+1}$ and $\ddot{\boldsymbol{y}}_{t+1}$ are the same as $\boldsymbol{\eta}_{t+1}$ and $\boldsymbol{y}_{t+1}$ but fail to include the first element of both vectors as the SDF is by construction unobservable.

will deliver the ML estimates of $\boldsymbol{\theta}$. Note that the (conditional) residuals used by the MLE program,

$$\begin{aligned}\ddot{\boldsymbol{\eta}}_{t+1} &\equiv \ddot{\boldsymbol{A}}_0\ddot{\boldsymbol{y}}_{t+1} - \ddot{\boldsymbol{\mu}} - \sum_{l=0}^{P-1}\ddot{\boldsymbol{A}}_l\ddot{\boldsymbol{y}}_{t+1-l} \\ &= \begin{bmatrix} \boldsymbol{I}_K & \boldsymbol{O}_{K\times n} \\ \boldsymbol{O}_{n\times K} & \boldsymbol{I}_n \end{bmatrix}\ddot{\boldsymbol{y}}_{t+1} - \begin{bmatrix} \boldsymbol{\varphi}_0 \\ -0.5\sigma_m^2(\boldsymbol{\gamma}) - 0.5diag_{2:n+1}(\boldsymbol{\Sigma}) - \boldsymbol{\Sigma}\boldsymbol{e}_1 \end{bmatrix} \\ &\quad - \sum_{l=0}^{P-1}\begin{bmatrix} \boldsymbol{\Phi}_l & \boldsymbol{O}_{K\times n} \\ \boldsymbol{O}_{n\times K} & \boldsymbol{O}_{n\times n} \end{bmatrix}\ddot{\boldsymbol{y}}_{t+1-l} \end{aligned} \tag{15}$$

imply the impossibility to concentrate the parameters of the log-likelihood function to separate the estimable elements of $\boldsymbol{\Omega}$ from those appearing in the conditional residuals $\boldsymbol{\eta}_{t+1}$.

In the case of the HMM extension introduced in Section 2.2, because the inverse of the matrix $\boldsymbol{A}_0$ still exists, and the corresponding Jacobian is unity, the log of the joint density function of the sample, conditioned on the initial values of the variables,

$$\begin{aligned} &\ln f\left(\boldsymbol{y}_1, \boldsymbol{y}_2, \ \ldots, \boldsymbol{y}_T, S_0, \ S_1, \ldots, S_T; \boldsymbol{\theta}\right) \\ &= -\frac{T}{2}\ln\det\left(\boldsymbol{\Sigma}\left(\boldsymbol{\gamma}_{S_{t+1}}\right)\right) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\left(\boldsymbol{A}_0\boldsymbol{y}_{t+1} - \boldsymbol{\mu}\left(\boldsymbol{\gamma}_{S_{t+1}}\right) - \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l}\right)' \\ &\quad \times\boldsymbol{\Sigma}\left(\boldsymbol{\gamma}_{S_{t+1}}\right)^{-1}\left(\boldsymbol{A}_0\boldsymbol{y}_{t+1} - \boldsymbol{\mu}\left(\boldsymbol{\gamma}_{S_{t+1}}\right) - \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l}\right), \end{aligned} \tag{16}$$

(up to an omitted constant term) is the same as the log-likelihood function for $\boldsymbol{\eta}_{t+1}$. It turns out that maximizing such a log joint density function with respect to the components of $\boldsymbol{\theta} \equiv [\{\boldsymbol{\gamma}_s\}_{s=1}^{J}, \ \boldsymbol{\varphi}_0, \ \{\boldsymbol{\Phi}_l\}_{l=0}^{P-1}, \ \{\boldsymbol{\Sigma}(\boldsymbol{\gamma}_s)\}_{s=1}^{J}]'$ will deliver the ML estimates of $\boldsymbol{\theta}$. Because the states are latent, an application of the iterative Expectation-Maximization (EM) algorithm in which the maximization is applied to

$$\begin{aligned} &-\frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{J}\Pr\left(S_t = s|\Psi_T\right)\ln\det\left(\boldsymbol{\Sigma}(\boldsymbol{\gamma}_s)\right) + \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{J}\Pr\left(S_t =: s|\Psi_T\right)\left(\boldsymbol{A}_0\boldsymbol{y}_{t+1} - \boldsymbol{\mu}(\boldsymbol{\gamma}_s) - \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l}\right)' \\ &\quad \times\boldsymbol{\Sigma}\left(\boldsymbol{\gamma}_{S_{t+1}}\right)^{-1}\left(\boldsymbol{A}_0\boldsymbol{y}_{t+1} - \boldsymbol{\mu}(\boldsymbol{\gamma}_s) - \sum_{l=0}^{P-1}\boldsymbol{A}_l\boldsymbol{y}_{t+1-l}\right), \end{aligned} \tag{17}$$

where $\{\Pr(S_t = s|\Psi_T)\}_{t=1}^{T}$ (s = 1, 2, ..., J) are the smoothed probabilities derived with the classical, Hamilton-Kim smoothing algorithm (Hamilton, 1994). Maximizing such a smoothed probability-weighted log joint density function delivers ML estimates of $\boldsymbol{\theta}$.

The assumption of normality of returns and finite number of states allows us to employ the Baum–Welch algorithm to estimate $\boldsymbol{\theta}$. The Baum–Welch algorithm is an expectation–maximization (EM) algorithm typically applied to HMM. Given an initial set of parameters $\boldsymbol{\theta}_0$ and the realized series of observations, $\{\boldsymbol{y}_t,\}$, the result of this algorithm always converges to a local maximum of the likelihood function. In the E-step, we compute the expected value of the $T+1$ latent Markov states (one at each point in the sample) given the observed data and the current, provisional estimates of the parameters. In the M-step, standard ML methods are used to update the unknown parameter estimates using an expanded data matrix with previous expectations as weights.[6] To protect against the perils of multiple, local maxima, as in Bae et al. (2014), we use randomly selected initial parameters and repeat the parameter estimation for each seed, where each new set of initial parameters is drawn from the results of the previous estimation run, assuming a normal distribution with mean equal to the estimate and variance equal to the square of the standard error obtained from the previous run of the algorithm.[7]

3. The data

We use a cross-section of eight portfolios of stocks, Treasuries, corporate bonds and commodity indices as test assets throughout. In particular, we consider the monthly excess returns of these portfolios over the one-month Treasury bill rate from January 1989 to December 2011.

The stock portfolio is proxied by the CRSP value-weighted index (VW CRSP), which includes all the firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ, obtained from Fama and French's data library. Treasuries are proxied by the 10-year constant maturity Treasury yields, that is, yields on actively traded non-inflation indexed issues adjusted to constant maturities, retrieved from the Federal Reserve of St. Louis FRED data repository. The two corporate bond portfolios are proxied by the Moody's seasoned Aaa corporate bond portfolio and the Moody's seasoned Baa corporate paper portfolio, respectively, obtained from the Federal Reserve of St. Louis FRED data repository. In order to transform yields into returns we use Shiller's (1979) approximation, which defines the holding period yield in terms of yield to maturity. The one-month Treasury Bill return series is obtained from the Ibbotson SBBI Classic Yearbook.

Similarly to research by Bae et al. (2014), the four commodity indices included among the test assets are all Standard & Poor's-Goldman Sachs (S&PGS) Spot Commodity Indices, namely the S&PGS Agriculture and Livestock, the S&PGS Precious Metals, the S&PGS Industrial Metals, and the S&PGS Energy indices. The S&PGS Spot Commodity Indices are built using front-end futures to exploit the proximity of traded future prices to spot prices. However, we are aware that an index of commodity spot prices simply tracks the evolution of the spot prices, and ignores all costs associated with the holding of physical commodities (storage, insurance, etc.). It is therefore an upper bound on the return that an investor in spot commodities would have earned in real time.

The macro-based pricing factors are built following an approach similar to Ludvingson and Ng (2009). We start from their rich database, containing 132 U.S. macroeconomic variables measured at monthly frequency (see Appendix B). We remove all financial return series and add three topical series: the Goldman Sachs Financial Conditions Index (GS-FCI), the Historical News-Based Policy Index (see Baker, Bloom, & Davis, 2012) and the Liquidity Factor of Pastor and Stambaugh (2003).[8] We end up with a set of

[6] Because the EM algorithm needs to store the JT entries of expectations over the latent space for each data pattern, computation time and storage increases exponentially with the number of observations, which makes this algorithm impractical. Indeed, for HMM, the special variant of the EM algorithm referred to as the forward–backward or Baum–Welch algorithm is needed because the model contains a very large number of entries in the joint posterior latent distribution generated by the T+1 latent variables. Baum–Welch algorithm circumvents the computation of this joint posterior distribution making use of the conditional independencies implied by the model, see Dias et al. (2015) for a discussion.

[7] Random draws of the starting parameter values that violated basic admissibility conditions (e.g., combinations of elements in the $\{\boldsymbol{A}_l\}$ matrices that made the VAR system non-stationary) were rejected.

[8] The GS Financial Stability Index is a weighted average of US real short-term interest rates, real long-term corporate bond yields, the real trade-weighted dollar index, and the ratio of equity market capitalization to nominal GDP (see Dudley and Hatzius, 2000). An increase in the GS-FCI indicates tightening of financial conditions, while a decrease indicates easing. The Historical News-Based Policy Index measures economic policy uncertainty in the U.S. by counting the number of articles published every month that contains the words "uncertain" or "uncertainty" in

Table 1
Summary statistics (January 1989–December 2011).
The following tables refer to all the series used in the parts of the paper based on the sample January 31, 1989–December 30, 2011. Panel A shows mean, median, standard deviation, skewness, excess kurtosis, and the p-value of the Jarque–Bera (JB) test. The JB statistic is used to test the hypothesis of normality of the series, hence a p-value lower than α% implies the rejection of the null hypothesis at α % confidence level. Panel B shows the correlations among the series.

Panel A

	Mean	Median	Std. dev.	Skewness	Excess Kurtosis	JB p-value
Factor 1	0.0000	0.0080	0.4385	−1.6940	4.8482	0.0000
Factor 2	0.0000	0.0014	0.3080	−0.7455	4.2518	0.0000
Factor 3	0.0000	−0.0001	0.2568	0.6517	2.5557	0.0000
HP Factor	0.0036	0.0049	0.0442	−0.1246	1.0115	0.0019
Basis Factor	0.0098	0.0109	0.0448	−0.0093	1.3934	0.0000
Momentum Factor	0.0088	0.0091	0.0503	0.3528	3.3293	0.0000
Agriculture and livestock	−0.0002	0.0003	0.0428	−0.0876	1.5607	0.0000
Precious metals	0.0030	−0.0024	0.0469	0.0980	1.2784	0.0001
Industrial metals	0.0008	−0.0007	0.0606	−0.1039	1.4521	0.0000
Energy	0.0071	0.0086	0.0904	0.3851	1.5595	0.0000
10Y treasury bonds	0.0023	0.0020	0.0019	0.0016	−0.7515	0.0389
Aaa corporate bonds	0.0033	0.0031	0.0019	−0.0005	−0.7615	0.0356
Baa corporate bonds	0.0042	0.0039	0.0019	0.0104	−0.7375	0.0437
Value-weighted equity CRSP	0.0054	0.0110	0.0447	−0.6165	1.0217	0.0000

Panel B

	Factor 1	Factor 2	Factor 3	AL	PM	IN	EN	Treasuries	Aaa bonds	Baa bonds	VW CRSP
Factor 1	1.00	0.00	0.00	0.08	−0.05	0.19	0.10	−0.08	−0.08	−0.07	0.09
Factor 2		1.00	0.00	0.08	0.24	0.16	0.27	0.03	0.04	0.06	0.04
Factor 3			1.00	0.00	−0.09	0.14	−0.04	0.12	0.13	0.15	0.10
Agriculture and livestock				1.00	0.29	0.31	0.16	0.12	0.12	0.13	0.22
Precious metals					1.00	0.26	0.21	0.21	0.21	0.21	0.03
Industrial metals						1.00	0.29	0.09	0.10	0.12	0.38
Energy							1.00	−0.01	0.00	0.01	0.09
10Y treasury bonds								1.00	1.00	1.00	−0.02
Aaa corporate bonds									1.00	1.00	−0.01
Baa corporate bonds										1.00	0.00
Value-weighted equity CRSP											1.00

112 variables and we use principal component analysis to summarize their covariance structure. We extract three, five or ten orthogonal factors that summarize 31.4%, 36%, and 52.9% of the variance, respectively. To gain insight on the economic meaning of the factors we investigate the sign of the factor loadings and the R-squares from univariate regressions of each of the factors on the macroeconomic variables (see Appendix C).

The first principal component (F1) is a business cycle, procyclical factor, the value of which increases with industrial production, “help wanted”, and employment growth, and decreases with changes in the unemployment rate. The second principal component (F2) can be defined as an inflation factor that positively loads with relatively high R-squares on most types of price indices (including consumers’, producers’, and personal consumption expenditure deflators). The third principal component (F3) can be defined as an inventory and new orders factor.

Table 1 shows summary statistics. The three macro factors are characterized by a mean equal to zero and by considerable volatility. This finding is explained by the fact that all factors are characterized by large and positive excess kurtosis and non-zero (and highly statistically significant, based on unreported tests) skewness. The evidence of non-normality (non-zero skewness and excess kurtosis) extends to all the portfolios considered, consistently with the fact that for all the series is possible to reject the null of normality in a Jarque–Bera (JB) test. However, bonds are characterized by negative excess kurtosis and essentially symmetric empirical distributions, and the rejection of the null of normality tends to be relatively weak, with p-values between one and five percent. Of course, the HMM methods developed in Section 2 can be considered as ways to capture and forecast such pervasive departures from normality, as discussed in Buckley, Saunders, and Seco (2008) and Dias et al. (2015). Interestingly, commodities show high return volatility not always associated to a high mean return: in fact, the Agriculture and livestock class displays a negative mean return despite a standard deviation close to that of stocks.

Table 1 also shows the correlations among the series. Interestingly, F2 is the pricing factor showing the highest correlations with the commodity series, implying that inflation tends to simultaneously correlate with all the commodity returns. This finding is consistent with a literature that considers commodities as inflation hedges (see, e.g., Erb & Harvey, 2006). The correlations among the test assets and the remaining pricing factors are generally low, except for Industrial Metals that significantly correlates with Precious Metals and Agriculture and Livestock. However, the range of correlations between pairs of commodities indices is wide, from 0.16 between agricultural and energy commodities to 0.31 between agricultural and industrial commodities.

4. Empirical results

4.1. Statistical model selection criteria

In Table 2, we conduct a formal specification search by comparing the values of three different information criteria—Akaike (AIC), Bayes–Schwarz (BIC), and Hannan–Quinn (HQIC)—to select the model that most likely represents the unknown data generating process. The table is organized around four panels. In the first panel, we present single-state models with three and five pricing factors and different VAR structures. In particular, with reference to the $\mathbf{\Phi}_1$ matrix shown in Eq. (9), we present model specifications that consist of a full VAR(1) where $\mathbf{\Phi}_1$ is a full matrix (each factor depends on its own lags, and on past values of both the other factors and of the test assets); a block factor VAR(1), where each

association with terms related to the economic cycle. Finally, Pastor–Stambaugh’s Liquidity Factor is a cross-sectional average of individual stock liquidity measures.

Table 2
Model selection.
In the table, SIC is the Bayes–Schwarz information criterion, AIC is the Akaike information criterion, and HQIC is the Hannan–Quinn information criterion. The Chi-squared test is a joint test applied to the Euler conditions (Eq. (2) in the text) of the null that they are simultaneously satisfied. We have boldfaced the best fitting model according to each of the information criteria and all models not rejected by the Chi-squared pricing test. HJ represents the Hansen–Jagannathan distance under the null of correct specification of the SDF. All models are specified with a VAR order of 1.

		Max Log-Likelihood	Avg. Log-Likelihood	No. Obs.	No. parameters	Saturation ratio	BIC	AIC	HQIC	Pricing RMSE	Chi-squared test	*p*-value	HJ distance
SINGLE STATE	3 factors, single state, full VAR model	8056.37	2.66	3025	40	75.63	−5.22	−5.30	−5.27	0.0096	149.14	0.000	6.832
	3 factors, single state, block Factor VAR	7964.23	2.63	3025	16	189.06	−5.22	−5.26	−5.24	0.0091	11,913	0.000	6.687
	3 factors, single state, diagonal Factor VAR	7828.44	2.59	3025	10	302.50	−5.15	−5.17	−5.16	0.0092	9392.8	0.000	6.687
	5 factors, single state, full VAR model	7537.22	2.11	3575	76	47.04	−4.04	−4.17	−4.13	0.0090	215.92	0.000	6.645
	5 factors, single state, block Factor VAR	7432.32	2.08	3575	36	99.31	−4.08	−4.14	−4.12	0.0092	10,932	0.000	6.691
	5 factors, single state, diagonal Factor VAR	7187.82	2.01	3575	16	223.44	−3.98	−4.01	−4.00	0.0092	8801.2	0.000	6.687
2 STATES	3 factors, 2 states, full VAR model	8228.45	2.72	3025	82	36.89	−5.22	−5.39	−5.33	**0.0054**	7.72	**0.461**	5.791
	3 factors, 2 states, block Factor VAR	8218.11	2.72	3025	34	88.97	−5.34	−5.41	−5.39	0.0093	1937.4	0.000	6.718
	3 factors, 2 states, diagonal Factor VAR	8080.99	2.67	3025	22	137.50	−5.28	−5.33	−5.31	0.0093	2080.0	0.000	6.726
	5 factors, 2 states, full VAR model	7866.04	2.20	3575	154	23.21	−4.05	−4.31	−4.22	0.0132	6.87	**0.551**	7.697
	5 factors, 2 states, block Factor VAR	7720.26	2.16	3575	74	48.31	−4.15	−4.28	−4.23	0.0092	2111.9	0.000	6.699
	5 factors, 2 states, diagonal Factor VAR	7513.10	2.10	3575	34	105.15	−4.13	−4.18	−4.16	0.0091	2454.9	0.000	6.670
3 STATES	**3 factors, 3 states, full VAR model**	8344.14	2.76	3025	126	24.01	−5.18	−5.43	−5.34	0.0075	**4.03**	**0.854**	6.000
	3 factors, 3 states, block Factor VAR	8301.35	2.74	3025	54	56.02	**−5.35**	−5.45	−5.41	0.0086	2663.5	0.000	6.778
	10 factors, 3 states, block Factor VAR	6987.68	1.41	4950	133	37.22	−2.59	−2.77	−2.71	0.0094	1474.6	0.000	NA
	3 factors, 3 states, diagonal Factor VAR	8182.93	2.71	3025	36	84.03	−5.31	−5.39	−5.36	0.0092	2857.3	0.000	6.712
	5 factors, 3 states, full VAR model	7978.98	2.23	3575	234	15.28	−3.93	−4.33	−4.19	0.0092	4.50	**0.809**	6.135
	5 factors, 3 states, block Factor VAR	7821.17	2.19	3575	114	31.36	−4.11	−4.31	−4.24	0.0093	1918.9	0.000	6.784
	5 factors, 3 states, diagonal Factor VAR	7609.88	2.13	3575	54	66.20	−4.13	−4.23	−4.19	0.0095	5871.4	0.000	6.794
4 STATES	3 factors, 4 states, full VAR model	8006.54	2.65	3025	172	17.59	−4.84	−5.18	−5.06	0.0179	14.35	0.073	9.742
	3 factors, 4 states, block Factor VAR	8376.91	2.77	3025	76	39.80	−5.34	**−5.49**	**−5.43**	0.0102	3085.2	0.000	6.849
	3 factors, 4 states, diagonal Factor VAR	7932.80	2.62	3025	52	58.17	−5.11	−5.21	−5.17	0.0099	3231.6	0.000	6.926
	5 factors, 4 states, full VAR model	7682.41	2.15	3575	268	13.34	−3.68	−4.15	−3.98	0.0110	7.52	**0.482**	7.076
	5 factors, 4 states, block Factor VAR	7593.78	2.12	3575	108	33.10	−4.00	−4.19	−4.12	0.0090	1409.6	0.000	6.636
	5 factors, 4 states, diagonal Factor VAR	7433.85	2.08	3575	68	52.57	−4.00	−4.12	−4.08	0.0099	2444.9	0.000	6.930

factor depends on its own lags and on past values of the pricing factors; and lastly, a diagonal factor VAR(1) where each factor depends only on its own past values, which is equivalent to stacking K different AR(1) processes (that may be cross-serially correlated).

In the remaining three panels we entertain non-linear, three- and five-factor HMM VAR(1) (i.e., MSVAR) models with two, three, and four regimes. Also in this case, we present three competing structures of the $\mathbf{\Phi}_1$matrix (as defined in Eq. (13)): full MSVAR, block factor MSVAR, and diagonal factor MSVAR. In addition, we also display the information criteria for a HMM model with ten factors to assess whether including a higher number of macro-related components (which are indeed able to explain more than 50% of the total variability of the series in the face of 31.4% and 36% of the three- and five-component sets) enhance the accuracy of the model. However, because a ten-factor HMM requires the estimation of a huge number of parameters, we limit this exercise to a three-state block MSVAR model, where "only" 133 parameters need to be estimated.[9]

The rationale behind information criteria is to provide a measure of statistical accuracy that strikes a balance between goodness of fit and parsimonious specifications of the model.[10] The most parsimonious BIC selects a three-state, three-factor block MSVAR(1) model, which requires the estimation of only 54 parameters and has a saturation ratio (number of observations per parameter) of 56. However, both the AIC and the BIC point towards a more richly parameterized four-state three-factor block MSVAR(1) model. Although this model requires the estimation of 76 parameters (i.e., 22 parameters more than its three-state counterpart selected by BIC), it has a saturation ratio of approximately 40, which is still considered acceptable in the literature.

Interestingly, all the criteria select HMM models, in which good and bad states are reflected in the dynamics of a non-linear SDF, over single-state models. In a statistical perspective, the only uncertainty concerns whether three or four regimes should be specified. Furthermore, additional factors do not seem to improve the statistical accuracy of the model. In particular, the ten-factor model does not provide any improvement of the fit to the data.

To provide an idea of the resulting estimates, Table 3 shows the empirical results for the four-states, three-factor block MSVAR(1) model. As a benchmark, Appendix D presents the estimates for the single-state counterparty of this model. For the sake of brevity, we do not report the estimates of the three-state, three factor block MSVAR(1) model that is selected by the BIC, but these are available upon request from the authors. In particular, Table 3 shows the AR(1) coefficients and the intercept for each of the factors in the VAR(1) that describes their dynamics, the loadings of the SDF on the factors and its intercept γ_0 in any of the four regimes, and the residual covariance matrix of the factors and test assets. Interestingly, in such model the majority of the coefficients is statistically significant at all standard levels of confidence. More precisely, the SDF strongly depends on the three factors and the corresponding loadings in all regimes are precisely estimated with small, practically zero p-values, except for the loading on the first factor (that we have interpreted as a business cycle proxy in Section 3), but only in the third regime.

The four regimes can be interpreted looking at their estimated (residual) variances and at the smoothed probabilities presented in Fig. 1. Regime 1 is characterized by high variances and a low implied duration (on average 3 months). Fig. 1 shows some spikes in the smoothed probabilities of this state in correspondence to the first half of the 1990s (when a short recession took place) and at the beginning of the new millennium, characterized by the burst of the dot-com bubble. Therefore, we interpret Regime 1 as a "pre-financialization crisis regime", referring to the fact that most of its occurrences tend to precede the process of "financialization" of commodities, i.e., the change in their relationships with traditional assets classes that has (allegedly) occurred over the last decade (see, e.g., Büyüksahin & Robe, 2014; Tang & Xiong, 2012). Regime 2 is instead a tranquil regime, characterized by a lower variance and a longer implied duration (approximately 10 months) vs. regime 1, consistently with the bulk of the empirical asset pricing literature (see e.g., Dias et al., 2015, and references therein), that describes the "good", state as highly persistent. Because regime 2 seems to characterize most of the time before 2000, we define it as the "pre-financialization bull regime". Interestingly, regimes 3 and 4 are similar to regimes 2 and 1, respectively, but almost exclusively occur after the new millenium. Regime 3 is less volatile and more persistent, with average duration of 5 months, than regime 2, while regime 4 is a deeply turbulent state that tends to characterize the period of 2008-2010, i.e., the recent financial crisis. Accordingly, we interpret these two states as the "post-financialization bull regime" and the "post-financialization crisis regime", respectively.

In Table 3, the estimated SDF loads negatively on the business cycle factor during the crises regimes 1 and 4. On the opposite, the SDF load coefficient is low but positive in regime 2 and not precisely estimated in regime 3. This has some intuitive sense: if we interpret it as a reflection of the marginal utility of future wealth (e.g., $m_{t+1} = U'(W_{t+1})/U'(W_t)$ where $U(\bullet)$ is a utility function), the SDF tends to be relatively low during economic booms and thus, in general, a positive shock to the business cycle in good times should decrease the SDF. However, the impact of a positive shock in the business cycle is much more pronounced during recession periods than in an already booming economy, when it can even be interpreted as a sign of overproduction and lead to an anticipation of a bubble burst. Not surprisingly, the single-state benchmark presented in Appendix D, that cannot empirically separate across different states, shows a precisely estimated but small, negative loading on the business cycle factor that intuitively "averages" across the regime-specific values in Table 3.

Finally, we observe that the intercept of the SDF is negative across all regimes, and this represents no problem given that (10) expresses the SDF in log-exponential form. However, the magnitude (in absolute terms) of this coefficient seems to have structurally increased after the "financialization" of the commodities kicked in, from −0.0014 and −0.0029 in regimes 1 and 2, to −0.0099 and −0.0074 in regimes 3 and 4, respectively. This suggests that, net of other systematic influences, state prices may have decreased as a result of the financialization process and this may reflect structural, systemic contagion risks across different asset markets. Not surprisingly, the SDF intercept of the single-state benchmark model is negative and equal to −0.0043, which is close to an average of the values observed in each of the four regimes in Table 3. This clearly shows the limitations of a single-state model.

4.2. Pricing performance

In this section, we compare the pricing performance of the competing SDFs presented in Table 2. Indeed, along with the information criteria, Table 2 reports the Root Mean Square Error (RMSE) in pricing, which measures the in-sample difference between the returns predicted by the model and the observed ones. More precisely, given the asset pricing model in (5), the RMSE is computed

[9] We have also estimated VAR and MSVAR models with more than one autoregressive lag but, due to the high number of parameters to be estimated, these models are never competitive with their one-lag counterparties. Therefore, also to save space, we refrain from including them in Table 2.

[10] All the information criteria are based on a formula built on −2 times the average log likelihood function adjusted by a penalty function, $\varphi(T)$, multiplied by the number of estimated parameters. In the AIC $\varphi(T) = 2$, while in the BIC $\varphi(T) = \ln T$; finally, in the HQIC $\varphi(T) = 2\ln(\ln T)$.

Table 3
Model estimates- three-factor block VAR(1), four regimes.

	Regime 1								Regime 2			
	Coefficient	Std. error	*t*-Statistic	*p*-value					Coefficient	Std. error	*t*-Statistic	*p*-value
Factor VAR coefficients, SDF loadings and estimated transition matrix												
[F1/F1] Coeff.	**0.4220**	0.028	15.327	0.000	**Transition matrix**				**0.3501**	0.044	7.911	0.000
[F1/F2] Coeff.	**−0.0662**	0.027	−2.437	0.015	0.6972	0.0544	0.0709	0.0840	0.0757	0.087	0.869	0.385
[F1/F3] Coeff.	**−0.2126**	0.035	−5.994	0.000	0.1715	0.8983	0.0000	0.0000	**−0.5066**	0.055	−9.258	0.000
F1: Intercept	**−1.2489**	0.136	−9.186	0.000	0.1312	0.0238	0.8002	0.2880	**1.3299**	0.157	8.455	0.000
[F2/F1] Coeff.	**0.0049**	0.035	0.139	0.889	0.0002	0.0235	0.1288	0.6280	**−0.0747**	0.021	−3.485	0.000
[F2/F2] Coeff.	**−0.5612**	0.039	−14.431	0.000	**Implied durations**				**−0.1975**	0.043	−4.632	0.000
[F2/F3] Coeff.	0.0369	0.049	0.753	0.451	Regime 1:		3.30		−0.0484	0.027	−1.814	0.070
F2: Intercept	**−0.4996**	0.127	−3.946	0.000	Regime 2:		9.84		**0.2989**	0.073	4.083	0.000
[F3/F1] Coeff.	**−0.3382**	0.023	−14.483	0.000	Regime 3:		5.01		**−0.2523**	0.027	−9.518	0.000
[F3/F2] Coeff.	**−0.1055**	0.024	−4.319	0.000	Regime 4:		2.69		−0.0360	0.049	−0.727	0.467
[F3/F3] Coeff.	**0.3997**	0.031	12.778	0.000	**Ergodic probs.:**				0.0620	0.032	1.962	0.050
F3: Intercept	**−0.5977**	0.101	−5.894	0.000	Regime 1:		0.182		**0.4463**	0.110	4.044	0.000
SDF: Loading on F1	**−0.0014**	0.0001	−26.265	0.000	Regime 2:		0.306		**0.0002**	0.0000	4.752	0.000
SDF: Loading on F2	**0.0006**	0.0001	11.099	0.000	Regime 3:		0.366		**−0.0016**	0.0001	−17.691	0.000
SDF: Loading on F3	**−0.0015**	0.0001	−20.178	0.000	Regime 4:		0.146		**0.0014**	0.0001	21.009	0.000
SDF: Intercept	**−0.0014**	0.0003	−4.774	0.000					**−0.0029**	0.0002	−12.226	0.000

	Regime 3				Regime 4			
	Coefficient	Std. error	*t*-Statistic	*p*-value	Coefficient	Std. error	*t*-Statistic	*p*-value
Factor VAR coefficients and SDF loadings								
[F1/F1] Coeff.	**0.6864**	0.041	16.816	0.000	**0.8857**	0.022	39.424	0.000
[F1/F2] Coeff.	−0.0561	0.051	−1.102	0.271	**0.2592**	0.027	9.597	0.000
[F1/F3] Coeff.	**−0.2928**	0.051	−5.768	0.000	**−0.1153**	0.044	−2.614	0.009
F1: Intercept	**0.4702**	0.128	3.673	0.000	**−0.5730**	0.191	−3.008	0.003
[F2/F1] Coeff.	**−0.1189**	0.026	−4.566	0.000	**−0.1162**	0.040	−2.887	0.004
[F2/F2] Coeff.	−0.0373	0.024	−1.533	0.125	**−0.1366**	0.041	−3.365	0.001
[F2/F3] Coeff.	−0.0132	0.025	−0.539	0.590	0.0192	0.066	0.291	0.771
F2: Intercept	**0.3016**	0.124	2.437	0.015	**−1.1753**	0.358	−3.279	0.001
[F3/F1] Coeff.	**−0.4456**	0.033	−13.590	0.000	**−0.2054**	0.022	−9.202	0.000
[F3/F2] Coeff.	0.0056	0.040	0.139	0.889	**0.1718**	0.026	6.607	0.000
[F3/F3] Coeff.	−0.0525	0.040	−1.312	0.190	**0.3054**	0.042	7.195	0.000
F3: Intercept	**0.4066**	0.109	3.741	0.000	**−0.9587**	0.191	−5.013	0.000
SDF: Loading on F1	−0.0001	0.0001	−0.842	0.400	**−0.0011**	0.0001	−10.943	0.000
SDF: Loading on F2	**−0.0027**	0.0001	−42.013	0.000	**−0.0024**	0.0001	−41.559	0.000
SDF: Loading on F3	**0.0010**	0.0001	13.303	0.000	**−0.0025**	0.0001	−21.225	0.000
SDF: Intercept	**−0.0099**	0.0004	−24.872	0.000	**−0.0074**	0.0009	−7.877	0.000

Continued on next page

Table 3
Continued.

	Residual covariance matrix of factors and test assets										
	(Regime 1 above; Regime 2 below)										
	Factor 1	Factor 2	Factor 3	Agr. and Livestock	Precious	Industrials	Energy	10Y Treasuries	Aaa Corporate	Baa Corporate	VW Equity CRSP
Factor 1	6.456|5.951	2.1495	2.5539	−0.0098	−0.0097	−0.0100	−0.0221	−0.0048	−0.0071	0.0429	0.0051
Factor 2	−0.59925	6.259|1.094	1.8393	−0.0006	−0.0006	−0.0011	−0.0277	−0.0263	0.0056	−0.0213	−0.0049
Factor 3	2.69437	−0.3575	3.363|3.002	−0.0059	−0.0057	−0.0062	−0.0022	−0.0125	−0.0018	0.0259	−0.0008
Agriculture and livestock	0.00650	−0.0023	0.00567	0.00003|0.00001	0.00003	0.00002	0.00002	0.00001	0.00000	−0.00011	−0.00002
Precious metals	0.00642	−0.0023	0.00560	0.00001	0.00003|0.00001	0.00002	0.00003	0.00000	−0.00001	−0.00011	−0.00001
Industrial Metals	0.00649	−0.0023	0.00569	0.00001	0.00001	0.00002|0.00002	0.00002	0.00001	0.00001	−0.00010	−0.00003
Energy	−0.01368	0.0005	−0.00146	−0.00001	0.00000	−0.00002	0.0020|0.0015	−0.00016	−0.00032	0.00026	−0.00136
10Y treasury bonds	0.00992	−0.0044	0.00116	0.00001	0.00001	0.00002	0.00008	0.00054|0.0013	0.00007	−0.00025	−0.00033
Aaa corporate bonds	−0.01413	0.0007	0.00175	0.00000	0.00000	0.00001	−0.00009	0.00005	0.0010|0.0011	−0.00036	0.00079
Baa corporate bonds	0.01123	−0.0002	0.00768	0.00002	0.00002	0.00002	−0.00002	0.00018	0.00039	0.0043|0.0014	0.00081
VW equity CRSP	−0.00630	0.0119	−0.02114	−0.00007	−0.00006	−0.00008	0.00017	0.00018	−0.00020	−0.00003	0.01142|0.0063
	(Regime 3 above; Regime 4 below)										
	Factor 1	Factor 2	Factor 3	Agr. and livestock	Precious	Industrials	Energy	10Y treasuries	Aaa corporate	Baa corporate	Baa corporate
Factor 1	4.733|9.577	1.41151	1.13496	−0.00245	−0.00244	−0.00241	0.00116	0.00547	0.00403	0.00842	−0.00349
Factor 2	−4.32624	5.874|38.56	−1.14499	−0.01730	−0.01729	−0.01732	−0.02381	−0.00111	0.01706	0.00362	0.05725
Factor 3	7.19695	−2.44773	3.564|9.660	0.00721	0.00721	0.00719	0.01139	0.00378	−0.00863	0.00042	−0.03138
Agriculture and livestock	−0.01906	−0.08629	−0.02670	0.00009|0.00034	0.00008	0.00009	0.00008	0.00002	−0.00006	0.00001	−0.00026
Precious metals	−0.01884	−0.08610	−0.02673	0.00034	0.00008|0.00035	0.00009	0.00008	0.00002	−0.00006	0.00001	−0.00025
Industrial metals	−0.01895	−0.08652	−0.02669	0.00034	0.00034	0.00010|0.00033	0.00008	0.00002	−0.00006	0.00001	−0.00028
Energy	−0.01291	−0.04177	0.01790	−0.00004	−0.00004	−0.00003	0.0018|0.0032	0.00054	0.00022	0.00128	−0.00008
10Y treasury bonds	−0.02481	−0.02036	0.00318	0.00002	0.00002	0.00002	0.00126	0.0017|0.0046	0.00017	0.00063	−0.00045
Aaa corporate bonds	−0.06913	0.06665	−0.06850	0.00014	0.00016	0.00012	0.00002	0.00279	0.0021|0.0058	0.00046	0.00070
Baa corporate bonds	0.03823	0.05599	0.04579	−0.00046	−0.00044	−0.00046	0.00262	0.00288	0.00243	0.0028|0.0081	0.00091
VW equity CRSP	−0.02005	0.20483	0.03212	−0.00080	−0.00081	−0.00080	0.00237	0.00308	0.00156	0.00483	0.0061|0.0093

Note: Significant conditional mean coefficients are boldfaced.

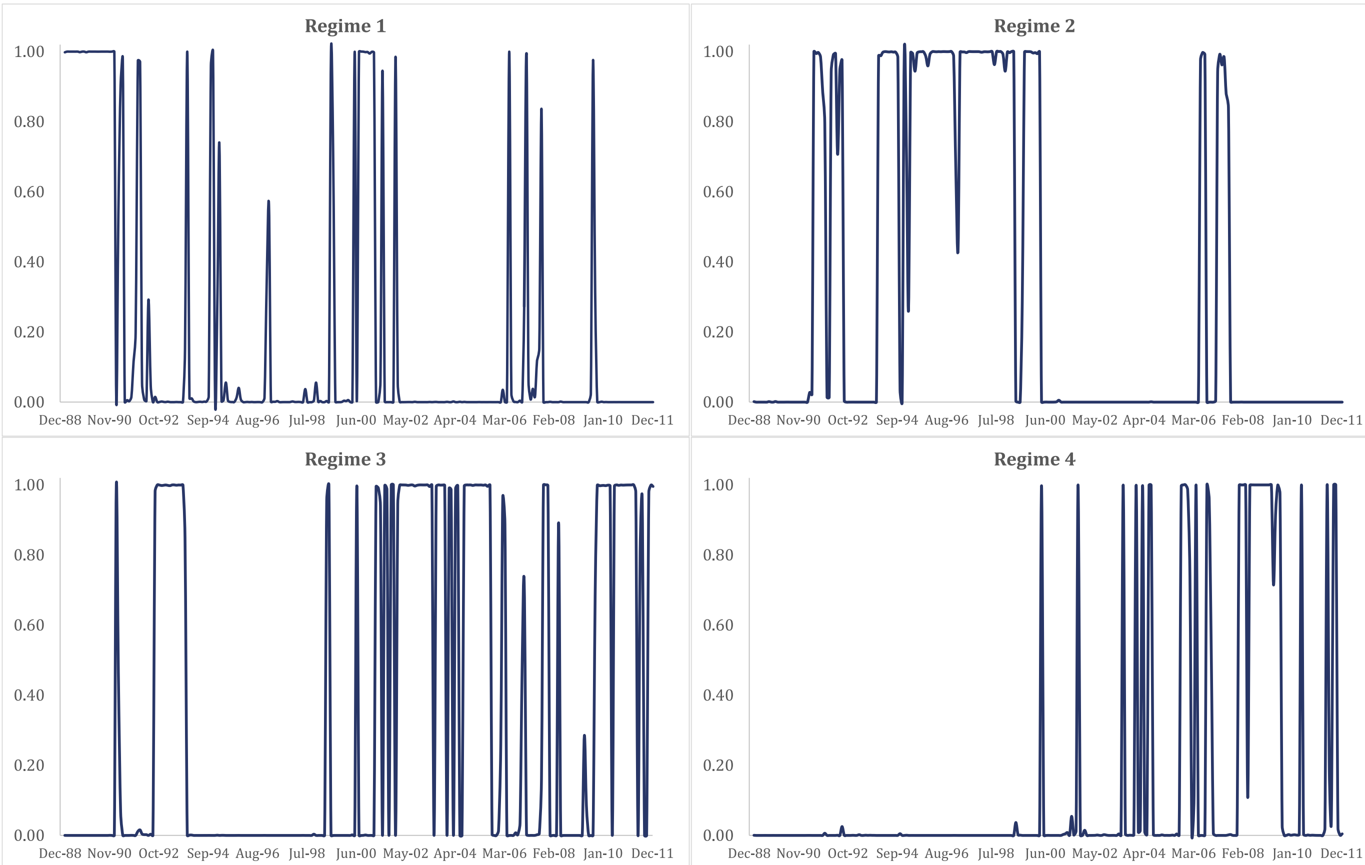

Fig. 1. Smoothed state probabilities from three-factor block VAR(1), four regimes.

as

$$RMSE \equiv \sqrt{\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{r}_{i,t+1}^{\mathcal{M}} - r_{i,t+1}\right)^2}, \tag{18}$$

where $\hat{r}_{i,t+1}^{\mathcal{M}}$ is the fitted asset i's return on the basis of model $\mathcal{M}$, i.e.,

$$\hat{r}_{i,t+1}^{\mathcal{M}} = -\hat{m}_{t+1}^{\mathcal{M}} - \frac{1}{2}\sigma_m^{2,\mathcal{M}} - \frac{1}{2}\sigma_i^2 - \sigma_{i,m}^{\mathcal{M}} \tag{19}$$

Clearly, the lowest the RMSE, the less the predicted values differ from the observed ones and the highest the pricing accuracy of the model. The two-states, three-factor full MSVAR turns out to be the top performing model according to the RMSE, being the one with the minimal average distance between observed and predicted values.

Moreover, to test whether or not the difference between fitted and observed values is statistically different from zero, we perform a two-tailed chi-squared test. In particular, the test statistic reported in Table 2 is defined as follows. Given the n Euler conditions expressed as

$$1 = E[M_{t+1}(1 + R_{i,t+1})|\Im_t], \qquad i = 1, 2, \ldots, n \tag{20}$$

a Wald-type test of the joint validity of the I conditions is:

$$\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}_{t+1}^{\mathcal{M}}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\right)'\hat{\boldsymbol{V}}^{-1} \times\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}_{t+1}^{\mathcal{M}}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\right), \tag{21}$$

where $\boldsymbol{\iota}_I$ is a $n \times 1$ vector of ones and $\mathbf{V}$ is the covariance matrix of the pricing errors, i.e.,

$$\hat{\boldsymbol{V}} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}_{t+1}^{\mathcal{M}}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\left(\boldsymbol{\iota}_I - \hat{M}_{t+1}^{\mathcal{M}}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)'. \tag{22}$$

Noticeably, under mild conditions of stationarity of both asset returns and the SDF and of existence of moments of the moment conditions, it turns out that

$$\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}_{t+1}^{\mathcal{M}}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\right)'\hat{\boldsymbol{V}}^{-1} \times\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}_{t+1}^{\mathcal{M}}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\right) \overset{T\to\infty}{\to} \chi_I^2. \tag{23}$$

Observing the p-values for the test, reported in Table 2 together with the test statistic, we notice that the null of equality between observed and fitted values cannot be rejected for five alternative HMM models: the two-state, three- and five-factor full MSVAR models, the three-state, three- and five-factor full MSVAR models, and the four-state, five-factor MSVAR model.

These results are interesting in several ways. First, we notice that in the single-state case, the null of no difference between predicted and observed returns is always rejected at any conventional level of significance (the p-values are equal to zero). This signals that the use of a non-linear SDFs really makes the difference when it comes to pricing performance. Second, the models that were selected by the information criteria are both rejected when their pricing performance is considered. Indeed, the hypothesis of accurate pricing performance (no difference between observed and predicted returns) is rejected for both the three-state, three-factor

Table 4
Reality check results.

Model	HJ-distance	P-value H_0: HJ = 0	T	Reality check p-value
3 factors, 2 states, full VAR(1)	5.7915	0.0001	−40.937	**0.9568**
3 factors, 3 states, full VAR(1)	6.0004	0.0001	40.937	0.8827
5 factors, 3 states, full VAR(1)	6.1350	0.0001	68.070	0.8395
5 factors, 4 states, block Factor VAR(1)	6.6365	0.0000	174.469	0.3889
5 factors, single state, full VAR(1)	6.6448	0.0000	176.297	0.3148

block MSVAR and the four-state, three-factor block MSVAR models. More generally, as far as the pricing performance is concerned, the full MSVAR models are strongly preferred over their more parsimonious block MSVAR counterparts. In fact, a full VAR structure, where each pricing factor also depends on the lags of the test asset returns, appears to be required to accurately predict future asset returns.

Because our empirical efforts are devoted to find an SDF that correctly prices a medium-sized cross-section of asset returns that includes commodities, it also seems natural to adopt Hansen and Jagannathan's (1997) distance measure to quantify (in the appropriate SDF mean-variance space) the distance between the candidate SDFs and what is required by our asset return data. In particular, HJ propose an in-sample measure of fit defined as a quadratic function of observed pricing errors (here, the n x 1 vector $\boldsymbol{\iota}_I - \hat{M}^{\mathcal{M}}_{t+1}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})$ for $t = 1, 2, \ldots, T-1$) weighted by the inverse of the second moment matrix of gross asset returns, $\boldsymbol{SM} \equiv E[(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})']$, which is positive definite by construction:[11]

$$HJ^{\mathcal{M}} \equiv \sqrt{\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}^{\mathcal{M}}_{t+1}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\right)'\boldsymbol{SM}^{-1}} \times \sqrt{\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\boldsymbol{\iota}_I - \hat{M}^{\mathcal{M}}_{t+1}(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})\right)\right)} \qquad (24)$$

The parameter of each SDF model enter $HJ^{\mathcal{M}}$ through $\hat{M}^{\mathcal{M}}_{t+1}$. Hansen and Jagannathan show that for a given set of parameters characterizing (24), $HJ^{\mathcal{M}}$ equals the maximum pricing error generated by the model. Obviously, lower values of $HJ^{\mathcal{M}}$ are then preferable. Such values are reported in the rightmost column of Table 2 and reveal that the minimal HJ distance in the SDF mean-variance space are achieved by two- and three-state HMM models. In particular, a three-factor, two-state full VAR(1) achieves a $HJ^{\mathcal{M}}$ value of 5.791. More generally, both three- and five-factor models with two and three regimes appear to minimize the distance from the SDF that the data imply.

For each of the models in Table 2 (i.e., as $\mathcal{M}$ changes), we also test the null hypothesis that $HJ^{\mathcal{M}} = 0$. However, the distribution of $HJ^{\mathcal{M}}$ under the null fails to be standard because the weighting matrix that appears in (24) is not optimal in the sense of Hansen (1982). Therefore, the p-value of the null $HJ^{\mathcal{M}} = 0$ has been computed using 10,000 simulations for a weighted sum of chi-squared distributions, exploiting Jagannathan and Wang's (1996) finding that such a weighted sum represents the asymptotic distribution of $(HJ^{\mathcal{M}})^2$. The results of such a test are not reported in Table 2 because all the corresponding p-values turned out to be essentially nil, an indication that even though some models may be more precise than others are, none of them could be characterized by insignificant overall violations of the set of Euler conditions on which our paper was based. This is of course unsurprising in the light of the results in the literature that tend to report that empirical SDF based on macroeconomic variables have a hard time fitting the cross-section of asset returns.

However, it may be also interesting to ask—given that all estimated SDFs appear to have been rejected—whether any models are at least superior, i.e., rejected with a statistically significant difference turning in their favor. Unfortunately, the distribution of $HJ^{\mathcal{M}}$ under the null hypothesis is not the same regardless of the model considered. Therefore, to make comparisons possible, we have selected the five models with the lowest HJ distance and applied the test proposed by Chen and Ludvingson (2009) and based on White's (2000) "reality check" method. For each of the five models, labeled as $i = 1, 2, 3, 4, 5$, the null hypothesis is:

$$H_o: \max_{j=2,3,4,5}\left[\left(HJ^i\right)^2 - \left(HJ^j\right)^2\right] \le 0 \qquad (25)$$

In words, under the null, the modeled labeled as $i = 1, 2, 3, 4, 5$, has a smaller (maximum) pricing error vs. all other models. Because th e alternative hypothesis is that $\max_{j=2,3,4,5}[(HJ^i)^2 - (HJ^j)^2] > 0$, this is a one-sided test. The value of the test statistic for model i is simply:

$$T_i \equiv \max_{j=2,3,4,5}\sqrt{T}\left[\left(HJ^i\right)^2 - \left(HJ^j\right)^2\right] \qquad (26)$$

The p-value of the test is computed considering the quantile of the distribution of T_i simulated according to the reality check procedure with a block bootstrap.

Table 4 shows that the 3-factor, full VAR(1) HMM with two regimes provides a superior pricing accuracy in the SDF space: when this model is used as a benchmark, the test statistic is negative (−40.9) and the null that the corresponding HJ distance is lower than all other models in the Table, cannot be rejected with a very high p-value in excess of 95%. Interestingly, also in this case, an unrestricted full VAR(1) dynamics is required for an HMM SDF to yield accurate pricing.[12]

Therefore a relatively simple, 3-factor, full VAR(1) HMM with two regimes is rejected in the HJ SDF space, but is at least the most accurate across all of our models. As already mentioned, this model has in fact the lowest possible pricing RMSE and the null hypothesis of correct pricing as expressed by a zero overall RMSE cannot be rejected by a chi-square test based on moment conditions and the covariance matrix of pricing errors. Table 5 shows the estimates for such a two-regime model. Also in this case, to save

[11] Clearly, $\boldsymbol{SM} \equiv E[(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})(\boldsymbol{\iota}_I + \boldsymbol{R}_{t+1})']$, must be estimated by some empirical, sample construct. In any event, such an estimator will differ from $\hat{\boldsymbol{V}}$ because based only on asset return data and not on pricing errors. The 10-factor model HJ distance statistic could not be computed because the number of factors exceeded the number of tests assets.

[12] We have also used Chen and Ludvigson (2009)-style reality check methodology to test whether we can reject the null of a smaller HJ distance of two-, three-, and four-state HMM SDF models vs. their single regime counterparts, given a fixed choice of number of macroeconomic factors. In the cases of two- and three-state HMM SDFs, we report that the null of smaller HJ distance vs. the single-state case cannot be rejected. Interestingly, the opposite occurs in the case of the four-state block VAR model. This confirms that both in the pricing and in the SDF spaces, an unrestricted full VAR dynamic structure plays a key role.

Table 5
Model estimates- three-factor full VAR(1), Two regimes.

	Regime 1						Regime 2			
	Coefficient	Std. error	*t*-Statistic	*p*-value			Coefficient	Std. error	*t*-Statistic	*p*-value
	Factor VAR coefficients, SDF loadings and estimated transition matrix									
[F1/F1] Coeff.	**0.5178**	0.044	11.757	0.000			**0.9524**	0.029	33.272	0.000
[F1/F2] Coeff.	−0.0093	0.068	−0.137	0.891			**0.3173**	0.030	10.478	0.000
[F1/F3] Coeff.	**−0.4335**	0.062	−7.046	0.000	**Transition matrix**		**−0.3387**	0.050	−6.781	0.000
F1: Intercept	**−2.3291**	0.838	−2.779	0.005	0.9328	0.3265	**−2.0790**	0.900	−2.309	0.021
[F2/F1] Coeff.	**−0.0871**	0.026	−3.384	0.001	0.0672	0.6735	**−0.3484**	0.037	−9.494	0.000
[F2/F2] Coeff.	**−0.2239**	0.036	−6.270	0.000			**−0.3094**	0.036	−8.505	0.000
[F2/F3] Coeff.	**0.0767**	0.032	2.381	0.017	**Implied durations**		**0.1776**	0.060	2.983	0.003
F2: Intercept	−0.8196	0.431	−1.901	0.057	Regime 1:	14.87	1.3048	1.074	1.215	0.224
[F3/F1] Coeff.	**−0.3077**	0.033	−9.253	0.000	Regime 2:	3.06	**−0.1368**	0.029	−4.707	0.000
[F3/F2] Coeff.	0.0049	0.051	0.095	0.924			**0.3051**	0.030	10.287	0.000
[F3/F3] Coeff.	**0.1083**	0.046	2.342	0.019			**0.3462**	0.049	7.108	0.000
F3: Intercept	**−1.7211**	0.630	−2.734	0.006	**Ergodic probs.:**		**5.5673**	0.878	6.338	0.000
SDF: Loading on F1	**−0.0002**	0.000	−2.412	0.016	Regime 1:	0.829	**−0.0006**	0.0001	−5.753	0.000
SDF: Loading on F2	**−0.0021**	0.000	−17.229	0.000	Regime 2:	0.171	**−0.0015**	0.0001	−17.103	0.000
SDF: Loading on F3	**0.0008**	0.000	6.245	0.000			**−0.0020**	0.0001	−15.773	0.000
SDF: Intercept	**−0.0049**	0.000	−14.864	0.000			−0.0006	0.0009	−0.681	0.496

Residual covariance matrix of factors and test assets

(Regime 1 above; Regime 2 below)

	Factor 1	Factor 2	Factor 3	Agr. and livestock	Precious	Industrials	Energy	10Y treasuries	Aaa corporate	Baa corporate	VW equity CRSP
Factor 1	5.134|8.322	0.5982	2.1685	−0.0008	−0.0007	−0.0008	−0.0009	0.0070	−0.0055	0.0167	−0.0073
Factor 2	−2.07911	2.798|19.232	0.0988	−0.0069	−0.0069	−0.0069	−0.0038	−0.0063	0.0045	0.0077	0.0252
Factor 3	4.97309	−1.2470	3.017|9.465	0.0018	0.0019	0.0018	0.0022	−0.0044	−0.0051	0.0052	−0.0192
Agriculture and livestock	−0.01494	−0.0254	−0.02358	0.00003|0.00019	0.00002	0.00003	0.00002	0.00003	0.00000	0.00000	−0.00012
Precious metals	−0.01476	−0.0250	−0.02345	0.00020	0.00002|0.00020	0.00003	0.00002	0.00003	0.00000	0.00000	−0.00011
Industrial metals	−0.01483	−0.0260	−0.02395	0.00019	0.00019	0.00003|0.00018	0.00001	0.00004	0.00001	0.00001	−0.00014
Energy	−0.05301	−0.0717	0.03395	−0.00007	−0.00007	−0.00006	0.0015|0.0041	0.00014	−0.00005	0.00048	−0.00023
10Y Treasury Bonds	−0.04994	0.0077	−0.00821	0.00001	0.00001	0.00001	0.00157	0.0014|0.0040	0.00014	0.00029	−0.00026
Aaa Corporate Bonds	−0.05562	0.1264	−0.06096	0.00015	0.00016	0.00014	0.00036	0.00259	0.0016|0.0051	0.00051	0.00045
Baa Corporate Bonds	0.01720	−0.0562	0.03608	−0.00028	−0.00025	−0.00028	0.00300	0.00278	0.00131	0.0028|0.0070	0.00067
VW Equity CRSP	−0.03731	0.0440	−0.00052	−0.00057	−0.00060	−0.00055	0.00192	0.00332	0.00160	0.00406	0.0076|0.0089

Note: Significant conditional mean coefficients are boldfaced.

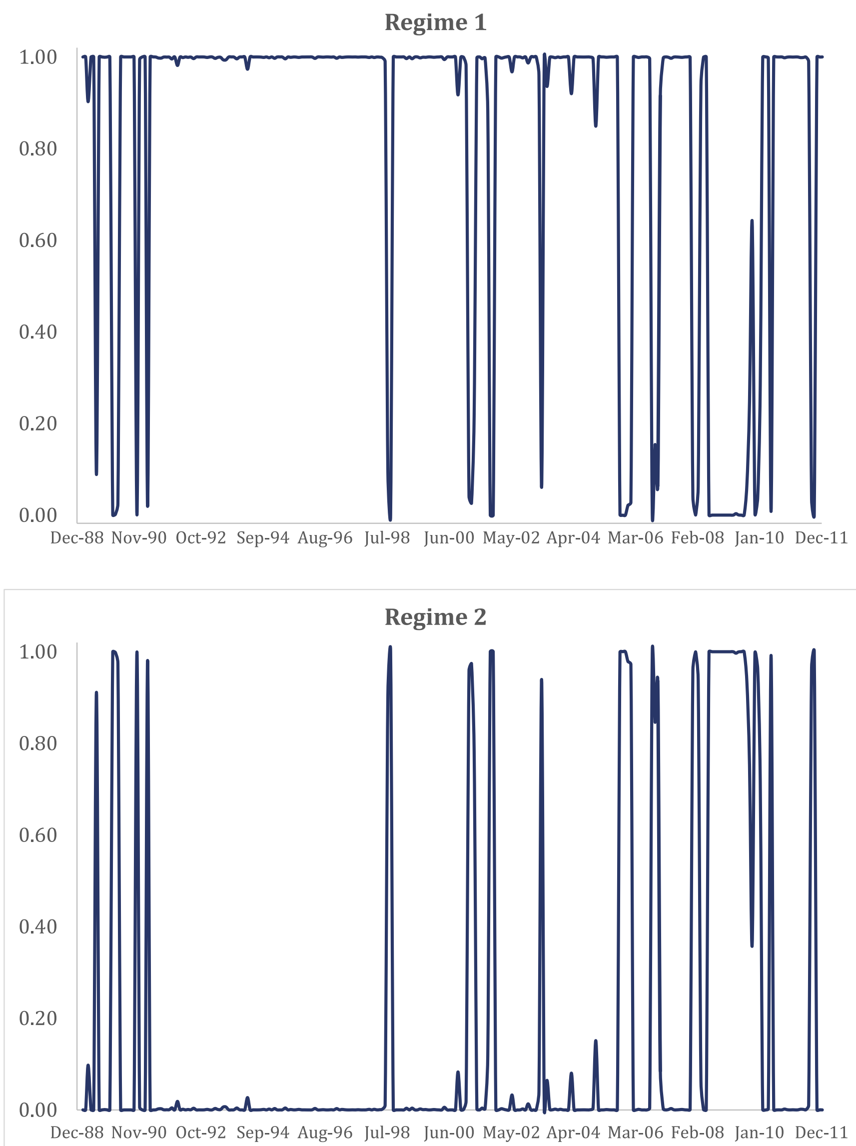

Fig. 2. Smoothed state probabilities from three-factor block VAR(1), two regimes.

space, we have omitted a large number of coefficients from the two 11 × 11 VAR(1), regime-specific matrices to focus only on key coefficients concerning the factors and the dynamics of the SDF. In general, most reported conditional mean coefficients are statistically significant and vary considerably across the two regimes estimated. This also concerns the VAR structure. For instance, the first, business-cycle related factor 1 is much more serially correlated in the second state (coefficient of 0.95) than in the first state (0.52).

Proceeding first with an interpretation of the two regimes, Fig. 2 shows that both Markov states are rather persistent, with the second able to single out crisis periods, such as the Summer 1998 Asian flu crisis, the 2001–2002 Enron and WorldCom scandals, the 2008–2009 great financial crisis, a few short bouts of European sovereign jitters, for instance in the Fall of 2011. These impressions are confirmed by Table 5, where both regimes are characterized by “stayer” probabilities on the main diagonal of the estimated transition matrix well in excess of 0.5 (their implicit durations are in fact 15 and 3 months). Even though it less persistent, regime 2 is no way irrelevant, as its ergodic, long-run probability is 17%, which seems a sensible assessment of the frequency of crisis periods in any long sample of US data. A look at the estimated, regime-specific covariance matrices confirms that estimated variances in the second state are considerably higher (up to 7 times), which is consistent with a turbulent regime characterization.

Finally, with reference to the estimated of the coefficients with which the SDF loads on the three macroeconomic factors, in Table 5, we note that although the sign of only one coefficient (on the third factor) switches across regimes, the absolute value of the coefficients is rather different across states. In particular, the overall level—as measured by the estimated intercept—of the SDF is higher under the crisis regime, which is sensible because during a crisis we do expect wealth to be lower, its certainty equivalent to be perceived as lower (due to increased risk), and hence marginal utility of future wealth to be higher. Moreover, the log-SDF loading (hence, the semi-elasticity of the stochastic discount) is roughly three times more sensitive to business cycle conditions under the crisis regime, which is also to be expected.

4.3. Matching sample moments with estimated SDFs

In this section, we assess to what extent a selection of competing SDFs are able to match the empirical mean, standard deviation, skewness, excess kurtosis, and pair-wise cross asset cor-

Table 6
Observed and implied moments – comparing different models.
The table reports the observed and the model-implied means, standard deviation, skewness, and excess kurtosis of different asset class returns. Implied values that have been boldfaced and underlined indicate that the model-implied moment falls within the 90% confidence interval formed around the corresponding sample moment.

		Lower bound 5%	Sample observed	Upper bound 95%	Single-regime, 3 Factors, Full VAR(1)	Two-regime, 3 Factors, Full VAR(1)	Four-regime, 3 Factors, Block VAR(1)
		Mean returns					
Macro factors	Factor 1	−0.4355	0.0000	0.4355	−7.6183	−1.3904	0.5493
	Factor 2	−0.3059	0.0000	0.3059	1.3638	**0.0825**	−**0.0874**
	Factor 3	−0.1040	0.0000	0.1040	2.6199	1.1234	−0.2028
Stocks and bonds	Aaa Corp.	0.0021	0.0023	0.0025	0.0051	0.0083	0.0046
	Baa Corp.	0.0031	0.0033	0.0035	0.0051	0.0083	0.0046
	10Y Treasuries	0.0040	0.0042	0.0044	0.0051	0.0083	0.0046
	VW CRSP Equity	0.0009	0.0054	0.0098	**0.0044**	**0.0095**	**0.0058**
Commodities	Agriculture and livestock	−0.0045	−0.0002	0.0040	0.0041	0.0091	0.0053
	Precious metals	−0.0038	0.0030	0.0098	**0.0039**	**0.0092**	**0.0055**
	Industrial metals	−0.0052	0.0008	0.0068	**0.0036**	0.0102	**0.0064**
	Energy	−0.0019	0.0071	0.0161	**0.0011**	**0.0121**	**0.0083**
		Standard deviation of returns					
Macro factors	Factor 1	4.0766	4.3852	4.6937	4.9518	**4.5825**	3.6785
	Factor 2	2.8634	3.0801	3.2968	**3.0230**	5.8184	**3.1802**
	Factor 3	0.9735	1.0471	1.1208	2.7631	5.9967	2.5256
Stocka and bonds	Aaa Corp.	0.0018	0.0019	0.0021	0.0028	0.0093	0.0064
	Baa Corp.	0.0018	0.0019	0.0021	0.0027	0.0092	0.0062
	10Y Treasuries	0.0018	0.0019	0.0021	0.0033	0.0095	0.0066
	VW CRSP Equity	0.0416	0.0447	0.0479	**0.0447**	**0.0447**	**0.0443**
Commodities	Agriculture and livestock	0.0398	0.0428	0.0458	**0.0428**	**0.0434**	**0.0429**
	Precious metals	0.0636	0.0684	0.0732	0.0467	0.0479	0.0471
	Industrial metals	0.0563	0.0606	0.0648	**0.0602**	**0.0601**	**0.0598**
	Energy	0.0841	0.0904	0.0968	**0.0903**	**0.0892**	**0.0893**
		Skewness					
Macro factors	Factor 1	−1.9452	−1.7033	−1.4613	0.0770	−0.1211	−0.2312
	Factor 2	−0.9915	−0.7495	−0.5076	−0.0023	−1.5819	−0.4689
	Factor 3	1.6964	1.9383	2.1803	−0.0086	**2.1427**	−0.1774
Stocka and bonds	Aaa Corp.	−0.2404	0.0016	0.2436	**0.0067**	1.5304	0.5880
	Baa Corp.	−0.2425	−0.0005	0.2414	**0.0092**	1.6212	0.2759
	10Y Treasuries	−0.2315	0.0105	0.2524	**0.0048**	1.3824	0.8507
	VW CRSP Equity	−0.8619	−0.6199	−0.3779	−0.0065	0.1041	−0.0387
Commodities	Agriculture and livestock	−0.3300	−0.0880	0.1539	**0.0070**	**0.1351**	**0.0329**
	Precious metals	0.0926	0.3346	0.5765	0.0072	**0.1350**	0.0438
	Industrial metals	−0.3465	−0.1045	0.1375	−**0.0078**	**0.0727**	−**0.0280**
	Energy	0.1453	0.3872	0.6292	0.0011	0.0111	−0.0179
		Excess Kurtosis					
Macro factors	Factor 1	4.4274	4.9593	5.4912	1.7618	5.0973	3.3287
	Factor 2	3.8200	4.3519	4.8838	1.5328	15.7805	**4.5131**
	Factor 3	46.2694	46.8013	47.3331	18.5047	9.3663	28.5483
Stocka and bonds	Aaa Corp.	−1.2751	−0.7432	−0.2113	−**0.5312**	3.1026	−0.0128
	Baa Corp.	−1.2853	−0.7534	−0.2215	−**0.5324**	3.4447	−0.1128
	10Y Treasuries	−1.2608	−0.7289	−0.1971	−**0.5226**	2.5976	0.1031
	VW CRSP Equity	0.5307	1.0625	1.5944	0.2034	**1.1319**	**0.9294**
Commodities	Agriculture and livestock	1.0796	1.6115	2.1433	0.4289	**1.5492**	**1.6788**
	Precious metals	2.4155	2.9474	3.4793	0.9636	**3.4047**	**2.5107**
	Industrial metals	0.9689	1.5008	2.0327	0.3857	**1.3416**	**1.6031**
	Energy	1.0783	1.6101	2.1420	0.4358	**1.1673**	**1.2445**

relations, with special emphasis on commodities. Specifically, we check whether each model-implied moment falls within the 90% confidence interval built around the sample estimate (i.e., whether it is lower than the 95% upper band and higher than the 5% lower band computed under standard assumptions). If this happens, we consider this sample moment to have been matched by a given SDF. The moments implied by an SDF are computed by Monte Carlo simulation, using 20,000 trials: as such they can be taken to represent population moments implied by any given SDF. In particular, we compare three of the competing model specifications discussed in Sections 4.1 and 4.2 and covered by Tables 2–5, namely, the four-state three-factor block MSVAR, which is the winning model according to the information criteria, the two-state, three-factor full MSVAR, the top performer in terms of pricing accuracy, and a benchmark single-state three-factor full VAR model.

Table 6 reports sample and model-implied means, standard deviations, skewness, and excess kurtosis coefficients of the returns on different asset classes. Implied values that fall within the 90% confidence interval have been boldfaced. As far as mean returns are concerned, no model clearly outperforms the others. While all SDFs match the mean return of the equity index, the four-regime model does particularly well at matching the means of the commodity series. Moreover, all models fail to reproduce the sample mean of bond returns.

Similar considerations apply to standard deviations, for which all the three models do match the volatility of all assets apart from bonds and precious metals. The situation is more heterogeneous for what concerns higher-order moments. The block MSVAR framework is quite weak at matching skewness, while both the two-state full MSVAR and the single-state SDFs can replicate four and five values out of eleven, respectively. Noticeably, the single-state model can generate the modest (close to zero) positive skewness of bonds, while the two-state model generates too much positive skewness. However, the two-state model outperforms its single-state counterpart as far as commodity series are concerned. Finally, as one would expect, both switching models are considerably better than the single-state SDF at yielding excess kurtosis, especially

Table 7
Observed and implied correlations – comparing different models.
The table reports the observed and the model-implied pairwise correlations for different asset class returns. Implied values that have been boldfaced and underlined indicate that the model-implied correlation falls within the 90% confidence interval formed around the corresponding sample moment.

	Lower bound 5%	Sample observed	Upper bound 95%	Single-regime, 3 factors, full VAR(1)	Two-regime, 3 factors, full VAR(1)	Four-regime, 3 factors, block VAR(1)
	Agricultural commodities and livestock					
Agric. and livestock – Aaa corp.	0.0233	0.1223	0.2212	**0.1252**	**0.1946**	**0.0670**
Agric. and livestock – Baa corp.	0.0370	0.1357	0.2345	**0.1168**	**0.1914**	**0.0600**
Agric. and livestock – Treasuries	0.0080	0.1071	0.2062	**0.1240**	**0.1971**	**0.0731**
Agric. and livestock – VW CRSP equity	0.1236	0.2209	0.3181	**0.2170**	**0.2231**	**0.2095**
Agric. and livestock – Precious metals	0.1969	0.2923	0.3876	**0.2905**	**0.3137**	**0.2808**
Agric. and livestock – Industrial metals	0.0605	0.1589	0.2573	0.3018	0.3024	0.2984
Agric. and livestock – Energy	0.2108	0.3057	0.4006	0.1476	0.1273	0.1319
	Precious metals					
Precious metals – Aaa corp.	0.0943	0.1921	0.2899	**0.1211**	**0.2397**	0.0884
Precious metals – Baa corp.	0.0984	0.1961	0.2938	**0.1333**	**0.2426**	0.0915
Precious metals – Treasuries	0.1001	0.1978	0.2955	**0.1085**	**0.2373**	0.0877
Precious metals – VW CRSP equity	−0.0735	0.0261	0.1258	**0.0254**	**0.0477**	**0.0203**
Precious metals – Industrial metals	0.1136	0.2111	0.3085	**0.2518**	**0.2632**	**0.2538**
Precious metals – Energy	0.1610	0.2573	0.3536	**0.2058**	**0.1926**	**0.1951**
	Industrial metals					
Industrial metals – Aaa corp.	−0.1106	−0.0109	0.0887	**0.0010**	**0.0558**	**0.0165**
Industrial Metals – Baa corp.	−0.0905	0.0092	0.1089	**0.0192**	**0.0609**	**0.0231**
Industrial metals – Treasuries	−0.1268	−0.0272	0.0724	**0.0072**	**0.0571**	**0.0191**
Industrial metals – VW CRSP equity	−0.0094	0.0899	0.1891	0.3747	**0.1693**	0.3748
Industrial metals – Energy	0.1913	0.2868	0.3823	**0.2802**	**0.2615**	**0.2726**
	Energy					
Industrial metals – Aaa corp.	−0.0077	0.0916	0.1908	−0.0841	−**0.0681**	−0.0938
Industrial metals – Baa corp.	0.0268	0.1257	0.2246	−0.0500	−0.0773	−0.0806
Industrial metals – Treasuries	−0.0113	0.0880	0.1873	−0.1036	−**0.0098**	−**0.0106**
Industrial metals – VW CRSP equity	0.2882	0.3804	0.4726	0.0900	0.0647	0.0821
	Lower bound 5%	Sample observed	Upper bound 95%	Single-regime, 3 factors, full VAR(1)	Two-regime, 3 factors, full VAR(1)	Four-regime, 3 factors, block VAR(1)
	Aaa corporate bonds					
Aaa corp. – Baa corp.	0.9645	0.9828	1.0012	0.9427	**0.9946**	**0.9890**
Aaa corp. – Treasuries	0.9784	0.9914	1.0045	0.9563	**0.9940**	**0.9883**
Aaa corp. – VW CRSP equity	−0.1113	−0.0116	0.0880	**0.0099**	0.1025	**0.0157**
	Baa corporate bonds					
Aaa corp. – Treasuries	0.9627	0.9817	1.0007	0.8184	**0.9789**	0.9578
Aaa corp. – VW CRSP equity	−0.0825	0.0172	0.1168	**0.0453**	**0.1133**	**0.0292**
	10-Year treasury bonds					
Aaa corp. – VW CRSP equity	−0.1071	−0.0075	0.0922	**0.2170**	**0.0096**	**0.0091**

in the case of commodity returns. Indeed, while the single-state model matches the small and negative excess kurtosis of the bond, the two- and four-state models reproduce the positive excess kurtosis of all the commodity returns.

The three SDF models seem to be equally good at reproducing cross-asset pair-wise correlations. Indeed, the single-state model matches 18 out of 28 cross-asset pair-wise correlations, while the four-state and the two-state models 17 and 19, respectively. Three plots in Appendix E graphically display where the model-implied correlations stand with respect to sample ones, focusing on the pair-wise correlations of commodities with traditional asset classes and among themselves, which are the most interesting sample moments in our analysis. Interestingly, the two-state model clearly outperforms its single-state benchmark for what concerns the pair-wise correlations between commodities and the other asset classes, with a hit ratio of 88% (compared to the 69% of the single-state model). This is not surprising considering that the relationships of commodities with traditional asset classes has changed in the last decade with the so called “financialization” of commodities and only a Markov switching SDF can capture this structural shift.

In conclusion, the two-state, three-factor full MSVAR model outperforms a simple single-state benchmark when it comes to estimate the moments of commodities and their correlations with the other asset classes. This supports previous findings in the literature. For example, Lombardi and Ravazzolo (2016) find that a bivariate Bayesian dynamic conditional correlation model, which can account for time variation in the correlation patterns, produces statistically more accurate density forecasts for equity and commodity returns, and gives large economic gains in an asset allocation exercise, relative to a benchmark random walk model. In addition, the model previously selected by the information criteria, namely the four-state three-factor block MSVAR, is outperformed by the two-regime three-factor full MSVAR. This is consistent with the better pricing performance of the latter already discussed in Section 4.2.

4.4. Comparisons with a commodity factor-based benchmark

As a final robustness check, we have also estimated one SDF-version of a benchmark pricing factor model proposed by BGR (2014). They propose a linear factor model based on three well-known, commodity-specific factors: Average, Carry, and Momentum. The average factor is the excess return of a long position in all available commodity futures; the carry factor is the return on an equally-weighted strategy that buys the five commodities that are most backwardated and shorts the ones that are most in contango at each point in time; the momentum factor is the return on an equally-weighted portfolio that is long in the five commodities with the highest returns over the previous six months and short the ones with the lowest returns over the previous six months. In our paper, for obvious comparability goals, we work with a single-state, SDF-type implementation of BGR's model in the sense that Average, Carry, and Momentum are used as the factors on which the SDF depends on. This strategy of trying to explain the cross-section of commodity returns using variables that capture the general conditions of the commodity market has recently become popular in a strand of the literature (see, e.g., Daskalaki and Skiadopoulos, 2011, Daskalaki, Kostakis, Skiadopoulos, 2014; Szymanowska et al., 2012; Yang, 2013).[13]

Empirically, the BGR's SDF yields (unreported, but available upon request) maximum likelihood estimates that are weakly statistically significant and characterized by uniformly positive loadings, which is consistent with the marginal utility of future wealth increasing as Average, Carry, and Momentum increase. However, the loadings of the log-SDF on Momentum and especially Carry are not estimated with sufficient accuracy (their p-values are 0.13 and 0.71, respectively). Despite the low statistical significance of the estimated loadings, the pricing performance of BGR's model is not completely amiss: for instance, the corresponding chi-square test for zero pricing errors delivers a 0.168 p-value. However, such a performance is dominated by all HMM factor-based models, regardless of the number of regimes specified. Moreover, also BGR is characterized by a relatively high and statistically significant HJ distance (13.20) that leads to a rejection of the model. Given that this had also occurred in the case of HMM models, we therefore proceed to use White's (2000) "reality check" method to test the null hypothesis that $[(HJ^{BGR})^2 - (HJ^{2\ REG})^2] \leq 0$, in words that a BGR-inspired SDF implies a a smaller (maximum) pricing error vs. two-state three-factor full VAR(1) model that had previously emerged as our "champion" of pricing performance. We find a one-sided test statistic $T_{2\text{REG-BGR}} = 2336$ that strongly rejects the null hypothesis of an inferior HJ distance by the BGR's benchmark vs. the HMM two-state model, with a p-value of essentially zero.

5. Conclusions

In this paper we have investigated whether it is possible to find a SDF that jointly prices the cross-section of eight portfolios of stocks, Treasuries, corporate bonds and commodities and replicates the first four empirical moments of (especially, correlations among) these assets. Importantly, besides being based on three through ten principal components estimated from a large set of macroeconomic indicators, our SDFs are further extended to include latent regime shifts governed by an ergodic and irreducible Markov chain.

We find that regime-switching models clearly outperform single-state ones both in term of statistical and pricing accuracy. However, while a four-state model is selected by standard information criteria, a two-state three-factor full VAR(1) model outperforms all others as far as the pricing accuracy is concerned. Finally, we notice that, although this model gives rather similar results to its single-state counterpart in terms of its ability to match sample moments, its Markov switching version outperforms the single-state model when the intra-commodity correlation are analyzed. This is not surprising because a literature has noticed that the relationships of commodities with traditional assets classes has changed over the last 15–20 years as a result of a so called "financialization" process, so that it is sensible that only a Markov switching SDF may be flexible enough to capture this structural shift.

Given our result that macro factor-based HMM models outperform an SDF implementation of BGR's framework, only based on commodity-specific factors, it would be interesting to further explore the econometric nature of the segmentation of commodities and to assess whether there exist some joint choice of factors (both macroeconomic and commodity-related) able to price the cross-section of returns of all assets. However, we will leave this question for future research.

[13] Of course, resorting to such a commodity factor-based SDF poses some logical issues even though one the key points made by Bakshi et al. is that together, their three factors appear to also forecast economic growth, the returns of government bonds and of equities. Note that Bakshi et al.'s paper is framed in terms of linear factor/regression representations and not in terms of a structural SDF estimation exercise. However, as explained in Cochrane (2008), there is a clear one-to-one mapping between linear factor representations and log-linear SDFs. Tabulated statistics for the time series of the three factors are available upon request. Note that our analysis spans only half of the overall 1970–2011 sample employed by BGR in their paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2017.07.045.

References

Alizadeh, A. H., Nomikos, N. K., & Pouliasis, P. K. (2008). A Markov regime switching approach for hedging energy commodities. *Journal of Banking and Finance, 32*, 1970–1983.
Asness, C. A., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum 'everywhere'. *Journal of Finance, 68*, 929–985.
Bae, G., Kim, W., & Mulvey, J. M. (2014). Dynamic asset allocation for varied financial markets under regime switching framework. *European Journal of Operational Research, 234*, 450–458.
Baker, S. R., Bloom, N., & Davis, S. J. (2012). "Has economic policy uncertainty hampered the recovery?" Chicago Booth Research Paper No. 12-06.
Bakshi, G., Gao, X., & Rossi, A. (2014). *A better specified model to explain the cross–section and time series of commodity returns*. University of Maryland Working Paper.
Basak, S., & Pavlova, A. (2013). Asset prices and institutional investors. *American Economic Review, 103*, 1728–1758.
Buckley, I., Saunders, D., & Seco, L. (2008). Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research, 185*, 1434–1461.
Büyükşahin, B., & Robe, M. A. (2014). Speculators, commodities and cross-market linkages. *Journal of International Money and Finance, 42*, 37–80.
Chen, X., & Ludvigson, S. C. (2009). Land of addicts? An empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics, 24*, 1057–1093.
Cochrane, J. H. (2008). *Asset pricing* (Revised ed.). Princeton, N.J.: Princeton University Press.
Daskalaki, C., & Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking and Finance, 36*, 2260–2273.
Daskalaki, C., Kostakis, A., & Skiadopoulos, G. (2014). Are there common factors in commodity futures returns? *Journal of Banking and Finance, 40*, 346–363.
De Roon, F. A., & Szymanowska, M. (2010). *The cross section of commodity futures returns*. Erasmus University Working Paper.
Dias, J., Vermunt, J., & Ramos, S. (2015). Clustering financial time series: New insights from an extended hidden Markov model. *European Journal of Operational Research, 243*, 852–864.
Dudley, W., & Hatzius, J. (2000). "The Goldman Sachs financial conditions index: The right tool for a new monetary policy regime." Goldman Sachs Global Economics Paper No. 44.
Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal, 62*, 69–97.
Gorton, G., & Rouwenhorst, K. G. (2006). "Facts and fantasies about commodity futures. *Financial Analyst Journal, 62*, 47–68.
Guidolin, M., & Timmermann, A. (2006). An econometric model of nonlinear dynamics in the joint distribution of stocks and bonds returns. *Journal of Applied Econometrics, 21*, 1–22.
Guidolin, M. (2011). Markov switching models in empirical finance. *Advances in econometrics, 27*, 1–86.
Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.
Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054.
Hansen, L. P., & Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance, 52*, 557–590.
Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance, 51*, 3–53.
Jensen, G. R., Mercer, J. M., & Johnson, R. R. (2002). Tactical asset allocation and commodity futures. *Journal of Portfolio Management*, 100–111.
Koijen, R. S. J., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2013). "Carry". Fama-Miller Center Working Paper.
Lee, H., & Yoder, J. (2007). A bivariate Markov regime switching GARCH approach to estimate time varying minimum variance hedge ratios. *Applied Economics, 39*, 1253–1265.

Lombardi, M., & Ravazzolo, F. (2016). On the correlation between commodity and equity returns: Implications for portfolio allocations. *Journal of Commodity Markets, 2*, 45–57.
Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies, 22*, 5027–5067.
Marroquìn Martinez, N., & Moreno, M. (2013). Optimizing bounds on security prices in incomplete markets. Does stochastic volatility specification matter? *European Journal of Operational Research, 225*, 429–442.
Pastor, L., & Stambaugh, R. L. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy, 111*, 642–685.
Shiller, R. J. (1979). The volatility of long-term interest rates and expectations models of the term structure. *Journal of Political Economy, 87*, 1190–1219.
Szymanowska, M., de Roon, F. A., Nijman, T. E., & Van den Goorbergh, R. (2012). An anatomy of commodity futures risk premia. *Journal of Finance, 69*, 453–482.
Tang, K., & Xiong, W. (2012). Index investments and the financialization of commodities. *Financial Analyst Journal, 68*, 54–74.
White, H. (2000). A reality check for data snooping. *Econometrica, 68*, 1097–1126.
Yang, F. (2013). Investment shocks and the commodity basis spread. *Journal of Financial Economics, 110*, 164–184.