WORKING PAPER SERIES

WORKING PAPER
N. 140
MAY 2020

Distilling Large Information Sets to Forecast Commodity Returns: Automatic Variable Selection or Hidden Markov Models?

By Massimo Guidolin and Manuela Pedio

This Paper can be downloaded without charge from The Social Science Research Network Electronic Paper Collection:
http://ssrn.com/abstract=3606933

Università Bocconi
BAFFI CAREFIN
Centre for Applied Research
on International Markets, Banking,
Finance and Regulation

Distilling Large Information Sets to Forecast Commodity Returns: Automatic Variable Selection or Hidden Markov Models?♣

Massimo Guidolin
Bocconi University, Baffi-CAREFIN Centre, and IGIER

Manuela Pedio
Bicocca University, Milan, and Baffi-CAREFIN Centre

April 2020

Abstract

We investigate the out-of-sample, recursive predictive accuracy for (fully hedged) commodity future returns of two sets of forecasting models, i.e., hidden Markov chain models in which the coefficients of predictive regressions follow a regime switching process and stepwise variable selection algorithms in which the coefficients of predictors not selected are set to zero. We perform the analysis under four alternative loss functions, i.e., squared and the absolute value, and the realized, portfolio Sharpe ratio and MV utility when the portfolio is built upon optimal weights computed solving a standard MV portfolio problem. We find that neither HMM or stepwise regressions manage to systematically (or even just frequently) outperform a plain vanilla AR benchmark according to RMSFE or MAFE statistical loss functions. However, in particular stepwise variable selection methods create economic value in out-of-sample mean-variance portfolio tests. Because we impose transaction costs not only ex post but also ex ante, so that an investor uses the forecasts of a model only when they increase expected utility, the economic value improvement is maximum when transaction costs are taken into account.

Key words: Backward and forward stepwise regressions; hidden Markov models, out-of-sample forecasting; commodity futures returns; mean-variance portfolios.

♣ We would like to thank Antonio Mauro for excellent research assistance.

1. INTRODUCTION

Commodity prices are known to be affected by a wide variety of systematic influences, which range from the standard, CAPM-style market portfolio, to macroeconomic factors representing prices and real economic activity to other variables that reflect both individual commodity and aggregate commodity market conditions. In fact, it is unclear which of these factors may outweigh any others in the balance of effects on commodity returns, especially in a forecasting perspective (see, e.g., Giampietro et al., 2017; Guidolin and Pedio, 2018). Yet, it seems crucial to try and understand whether commodities may represent a separate (better, segmented), *special* asset class or they may be priced by the same factors that price all asset classes, chiefly worldwide stocks and bonds. The underlying conjecture is that if the fundamental pricing measure (often known as the stochastic discount factor, see Cochrane, 2009) need to correctly predict the cross section of commodity prices just depends (or itself predicted by) systematic, aggregate influences in a flexible framework, this would be evidence of the integration of commodities with the rest of the asset classes, while in case this measure were to be predicted by additional variables only typical of the commodity space (or typical of each single commodity), then we would obtain prima facie evidence of the segmentation of commodities (see also the discussion in Giampietro et al., 2017). This question has become increasingly important as major institutional investors have entered the space of commodity investing cumulating positions in commodities in addition to traditional equities and bonds (see Hamilton e Wu, 2015; Tang and Xiong, 2012) and it naturally leads researchers to investigate the predictive performance of large numbers of predictors.

In this paper we tackle exactly these questions and systematically study the out-of-sample, recursive predictive accuracy for (fully hedged) commodity future returns of two sets of forecasting models, i.e., hidden Markov chain (henceforth HMM) models in which the coefficients of predictive regressions follow a regime switching process and stepwise variable selection algorithms in which the coefficients of predictors not selected are set to zero. In essence, in a HMM model the predictive coefficients are allowed to take different values in different regimes and the transition between them is governed by a simple Markov chain process in which there is potentially (this is estimable) memory for the most recent values taken by the coefficients. On the contrary, stepwise regression is an *automatic* variable selection procedure, which chooses from a set of candidates the explanatory variables that are, jointly, the most relevant. In particular, unidirectional forward (backward) methods start with no (all the) predictors in the model and progressively testing the inclusion (exclusion) of each variable under a chosen model-fit criterion, adding (excluding) the variable (if any) whose inclusion gives the most (the least) statistically significant improvement of the fit, and repeating this process until none of the remaining variables improves (fails to improve) the model to a statistically significant extent. Therefore while HMM does not impose that predictors always forecast in the same way but by allowing flexibility expands the size of the models by increasing the number of coefficients, stepwise methods prevent over-parameterization and unhealthy model complexity by preventing predictors to be including or purging a given initial model by the least relevant predictors, thus imposing boundaries on over-fitting and "model noise". However, both approaches—and this is what makes it interesting to formally compare them—share a automatic nature, in the sense that in HMM the time variation in predictive coefficients results from endogenous estimation of the Markov chain properties, given the number of pre-specified hidden regimes; under stepwise algorithm, the rules of inclusion/exclusion of predictors are assigned, given a pre-determined criterion (such as information criteria, Wald test statistics, coefficient p-values, R-squares, etc.).

Our research design is rather simple: in a typical, recursive backtesting exercises that spans a January 2004 – May 2018 sample, we systematically compare the out-of-sample (henceforth, OOS) forecasting performance of a large set of HMM vs. stepwise predictive regressions. Both forecasting approaches are combined with dimension reduction techniques used to summarize the information content of as many as 132 macroeconomic variables (that include prices, employment, output, interest rates, monetary and credit aggregates, housing statistics, sales, international trade and exchange rate statistics, etc.) in a much smaller set of (eight) representative principal component factors, following Ludvigson and Ng (2009) and Stock and Watson (2006). These macro factors are supplemented with a standard set of commodity factors (hedging pressure, the basis reflecting front-end backwardation vs. contango, net trading, and momentum) implemented in two ways, at the aggregated level (as in Guidolin and Pedio, 2018) and at the individual commodity level, which instead appears to be new to the literature. A classical, parsimonious AR(1) benchmark allows us to measure the relative pay-off to model and algorithm complexity. All models are tested both in their own, off-the-shelf representation as well as considering their intersection with simple algorithms used to classify periods as low- vs- high volatility ones (some would argue, risk-on vs. risk-off periods) on the basis of VIX (option-implied future expected variance) data, to provide an additional ground of time variation useful to test an additional dimension of model instability.

Because we want our empirical findings to be relevant to a general public of forecasters and asset managers, we test the relative OOS performance of HMM and of alternative variable selection algorithms with reference to a long, January 1989 – May 2018 sample of 14 commodity futures returns series (cocoa, coffee, corn, cotton, gold, orange juice, light crude oil, live cattle, platinum, sugar, silver, soy, timber, and wheat). In fact, while Jan. 1989 – December 2003 is just used for algorithmic training purposes, Jan. 2003- May 2018 provides our forecast testing sample. In fact, we perform our recursive tests using four alternative loss functions, two of them directly drawn from the toolkit of portfolio managers and of financial economics. The first two loss functions are typical of the statistical literature and consist of the squared and the absolute value losses, which imply that the root-mean-squared forecast error (RMSFE) and the mean absolute forecast error (MAFE) are used as summaries of OOS predictive performance, respectively. The remaining two loss functions are the realized portfolio Sharpe ratio and MV utility, when the portfolio is built upon optimal weights computed solving a standard MV portfolio problem, as in DeMiguel et al. (2007) and Rapach and Zhou (2013). Although the two loss functions are traditional and largely understood, they carry the nature of value functions of the underlying optimization problem and as such they are not only of a more applied, practical nature, but also more complex.

We report three key findings. First, neither HMM nor stepwise regressions manage to systematically (or even just frequently) outperform a plain vanilla autoregressive benchmark according to RMSFE or MAFE statistical loss functions. In this—after one realizes that at the cost of just one additional parameter, an AR(1) model nests a constant mean return model—we fall very close to typical literature (see the summary in Rapach and Zhou, 2013) on stock return predictability that has been reporting (with rare but notable exceptions, see e.g., Dangl and Halling, 2012; Henkel et al., 2011) that is rather hard for complex statistical models to outperform even a simple historical mean return predictor. Second, stepwise regressions and especially HMM show some evidence of predictive power relative to simple benchmarks only in low-volatility regimes. At first blush, this finding appears to be unique to the commodity space since in the case of equities the general result (see Henkel et al., 2011) appears instead to be that asset returns become more predictable during times of market distress. However, deeper scrutiny reveals that historically, the first stage of phases of stock and bond market distress have represented low-volatility regimes for a number of commodities, especially precious metals, so the mapping is unclear. In any event, when we average the predictive performances over low- vs. high-volatility regimes, we find that it remains hard for both HMM and stepwise methods to forecast well, in relative terms. Interestingly, these two results do not seem to strongly depend on whether commodity factors are included in the analysis or not, and there is a generalized, mild evidence that simpler models including only macroeconomic effects may be “rich enough” to reveal most of the forecastability in the data.

Third, and this is particularly evident in the case of models recursively built going stepwise “simple to general” to avoid over-parameterizations, complex predictive models create economic value in OOS MV portfolio tests. Interestingly, because we impose transaction costs not only ex post but also ex ante, so that an investor will use the forecasts of a model only when they increase expected utility, contrary to naïve priors, the increase in realized Sharpe ratios caused by and the fees an investor would be ready to pay to access to HMM and stepwise regression models is maximum under plausibly sized, proportional transaction costs. These results turn out to be not depend on the specific coefficient of risk aversion adopted and on the fact that the asset menu would also include stocks and bonds, in addition to commodities. Therefore we find evidence in the commodity space of a substantial misalignment between the typical, statistical loss functions used in applied forecasting work—under which either HMM or stepwise algorithms are of limited, at best regime- and commodity-specific value—and the most used loss function in applied portfolio management work, mean-variance realized utility, consistently with earlier evidence in the equity space provided by Cenesizoglu and Timmermann (2012) and Dal Pra et al. (2018). On the one hand, this illustrates that sophisticated quantitative techniques are likely to deserve attention in the practice of commodity return forecasting and in the associated academic literature. On the other hand, this stresses that the true value of such complex techniques may really taken to the surface only by sufficiently well-articulated loss functions. Moreover, under economic loss functions, there is more robust evidence that a portfolio manager will need to have a tight grip on the predictive contents of commodity factors to optimize portfolios expanded to include commodities, consistently with earlier findings by Guidolin and Pedio (2018).

Our paper contributes to at least two distinct literatures. First, there is a traditional literature on forecasting commodity returns, using either classical prediction models widespread in finance (see, e.g., Bodie and Rosansky, 1980; Breeden, 1980) or models mostly based on commodity-specific predictors, such as the hedging pressure, the basis, and net trading (see e.g., Acharya et al., 2013; Bakshi et al., 2017; Bessembinder and Chan, 1992; de Roon et al., 2000; de Roon and Szymanowska, 2010; Gorton et al., 2008; Gospodinov and Ng, 2013; Hirshleifer, 1988; Khan et al., 2008; Yang, 2013). More recently, a few attempts have been made to relate commodity returns to macroeconomic factors that are well known to explain the cross section of bond and stock returns, see, e.g., Ahmed e Tsvetanov (2016), Gargano and Timmermann (2014) and Giampietro et al. (2018). Although this does not represent our main focus, we contribute to this literature by implicitly testing whether and how commodity-specific vs. macroeconomic factors may be needed relative to simple benchmarks, thus extending previous results in Guidolin and Pedio (2018) to HMM models and to longer time series.

More interestingly, there is a growing literature that has investigated the application of novel statistical modelling techniques to commodity data. For instance, Date et al. (2013) study commodity return forecasting under HMMs and Panella et al. (2012) have studied the application of neural networks. We are not aware of any direct comparison between variable selection and non-linear, time-varying coefficients models, even though in the presence of a multiplicity of predictors this seems important.

The rest of this paper is organized as follows. Section 2 describes the methodologies used in the paper and in particular the three statistical building blocks of our efforts, i.e., automatic variable selection algorithms, HMM specification and estimation, and dimensionality reduction techniques through principal component analysis. In this section, we also formally describe our backtesting framework. Section 3 describes that data, the macroeconomic factors, and the way we have built from microeconomic, commodity-specific information provided by the US Federal Trade Commission our commodity factors.[1] Section 4 reports detailed results of realized predictive performances under statistical loss functions. Section 5 repeats the backtesting exercise in section 4 under economic loss functions and discusses the reasons for the strikingly heterogeneous findings. Section 6 concludes.

2. Methodology

2.1. Automatic Variable Selection Methods and Stepwise Regressions

Consider a typical time t supervised learning problem in which inputs ($\boldsymbol{Z}_T$) are used to learn (forecast) an outcome at time $T + H$ (R_{t+H}), which in our case corresponds to commodity j futures returns and for which historical data on $(R_1, ..., R_T)'$ and other predictors are available. Denote by $\boldsymbol{W}_t = (w_{1t}, ..., w_{Mt})'$ a set of M “must have” (we shall call them “always included”) predictors and by $\boldsymbol{X}_t = (x_{1t}, ..., x_{Nt})'$ a set of N potential predictors, which are lagged, predetermined regressors chosen with the forecast horizon H in mind. Let $\boldsymbol{Z}_t = (\boldsymbol{W}'_t \boldsymbol{X}'_t)'$. A predictive regression that includes all available predictors is

$$R_t^j = \boldsymbol{W'}_t\boldsymbol{\alpha}^j + \boldsymbol{X'}_t\boldsymbol{\beta}^j + \epsilon_t^j, \qquad (1)$$

where ϵ_t^j is a white noise disturbance with variance $(\sigma^j)^2$. The best linear unbiased H-period forecast, given information up to time T, is the linear projection

$$R_{T+H|t}^j = \boldsymbol{W'}_{T+H|t}\boldsymbol{\alpha}^j + \boldsymbol{X'}_{T+H|t}\boldsymbol{\beta}^j, \qquad (2)$$

which is however infeasible because $\boldsymbol{\delta}^j = ((\boldsymbol{\alpha}^j)' \ (\boldsymbol{\beta}^j)')'$ is unknown. Assuming that the time series $\boldsymbol{Z}_t$ does not contain any linearly dependent predictor, $\boldsymbol{\delta}^j$ can be replaced by its least squares estimates so that the resulting forecast is $\hat{R}_{T+H|t}^j = \boldsymbol{W'}_{T+H|t}\hat{\boldsymbol{\alpha}}^j + \boldsymbol{X'}_{T+H|t}\hat{\boldsymbol{\beta}}^j$ (assuming the predictor vectors $\boldsymbol{W}_{T+H|t}$ aand $\boldsymbol{X}_{T+H|t}$ are available, for instance because pre-determined). Although the least squares estimate

[1] As typical of the literature, while commodity futures returns cannot be assigned to a country (although most negotiations occur in venues that are wrapped within the operations of US legal entities), our macroeconomic and commodity-specific data are of US origins.

of $\boldsymbol{\delta}^j$ is root-T consistent for $\boldsymbol{\delta}^j$, the mean square forecast error is increasing in $dim(\boldsymbol{\beta}^j)$ for given $dim(\boldsymbol{\alpha}^j)$, and not every potentially important predictor is actually relevant: retaining the weak predictors can introduce unwarranted sampling variability to the prediction.

Following Ng (2013), let $\mathcal{A}$ be an index set containing the positions of the variables deemed empirically relevant, so that $X_{\mathcal{A}}$ is the active set of predictors. Let $\widehat{\boldsymbol{\beta}}^j_{\mathrm{A}}$ be a Nx1 vector of estimates whose m-th element is zero if the corresponding regressor's index is not in $\mathcal{A}$, and equal to the least square estimates otherwise. At this point, two possible paths can be followed. In the first case, only a small subset of $\boldsymbol{X}_t$ with "significant" predictive power is used (when the number of non-zero entries in the population coefficient vector $\boldsymbol{\beta}^j$ is much smaller than the dimension of $\boldsymbol{\beta}^j$, we say that the model is *sparse*); in the second case, we use information in a large number of predictors, however small the contribution of each series is in predicting the commodity returns, and the resulting model is *dense*.

Stepwise regression is a classical data mining method of sparse specification of linear models in which the choice of predictive variables is carried out by an automatic procedure. In each step, a variable is considered for addition to or subtraction from the set of explanatory variables derived from an earlier iteration, on the basis of some pre-specified criterion. This may take the form of a sequence of *F*- or *t*-tests, but other criteria are possible, such as the adjusted R^2, information criteria (such as Akaike's or the Bayesian criterion), Mallows's Cp, etc. In inferential applications—which, we stress, do *not* represent in any way a goal of our empirical application—the widespread (mal-)practice of fitting the final selected model followed by reporting the estimates and confidence intervals without adjusting them to take the model pre-test process into account has led to calls to stop using stepwise algorithms altogether (see, e.g., Butler, 1984; Grechanovsky and Pinsker, 1995; Smith 2018), or at least to appropriately adjust the outputs related to inference and hypothesis testing (e.g., the standard errors, see e.g., Chatfield, 1995).[2] One of the main issues with stepwise regression is that it searches a large space of possible models. Hence it is prone to overfitting the data, i.e., it will often fit much better in sample than it does on new, out-of-sample data (see Smith, 2018). In our application, we use instead

[2] The very tests on which the stepwise algorithms are based are themselves invalid, since they are based on the same data, see Rencher and Pun (1980). For instance, Wilkinson and Dallal (1981) computed multiple correlation coefficients obtained by stepwise methods by simulation and showed that a final regression obtained by forward selection and reported by an F-test to be significant at 0.1%, was in fact only significant at 5%. Moreover, when estimating the degrees of freedom, the number of the candidate independent variables from the best fit selected may be smaller than the total number of final model variables, causing the fit to appear better than it is when adjusting the R^2 for the number of degrees of freedom. This is because one ought to consider how many degrees of freedom have been used up in the entire model, not just count the number of independent variables in the final fit.

stepwise methods to provide and test genuine, OOS forecasts and therefore we avoid altogether these classical problems.[3]

Under a classical stepwise regression design, there are three alternative approaches: *forward selection*, which involves starting with no variables in the model, testing the addition of each variable using a chosen fit criterion, adding the variable (if any) whose inclusion gives the most statistically significant improvement of the fit, and repeating this process until none improves the model to a statistically significant extent; *backward elimination*, which involves starting with all candidate variables, testing the deletion of each variable using a chosen model fit criterion, deleting the variable (if any) whose loss gives the most statistically insignificant deterioration of the model fit, and repeating this process until no further variables can be deleted without a statistically significant loss of fit; *bi-directional elimination*, a combination of the above, testing at each step for variables to be included and/or excluded.[4] In this paper, we pursue and compare both forward and backward designs in predictive applications, which avoids any issues with inference and hypothesis testing, while we do not experiment with bi-direction elimination, because its nature is complex and we know little (even less than with forward/backward) from the statistical literature about its properties, even though logically they should imply pros and cons that are intermediate vs. backward and forward stepwise methods.

As for the fit criteria commonly employed in stepwise regressions, a large class of information criteria (henceforth, IC) determine the size of a model, in terms of $dim(\boldsymbol{\beta}^j)$, by finding:

$$\hat{p}_{IC}^j = \underset{p=1,2,\dots,p_{max}}{\operatorname{argmin}} \left[\ln(\hat{\sigma}_p^j)^2 + p\frac{C_T}{T}\right], \tag{3}$$

where $p_{max} \leq dim(\boldsymbol{\beta}^j)$ is the maximum number of variables considered, $(\hat{\sigma}_p^j)^2$ is the sum of squared residuals scaled by T, and C_T/T is a term that penalizes model complexity in favor of parsimony. Different choices for C_T deliver different information criteria.[5]

[3] Stepwise algorithms are "greedy" because the locally optimal choices made at each stage may not be globally optimal. They perform what is known as "hard thresholding": a variable is either in or out of the predictor set. A resulting, undesirable feature is that a regressor set selected from N available predictors may depart from the one chosen when N is increased or decreased slightly, so that hard thresholding is sensitive to small changes in the data because of discreteness of the decision rule (the *bouncing beta* problem), see Fan and Li (2001).

[4] For instance, in Efroymson's algorithm, at each stage in the process, after a new variable is added, a test is applied to check whether any variables can be deleted without appreciably increasing the residual sum of squares. The procedure terminates when the measure is (locally) maximized, or when the available improvement falls below some threshold.

[5] All ARMAX model of order p must be estimated using $T - p_{max}$ observations even if $p < p_{max}$. This is necessary for the goodness of fit component of the information criteria to not depend on the complexity component of the criteria.

Different values for C_T have been proposed but the most widely used ones are probably ln T and 2. The BIC (Bayesian Information Criterion) of Schwarz (1978) assigns a non-zero prior probability to a model of small dimension. Maximizing an approximation to the posterior probability of the model is equivalent to minimizing the IC in (3) with $C_T = \ln T$. Akaike (1970) has proposed to measure adequacy by the final prediction error (FPE, also called mean squared forecast error), $E[(R^j_{T+H|t} - \hat{R}^j_{T+H|t})^2]$, which in large samples can be approximated as

$$E[(R^j_{T+H|t} - \hat{R}^j_{T+H|t})^2] \cong \exp\left((\hat{\sigma}^j_p)^2\right)\exp\left(\frac{2p}{T}\right), \tag{4}$$

so that maximizing FPE or $(\hat{\sigma}^j_p)^2 + 2p/T$ are equivalent, and this reveals that Akaike's criterion is equivalent to (3) when $C_T = 2$. In addition to the FPE, $C_T = 2$ can also be motivated from perspective of the Kullback–Leibler (KL) distance. Following Cavanaugh (1997), the KL distance between the candidate model parameterized by $\boldsymbol{\beta}^j_{\mathrm{A}}$ and the true model with density g is

$$D\left(\boldsymbol{\alpha}^j, \boldsymbol{\beta}^j_{\mathrm{A}}\right) = D\left(\boldsymbol{\delta}^j_{\mathrm{A}}\right) = E_g[-2lnL(R^j_1, R^j_2, \dots, R^j_T; \boldsymbol{\delta}^j_{\mathrm{A}})], \tag{5}$$

where $E_g[\cdot]$ denotes expectation taken with respect to the true density AND $L\left(R^j_1, R^j_2, \dots, R^j_T; \boldsymbol{\delta}^j_{\mathrm{A}}\right) = L(\boldsymbol{\delta}^j_{\mathrm{A}}; R^j_1, R^j_2, \dots, R^j_T)$ is the likelihood of the candidate model. While $\boldsymbol{\delta}^j_{\mathrm{A}}$ can be estimated from the data, the KL still cannot be used to evaluate models without knowledge of g. Akaike (1974) considers the expectation of KL when the candidate models nest the true model parameterized by $\boldsymbol{\delta}^j$:

$$\begin{aligned} E_{true}\left[D\left(\hat{\boldsymbol{\delta}}^j_{\mathrm{A}}\right)\right] &= E_{true}\left[-2lnL\left(\hat{\boldsymbol{\delta}}^j_{\mathrm{A}}; R^j_1, R^j_2, \dots, R^j_T\right)\right] \\ &+ \left\{E_{true}\left[-2lnL\left(\boldsymbol{\delta}^j; R^j_1, R^j_2, \dots, R^j_T\right)\right] - E_{true}\left[-2lnL\left(\hat{\boldsymbol{\delta}}^j_{\mathrm{A}}; R^j_1, R^j_2, \dots, R^j_T\right)\right]\right\} \\ &+ \left\{E_{true}\left[D\left(\hat{\boldsymbol{\delta}}^j_{\mathrm{A}}\right)\right] - E_{true}\left[-2lnL\left(\boldsymbol{\delta}^j; R^j_1, R^j_2, \dots, R^j_T\right)\right]\right\}, \end{aligned} \tag{6}$$

Because the second order expansion of each of the last two terms is a likelihood ratio statistic, which can be approximated by p since the expected value of a χ^2 random variable with p degrees of freedom is p, the expected KL suggests to select the best model minimizing:

$$-2lnL\left(\hat{\boldsymbol{\delta}}^j_{\mathrm{A}}; R^j_1, R^j_2, \dots, R^j_T\right) + 2p. \tag{7}$$

In the least squares case this further simplifies to $T(\hat{\sigma}^j_p)^2 + 2p$, which returns Akaike's criterion in (4).[6]

---

[6] Phillips and Ploberger's (1996) is a generalization of the BIC based on a data-dependent term in place of p as a measure of complexity. In fact, most other criteria proposed in the literature have been shown to be related to the AIC or the BIC. For example, Rissanen (1986b) suggested using a predictive principle that minimizes the accumulative squares of prediction errors. Wei (1992) shows that the resulting model selection rule is asymptotically equivalent to the BIC for ergodic models. Rissanen (1986a) uses coding theory to choose a model with the minimum description length (MDL).The MDL of a fitted model has a component that depends on complexity, and another that depends on the fit. As discussed in Stine (2004), the MDL behaves like the AIC for some choice of coding parameters and the BIC for special choice of the prior.

Of course, stepwise algorithms may also be implemented via sequential t- or F-testing, the latter when sets of predictors have to be tested. For instance, in a general-to-specific (backward) approach, we start from the largest model and check if the coefficient on the least (set of) significant variable(s) is zero at some prescribed significance level. If it is not significant, the model is re-estimated dropping that (set of) variable(s) and next least significant (set of) predictor(s) is then tested on its turn. The test on the least significant (set of) variable(s) is repeated until the estimated coefficient on the least significant variable is found to be significant. The crucial parameter in a sequential testing procedure is the size of the test. If the size is too small, the critical value will be large and few variables will be selected.[7]

However, information criteria may achieve consistency in model selection procedures, while sequential Wald-type tests generally cannot achieve that, as they are often based on fixed, large-sample approximations. Shibata (1980) considers selecting the lag order of infinite order Gaussian autoregressions and assumes that the data used for estimation are independent of those used in forecasting. Using a FPE minimization loss, he shows that the (finite) p selected by the AIC is efficient in the sense that no other selection criterion achieves a smaller conditional mean squared prediction error asymptotically. Lee and Karagrigoriou (2001) obtain similar results for non-Gaussian autoregressions. However, Ing and Wei (2003) extend the analysis to allow the sample used for prediction to overlap with that used in estimation. The issue is that while $C_T = 2$ will find the best model amongst the incorrect ones, the dimension of the selected model tends to be unnecessarily large and AIC is understood to fall short when it comes to consistent model selection.[8] In fact, when it comes to consistent model selection, results tend to favor a C_T that increases with T and Geweke and Meese (1981) showed in a stochastic regressors setup that this condition is necessary for consistent model selection.[9] Speed and Yu (1993) show that the BIC with $C_T = ln\ T$ is desirable for prediction and the asymptotic efficiency of the BIC is proven in Shao (1997): $C_T = ln\ T$ is both consistent and optimal for prediction of finite dimensional (parametric) models with observed regressors.

---

[7] Information criteria can also be seen from a stepwise testing perspective. The AIC and BIC choose a test size that corresponds to critical values of $2^{0.5}$ and $ln\ T$, respectively. Conversely, seen from the perspective of information criteria, a two-tailed five percent t test corresponds to a C_T of 1.96.

[8] For instance, Shibata (1976) shows that AIC has a non-zero probability of over-parameterizing finite order autoregressions. Shibata (1984) considers a generalized final prediction error that replaces C_T =2 with some other value, and obtains that such a value needs to exceed one for prediction efficiency but simulations suggest that approximate efficiency is still low when C_T = 2.

[9] A model selection procedure is said to be consistent if the probability of selecting the true model approaches one as the sample size increases. Interestingly, this is different from consistent estimation of a regression function and consistent model selection can conflict with the objective of mean squared prediction accuracy: while the parameter estimates may be biased when the selected model is too small, the parameter estimates will not be efficient if the model is too large.

Our implementation of the automatic variable selection methods is based on rather straightforward uni-directional *forward* and *backward* methods. Because of its good (albeit imperfect) theoretical properties and because it can also be motivated from perspective of the Kullback–Leibler (KL) distance minimization, our algorithms are all based on the minimization of the AIC in (4). We have also tried to use as a criterion individual Wald tests (i.e., including predictors that imply the smallest p-values from standard t tests and until these p-values are inferior to 0.10 and dropping variables that have the largest p-values and removing until their p-values exceed 0.10) but found qualitatively similar results that are available upon request.[10]

2.2. Dimension Reduction Methods

One problem with information criteria when there is a large set of predictors with no natural ordering is that enumeration of 2N predictive regressions is necessary. When N exceeds 10, hundreds of thousands if not millions of models ought to be estimated and compared to implement a stepwise regression approach. A different approach is to use all data available intelligently.[11] A popular technique that combines the potentially relevant predictors $\mathbf{X}_t$ into new predictors is principal components. By definition, the TxN principal components (henceforth, PCs) of the matrix of predictors $\mathbf{X}$ are defined as

$$\boldsymbol{X}_{PC} = \boldsymbol{U}_X \boldsymbol{D}_X, \tag{8}$$

where $\boldsymbol{D}_X$ is a NxN diagonal matrix that collects the N eigenvalues of $\mathbf{X}$ each ordered from high to low on the main diagonal, and $\boldsymbol{U}_X$ is a TxN orthogonal matrix spanning the column space of $\mathbf{X}$. Because of (8), the j-th PC $\boldsymbol{X}_{PC,j}$ (j = 1, 2, …, N) s the linear combination of $\mathbf{X}$ that captures the j-th largest variation in $\mathbf{X}$. A principal component regression replaces the TxN predictor matrix $\mathbf{X}$ with a Txr_{x} sub-matrix of PCs, $\boldsymbol{X}_{PC,1:r_{\mathrm{X}}}$ to estimate:

$$R^j_{T+H|t} = \boldsymbol{W'}_{T+H|t}\boldsymbol{\alpha}^j + \boldsymbol{X'}_{PC,1:r_{\mathrm{X}}}\boldsymbol{\beta}^{\mathrm{j}}_{PC} + \epsilon^j_{t+H|t}, \tag{9}$$

Principal component analysis is an unsupervised dimension reduction technique.[12] Ng (2013) characterizes this technique as a way to shrink the OLS coefficient vector away from directions in the

---

[10] In preliminary tests (based only on the macro factors), stepwise methods applied using the BIC criterion have led to results were indistinguishable from the AIC-based ones and therefore they were not pursued further.

[11] One could also use a subset of the predictors at a time and then combine the forecasts produced by the different subset of predictors, the so-called method of model averaging pioneered by Bates and Granger (1969) and reviewed in Timmermann (2006). Because stepwise regressions consider all (combinations of) predictors jointly, to perform a comparison with HMM resting on equal grounds, in this paper we eschew an extension to compare with forecast combination methods.

[12] In contrast,(static) factor analysis assumes that the data have a specific structure but Tipping and Bishop (1999) have shown that when T is small and N is large, a PC regression model can be seen as a Gaussian latent

predictor space that have low variance, by using rectangular (on/off) weights to truncate small eigenvalues in the space of predictors to exactly zero.[13]

2.3. One Encompassing Framework

Pulling together all the predictive approaches listed above, in this paper we recursively estimate and backtest the predictive performance of the following regressions (for j = 1, 2, …, N):

$$R^j_{t+1|t} = \kappa^j_t + \sum_{i=1}^{r_X} \beta^j_{i,t} PC_{i,t} + DS^j_t \left(\sum_{k=1}^{K} \gamma^j_{k,t} CS^j_{k,t}\right) + DA^j_t \left(\sum_{k=1}^{K} \varrho^j_{k,t} CA_{k,t}\right) + \epsilon^j_{t+1|t}, \quad (10)$$

where $\epsilon^j_{t+1|t}$ IID $N(0, (\sigma^j_\epsilon)^2)$, $PC_{i,t}$ (i = 1, 2, … r_X) are the selected PCs from macroeconomic factors, $CS^j_{k,t}$ are commodity-specific predictors measured for each commodity, DS^j_t is a dummy that takes a value of one when the commodity-specific predictors are present and zero otherwise, $CA_{k,t}$ are commodity-specific predictors *measured in the aggregate across all commodities*, and $DA_{t,j}$ is a dummy that takes a value of one when the aggregate, commodity-relevant predictors are present and zero otherwise. Note that all the estimable coefficients, i.e., $\beta^j_{i,t}$ (i = 1, 2., r_X), $\gamma^j_{k,t}$ (k = 1, 2, …, K), and $\rho^j_{k,t}$ (k = 1, 2, …, K) are both specific to each predictive regression but also potentially time-varying, according to two potential schemes: a regime switching, HMM model to be specified in what follows, or the recursive application of a stepwise predictor selection scheme.[14,15]

By imposing alternative restrictions on the structure of (10), it is possible to obtain a number of models, whose recursive forecasting performance is subject to backtesting:

---

variable model that is closely related to factor analysis. While a factor interpretation is not necessary to motivate the use of PCs as predictors, a few results are available when a factor structure is imposed, see Stock and Watson (2002).

[13] There has always been a disagreement as to whether one should reduce the dimension of **X** on the basis of the marginal distribution of **X**, or the conditional distribution of R^j_{T+H} given **X**. For instance, Kiers and Smilde (2007) argue that aiming to explain both the predictors and the endogenous variable will be better able to yield models that predict well. Techniques, such as Partial Least Squares Regressions (PLS), exist to estimate principal components in the form of factors that explicitly reflect when our objective is to forecast a specific target. Factors that have good explanatory power for **X** may not be good predictors for R^j_{T+H}. However, there is evidence in the literature that targeting R^j_{T+H} does not necessarily give better finite sample properties and provide only limited control over the degree of shrinkage obtained.

[14] Technically, the coefficients are time-varying and predicted for time t+1 on the basis of the information set at time t, for instance $\beta^j_{i,t+1|t}$ or $\beta^j_{i,Q_{t+1}}$. However, we simplify the notation and simply write $\beta^j_{i,t}$ and hence β^j_{i,Q_t}, aligning the pedix to the information set.

[15] However, in all cases when a HMM drives the predictive regression, the residual variance is specified as homoskedastic. On the one hand, we are interested in the presence of regimes in the predictability patterns concerning mean commodity returns and not in forecasting their variance (a task that is interesting in itself, but remote from our goals in this paper). On the other hand, such a design allows us to perform an informative comparisons with stepwise regression methods that are applied on a regime-specific basis when the regimes is provided by a HMM model applied to observable financial market variance, as proxied by the VIX.

0. The benchmark is represented by a simple AR(1) model for commodity returns:[16]

$$R_{t+1|t}^{j} = \kappa^{j} + \phi^{j} R_{t}^{j} + \epsilon_{t+1|t}^{j}, \tag{11}$$

1. When $DS_t^j = DA_t^j = 0$ and $\kappa_t^j = \kappa^j$, $\beta_{i,t}^j = \beta_i^j$, we obtain a constant coefficient predictive regression model that contains only macro factors as predictors.
2. When $DS_t^j = DA_t^j = 0$ and the parameters (κ_t^j and $\beta_{i,t}^j$) are driven by a standard (first-order, irreducible, time homogenous, ergodic) Markov state variable, we have a HMM macro factor model.
3. When $DS_t^j = DA_t^j = 0$ and the parameters (κ_t^j and $\beta_{i,t}^j$) are recursively selected by backward/forward stepwise regressions, we have a stepwise regression macro factor model.
4. When $DA_t^j = 0$ and the parameters (κ_t^j, $\beta_{i,t}^j$, and $\gamma_{k,t}^j$) are driven by a standard (first-order, irreducible, time homogenous, ergodic) Markov state variable, we have a HMM model that includes macro- as well as commodity-specific factors.
5. When $DS_t^j = 0$ and the parameters (κ_t^j, $\beta_{i,t}^j$, and $\varrho_{k,t}^j$) are driven by a standard (first-order, irreducible, time homogenous, ergodic) Markov state variable, we have a HMM model that includes macro- as well as aggregate (across all commodities) commodity-specific factors.
6. When $DA_t^j = 0$ and the parameters (κ_t^j, $\beta_{i,t}^j$, and $\gamma_{k,t}^j$) are recursively selected by backward/forward stepwise regressions, we have a stepwise regression model that includes macro- as well as commodity-specific factors; we also implement a version in which the stepwise selection applies only to macro factors, but the commodity-specific factors are always included.
7. When $DS_t^j = 0$ and the parameters (κ_t^j, $\beta_{i,t}^j$, and $\varrho_{k,t}^j$) are recursively selected by backward/forward stepwise regressions, we have a stepwise regression model that includes macro- as well as aggregate (across all commodities) commodity-specific factors; we also implement a version in which the stepwise selection applies only to macro factors, but the aggregate commodity-specific factors are always included.[17]

Besides fully-fledged HMM models (as in 2, 4, and 5),

$$R_{t+1|t}^{j} = \kappa_t^{j} + \sum_{i=1}^{r_X} \beta_{i,Q_t}^{j} PC_{i,t} + \sum_{k=1}^{K} \gamma_{k,Q_t}^{j} CS_{k,t}^{j} + \sum_{k=1}^{K} \varrho_{k,Q_t}^{j} CA_{k,t} + \epsilon_{t+1|t}^{j}, \tag{12}$$

---

[16] As commented in section 2.3, this standard benchmark in the forecasting literature should not be confused with our definition of the commodity-specific momentum factor that also employs AR(p) models.

[17] The implementation in which either the commodity-specific or the aggregated commodity factors are restricted to enter all the models experimented avoids that we introduce an asymmetry between the macro factors and the commodity predictors, in the sense of always keeping the macro factors in the analysis.

(Q_t is a two state HMM with constant transition matrix similar to (14) specified below), we also implement a simpler, volatility-driven regime switching aiming at boosting the predictive performance of the stepwise regression forecasts (of course, to apply such regime switching boosting to explicit HMM models would be just awkward). It is well-known that volatility regimes ought to be taken into account in predictability tests applied to financial returns (see, e.g., Rapach e Zhou, 2013). In particular, we adopt the VIX, computed from the CBOE from the price of puts and calls written on the S&P 500 index as market portfolio volatility proxy with reference to the sample entertained in the paper. Similarly to Papanicolaou and Sircar (2014), we fit on the VIX a two-state HMM that separates the low- from the high-volatility regimes using a simple Markov switching AR(1) process,

$$VIX_t = v_{S_t} + \varphi VIX_{t-1} + \varsigma_t, \tag{13}$$

with ς_{t+1} IID $\mathrm{N}(0, \sigma^2)$. S_t is a latent two-state Markov state variable that is driven by the transition matrix:

$$P = \begin{bmatrix} P(S_t = low\ vol | S_{t-1} = low\ vol) & P(S_t = high\ vol | S_{t-1} = low\ vol) \\ P(S_t = low\ vol | S_{t-1} = high\ vol) & P(S_t = high\ vol | S_{t-1} = high\ vol) \end{bmatrix} = \begin{bmatrix} p_{ll} & p_{hl} \\ p_{lh} & p_{hh} \end{bmatrix}, \tag{14}$$

where $p_{il} + p_{ih} = 1$ for i = l, h. The model in (13) implies that the unconditional mean VIX in each regime is $E[VIX_t | S_t] = v_{S_t}/(1-\varphi)$. Figure 1 shows the one-month ahead predicted probabilities that the VIX is a low-volatility regime (of course, the plot for the high volatility regime is specular, as the two predicted probabilities must sum to one by construction). The figure clearly shows that our sample may be roughly sub-divided in four sub-periods. January 1989 – December 1997, January 2004 – December 2007, and January 2013 – May 2018 are dominated by a low mean level of the VIX; January 1998 – December 2003 and January 2008 – December 2012 are instead clearly high-VIX sub-samples.

In correspondence to each month t in our backtest period, the forecasts are obtained exploiting the predicted probabilities from the estimated HMM model in (13) on VIX data, in two alternative ways:

1. By ex-ante classifying whether the forecasts will be originated by different stepwise predictive regressions recursively estimated on two sub-samples of data, i.e.,

$$R^j_{t+1|t} = \kappa^j_{t,l} + \sum_{i=1}^{r_X} \beta^j_{i,l} PC_{i,t} + DS^j_t \left(\sum_{k=1}^{K} \gamma^j_{k,l} CS^j_{k,t} \right) + DA^j_t \left(\sum_{k=1}^{K} \varrho^j_{k,l} CA_{k,t} \right) + \epsilon^j_{t+1|t} \quad t \in Low\,Vol\ data$$
$$(t \ \ni \ P_t(S_{t+1} = low\ vol) > 0.5) \tag{15}$$

$$R^j_{t+1|t} = \kappa^j_{t,h} + \sum_{i=1}^{r_X} \beta^j_{i,h} PC_{i,t} + DS^j_t \left(\sum_{k=1}^{K} \gamma^j_{k,h} CS^j_{k,t} \right) + DA^j_t \left(\sum_{k=1}^{K} \varrho^j_{k,h} CA_{k,t} \right) + \epsilon^j_{t+1|t} \quad t \in High\,Vol\ data$$
$$(t \ \ni \ P_t(S_{t+1} = low\ vol) \leq 0.5) \tag{16}$$

   where $P_t(S_{t+1} = low\ vol)$ and $P_t(S_{t+1} = high\ vol)$ are the predicted probabilities of each state.

2. By computing forecasts that are weighted by the HMM predicted state probabilities,

$$
\begin{aligned}
R^j_{t+1|t} = {} & \left[P_t(S_{t+1}\text{=low vol})\kappa^j_{t,l} + P_t(S_{t+1}\text{=high vol})\kappa^j_{t,l}\right] + P_t(S_{t+1}\text{=low vol}) \sum_{i=1}^{r_X} \beta^j_{i,l} PC_{i,t} \\
& + P_t(S_{t+1}\text{=high vol}) \sum_{i=1}^{r_X} \beta^j_{i,h} PC_{i,t} \\
& + DS^j_t \left[P_t(S_{t+1}\text{=low vol}) \sum_{k=1}^{K} \gamma^j_{k,l} CS^j_{k,t} + P_t(S_{t+1}\text{=high vol}) \sum_{k=1}^{K} \gamma^j_{k,h} CS^j_{k,t}\right] \\
& + DA^j_t \left[P_t(S_{t+1}\text{=low vol}) \sum_{k=1}^{K} \varrho^j_{k,l} CA_{k,t} + P_t(S_{t+1}\text{=high vol}) \sum_{k=1}^{K} \varrho^j_{k,h} CA_{k,t}\right] + \epsilon^j_{t+1|t} \quad (17)
\end{aligned}
$$

2.4. The Recursive, Pseudo-Out of Sample Backtesting Exercise

For each of the commodities futures return series listed below, we conduct recursive pseudo OOS exercises in the spirit of Rapach and Zhou (2013). Each of our models—as defined by a selection of predictors (within the list that includes the macro PCs, the commodity-specific predictors, and the commodity-specific variables aggregated across the cross-section of commodities)—is estimated recursively on an expanding window of data starting with the January 1989-December 2003 subsample. We then proceed to perform estimation on the sub-samples ending in 2004:01, 2004:02, and so forth, until the last subsample ending in April 2018. This simulates the sequence of estimations that a sector analyst or a commodity futures traders would have historically performed. Interestingly, because our backtest concerns a 2004-2018 sample, we have no concerns that any of the forecasting frameworks we entertain would have been developed only after 2004. Yet, the long data sample spanning almost three decades worth of monthly data is sufficient to estimate even those econometric models that involve a large number of parameters, such as the model in (15), when stepwise predictive regression estimation is performed over two alternative sub-samples. In fact, even the very first sub-sample (1989-2003) includes 288 observations that can easily accommodate the estimation of up to 15 x 2 (the number of regimes) = 30 parameters in the conditional predictive mean function. Moreover, our sample contains the entire financial crisis and the run-up—also involving commodity prices between late 2007 and 2009—of asset prices leading to it.[18] In correspondence to each month between December 2003 and April 2018, we produce one-month ahead return forecasts for each of the

[18] Although not uncommon in empirical finance, we do not entertain moving window, rolling regression based on a fixed number of observations because such data schemes are justified only by the presence of regimes and/or structural instability. In our application, the potential for regime or breaks is dealt with by performing afresh stepwise model selection as the sample expands and, even more formally, by estimating HMM models that do imply regime switching dynamics in predictability relationships.

commodities under investigation. For example, for the subsample ending with 2003:12 we compute forecasts for 2004:01, for the sub-sample ending in 2004:01, we produce forecasts for 2004: 02, etc., up until 2018:04 when we compute the last set of predictions referring to 2018:05.

3. The Data

3.1. The Commodity Futures Return Series

We consider monthly return series computed from the last-trading-day-of-the-month settlement prices for 14 commodity futures contracts. The data are collected from Thomson Reuters Eikon over a January 1989 – May 2018 sample. The commodities are light crude oil, seven agricoltural products (cocoa, coffee, corn, cotton, sugar, soy, and wheat), three precious metals (gold, silver, and platinum), two soft commodities (orange juice and timber), and live cattle. Following common practice in the literature (see, e.g., Gorton and Rouwenhorst, 2006; Basu and Miffre, 2013; Fuertes et al., 2015) we just focus on fully collateralized positions. This carries two implications. First, we exclude that investors be allowed to operate on margin, thus using leverage; while this approach bound from above the returns that an investor may achieve, it has the advantage to make investments in commodities directly comparable with the investments in other asset classes, which usually require an initial money outflow. This is of course crucial to our portfolio exercise. Second, the lack of a margin system limits the possibility of any unintentional liquidation (due to insufficient collateral) of the position before the end of the investor's holding period, which again makes commodity positions directly comparable to those in stocks or bonds.

Because of the full-collateralization assumption, we can calculate the return on a commodity *j* future position simply as

$$R_{t+1}^{j} = \frac{F_{t+1}^{j}}{F_{t}^{j}} - 1. \tag{18}$$

As it is commonly acknowledged when it comes to using futures price data, calculating returns is made more complicated by the fact that the "*front-end*" contract, the one close to expiry (which is in general also the most liquid) is usually to be replaced before its natural maturity because of a need to hold the position over time (avoiding that delivery of the commodity is triggered). In fact, especially in the case of commodities for which physical delivery is possible, to avoid delivery the traders typically close their positions before the "*First Notice Day*", i.e., the first day from which the Exchange may dispose the physical delivery of the underlying commodity, and hence before the "*Last Trading Day*". To identify the *First Notice Day* for all contracts considered, we have resorted to the official trading calendar of the relevant exchange. Moreover, to try and forecast when it is most likely that trader would shift out of expiring contracts and into the second-to-expiry contract (i.e., before *First Notice Day*), we adopt the methodology proposed by Bakshi et al. (2017): because an investor aims at avoiding the delivery of the commodity, we assume that she would take a position in the futures contract with the second closest maturity on the last business day of each month t, when the contract's First Notice Day occurs after the end of month t+1.

In table A1, in the on-line Appendix, we report the descriptive statistics for the returns on the 14 commodity futures under investigation. Interestingly, all the return series imply similar sample means standard deviations, ranging from 0.2% per month in the case of live cattle to 0.76% in the case of light crude oil (from 4.5% per month in the case of gold to 10.3% in the case of coffee, when it comes to sample standard deviations). While all series display a positive mean, a few are characterized by a negative median return (coffee, silver, and wheat), which is possible thanks to a rather large, precisely estimated positive skewness, i.e. their "normal" price change is negative, apart from a few occasional, large price spikes. Only live cattle and soybeans returns are characterized by negative skewness. Finally, all series display positive excess kurtosis, which—jointly with the widespread evidence of non-zero skewness—implies that returns are strongly non-normal.

3.2. Macroeconomic factors

Because we aim at representing broad categories of economic activity after reducing the dimensionality of the problem, we consider the 132 macroeconomic and financial variables made available by Ludvigson and Ng (2009), with reference to a sample starting in January 1989 and updated up to May 2018. The series represent wide classes of general macroeconomic U.S. (with a few exceptions, such exchange rates and international trade indicators) variables: output and income, labour market, housing, consumption, orders and inventories, money and credit, stock market, bond and exchange rate, and prices series.[19] Notably, the 132 series include a few financial and monetary variables—with the exception of commodity price indices, of course (to avoid duplicating spurious autoregressive structures)—because these have been proven to reflect or even anticipate business cycle developments. All variables are made stationary by applying appropriate (mostly log-differencing) transformations before extracting the PCs.

To reduce the number of parameters to be estimated, and because these explain more than 50% of the total variance of the 132 variables under consideration (to be precise, they explain 52.2% of it), we

[19] Also because these data are publicly available, a Reader may find a detailed description of the variables and of the sources from which they were collected in an appendix available upon request from the Authors.

have opted to include in the empirical exercise only the first eight PCs (plus the third power of the first principal component, following Guidolin and Pedio, 2018) to capture any non-linear effects, i.e., $r_{\mathrm{x}} = 9$ and $\boldsymbol{X}_{PC,1:9}$ replaces $\boldsymbol{X}_{PC}$ that in principle ought to include 133 columns. Interestingly, the first four PCs plus the eigth component, explain by themselves 41% of the total variance of the 132 original series, confirming that there is a low dimensional vector of shocks that drives the state of the economy. In particular, the first PC mostly reflects (i.e., it weighs heavily on) measures of employment and output (e.g., non-farm payroll employment and manufacturing output) as well as indicators of capacity utilizations and new manufacturing orders. To confirm its nature of an output growth factor, the first PC shows modest correlation with prices and financial variables. The second PC reflects the dynamics of interest rate spreads, well-known to be key business cycle indicators, and in fact is characterized by a strikingly high correlation of almost 70% with the junk spread between short term, Baa corporate rates and the Fed Funds Rate. The third and fourth PCs instead overweigh the variables related to inflation and prices and—by contrast with the first PC—show modest correlations with measures of output and employment. Because the third and fourth PCs—that are by construction orthogonal to each other and hence reflect different information but command a roughly similar interpretation (yet, e.g., the third PC loads heavily on consumer price indices, while the fourth on specific production price sub-indices)—represent inflation factors, they also reflect information in the changes in nominal interest rates, especially 5-year Treasury note yields. The eight PC loads heavily on measures of aggregate stock market capitalization (whose log first differences are of course akin to the standard notion of market portfolio return), and appears to be a sort of orthogonalized CAPM-style factor.[20]

4.3. Commodity factors

We compute the commodity factors—both specific to each single commodity and built by aggregation over the cross-section of commodities and for the entire market—from micro-level, disaggregated data provided by the US Commodity Futures Trading Commission (CFTC). In particular, the data have been extracted from the weekly "*Commitment of Trader*" (COT) reports, which detail for each traded futures contract the aggregate number and value of long and short positions by trader type, i.e., differentiating across "*commercials*", "*non-commercials*", and "*non-reportables*". Such statistics are compiled at weekly frequency on Tuesdays and published three days later, in correspondence to the market close on Friday. The CFTC classifies a trader as "commercial" when she uses the futures contracts for hedging

[20] Unfortunately, even though they are statistically relevant and logically key to bringing the fraction of explained variance to exceed 50%, the fifth, sixth, and seventh PCs fail to have a clear interpretation, as they weakly load with coefficients that have heterogeneous signs and large sub-sets of the starting 132 variables.

purposes as defined by Rules 1.3 (z) CFTC, 17 CFR 1.3 (z). "Non-commercials" are agents that cannot justify their trades with hedging goals but whose nature can be identified (e.g., but not exclusively, these are financial institutions). Based on these CFTC statistics, we compute four commodity factors: the *hedging pressure* (HP), the *basis*, the momentum factor (these three factors follow Daskalaki et al., 2013), and the *net trading* (NT) factor (as in Kang et al., 2018). Even though we follow the literature that we cite, we briefly comment on the definitions of these four factors in what follows. Because the commodity factors are only four (eight when their aggregate is considered), it did not make sense to try and build PCs for these factors too.

The HP factor (see also Basu and Miffre, 2013) is measured as the difference between the negative and the positive hedge positions. HP for commodity *j* at time *t* is computed as the ratio between the number of hedged short positions minus the number of hedged long positions, divided by total number of hedgers in that commodity futures at time *t*:

$$HP_{j,t} = \frac{\#shorthedgeposition_{j,t} - \#longhedgeposition_{j,t}}{Total\#hedgeposition_{j,t}} \tag{19}$$

The basis factor is defined as the difference between the return of a portfolio of all futures (for a given commodity *j*) with a positive basis and a portfolio of all futures with a negative basis:

$$Base_{j,t} = \frac{F_{j,t} - F_{j,t+1}}{F_{j,t}} \tag{20}$$

where $F_{j,t}$ is the price of the futures closer to expiry and $F_{j,t+1}$ is the price of the futures with next closest maturity for commodity *j*; the basis is positive when the futures price increases with more distant expiry dates.

The NT factor for a commodity futures contract *j* is defined as the difference between the number of long positions at time *t* and the number of long positions on the same contract at time *t*-1, divided by the *open interest* (OI) on the same contract at time *t*, where the latter is a measure of the number of contracts that have been issues/originated by time *t*:

$$NT_{j,t} = \frac{\#netlongeposition_{j,t} - \#netlongposition_{j,t-1}}{OI_{j,t-1}} \tag{21}$$

A high $NT_{j,t}$ indicates that net positions are growing relative to the total number of open futures contracts open for trading. Finally, because we study commodity returns at the disaggregated level of each individual contract and not of portfolios, so that a cross-section of commodity returns cannot be defined, we characterize momentum (see de Groot et al., 2014) as a family of autoregressive processes AR(*p*), with *p* = 1, …, 12 estimated on each individual futures return series.

Once we have constructed the factors at the individual commodity level, we also proceed to build aggregate commodity factors with reference to all futures and all commodities. As far as the HP and NT factors go, we use the formulas in (19) and (21) but we apply them to all traded futures as reported by the COT reports (i.e., the data concern *#net long/short position_aggregated, #long/short hedging position_aggregated*, and the overall open interest). As for the basis factor, we compute a simple average of the basis factors associated to each of the 14 commodities under investigation. Finally, when it comes to momentum, we simply fit a family of AR(p) models (p = 1, …, 12) to the simple average of the returns on the 14 commodity futures series, whose expiry dates have been selected according to the criteria illustrated in section 3.1.[21]

Table A1 in an on-line Appendix show summary statistics for the four individual commodity as well as the aggregate commodity factors. However, the interpretability of these statistics is limited. Means and medians of hedging pressure are structurally positive—this means that short hedged positions tend to exceed the long ones. However, most HP series all display negative skewness, to indicate that there are episodes/sub-periods in which long hedging positions do prevail. Mean and median basis factors are instead predominantly negative, which indicates that typically the futures term structure is upward sloping (in contango) as the maturity grows. In this case, the prevailing positive skewness indicates that however, at least episodically, commodity futures markets turn into backwardation. Finally, the net trading factor tends to display means and medians that are small and positive, an indication that trading has grown slowly relative to open interest.

4. The statistical predictive performance

Table 1 reports the RMSFE results, i.e., the estimates of

$$RMSFE^{j}(\mathcal{M}) = \sqrt{\frac{\sum_{t=1}^{172}\left(\hat{R}^{j}_{t+1|t}(\mathcal{M})-R^{j}_{t+1}\right)^{2}}{172}} \tag{22}$$

(where j indices different commodities and $\mathcal{M}$ refers to a specific model as defined by the selection of predictors and/or of a framework of variable selection vs. dimension reduction) from the OOS backtest exercise when the existence of (volatility) regimes in the data is used to estimate alternative stepwise regressions separately applied to low vs. high volatility regimes (as identified in Figure 1) and by

[21] While in the case of individual commodities, identifying momentum with a predicted return from a AR process is the only possibility we have, in the aggregate case, we may have explored building a long-short portfolio based on the lagged performance of alternative futures and commodities (e.g., going long in the five best performing commodities over the past 12 months while shorting the worst five performing commodities over the same span of time). However, to preserve consistency across alternative experiments and also because the commodity-specific and aggregate commodity factors have been occasionally combined together, we have refrained from performing tis robustness check.

applying the forecasting equations (15)-(16) when the coefficients are time-varying because backward of forward stepwise methods can set some of them to zero, when a predictor ends up not being included at time t. The different panels of the table refer to the models 0, 1, 3, 6, and 7 described in Section 3.3. Models 2, 4, and 5 are instead HMM models that correspond to the forecasting framework in equation (17) and therefore shall be examined later. In other words, Table 1 simply combines alternative subsets of variables using stepwise regression to perform variable selection, and a crude HMM analysis applied to mean of the VIX to classify alternative regimes, as often done by practitioners (see, e.g., Cano, 2019). However in Baseline Panel at the top of the table, we find the OOS RMSFEs for the "kitchen sink model" in (10) in which all the variables are always included and not stepwise regression methods have been applied. Also the kitchen sink model has been however used to produce forecasts using equations (15) and (16), i.e., it may also benefit from exploiting the information in VIX regimes.

Table 1 delivers only one stark result: while in quiet periods (Jan. 2004 – Dec. 2007, Jan. 2013 – May 2018), it is possible for most commodities to outperform a simple AR(1) benchmark, in the high volatility regime (Jan. 2008 – Dec. 2012), this cannot be achieved. Moreover, the table shows that while stepwise regressions are generally useful in predictive exercises, there is no clear outperforming model in three dimensions: (i) predicting more accurately vs. a simple AR(1) benchmark in the low volatility regime and (ii) which variables ought to enter the predictive regressions, and (iii) whether backward or forward algorithms turn out to be more promising. For instance, for 10 commodities out of 14, in the low volatility regime, stepwise regressions forecast better than the AR(1) benchmark; in fact, in the case of crude oil future returns, this occurs in both regimes and the RMSFE is between 6 and 7% per month vs. 8-14% under a kitchen sink model and 8.6% from an AR(1). The table shows that no combination of variables turns out to dominate the others so that the question as to whether commodity-specific factors ought to be included or not fails to find a precise answer based on this crude, VIX regime-classification driven analysis. Given that in 12 occasions out of 28 (as defined by the combination of underlying commodity and volatility regime) the framework yielding superior predictions is different from the benchmark, we record that the "winner" is represented by a model that includes the nine PCs in 5 occasions and by a model that includes the macro factors and aggregated commodity factors in 3 cases; in the remaining 4 cases, commodity-specific information is included in the best model, one way or the other. Hence, our claim that all in all, it remains unclear whether commodity predictors may be needed in a forecasting perspective. Finally, out of 11 combinations in which stepwise predictive regressions lower the RMSFE vs. the benchmark, eight such cases involve forward, simple-to-general algorithms, which can be seen as progressive expansions of the benchmark AR(1) to improve its predictive yield. However, these empirical findings do echo the classical result in forecasting that simpler, even minimal predictive models characterized by few parameters—or at least construction algorithms that start out with essential models and expand them on a simple-to-general path therefore limiting model complexity in a data-driven fashion—turn out to be superior vs. more richly parameterized frameworks that however fit the data better.

Table 2 repeats the same backtesting OOS exercises in Table 1 when the conditioning on the regimes occurs using a consistently specified HMM framework, as in equations (12) and (17).[22] Because all forecasts are in this case weighted by the predicted state probabilities derived from the commodity futures return series, the results are no longer depend on whether volatility is "high" or "low". Of couse, this enhances the readability of results. However, the distinction between backward and forward stepwise algorithms remains. Compared to Table 1, now the AR(1) benchmark appears to be less dominant. In the case of six underlying commodities out of 14, a HMM implementation of our predictive framework yields competitive RMSFE results. In three of such six cases, the lowest OOS RMSFE was achieved by mixing HMM with forward-stepwise model building when the predictive model employed only the 9 macro PCs (this occurred for gold, orange juice, and sugar); however, in the three remaining cases (coffee, crude oil, and silver), the best performing model was either including the individual commodity specific factors, or all the predictors jointly, through the HMM kitchen sink implementation. Therefore, even from Table 2 it remains unclear whether commodity factors may be need on top of the macro factors in a predictive perspective: while it seems safer to include them because it is not their inclusion that may cause a more complex model to not outperform the AR(1) benchmark, in three cases out of six their presence hurts predictive performance, which confirms earlier results in Guidolin and Pedio (2018), without the HMM component. However, Table 2 does give a strong implication on one account: the presence of coherently formulated HMM components generally improves forecasting power. This can be inferred by comparing the RMSFE in Table 2 and noting that they are lower vs. Table 1 and from the fact the in 12 out of 14 cases, a HMM kitchen sink forecast model outperforms a single-state, constant coefficient (linear) predictive regression. For instance, in the case of gold futures returns, switching from a standard predictive regression to one with Markov regimes in the mean coefficients lowers the RMSFE from 7.1 to 6.7% per month; however, these accuracy metrics are inferior vs. those guaranteed by the AR(1) benchmark, 5.6% per month. However, when the 9 macro factors are brought at work using a HMM, forward stepwise regression approach that combines both dimension reduction and automatic selection techniques, RMSFE is further reduced to 5.2%. The message is that either a model is kept to zero complexity and simple ARMA may work, or the complexity

[22] The framework is consistent because the prediction of the hidden regimes are based on the very predictive regression to which the regime probabilities are then applied to in order to compute the forecasts.

must compound in appropriate ways for the predictive power to be enhanced.

One limitation of Tables 1 and 2 is that they fail to actually compare the realized OOS RMSFE from "pure" HMM implementations of the predictive regression vs. stepwise selection algorithms, so that to deal with the question is model selection better than time-variation in coefficients remains unnecessarily difficult. Table 3, with reference to realized RMSFEs, addresses this comparison: panels A and C report results for pure HMM implementations, in which all variables are included but their coefficients are allowed to follow a first-order Markov process with two regimes.[23] In overall terms, the table gives the same evidence as Table 2: in six out of 14 cases, HMM and/or stepwise regression models outperform an AR(1) benchmark, which given the complexity of the former set of models, may be thought of as a lukewarm turn out. However, as for the specific question raised above, out of the six cases, in two of them it is the HMM framework that minimizes the RMSFE, with reference to coffee and silver futures returns. Interestingly, it is a full-model HMM representation that achieves this result, which supports a good forecasting power for HMM. However, two out of six cases is far from striking performance.

Table 4 performs the same backtest analysis as in Table 4 but reports the mean absolute forecast error (MAFE), i.e.,

$$MAFE^{j}(\mathcal{M}) = \sqrt{\frac{\sum_{t=1}^{172} |\hat{R}_{t+1|t}^{j}(\mathcal{M}) - R_{t+1}^{j}|}{172}}, \tag{23}$$

which uses an absolute value loss function. The results in Table 4 are somewhat more favorable to the cause of complex, HMM and variable selection models vs. Table 3. In this case, for seven out of 14 underlying commodities, the AR(1) benchmark is outperformed in our backtests. Interestingly, in five out of seven such cases (coffee, corn, orange juice, silver, and soybeans), we find that the lowest achievable MAFE is returned by HMM models that include both the macro and the aggregated commodity factors. This implies that the individual commodity factors are hardly relevant in a forecasting performance; the same can be said (with the exception of crude oil and gold) about stepwise predictive regressions that appears to be less effective under an absolute value loss function. This implies that while AR(1) tends to minimize the impact of extreme forecast errors relatively to HMM and stepwise regressions, this property is of lower relevance when the squared loss function is replaced by an absolute value one.[24]

---

[23] By construction, panel C in which all predictors enter the HMM framework coincides with the second row of Table 2. We have also tried HMM specifications including only aggregated commodity factors, but these turned out not deliver a superior OOS performance vs. panel C of Table 3 and were therefore omitted to save space.

[24] Interestingly, Table A2 in the on-line Appendix shows another intriguing results: when separate models are estimated over low vs. high-VIX regimes, under a MAE loss function for as many as eleven underlying

Tables 5 and 6 provide an additional summary of the HMM predicted probability-weighted results in Tables 3 and 4: for the cases of the RMSFE (Table 5) and MAFE (Table 6) loss functions, we build equally-weighted summaries of the OOS predictive performances of HMM, stepwise selection methods, and the AR(1) benchmark. The equal weighting is applied in an extremely crude way: for each underlying commodities we simply average predictive performance measures across all model specifications and backward vs. forward of the stepwise methods. Although this choice suffers from obvious limitations, the point of these two very simple tables is to tease out the ex-ante, expected predictive performance that would accrue to a forecaster who would not have access to any of the backtesting results in Tables 1-4 and would randomly picky any of the models in the set including HMM, stepwise regressions, and the AR(1) benchmark. Consistently with earlier results, in both tables the "average" HMM and stepwise models show considerable difficulties at outperforming the benchmark. For only two commodities (corn and crude oil) out of 14 and only in the case of the MAFE, the benchmark may be outperformed by randomly picking a more complex model. This implies that even though our recursive OOS exercise has been carefully performed avoiding any hindsight bias, a forecaster that would be forced to pick either a HMM or a stepwise algorithm would a very slim chance to predict more accurately than a simple, but parsimonious AR(1) model. Interestingly, there is also some evidence that—at least for most commodities—stepwise regressions outperform HMM forecasts.

5. Asset allocation performances

At a superficial level, the OOS evidence in section 4 seems to place a final world on the issue of whether or not either a combination of dimensionality reduction and automatic variable selection techniques or HMM models (or both) may deliver more accurate recursive forecasts of commodity returns vs. an off-the-shelf, rudimental AR(1) model. However, considerable literature has cumulated on applications and theoretical reasons for why (relatively) inaccurate forecast models in a *statistical perspective* may turn out to yield portfolio decisions that yield positive OOS *economic value* in backtesting exercises (see Leitch and Tanner, 1991; Abhyankar et al., 2012; Dal Pra et al., 2018; Gao and Nardari, 2018). Of course, the reason for such findings is that when in a recursive forecasting exercise a given loss function is replaced by a (sufficiently different) new loss function, the OOS rankings across models may be radically affected. Therefore in this section, we perform an additional OOS test, in the form of a

---

commodities, we find that stepwise regression models yield a forecasting performance that is superior to the AR(1) benchmark. In particular, models based on macro factors only turn out to be relatively successful. However, Table A3 in the Appendix, again under a MAE loss, shows that—similarly to Table 3—HMM and stepwise models are considerably less successful when the forecasts are weighted by the predicted state probabilities under HMM.

recursive mean-variance (MV) asset allocation exercise that uses the forecasts from the range of predictive models analyzed in section 4 to compute optimal portfolio shares and proceeds to find their implied realized performances. Such an exercise appears to be particularly sensible in the light of the growing interest by investors in including commodities in their portfolios because of their diversification benefits (see, e.g., Chong and Miffre, 2010; Lombardi and Ravazzolo, 2016; Henriksen et al., 2018), even though our primary goal is not to re-assess whether the inclusion of commodities in an otherwise standard portfolio may generate economic value (see the negative evidence in Daskalaki and Skiadopoulos, 2011; Yan and Garcia, 2017).

Our objective is to perform a systematic comparison of the realized (risk-adjusted) OOS performance of portfolios that exploit forecasts of futures returns estimated with HMM vs. stepwise regressions. In all cases, the models may including macroeconomic, commodity-specific, as well as aggregated commodity factors. We use a robust MV framework that over short investment intervals is known to well approximate many other types of utility-based state preference portfolio modes. Besides the commodity future contracts already entertained above, our asset menu includes the S&P 500 (total return index, a proxy for the equity market), US 10-Year Treasuries, and 30-day T-bills (to proxy cash investments). We include equity, bonds, and cash on the top of our 15 commodity futures strategies for the sake of realism, to simulate the strategic asset allocation decisions of a US investor over time. To retain symmetry and obtain a fair "playing field" across assets, we apply to stock and bond returns the same predictive models (for instance, in terms of whether only macro variables are to be included, as well as in terms of the backward or forward stepwise approaches implemented) applied to commodity futures returns. For simplicity, we assume instead that the 1-month T-bill rate is a constant and known in advance.

Let $\mu_{i,t+1}$ be the predicted (mean) return and $\sigma^2_{i,t+1}$ the variance on asset *i*. An investor allocates her wealth at time t according to weights $\boldsymbol{\omega}_t$, such that $\sum_{i=1}^{16} \omega_{i,t} = 1$ holds (16 is the number of risky assets available to the investor) and that she cares (at least locally, i.e., for a one-period investment horizon) only about the conditional mean and variance of her portfolio returns, so that she maximizes

$$\max_{\boldsymbol{\omega}_t} \mu_{p,t+1} - \frac{\gamma}{2}\sigma^2_{P,t+1} \tag{24}$$

where γ represents her risk aversion coefficient. If we indicate the risk-free rate as r^f_{t+1}, we have that, at any point in time, the portfolio conditional expected return and variance are $\mu_{p,t+1} = r^f_{t+1} + (\boldsymbol{\mu}_{t+1} - r^f_{t+1}{'}\mathbf{1})'\boldsymbol{\omega}_t$ and $\sigma^2_{P,t+1} = \boldsymbol{\omega}_t{'}\boldsymbol{\Sigma}_{t+1}\boldsymbol{\omega}_t$, respectively, where $\boldsymbol{\Sigma}_{t+1}$ is the covariance matrix of asset returns predicted at time t for time $t+1$.[25] This leads to classical, unconstrained program

$$\max_{\boldsymbol{\omega}_t} \; r^f_{t+1} + (\boldsymbol{\mu}_{t+1} - r^f_{t+1}{}'\mathbf{1})'\boldsymbol{\omega}_t \; -\frac{\gamma}{2}\boldsymbol{\omega}_t{}'\boldsymbol{\Sigma}_{t+1}\boldsymbol{\omega}_t \qquad s.t.\; \boldsymbol{\omega}_{t+1}{}'\mathbf{1} = 1, \tag{25}$$

whose first-order condition (also sufficient because of the concavity of the objective) leads to the optimal weights:

$$\widehat{\boldsymbol{\omega}}_t = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}_{t+1}(\boldsymbol{\mu}_{t+1} - r^f_{t+1}{}'\mathbf{1}) \tag{26}$$

No short sale constraints are imposed as most systematic traders are allowed to write futures. Following DeMiguel et al. (2007), we build static one-period optimal portfolios over time, considering expanding windows of data, starting from the beginning of our backtesting sample and including all data until the forecast origin. Next, we calculate the corresponding portfolio mean return, realized mean-variance utility and Sharpe ratio for each time t. The recursive exercise is initialized with reference to January 1989 - December 2003 to produce a January 2004 portfolio and then iterated 172 times until the last estimation sample, January 1989-April 2018, to produce a portfolio for May 2018. Because in this paper we do not compute forecasts of second-order moments, we use historical data on the same expanding window described above to estimate the sample covariance matrix.

For the sake of robustness, we use three different values of the risk aversion coefficient γ (0.10, 0.25, and 0.5). To enhance the realism of the asset allocation exercise, we consider the transaction costs implied by the rebalancing the portfolio at any time t. Crucially, instead of just applying such costs on an ex-post basis, we account for the fact that an investor will optimally rebalance her portfolio if and only if she can get an advantage in terms of an increased risk-adjusted expected return from the adjusted portfolio, i.e., when a portfolio re-shuffling $\Delta\boldsymbol{\omega}_{i,t+1}$ increases ex-ante, anticipated MV utility. In our set up, this implies that she will change portfolio weights at time t only if the transaction costs of rebalancing do not exceed the portfolio expected returns at $t+1$. Therefore, we also solve the portfolio problem under the following, additional condition

$$\sum_{i=1}^{n}\left|\Delta\boldsymbol{\omega}_{i,t}\right| \times \; tc \; \leq \sum_{i=1}^{n}\boldsymbol{\mu}_{t+1}\widetilde{\boldsymbol{\omega}}_{i,t} \tag{27}$$

where tc is a proportional transaction cost, $\Delta\boldsymbol{\omega}_{i,t+1}$ represents a hypothetical change in weights in case of rebalancing and $\widetilde{\boldsymbol{\omega}}_{i,t+1} \equiv \boldsymbol{\omega}_{i,t} + \Delta\boldsymbol{\omega}_{i,t+1}$ represents the hypothetical weights at $t+1$ in case of rebalancing. If (27) fails to hold, then an investor would not rebalance between t and $t+1$, as the implied

[25] In line with the empirical literature on asset return predictability (see Bossaerts and Hillion, 1999), we approximate the conditional covariance matrix with an expanding, sample estimator applied to the residuals of the conditional mean equation estimates that classifies the data according to whether they are drawn from either a low- or a high-variance regime, see Figure 1.

costs are higher than the resulting expected benefits and therefore the solution to (25)-(27) becomes $\tilde{\boldsymbol{\omega}}_{i,t+1} = \tilde{\boldsymbol{\omega}}_{i,t}$, which creates an interesting, path-dependent, no trade region (see Mei and Nogales, 2018). As for the level of the transaction costs, we use a simplified set up in which there are only proportional transaction costs. In reality, an investor willing to rebalance her portfolio, would pay two types of transaction costs: fixed costs to access the market (or infrastructure costs) and liquidity costs. Because it is difficult to make assumptions on the first type of costs, as they are likely to depend on the exact nature of an investor (e.g., whether she is an institutional investor and of what size), we can reasonably assume the liquidity costs being close to the bid-ask spread as a percentage of the mid-price and therefore being proportional to the amount transacted. Therefore, to estimate the parameter *tc*, we have collected daily best ask and best bid prices for commodity futures contracts, for the period Jan. 1996 - Dec. 2017. For each commodity, we estimate a daily time series for *tc* as the bid-ask spread as a percentage of the mid-price. Next, we compute the average *tc* as the *grand* average of all such daily values. We get an average value of the transaction costs across all commodities equal to approximately 0.093%. Therefore, also because such estimates are considerably variable over time and across different commodities, to simplify we set *tc* = 0.1%. When the constraint (27) is added to the general mean-variance program in (25), the solution must be performed numerically using a non-linear optimization algorithm.

Tables 7 (with *tc* = 0) and 8 (with *tc* = 0.1%) report backtest results for optimal asset allocation strategies. Two results emerge with strength. First, even when models cannot improve the OOS forecasting accuracy for a majority of underlying commodities, the resulting predictions may generate positive, high economic value when combined in an optimal portfolio strategy. Second, transaction costs matter, in the sense that the best performing model(s) in Table 8 sensibly differs from the one(s) in Table 7. However, in both cases, in the metrics that matter the most—i.e., realized MV utility and Sharpe ratios—it remains the case that an AR(1) model that was practically very hard to beat in Tables 1-6, turns out to be regularly dominated by other models. This result admits only one exception: when transaction costs are not accounted for, the AR(1) model yields the lowest realized portfolio standard deviation, i.e., it represents the least risky strategy that however exploits predictability. Unfortunately, AR(1) also tends to return negative or very modest realized mean returns.

In more detail, Table 7 shows that a stepwise, forward predictive regression that includes all commodity specific factors and—if required by the recursive AIC—also macro and aggregated commodity factors, always delivers the highest realized MV objective across all choices for the risk aversion parameter. In the case of $\gamma = 0.1$ and 0.5 this is also the case in a Sharpe ratio dimension, while for $\gamma = 0.25$, the best performing model obtains under HMM when the predictors are macro PCs and aggregated commodity factors. In general, the outperformance of the former model obtains because the resulting portfolios deliver the highest realized means.[26] For instance, when $\gamma = 0.5$, the best stepwise regression delivers an annualized Sharpe ratio of 0.355 which outperforms the 0.245 delivered by the strongest HMM framework, while the benchmark returns a negative Sharpe ratio; such a statistic derives from an annualized mean of approximately 4.4% and a standard deviation of 12.5%, which of course compares favourably with the benchmark that leads to a negative mean performance. The implied certainty equivalent return (CER)—often interpreted in the literature, see e.g., Rapach and Zhou (2013)—of switching from the benchmark to the best performing stepwise model is massive, 4.3% per year; the analogue CER for switching from the best performing HMM to the stepwise regression model is instead another hefty 2%. Interestingly, in Tables 1-4, these models and in particular the forward stepwise one were hardly ever leading to the most accurate forecasts among the available models. This represents an additional—because coming from the commodity return predictability space—powerful warning of the fact that statistical forecasting power should not be confused with the ability to return positive economic value. Tables A4-A8 (panels A of each) in an on-line Appendix show that the best performing model in Table 7 is on average mostly invested (between 80 and 85%) in US 10-year Treasury notes while the stock market weight is on average negative, although this hides considerable variation. Of course, this is relatively unsurprising as between 2004 and 2018, government bonds have yielded very high average excess returns, both because long-term rates have dropped to levels below 3% on several occasions, and because short-term rates have been practically zero between 2009 and 2015. The position in commodity futures are therefore long in net terms on average, in the order of 14-18% of the total, but with a few notable short positions (silver, soybeans, wheat) the go to finance even larger long positions, occasionally exceeding 5% of the total (corn, gold, live cattle), which is rather massive in relative terms. The position in cash is instead rather limited (5% at most), also as a result of the stabilization/diversification benefits of commodities.

Table 8 reports results structured in the same way as we do in Table 7, when sensible transaction costs are imposed. Interestingly, the simple two-parameter AR(1) benchmark loses any specialty and three models emerge to our attention. On the one hand, and across all values for γ, a forward stepwise algorithm based on macro PCs and individual commodity-specific factors always returns the lowest realized portfolio volatility. On the other hand, for $\gamma = 0.1$ and 0.25, the model generating the top

[26] Such a misalignment between MV and Sharpe ratio maximization is clearly possible because while ex-ante the two-fund separation theorem guarantees that these two operations lead to identical results, on an ex-post, realized basis, optimizing $\mu_{p,t+1} - \frac{\gamma}{2}\sigma^2_{P,t+1}$ subject to constraints that involve the riskless asset vs. $(\mu_{p,t+1} - r^f_{t+1})/\sqrt{\sigma^2_{P,t+1}}$ will deliver different results.

realized Sharpe ratio and MV utility is a stepwise, forward regression that includes the macro PCs and aggregate commodity factors when the latter are always included. This represents the strongest evidence, which is otherwise rather mild, in favour of commodity factors reported in our paper.[27] For instance, when $\gamma = 0.25$, this model delivers an annualized Sharpe ratio of 0.665 which outperforms the 0.336 delivered by the best HMM, while the benchmark AR(1) returns a Sharpe ratio of 0.296; such a statistic derives from an annualized mean of 7.8% and a standard deviation of 11.7%; the implied certainty equivalent return (CER) of switching from the benchmark to the best performing stepwise model is another massive 6.2% per year; the corresponding CER for switching from the best performing HMM to the stepwise regression model is an equally sizeable 4.4%. Finally, we note that when compared to Table 7, in Table 8 most statistics concerning the realized performances increase: this is a result of the fact that transaction costs are applied on an ex-ante basis and therefore they prevent the implementation of portfolio strategies that would fail being value-enhancing.[28]

6. Conclusions

In this paper, we have performed a systematic, out-of-sample comparison of the performance of two general sets of forecasting methods, i.e., hidden Markov model (often known as Markov switching models) vs. stepwise predictive regressions, with reference to an important application to commodity futures returns. Loosely speaking, both HMM and stepwise methods represent variable selection methods, even though HMM simply allows coefficients to take different vales (zero is included) in different regimes. Both forecasting approaches are combined with dimension reduction techniques used to summarize the information content of as many as 132 macroeconomic variables in a much smaller set of representative principal component factors. These macro factors are supplemented with a standard set of commodity factors (hedging pressure, basis, net trading, and momentum) implemented in two ways, at the aggregated level (as in Guidolin and Pedio, 2018) but also the individual commodity level.

---

[27] However, when $\gamma = 0.5$, the highest economic value is returned by a rather different model, i.e., a backward stepwise regression that includes macro PCs and aggregate commodity-specific factors. Even though all models tend to yield much higher Sharpe ratios when risk aversion is high and this results from the fact that the resulting portfolios imply low realized volatilities, the intuition for this switch in top performance is elusive.

[28] Interestingly, an on-line Appendix shows that when transaction costs are brought into the picture, the resulting optimal asset allocations turn less extreme in the sense that the demands for long-term government bond declines to between 70 and 85% while a positive weight to equities appears. As far as commodity futures are concerned, transaction costs reduce the average commitment to many commodities and for them—at least in the case of forward stepwise forecasts—one finds evidence of an interesting strategy long gold and short platinum.

The novelty of our research design is to combine our OOS backtesting efforts under four alternative loss functions. The first two functions are typical of the statistical forecasting literature and consist of the squared and the absolute value losses, which imply that RMSFE and MAFE are used as summaries, respectively. The remaining two loss functions are the realized, portfolio Sharpe ratio and MV utility, when the portfolio is built upon optimal weights computed solving a standard MV portfolio problem. Although the two loss functions are traditional and largely understood, they carry the nature of value functions of the underlying optimization problem and as such they are not only of a more applied, practical nature, but also more complex.

Our key finding is striking: while neither HMM nor stepwise regressions manage to systematically (or even just frequently) outperform a plain vanilla AR benchmark according to RMSFE or MAFE statistical loss functions, they—in particular this is evident in the case of models recursively built going "simple to general" to avoid over-parameterizations—create economic value in OOS MV portfolio tests. Interestingly, because we impose transaction costs not only ex post but also ex ante, so that an investor will use the forecasts of a model only when they increase expected utility, contrary to naïve priors, the improvement is maximum under plausible, proportional transaction costs.

A number of extensions are of course possible. In this paper, we have not dealt with the performance of regularization methods (such as ridge regressions or, more generally, least absolute shrinkage selection operators, LASSO). On the one hand, as discussed by Ng (2013) the opposition between stepwise regressions based on information criteria and regularization methods is illusory, as all general case relate to their own special cases: stepwise regressions based on information criteria are indeed cases of L_0 norm regularization methods (i.e., when the shrinkage does not depend on the estimated regression coefficients, but only on their number). Moreover, work by Efron et al. (2004) on least angle regressions (LAR, which is a generalization of LASSO) motivates LASSO as a forward stage-wise regression. On the other hand, it would be interesting LASSO-type methods have witnessed a powerful upward trend and their exploration, also within the domain of the financial economics of commodities seems to be highly justified.

References

Abhyankar, A., Basu, D., and Stremme, A. (2012). The optimal use of return predictability: An empirical study. *Journal of Financial and Quantitative Analysis*, 47, 973-1001.

Acharya, V. V., Lochstoer, L. A., and Ramadorai, T. (2013). Limits to arbitrage and hedging: Evidence from commodity markets. *Journal of Financial Economics*, 109, 441-465.

Ahmed, S., and Tsvetanov, D. (2016). The predictive performance of commodity futures risk factors. *Journal of Banking and Finance*, 71, 20–36.

Akaike, H., 1970. Statistical predictor identification. *Annals of Institute of Statistical Mathematics* 22, 203–217.

Bates, J. M., and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20, 451-468.

Bakshi, G ., Gao, X., e Rossi, A. (2017). Understanding the sources of risk underlying the cross section of commodity returns. *Management Science*, 65, 619-641.

Basu, D., and Miffre, J. (2013). Capturing the risk premium of commodity futures: The role of hedging pressure. *Journal of Banking and Finance*, 37, 2652-2664.

Bessembinder, H., and Chan, K. (1992). Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics*, 32, 169-193.

Bodie, Z., and Rosansky, V. I. (1980). Risk and return in commodity futures. *Financial Analysts Journal*, 36, 27-39.

Bossaerts, P., and Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies*, 12, 405-428.

Breeden, D. T. (1980). Consumption risk in futures markets. *Journal of Finance*, 35, 503-520.

Butler, R. W. (1984). The significance attained by the best-fitting regressor variable. *Journal of the American Statistical Association*, 79, 341-348.

Cano, G. (2019). What market volatility has meant for factors. MSCI Research (https://www.msci.com/www/blog-posts/what-market-volatility-has/01253098146).

Cavanaugh, J., 1997. Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics and Probability Letters*, 33, 201–208.

Cenesizoglu, T., and Timmermann, A. (2012). Do return prediction models add economic value?. *Journal of Banking and Finance*, 36, 2974-2987.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society: Series A*, 158, 419-444.

Chong, J., and Miffre, J. (2010). Conditional correlation and volatility in commodity futures and traditional asset markets. *Journal of Alternative Investments*, 12, 61-75.

Cochrane, J. H. (2009). *Asset pricing*: Revised edition. Princeton University Press.

Dal Pra, G., Guidolin, M., Pedio, M., e Vasile, F. (2018). Regime shifts in excess stock return predictability: an out-of-sample portfolio analysis. *Journal of Portfolio Management*, 40, 10-24.

Dangl, T., and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106, 157-181.

Daskalaki, C., e Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all? New evidence. *Journal of Banking and Finance*, 35, 2606–2626.

Daskalaki, C., Kostakis, A., and Skiadopoulos, G. (2013). Are there common factors in individual commodity futures returns? *Journal of Banking and Finance*, 40, 346–363.

Date, P., Mamon, R., and Tenyakov, A. (2013). Filtering and forecasting commodity futures prices under an HMM framework. *Energy Economics*, 40, 1001-1013.

de Groot, W., Karstanje, D., and Zhou W. (2014). Exploiting commodity momentum along the futures curves. *Journal of Banking and Finance*, 48, 79–93.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22, 1915-1953.

de Roon, F.A., Nijman, T.E., and Veld, C. (2000). Hedging pressure effects in futures markets. *Journal of Finance*, 55, 1437–1456.

Efron, B., Hastie, T., Johnstone, and I., Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*. 32, 407–499.

Fuertes, A. M., Miffre, J., and Fernandez-Perez, A. (2015). Commodity strategies based on momentum, term structure, and idiosyncratic volatility. *Journal of Futures Markets*, 35, 274-297.

Gao, X., & Nardari, F. (2018). Do commodities add economic value in asset allocation? New evidence from time-varying moments. Journal of Financial and Quantitative Analysis, 53(1), 365-393.

Gargano, A., and Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting*, 30, 825–843.

Geweke, J., and Meese, R. (1981). Estimating regression models of finite but unknown order. *International Economic Review* 23, 55–70.

Giampietro, M., Guidolin, M., and Pedio, M. (2018). Estimating stochastic discount factor models with hidden regimes: Applications to commodity pricing. *European Journal of Operational Research*, 265, 685-702.

Gorton, G., and Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, 62, 47-68.

Gorton, G. B., Hayashi, F., and Rouwenhorst, K. G. (2008). The fundamentals of commodity futures returns. *Review of Finance*, 17, 35-105.

Gospodinov, N., and Ng, S. (2013). Commodity prices, convenience yields, and inflation. *Review of Economics and Statistics*, 95, 206-219.

Grechanovsky, E., and Pinsker, I. (1995). Conditional p-values for the F-statistic in a forward selection procedure. *Computational Statistics and Data Analysis*, 20, 239-263.

Guidolin, M., and Pedio, M. (2018). Forecasting commodity futures returns: an economic value analysis of macroeconomic vs. specific factors. IGIER Bocconi Working Paper No. 86.

Hamilton, J. D., and Wu, J. C. (2015). Effects of index-fund investing on commodity futures prices. *International Economic Review*, 56, 187-205.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99, 560-580.

Henriksen, T. E. S., Pichler, A., Westgaard, S., and Frydenberg, S. (2018). Can commodities dominate stock and bond portfolios? *Annals of Operations Research*, 282, 155-177.

Hirshleifer, D. (1988). Residual risk, trading costs, and commodity futures risk premia. *Review of Financial Studies*, 1, 173-193.

Ing,C., and Wei,C. (2003). On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis*, 85, 130–155.

Kang, W., Rouwenhorst, K. G., and Tang, K. (2018). "A tale of two premiums: the role of hedgers and speculators in commodity futures markets". Yale International Center for Finance, WP No. 14-24.

Khan, S., Khokher Z., and Simin T. (2008). Expected commodity futures returns. Working paper, North Carolina State University.

Kiers, H. A., and Smilde, A. K. (2007). A comparison of various methods for multivariate regression with highly collinear variables. *Statistical Methods and Applications*, 16, 193-228.

Lee, S., and Karagrigoriou, A. (2001). An asymptotically optimal selection of the order of a linear process. *Sanklya, The Indian Journal of Statistics*, Series A, 63, 93-106.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review*, 81, 580-590.

Lombardi, M. J., and Ravazzolo, F. (2016). On the correlation between commodity and equity returns: implications for portfolio allocation. *Journal of Commodity Markets*, 2, 45-57.

Ludvigson, S. C., and Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, 22, 5027-5067.

Mei, X., and Nogales, F. J. (2018). Portfolio selection with proportional transaction costs and predictability. *Journal of Banking and Finance*, 94, 131-151.

Ng, S. (2013). Variable selection in predictive regressions. In *Handbook of Economic Forecasting*, G. Elliott, C. Granger, and A. Timmermann (eds.), vol. 2, pp. 752-789, Elsevier.

Panella, M., Barcellona, F., and D'ecclesia, R. L. (2012). Forecasting energy commodity prices using neural networks. *Advances in Decision Sciences*, 2012.

Papanicolaou, A., and Sircar, R. (2014). A regime-switching Heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance*, 14, 1811-1827.

Phillips, P. C., and Ploberger, W. (1996). An asymptotic theory of Bayesian inference for time series. *Econometrica*, 64, 381-412.

Rapach, D., E., and Zhou, G. (2013). Forecasting stock returns. *Handbook of Economic Forecasting*, G. Elliott, C. Granger, and A. Timmermann (eds.), vol. 2, 328-383, Elsevier.

Rencher, A. C., and Pun, F. C. (1980). Inflation of R2 in best subset regression. *Technometrics*, 22, 49-53.

Rissanen, J. (1986a). Modeling the shortest data description. *Automatica*, 14, 465–471.

Rissanen, J. (1986b). A predictive least squares principle. *IMA Journal of Mathematics Control Information*, 3, 211–222.

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

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criteria. *Biometrika*, 63, 117–126.

Shibata, R. (1980). Asymptotic efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, 8, 147–164.

Shibata, R. (1984). Approximate efficiency of a selection procedure for the number of regression variables. *Biometrika*, 71, 43–49.

Smith, G. (2018). Step away from stepwise. *Journal of Big Data*, 5:32.

Stock, J. H., and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, 1167-1179.

Stock, J. H., and Watson, M. W. (2006). Forecasting with many predictors. *Handbook of Economic Forecasting*, G. Elliott, C. Granger, and A. Timmermann (eds.), vol. 1, 515-554, Elsevier.

Speed, T., and Yu, B. (1993=. Model selection and prediction: normal regression. *Annals of Institute of Statistical Mathematics*, 45 (1), 35–54.

Stine, R., 2004. Model selection using information theory and the MDL principle. *Sociological Methods and Research*, 33, 230–260.

Tang, K., and Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, 68, 5, 54-74.

Timmermann, A. (2006). Forecast combinations. In: *Handbook of Forecasting,* vol. 1. Elsevier, Amsterdam, pp. 135–196.

Tipping, M. E., and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B*, 61, 611-622.

Wei, C. (1992). On predictive least squares principle. *Annals of Statistics*, 20, 1-42.

Wilkinson, L., and Dallal, G. E. (1981). Tests of significance in forward selection regression with an F-to-enter stopping rule. *Technometrics*, 23, 377-380.

Yan, L., and Garcia, P. (2017). Portfolio investment: Are commodities useful? *Journal of Commodity Markets*, 8, 43-55.

Yang, F. (2013). Investment shocks and the commodity basis spread. *Journal of Financial Economics*, 110, 164-184.

Table 1

Root Mean Square Error Forecast Performances – Classification by Low vs. High Risk Aversion Samples

The table reports the RMSFE implied by the forecasts obtained from the 12 stepwise (forward vs. backward) predictive regression models recursively implemented in our analysis. In particular, the table distinguishes between the RMSFE obtained in periods of low vs. high market volatility (a proxy for aggregate risk aversion). Specifically, the periods of low risk aversion are identified with January 2004 – December 2007 and January 2013 – May 2018, while the high risk aversion sub-sample refers to Januart 2008 - December 2012. For both the low and high risk-aversion sub-samples, we have boldfaced the forecasting model that yields the minimum RMSFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Baseline	**Kitchen Sink Predictive Regression**													
Low	0.0798	0.0913	0.0787	0.1041	0.0865	0.0759	0.0913	**0.0644**	0.0545	0.0838	0.0821	0.0948	0.0985	0.0964
High	0.1396	0.1420	0.1176	0.2111	0.1587	0.1523	0.1515	0.1471	0.0875	0.1710	0.3360	0.1096	0.1755	0.1403
Panel A	**Macro Principal Components Only**													
Forward - Low	0.0768	0.0748	0.0764	0.0825	0.0918	**0.0703**	**0.0707**	0.0698	0.0454	0.0801	0.0786	**0.0842**	**0.0755**	0.0758
Forward - High	0.1142	0.1324	0.1149	0.1264	0.1006	0.1214	0.1113	0.1125	0.0655	0.1457	0.1498	0.1043	0.0973	0.0925
Backward - Low	0.0775	0.0749	0.0762	**0.0823**	0.0902	0.0703	0.0709	0.0695	0.0455	0.0820	0.0785	0.0842	0.0755	0.0752
Backward - High	0.1166	0.1407	0.1146	0.1286	0.1023	0.1620	0.1125	0.1139	0.0726	0.1458	0.1488	0.1057	0.1130	0.0926
Panel B	**Macro Principal Components + Commodity-Specific Factors (Always Included)**													
Forward - Low	0.0823	0.0741	0.0788	0.0862	0.0990	0.0791	0.0755	0.0661	0.0493	0.0905	0.0869	0.0946	0.0813	0.1039
Forward - High	0.1182	0.1565	0.1140	0.1549	0.1339	0.1609	0.1180	0.1290	0.0824	0.1766	0.8282	0.1151	0.1247	0.1354
Backward - Low	0.0822	0.0736	0.0804	0.0861	0.1003	0.0798	0.0752	0.0694	0.0481	0.0912	0.0938	0.0944	0.0801	0.0984
Backward - High	0.1197	0.1780	0.1298	0.1666	0.1244	0.1659	0.1258	0.1414	0.0849	0.1768	0.7025	0.1270	0.1523	0.1538
Panel C	**Macro Principal Components + Commodity-Specific Factors**													
Forward - Low	0.0707	**0.0725**	0.0773	0.0846	0.0917	0.0728	0.0720	0.0682	0.0472	0.0829	0.0833	0.0900	0.0797	0.0833
Forward - High	0.0648	0.1526	0.1029	0.1468	0.1159	0.1898	0.1126	0.1214	0.0762	0.1510	0.1539	0.1086	0.1182	0.1117
Backward - Low	**0.0696**	0.0731	0.0780	0.0841	0.0925	0.0746	0.0717	0.0693	0.0471	0.0836	0.0840	0.0946	0.0791	0.0836
Backward - High	**0.0634**	0.1598	0.1186	0.1632	0.1248	0.1616	0.1120	0.1228	0.0825	0.1707	0.7496	0.1171	0.1448	0.1228
Panel D	**Macro Principal Components + Aggregated Commodity-Specific Factors (Always Included)**													
Forward - Low	0.0851	0.0775	0.0894	0.0877	0.0971	0.0750	0.0788	0.0759	0.0661	0.0836	0.0792	0.0964	0.0852	0.1154
Forward - High	0.2179	0.1829	0.1810	0.2229	0.1574	0.1652	0.2498	0.1851	0.0958	0.2103	0.2792	0.1316	0.1854	0.1140
Backward - Low	0.0846	0.0780	0.0786	0.0873	0.0960	0.0750	0.0793	0.0767	0.0603	0.0837	0.0792	0.0960	0.0858	0.0909
Backward - High	0.2113	0.1822	0.1741	0.2233	0.1614	0.1549	0.2234	0.2726	0.1169	0.2345	0.2644	0.1312	0.1884	0.1139
Panel E	**Macro Principal Components + Aggregated Commodity-Specific Factors**													
Forward - Low	0.0789	0.0786	**0.0757**	0.0852	0.0929	0.0750	0.0744	0.0724	**0.0448**	**0.0796**	0.0782	0.0959	0.0807	0.0770
Forward - High	0.1265	0.1593	0.1107	0.1990	0.1444	0.1374	0.1220	0.1157	0.0761	0.1428	0.1701	0.1205	0.1212	0.0969
Backward - Low	0.0795	0.0780	0.0767	0.0864	0.0909	0.0735	0.0746	0.0734	0.0458	0.0802	0.0782	0.0951	0.0850	0.0760
Backward - High	0.1221	0.1589	0.1395	0.2050	0.1529	0.1340	0.1521	0.1691	0.0813	0.1700	0.1750	0.1233	0.1307	0.0979
Panel F	**Macro Principal Components + Commodity-Specific Factors + Aggregated Commodity-Specific Factors (Always Included)**													
Forward - Low	0.0848	0.0795	0.0773	0.0945	0.1041	0.0815	0.0844	0.0746	0.0475	0.0953	0.0827	0.1038	0.0884	0.0855
Forward - High	0.2278	0.1739	0.1580	0.2322	0.1471	0.1722	0.2192	0.2086	0.1152	0.1963	0.5971	0.1322	0.1962	0.1467
Backward - Low	0.0862	0.0788	0.0801	0.0939	0.1064	0.0838	0.0850	0.0738	0.0481	0.1841	0.0821	0.1085	0.0882	0.0865
Backward - High	0.2264	0.1866	0.1607	0.2233	0.1663	0.1754	0.2242	0.3524	0.1360	0.2070	0.5627	0.1742	0.2487	0.1956
Panel G	**Benchmark AR(1)**													
AR(1)	0.0856	**0.0867**	**0.0792**	**0.0946**	**0.0882**	**0.0799**	**0.0861**	**0.0800**	**0.0522**	**0.0955**	**0.0723**	**0.0916**	**0.0799**	**0.0463**

Table 2

Root Mean Square Error Forecast Performances – Weighting on the Basis of Predicted Risk-Aversion Regime Probabilities

The table reports the RMSFE implied by the forecasts obtained from the 12 stepwise (forward vs. backward) predictive regression models recursively implemented in our analysis. In particular, the table obtains the forecasts by weighting (over the entire OOS period, January 2004 – May 2018) the regime-specific stepwise regression forecasts by the predicted regime probabilities obtained from a two-state HMM for the VIX index. In the different p anels, we have boldfaced the forecasting model that yields the minimum RMSFE. In the first panel, we underscore the kitchen sink model (including all predictors) that delivers the lowest RMSFE when the HMM implementation is compared with the single-state one (linear predictive model).

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
	Kitchen Sink Predictive Regression													
Baseline	*0.0878*	0.1166	0.0981	0.1576	0.1226	0.1141	0.1214	0.1058	0.0710	0.1274	0.2090	*0.1022*	0.1370	0.1183
HMM	0.0924	*0.1123*	*0.0958*	*0.1160*	***0.0856***	*0.1049*	*0.0923*	*0.0955*	*0.0673*	***0.0913***	*0.1532*	0.1060	*0.0810*	*0.1033*
Panel A	**Macro Principal Components Only**													
Forward	0.0830	0.0920	0.0846	0.0973	0.0925	0.0836	**0.0858**	0.0821	**0.0516**	0.0962	0.0832	**0.0911**	0.0824	0.0829
Backward	0.0817	0.0918	0.0845	0.0976	0.0916	0.0847	0.0860	0.0820	0.0531	0.0972	0.0830	0.0913	0.0843	0.0825
Panel B	**Macro Principal Components + Commodity-Specific Factors (Always Included)**													
Forward	0.0841	0.0956	0.0838	0.1019	0.0981	0.0926	0.0880	0.0824	0.0546	0.1032	0.2508	0.0975	0.0901	0.0971
Backward	0.0840	0.0936	0.0867	0.1032	0.0974	0.0957	0.0887	0.0860	0.0545	0.1035	0.2128	0.0978	0.0909	0.0977
Panel C	**Macro Principal Components + Commodity-Specific Factors**													
Forward	0.0733	0.0916	0.0824	0.1019	0.0938	0.0899	0.0862	0.0818	0.0546	0.0972	0.0853	0.0943	0.0883	0.0869
Backward	**0.0723**	0.0916	0.0849	0.1023	0.0934	0.0918	0.0862	0.0830	0.0536	0.0993	0.2267	0.0973	0.0892	0.0881
Panel D	**Macro Principal Components + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0989	0.0982	0.1038	0.1159	0.0996	0.0871	0.1144	0.0931	0.0662	0.1144	0.1080	0.1008	0.1032	0.1051
Backward	0.0980	0.0984	0.0971	0.1159	0.0996	0.0858	0.1100	0.1151	0.0696	0.1146	0.1041	0.1000	0.1036	0.0922
Panel E	**Macro Principal Components + Aggregated Commodity-Specific Factors**													
Forward	0.0826	0.0950	0.0822	0.1131	0.0969	0.0871	0.0890	0.0824	0.0524	0.0972	0.0896	0.0994	0.0876	0.0842
Backward	0.0802	0.0951	0.0887	0.1162	0.0953	0.0831	0.0944	0.0874	0.0541	0.0978	0.0891	0.0986	0.0907	0.0837
Panel F	**Macro Principal Components + Commodity-Specific Factors + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0986	0.0963	0.0911	0.1185	0.0987	0.0914	0.1151	0.0945	0.0590	0.1096	0.1896	0.1037	0.1062	0.0914
Backward	0.0957	0.1001	0.0937	0.1193	0.1029	0.0934	0.1108	0.1347	0.0674	0.1600	0.1939	0.1116	0.1146	0.0990
Panel G	**Benchmark AR(1)**													
AR(1)	0.0856	**0.0867**	**0.0792**	**0.0946**	0.0882	**0.0799**	0.0861	**0.0800**	0.0562	0.0955	**0.0723**	0.0916	**0.0799**	**0.0463**

Table 3

Root Mean Square Error Forecast Performances – Hidden Markov Chain vs. Stepwise Regression Models

The table reports the RMSFE implied by the forecasts obtained from the three HMM and the 12 stepwise (forward vs. backward) predictive regression models recursively implemented in our analysis. In particular, the table obtains the forecasts by weighting (over the entire OOS period, January 2004 – May 2018) the regime-specific stepwise regression forecasts by the predicted regime probabilities obtained from a two-state HMM for the VIX index. In the different p anels, we have boldfaced the forecasting model that yields the minimum RMSFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Panel A	**HMM - Macro Principal Components Only**													
	0.1014	0.0970	0.1021	0.1069	0.2446	0.0909	0.1036	0.0958	0.1088	0.1215	0.0968	0.1213	0.0857	0.0472
Panel B	**HMM - Maco Principal Components + Commodity-Specific Factors**													
	0.0871	0.1025	0.1430	0.1280	0.1153	0.1151	0.1040	0.0966	0.0989	0.3136	0.0921	0.1084	0.1074	0.0515
Panel C	**HMM - Macro Principal Components + Aggregated Commodity-Specific Factors**													
	0.0924	0.1123	0.0958	0.1160	**0.0856**	0.1049	0.0923	0.0955	0.0673	**0.0913**	0.1532	0.1060	0.0810	0.1033
Panel D	**Stepwise Regression - Macro Principal Components Only**													
Forward	0.0830	0.0920	0.0846	0.0973	0.0925	0.0836	**0.0858**	0.0821	**0.0516**	0.0962	0.0832	**0.0911**	0.0824	0.0829
Backward	0.0817	0.0918	0.0845	0.0976	0.0916	0.0847	0.0860	0.0820	0.0531	0.0972	0.0830	0.0913	0.0843	0.0825
Panel E	**Stepwise Regression - Maco Principal Components + Commodity-Specific Factors (Always Included)**													
Forward	0.0841	0.0956	0.0838	0.1019	0.0981	0.0926	0.0880	0.0824	0.0546	0.1032	0.2508	0.0975	0.0901	0.0971
Backward	0.0840	0.0936	0.0867	0.1032	0.0974	0.0957	0.0887	0.0860	0.0545	0.1035	0.2128	0.0978	0.0909	0.0977
Panel F	**Stepwise Regression - Macro Principal Components + Commodity-Specific Factors**													
Forward	0.0733	0.0916	0.0824	0.1019	0.0938	0.0899	0.0862	0.0818	0.0546	0.0972	0.0853	0.0943	0.0883	0.0869
Backward	**0.0723**	0.0916	0.0849	0.1023	0.0934	0.0918	0.0862	0.0830	0.0536	0.0993	0.2267	0.0973	0.0892	0.0881
Panel G	**Stepwise Regression - Principal Components + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0989	0.0982	0.1038	0.1159	0.0996	0.0871	0.1144	0.0931	0.0662	0.1144	0.1080	0.1008	0.1032	0.1051
Backward	0.0980	0.0984	0.0971	0.1159	0.0996	0.0858	0.1100	0.1151	0.0696	0.1146	0.1041	0.1000	0.1036	0.0922
Panel H	**Stepwise Regression - Macro Principal Components + Aggregated Commodity-Specific Factors**													
Forward	0.0826	0.0950	0.0822	0.1131	0.0969	0.0871	0.0890	0.0824	0.0524	0.0972	0.0896	0.0994	0.0876	0.0842
Backward	0.0802	0.0951	0.0887	0.1162	0.0953	0.0831	0.0944	0.0874	0.0541	0.0978	0.0891	0.0986	0.0907	0.0837
Panel I	**Stepwise Regression - Principal Components + Commodity-Specific Factors + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0986	0.0963	0.0911	0.1185	0.0987	0.0914	0.1151	0.0945	0.0590	0.1096	0.1896	0.1037	0.1062	0.0914
Backward	0.0957	0.1001	0.0937	0.1193	0.1029	0.0934	0.1108	0.1347	0.0674	0.1600	0.1939	0.1116	0.1146	0.0990
Panel J	**Benchmark AR(1)**													
AR(1)	0.0856	**0.0867**	**0.0792**	**0.0946**	0.0882	**0.0799**	0.0861	**0.0800**	0.0522	0.0955	**0.0723**	0.0916	**0.0799**	**0.0463**

Table 4

Mean Absolute Error Forecast Performances – Hidden Markov Chain vs. Stepwise Regression Models

The table reports the MAFE implied by the forecasts obtained from the three HMM and the 12 stepwise (forward vs. backward) predictive regression models recursively implemented in our analysis. In particular, the table obtains the forecasts by weighting (over the entire OOS period, January 2004 – May 2018) the regime-specific stepwise regression forecasts by the predicted regime probabilities obtained from a two-state HMM for the VIX index. In the different p anels, we have boldfaced the forecasting model that yields the minimum MAFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Panel A	**HMM - Macro Principal Components Only**													
	0.0691	0.0710	0.0673	0.0766	0.0899	0.0675	0.0786	0.0691	0.0511	0.0809	0.0588	0.0833	0.0656	0.0359
Panel B	**HMM - Macro Principal Components + Commodity-Specific Factors**													
	0.0656	0.0718	0.0757	0.0875	0.0862	0.0767	0.0781	0.0697	0.0510	0.1198	0.0611	0.0825	0.0763	0.0409
Panel C	**HMM - Principal Components + Aggregated Commodity-Specific Factors**													
	0.0913	**0.0579**	**0.0509**	0.0861	**0.0641**	0.0881	0.0719	0.0627	0.0442	**0.0667**	0.0619	**0.0701**	0.0735	0.0405
Panel D	**Stepwise Regression - Macro Principal Components Only**													
Forward	0.0653	0.0699	0.0628	0.0734	0.0700	0.0658	0.0678	0.0627	**0.0404**	0.0754	0.0672	0.0732	0.0645	0.0656
Backward	0.0638	0.0699	0.0626	0.0735	0.0692	0.0660	0.0684	0.0628	0.0410	0.0757	0.0671	0.0733	0.0656	0.0652
Panel E	**Stepwise Regression - Macro Principal Components + Commodity-Specific Factors (Always Included)**													
Forward	0.0662	0.0711	0.0653	0.0759	0.0773	0.0709	0.0694	0.0619	0.0417	0.0815	0.1064	0.0781	0.0704	0.0752
Backward	0.0664	0.0697	0.0676	0.0762	0.0765	0.0733	0.0696	0.0642	0.0413	0.0822	0.1009	0.0784	0.0716	0.0748
Panel F	**Stepwise Regression - Macro Principal Components + Commodity-Specific Factors**													
Forward	0.0585	0.0692	0.0643	0.0751	0.0705	0.0680	0.0686	**0.0617**	0.0421	0.0764	0.0688	0.0751	0.0693	0.0687
Backward	**0.0571**	0.0691	0.0657	0.0748	0.0705	0.0696	0.0689	0.0624	0.0416	0.0782	0.1003	0.0784	0.0701	0.0696
Panel G	**Stepwise Regression - Macro Principal Components + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0706	0.0733	0.0715	0.0793	0.0740	0.0681	0.0839	0.0685	0.0485	0.0818	0.0748	0.0791	0.0732	0.0814
Backward	0.0696	0.0735	0.0689	0.0792	0.0739	0.0683	0.0813	0.0729	0.0482	0.0814	0.0736	0.0787	0.0731	0.0731
Panel H	**Stepwise Regression - Macro Principal Components + Aggregated Commodity-Specific Factors**													
Forward	0.0650	0.0721	0.0631	0.0783	0.0716	0.0694	0.0704	0.0636	**0.0404**	0.0756	0.0696	0.0782	0.0681	0.0662
Backward	0.0642	0.0719	0.0660	0.0808	0.0706	0.0669	0.0732	0.0656	0.0417	0.0755	0.0707	0.0780	0.0701	0.0660
Panel I	**Stepwise Regression - Macro Principal Components + Commodity-Specific Factors + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0711	0.0730	0.0674	0.0857	0.0743	0.0719	0.0863	0.0697	0.0434	0.0856	0.0928	0.0824	0.0754	0.0706
Backward	0.0696	0.0735	0.0701	0.0833	0.0789	0.0732	0.0848	0.0732	0.0449	0.0928	0.0896	0.0860	0.0791	0.0750
Panel J	**Benchmark AR(1)**													
AR(1)	0.0682	0.0676	0.0603	**0.0711**	0.0647	**0.0635**	**0.0674**	**0.0610**	0.0408	0.0748	**0.0492**	0.0727	**0.0635**	**0.0356**

Table 5

Root Mean Square Error Forecast Performances –Markov Switching vs. Stepwise Regression Models, Equally Weighted

The table compares the RMSFE stastics implied by an equally-weighted average of the forecasts obtained from the three HMM and the 24 stepwise (forward and backward) predictive regression models recursively implemented in our analysis. In particular, the table obtains the forecasts by weighting (over the entire OOS period, January 2004 – May 2018) the regime-specific stepwise regression forecasts by the predicted regime probabilities obtained from a two-state HMM for the VIX index. In the different panels, we have boldfaced the forecasting model that yields the minimum RMSFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Panel A	**HMM**													
	0.0936	0.1039	0.1136	0.1170	0.1485	0.1036	0.1000	0.0960	0.0916	0.1754	0.1141	0.1119	0.0914	0.0673
Panel B	**Stepwise Regression Models**													
	0.0860	0.0949	0.0886	0.1086	0.0966	0.0889	0.0962	0.0920	0.0576	0.1075	0.1430	0.0986	0.0943	0.0909
Panel C	**Benchmark AR(1)**													
AR(1)	**0.0856**	**0.0867**	**0.0792**	**0.0946**	**0.0882**	**0.0799**	**0.0861**	**0.0800**	**0.0522**	**0.0955**	**0.0723**	**0.0916**	**0.0799**	**0.0463**

Table 6

Mean Absolute Error Forecast Performances –Markov Switching vs. Stepwise Regression Models, Equally Weighted

The table compares the Mfe stastics implied by an equally-weighted average of the forecasts obtained from the three HMM and the 24 stepwise (forward and backward) predictive regression models recursively implemented in our analysis. In particular, the table obtains the forecasts by weighting (over the entire OOS period, January 2004 – May 2018) the regime-specific stepwise regression forecasts by the predicted regime probabilities obtained from a two-state HMM for the VIX index. In the different p anels, we have boldfaced the forecasting model that yields the minimum RMSFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Panel A	**HMM**													
	0.0753	**0.0669**	0.0646	0.0834	0.0801	0.0774	0.0762	0.0672	0.0488	0.0891	0.0606	0.0787	0.0718	0.0391
Panel B	**Stepwise Regression Models**													
	0.0656	0.0713	0.0663	0.0780	0.0731	0.0693	0.0744	0.0658	0.0429	0.0802	0.0818	0.0782	0.0709	0.0709
Panel C	**Benchmark AR(1)**													
AR(1)	0.0682	0.0676	**0.0603**	**0.0711**	**0.0647**	**0.0635**	**0.0674**	**0.0610**	**0.0408**	**0.0748**	**0.0492**	**0.0727**	**0.0635**	**0.0356**

Table 7

Realized Portfolio Performance of Mean-Variance Strategies Based on HMM vs. Stepwise Regression Forecasts (No Transaction Costs)

The table reports a range of realized OOS performance measures obtained from recursive mean-variance portfolio optimization strategies over a January 2004 – May 2018 sample. The 16 models reported are obtained as three HMM with regimes plus the best 12 stepwise predictive regressions selected on the basis of their RMSFE, and the AR(1) benchmark. We boldface the highest realized performances.

	$\gamma = 0.1$				$\gamma = 0.25$				$\gamma = 0.5$			
	Mean Return	**Std. Dev.**	**Sharpe Ratio**	**Realized MV**	**Mean Return**	**Std. Dev.**	**Sharpe Ratio**	**Realized MV**	**Mean Return**	**Std. Dev.**	**Sharpe Ratio**	**Realized MV**
HMM Macro Factors	0.0006	0.0198	0.0962	0.0005	0.0001	0.0190	0.0258	0.0001	0.0002	0.0182	0.0328	0.0001
HMM Macro and Commodity-Specific Factors	0.0020	0.0388	0.1813	0.0020	-0.0002	0.0351	-0.0246	-0.0004	0.0019	0.0268	0.2446	0.0017
HMM Macro and Aggregate Commodity Factors	0.0026	0.0429	0.2095	0.0025	0.0020	0.0185	**0.3660**	0.0019	0.0012	0.0213	0.1977	0.0011
Stepwise PCs - Forward	0.0001	0.0181	0.0035	0.0001	-0.0001	0.0183	-0.0109	-0.0001	0.0002	0.0166	0.0318	0.0001
Stepwise PCs - Backard	-0.0003	0.0220	-0.0409	-0.0003	-0.0002	0.0202	-0.0339	-0.0002	0.0002	0.0182	0.0291	0.0001
Stepwise PCs and Commodity-Specific Factors (always incl.) - Forward	**0.0032**	0.0376	**0.2932**	**0.0031**	**0.0032**	0.0369	0.3000	**0.0030**	**0.0037**	0.0361	**0.3548**	**0.0034**
Stepwise PCs and Commodity-Specific Factors (always incl.) - Backward	-0.0025	0.0388	-0.2229	-0.0026	-0.0024	0.0379	-0.2198	-0.0026	-0.0027	0.0381	-0.2442	-0.0030
Stepwise PCs and Commodity-Specific Factors - Forward	-0.0021	0.0291	-0.2505	-0.0021	-0.0005	0.0312	-0.0527	-0.0006	0.0011	0.0246	0.1578	0.0010
Stepwise PCs and Commodity-Specific Factors - Backward	-0.0016	0.0274	-0.1975	-0.0016	-0.0016	0.0271	-0.1988	-0.0016	-0.0017	0.0253	-0.2266	-0.0018
Stepwise PCs and Aggregate Commodity Factors (always incl.) - Forward	-0.0030	0.0334	-0.3088	-0.0030	-0.0024	0.0322	-0.2582	-0.0025	-0.0024	0.0322	-0.2577	-0.0026
Stepwise PCs and Aggregate Commodity Factors (always incl.) - Backward	-0.0020	0.0329	-0.2129	-0.0021	-0.0020	0.0361	-0.1890	-0.0021	-0.0007	0.0322	-0.0796	-0.0010
Stepwise PCs and Aggregate Commodity Factors - Forward	-0.0010	0.0179	-0.2020	-0.0011	-0.0007	0.0184	-0.1408	-0.0008	-0.0008	0.0183	-0.1520	-0.0009
Stepwise PCs and Aggregate Commodity Factors - Backward	0.0002	0.0319	0.0259	0.0002	0.0002	0.0319	0.0203	0.0001	-0.0005	0.0320	-0.0594	-0.0008
Stepwise PCs and Aggregate Commodity-Specific Factors (always incl.) - Forward	0.0019	0.0427	0.1550	0.0018	0.0005	0.0417	0.0400	0.0003	0.0011	0.0337	0.1151	0.0008
Stepwise PCs and Aggregate Commodity-Specific Factors (always incl.) - Backward	-0.0073	0.1213	-0.2095	-0.0081	0.0026	0.0370	0.2408	0.0024	0.0009	0.0341	0.0881	0.0006
AR(1) Benchmark	-0.0001	**0.0154**	-0.0296	-0.0001	-0.0001	**0.0154**	-0.0298	-0.0002	-0.0001	**0.0154**	-0.0302	-0.0002

Table 8

Realized Portfolio Performance of Mean-Variance Strategies Based on HMM vs. Stepwise Regression Forecasts (With Transaction Costs)

The table reports the realized OOS performance measures obtained from recursive mean-variance portfolio optimization strategies over a January 2004 – May 2018 sample. The 16 models reported are obtained as three HMM with regimes plus the best 12 stepwise predictive regressions selected on the basis of their RMSFE, and the AR(1) benchmark. Transaction costs are not only imputed ex-post as a function of the portfolio trades executed but also ex-ante in the sense that only monthly portfolio changes that are expected ex-ante to increase mean-variance utility in excess of the loss caused by the transaction costs are executed. We boldface the highest realized performances.

	$\gamma = 0.1$				$\gamma = 0.25$				$\gamma = 0.5$			
	Mean Return	**Std. Dev.**	**Sharpe Ratio**	**Realized MV**	**Mean Return**	**Std. Dev.**	**Sharpe Ratio**	**Realized MV**	**Mean Return**	**Std. Dev.**	**Sharpe Ratio**	**Realized MV**
HMM Macro Factors	0.0002	0.0193	0.0330	0.0002	-0.0020	0.0218	-0.3101	-0.0020	-0.0016	0.0214	-0.2637	-0.0017
HMM Macro and Commodity-Specific Factors	0.0029	0.0333	0.3007	0.0028	0.0011	0.0211	0.1764	0.0010	0.0012	0.0213	0.1977	0.0012
HMM Macro and Aggregate Commodity Factors	0.0024	0.0473	0.1727	0.0022	0.0027	0.0277	0.3362	0.0026	0.0021	0.0194	0.3697	0.0020
Stepwise PCs - Forward	0.0027	0.0306	0.3008	0.0026	0.0022	0.0306	0.2445	0.0020	0.0021	0.0304	0.2447	0.0019
Stepwise PCs - Backard	0.0026	0.0363	0.2506	0.0026	0.0025	0.0360	0.2388	0.0023	0.0029	0.0346	0.2923	0.0026
Stepwise PCs and Commodity-Specific Factors (always incl.) - Forward	0.0022	0.0541	0.1378	0.0020	0.0030	0.0530	0.1942	0.0026	0.0015	0.0506	0.1035	0.0009
Stepwise PCs and Commodity-Specific Factors (always incl.) - Backward	-0.0015	0.0352	-0.1495	-0.0016	-0.0010	0.0341	-0.1051	-0.0012	-0.0004	0.0347	-0.0373	-0.0007
Stepwise PCs and Commodity-Specific Factors - Forward	0.0011	**0.0089**	0.4356	0.0011	0.0010	**0.0082**	0.4170	0.0010	0.0010	**0.0083**	0.4304	0.0010
Stepwise PCs and Commodity-Specific Factors - Backward	0.0017	0.0248	0.2440	0.0017	0.0029	0.0214	0.4731	0.0029	0.0011	0.0217	0.1726	0.0010
Stepwise PCs and Aggregate Commodity Factors (always incl.) - Forward	0.0013	0.0213	0.2140	0.0013	0.0009	0.0240	0.1344	0.0009	0.0009	0.0240	0.1266	0.0007
Stepwise PCs and Aggregate Commodity Factors (always incl.) - Backward	-0.0003	0.0219	-0.0431	-0.0003	-0.0003	0.0218	-0.0408	-0.0003	0.0002	0.0238	0.0262	0.0000
Stepwise PCs and Aggregate Commodity Factors - Forward	0.0011	0.0198	0.1996	0.0011	0.0010	0.0194	0.1774	0.0009	0.0008	0.0169	0.1577	0.0007
Stepwise PCs and Aggregate Commodity Factors - Backward	-0.0023	0.0447	-0.1746	-0.0024	-0.0022	0.0446	-0.1690	-0.0024	0.0012	0.0164	0.2497	0.0011
Stepwise PCs and Aggregate Commodity-Specific Factors (always incl.) - Forward	**0.0066**	0.0410	**0.5590**	**0.0065**	**0.0065**	0.0337	**0.6651**	**0.0063**	0.0063	0.0338	0.6452	0.0060
Stepwise PCs and Aggregate Commodity-Specific Factors (always incl.) - Backward	0.0066	0.0410	0.5590	0.0065	0.0065	0.0337	0.6651	0.0063	**0.0073**	0.0325	**0.7738**	**0.0070**
AR(1) Benchmark	0.0011	0.0133	0.2958	0.0011	0.0011	0.0133	0.2962	0.0011	0.0011	0.0132	0.2975	0.0011

Figure 1

The VIX index and the predicted probabilities from a two-state HMM

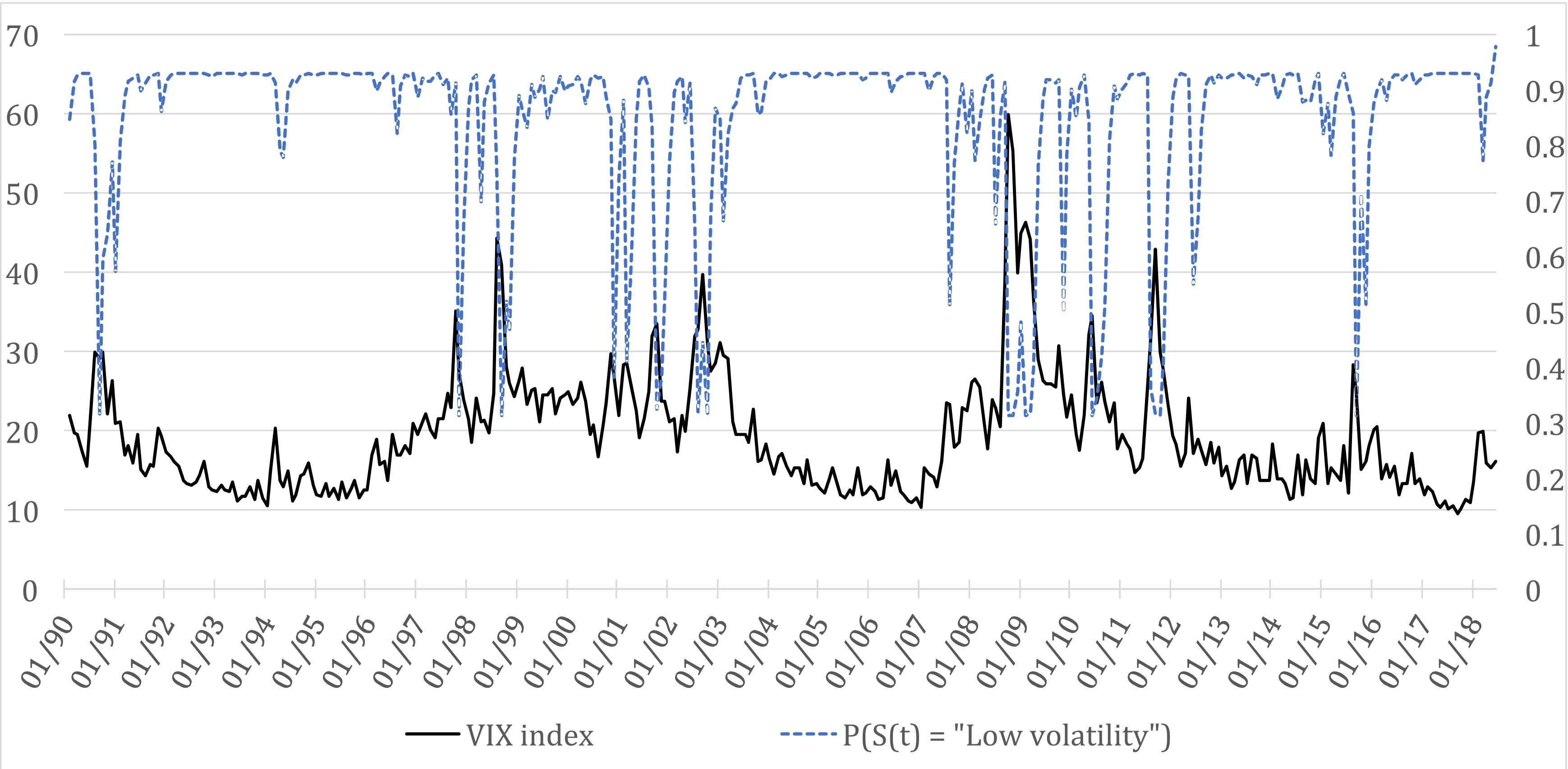

Table A1

Descriptive statistics for commodity futures returns and the factors

Panel A

	Commodity Futures Returns				
	Mean	**Median**	**Standard Deviation**	**Skewness**	**Excess Kurtosis**
Light Crude Oil	0.0076	0.0085	0.0863	0.2767	2.3781
Corn	0.0037	0.0000	0.0745	0.1767	0.7107
Soybeans	0.0030	0.0000	0.0686	-0.2925	1.1918
Wheat	0.0039	-0.0014	0.0808	0.6681	2.6618
Coffee	0.0045	-0.0095	0.1030	1.0986	3.3857
Cocoa	0.0049	0.0000	0.0832	0.4364	0.8958
Sugar	0.0041	-0.0010	0.0860	0.3372	0.9646
Cotton No.2	0.0042	0.0054	0.0756	0.1088	0.4408
Gold	0.0043	0.0003	0.0446	0.1784	1.2249
Silver	0.0060	-0.0017	0.0798	0.1315	0.9535
Platinum	0.0037	0.0032	0.0614	0.2097	7.5443
Orange Juice	0.0041	0.0047	0.0904	0.5444	1.1074
Lumber	0.0062	0.0029	0.0784	0.3930	0.3594
Live Cattle	0.0019	0.0028	0.0446	-0.7534	3.4675

Panel B

	Hedging Pressure				
	Mean	**Median**	**Standard Deviation**	**Skewness**	**Excess Kurtosis**
Light Crude Oil	0.0550	0.0421	0.1002	0.1521	-0.4743
Corn	0.0167	0.0170	0.1322	-0.1640	-0.7475
Soybeans	0.1015	0.1277	0.1633	-0.2676	-0.7006
Wheat	0.0177	-0.0116	0.1555	0.4744	-0.7753
Coffee	0.1338	0.1351	0.1588	-0.0382	-1.0892
Cocoa	0.1344	0.1416	0.1596	-0.1174	-0.5712
Sugar	0.1656	0.1929	0.1888	-0.0703	-0.6391
Cotton No.2	0.0905	0.1114	0.2255	-0.3090	-0.8422
Gold	0.2051	0.3013	0.2801	-0.6206	-0.9246
Silver	0.3971	0.4060	0.1561	-0.1283	-0.5484
Platinum	0.4553	0.5010	0.2346	-0.6685	-0.3273
Orange Juice	0.2035	0.2385	0.2500	-0.4662	-0.4220
Lumber	0.0735	0.0521	0.2000	0.0721	-0.7776
Live Cattle	0.0638	0.0517	0.1066	0.0835	-0.9132

Table A1 (continued)
Descriptive statistics for commodity futures returns and the factors

Panel C	Basis				
	Mean	**Median**	**Standard Deviation**	**Skewness**	**Excess Kurtosis**
Light Crude Oil	-0.0011	-0.0024	0.0197	-0.3821	3.6552
Corn	-0.0197	-0.0239	0.0316	3.4492	19.9205
Soybeans	-0.0014	-0.0070	0.0223	2.6293	8.8736
Wheat	-0.0211	-0.0258	0.0313	1.6318	4.1943
Coffee	-0.0164	-0.0206	0.0301	1.7687	5.5912
Cocoa	-0.0113	-0.0104	0.0202	0.0490	1.2644
Sugar	-0.0008	-0.0048	0.0496	-0.0433	4.3632
Cotton No.2	-0.0104	-0.0150	0.0401	2.0345	12.2371
Gold	-0.0025	-0.0023	0.0019	-0.6186	-0.2280
Silver	0.0010	-0.0024	0.0519	13.1632	172.8415
Platinum	0.0293	-0.0007	0.1687	5.5658	29.3783
Orange Juice	-0.0086	-0.0093	0.0292	0.6862	0.6978
Lumber	-0.0215	-0.0174	0.0459	-0.0647	0.5445
Live Cattle	0.0029	0.0003	0.0356	0.4074	0.0202

Panel D	Net Trading				
	Mean	**Median**	**Standard Deviation**	**Skewness**	**Excess Kurtosis**
Light Crude Oil	0.0086	0.0058	0.0510	0.2517	1.3783
Corn	0.0069	0.0026	0.0758	-0.0493	8.3574
Soybeans	0.0080	0.0096	0.0820	-0.3949	4.2786
Wheat	0.0080	0.0015	0.0894	0.0856	5.2694
Coffee	0.0116	0.0210	0.0975	0.3201	0.6292
Cocoa	0.0082	0.0095	0.0760	0.0294	0.8178
Sugar	0.0082	0.0055	0.0767	-0.1195	0.2702
Cotton No.2	0.0117	0.0117	0.1012	0.2508	1.5604
Gold	0.0070	-0.0006	0.0943	0.5189	1.2594
Silver	0.0061	0.0011	0.0803	0.7065	2.0992
Platinum	0.0108	0.0107	0.1081	0.4331	0.8757
Orange Juice	0.0067	0.0040	0.1109	0.4007	0.7497
Lumber	0.0104	-0.0026	0.1481	1.9518	12.6281
Live Cattle	0.0083	0.0041	0.0746	0.2013	0.2940

Panel E	Aggregate Commodity Specific Factors				
	Mean	**Median**	**Standard Deviation**	**Skewness**	**Excess Kurtosis**
Hedging Pressure	0.0975	0.1002	0.0772	-0.2829	-0.3088
Basis	-0.0055	-0.0073	0.0143	2.0651	7.4183
Net Trading	0.0051	0.0045	0.0414	-1.6270	16.6478

Table A2

Mean Absolute Square Error Forecast Performances – Classification by Low vs. High Risk Aversion Samples

The table reports the MAFE implied by the forecasts obtained from the 12 stepwise (forward vs. backward) predictive regression models recursively implemented in our analysis. In particular, the table distinguishes between the MAFE obtained in periods of low vs. high market volatility (a proxy for aggregate risk aversion). Specifically, the periods of low risk aversion are identified to correspond to January 2004 – December 2007 and January 2013 – May 2018, while the high risk aversion sub-sample refers to Januart 2008 - December 2012. For both the low and high risk-aversion sub-samples, we have boldfaced the forecasting model that yields the minimum MAFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Panel A	**Macro Principal Components Only**													
Forward - Low	0.0613	0.0592	0.0547	0.0644	0.0688	0.0555	**0.0583**	0.0544	0.0356	0.0638	0.0653	**0.0669**	**0.0610**	0.0625
Forward - High	0.0829	0.0952	0.0872	0.0932	0.0801	0.0971	0.0866	0.0842	0.0494	0.1084	0.0916	0.0874	0.0721	0.0701
Backward - Low	0.0615	0.0597	**0.0545**	**0.0641**	0.0672	**0.0555**	0.0584	0.0543	0.0356	0.0645	0.0655	0.0670	0.0610	0.0618
Backward - High	0.0817	0.1019	0.0869	0.0947	0.0826	0.1077	0.0894	0.0861	0.0517	0.1077	0.0890	0.0884	0.0847	0.0703
Panel B	**Macro Principal Components + Commodity-Specific Factors (Always Included)**													
Forward - Low	0.0644	0.0574	0.0605	0.0665	0.0791	0.0630	0.0610	**0.0510**	0.0378	0.0716	0.0702	0.0745	0.0659	0.0789
Forward - High	0.0902	0.1192	0.0866	0.1139	0.0976	0.1139	0.0918	0.0990	0.0572	0.1378	0.4112	0.0969	0.0972	0.0984
Backward - Low	0.0646	**0.0570**	0.0625	0.0660	0.0801	0.0642	0.0608	0.0542	0.0369	0.0723	0.0735	0.0741	0.0652	0.0736
Backward - High	0.0923	0.1260	0.0971	0.1267	0.0923	0.1175	0.0934	0.1059	0.0587	0.1385	0.3565	0.1059	0.1162	0.1134
Panel C	**Macro Principal Components + Commodity-Specific Factors**													
Forward - Low	0.0563	0.0579	0.0585	0.0651	0.0682	0.0575	0.0597	0.0523	0.0369	0.0658	0.0679	0.0701	0.0650	0.0671
Forward - High	0.0517	0.1089	0.0827	0.1080	0.0858	0.1129	0.0894	0.0905	0.0568	0.1214	0.0906	0.0893	0.0912	0.0868
Backward - Low	0.0549	0.0581	0.0593	0.0644	0.0693	0.0598	0.0594	0.0523	0.0365	0.0669	0.0675	0.0745	0.0647	0.0673
Backward - High	**0.0507**	0.1129	0.0918	0.1179	0.0930	0.1102	0.0883	0.0943	0.0586	0.1346	0.3724	0.0981	0.1083	0.0908
Panel D	**Macro Principal Components + Aggregated Commodity-Specific Factors (Always Included)**													
Forward - Low	0.0675	0.0630	0.0620	0.0691	0.0749	0.0596	0.0649	0.0589	0.0461	0.0665	0.0644	0.0761	0.0658	0.0908
Forward - High	0.1292	0.1305	0.1195	0.1373	0.1172	0.1213	0.1761	0.1295	0.0708	0.1473	0.1526	0.1013	0.1174	0.0836
Backward - Low	0.0668	0.0633	0.0585	0.0687	0.0736	0.0597	0.0651	0.0600	0.0433	0.0663	0.0645	0.0760	0.0656	0.0754
Backward - High	0.1216	0.1304	0.1167	0.1394	0.1195	0.1205	0.1539	0.1517	0.0747	0.1576	0.1478	0.1014	0.1187	0.0830
Panel E	**Macro Principal Components + Aggregated Commodity-Specific Factors**													
Forward - Low	0.0625	0.0628	0.0563	0.0661	0.0686	0.0608	0.0614	0.0568	**0.0350**	0.0630	0.0653	0.0749	0.0631	0.0635
Forward - High	0.0896	0.1122	0.0889	0.1258	0.1076	0.1047	0.0973	0.0869	0.0536	0.1128	0.1091	0.0946	0.0950	0.0713
Backward - Low	0.0631	0.0621	0.0582	0.0679	0.0676	0.0586	0.0618	0.0573	0.0359	**0.0630**	0.0654	0.0746	0.0661	0.0631
Backward - High	0.0854	0.1127	0.0962	0.1322	0.1147	0.1081	0.1132	0.1123	0.0577	0.1342	0.1163	0.0985	0.1000	0.0726
Panel F	**Macro Principal Components + Commodity-Specific Factors + Aggregated Commodity-Specific Factors (Always Included)**													
Forward - Low	0.0672	0.0615	0.0602	0.0734	0.0787	0.0650	0.0700	0.0588	0.0370	0.0760	0.0658	0.0829	0.0689	0.0688
Forward - High	0.1309	0.1352	0.1131	0.1714	0.1203	0.1244	0.1515	0.1536	0.0796	0.1532	0.3081	0.1029	0.1339	0.1055
Backward - Low	0.0678	0.0595	0.0626	0.0728	0.0815	0.0662	0.0702	0.0576	0.0377	0.0893	0.0641	0.0858	0.0690	0.0698
Backward - High	0.1376	0.1409	0.1157	0.1536	0.1390	0.1301	0.1587	0.1785	0.0879	0.1602	0.2736	0.1257	0.1548	0.1367
Panel G	**Benchmark AR(1)**													
AR(1)	0.0682	0.0676	0.0603	0.0711	**0.0647**	0.0635	0.0674	0.0610	0.0408	0.0748	**0.0492**	0.0727	0.0635	**0.0356**

Table A3

Mean Absolute Error Forecast Performances – Weighting on the Basis of Predicted Risk-Aversion Regime Probabilities

The table reports the MAFE implied by the forecasts obtained from the 12 stepwise (forward vs. backward) predictive regression models recursively implemented in our analysis. In particular, the table obtains the forecasts by weighting (over the entire OOS period, January 2004 – May 2018) the regime-specific stepwise regression forecasts by the predicted regime probabilities obtained from a two-state HMM for the VIX index. In the different p anels, we have boldfaced the forecasting model that yields the minimum MAFE.

	Light Crude Oil	Corn	Soybeans	Wheat	Coffee	Cocoa	Sugar	Cotton No.2	Gold	Silver	Platinum	Orange Juice	Lumber	Live Cattle
Panel A	**Macro Principal Components Only**													
Forward	0.0653	0.0699	0.0628	0.0734	0.0700	0.0658	0.0678	0.0627	**0.0404**	0.0754	0.0672	0.0732	0.0645	0.0656
Backward	0.0638	0.0699	0.0626	0.0735	0.0692	0.0660	0.0684	0.0628	0.0410	0.0757	0.0671	0.0733	0.0656	0.0652
Panel B	**Macro Principal Components + Commodity-Specific Factors (Always Included)**													
Forward	0.0662	0.0711	0.0653	0.0759	0.0773	0.0709	0.0694	0.0619	0.0417	0.0815	0.1064	0.0781	0.0704	0.0752
Backward	0.0664	0.0697	0.0676	0.0762	0.0765	0.0733	0.0696	0.0642	0.0413	0.0822	0.1009	0.0784	0.0716	0.0748
Panel C	**Macro Principal Components + Commodity-Specific Factors**													
Forward	0.0585	0.0692	0.0643	0.0751	0.0705	0.0680	0.0686	0.0617	0.0421	0.0764	0.0688	0.0751	0.0693	0.0687
Backward	**0.0571**	0.0691	0.0657	0.0748	0.0705	0.0696	0.0689	0.0624	0.0416	0.0782	0.1003	0.0784	0.0701	0.0696
Panel D	**Macro Principal Components + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0706	0.0733	0.0715	0.0793	0.0740	0.0681	0.0839	0.0685	0.0485	0.0818	0.0748	0.0791	0.0732	0.0814
Backward	0.0696	0.0735	0.0689	0.0792	0.0739	0.0683	0.0813	0.0729	0.0482	0.0814	0.0736	0.0787	0.0731	0.0731
Panel E	**Macro Principal Components + Aggregated Commodity-Specific Factors**													
Forward	0.0650	0.0721	0.0631	0.0783	0.0716	0.0694	0.0704	0.0636	**0.0404**	0.0756	0.0696	0.0782	0.0681	0.0662
Backward	0.0642	0.0719	0.0660	0.0808	0.0706	0.0669	0.0732	0.0656	0.0417	0.0755	0.0707	0.0780	0.0701	0.0660
Panel F	**Macro Principal Components + Commodity-Specific Factors + Aggregated Commodity-Specific Factors (Always Included)**													
Forward	0.0711	0.0730	0.0674	0.0857	0.0743	0.0719	0.0863	0.0697	0.0434	0.0856	0.0928	0.0824	0.0754	0.0706
Backward	0.0696	0.0735	0.0701	0.0833	0.0789	0.0732	0.0848	0.0732	0.0449	0.0928	0.0896	0.0860	0.0791	0.0750
Panel G	**Benchmark AR(1)**													
AR(1)	0.0682	**0.0676**	**0.0603**	**0.0711**	**0.0647**	**0.0635**	**0.0674**	**0.0610**	0.0408	**0.0748**	**0.0492**	**0.0727**	**0.0635**	**0.0356**

Table A4

Opimal Asset Allocation – Hidden Markov Models

	No Transaction Costs			With Transaction Costs		
	Macro Principal Components Only					
	γ = 0.1	**γ = 0.25**	**γ = 0.5**	**γ = 0.1**	**γ = 0.25**	**γ = 0.5**
Average Realized Return	0.0001	-0.0001	-0.0001	0.0006	-0.0014	-0.0012
Realized Standard Deviation	0.0275	0.0267	0.0262	0.0301	0.0306	0.0304
Annualized Sharpe Ratio	0.0067	-0.0106	-0.0128	0.0660	-0.1631	-0.1397
Realized Average MV Utility	0.0000	-0.0002	-0.0003	0.0005	-0.0016	-0.0015
	Average Portfolio Weights					
Light Crude Oil	0.0145	0.0113	0.0081	0.0319	0.0254	0.0195
Corn	-0.0037	-0.0060	-0.0006	-0.0425	-0.0468	-0.0376
Soybeans	0.0008	0.0103	0.0081	-0.0164	-0.0014	-0.0028
Wheat	0.0034	0.0014	-0.0016	0.0650	0.0635	0.0579
Coffee	-0.0031	-0.0050	-0.0060	0.0011	-0.0023	-0.0049
Cocoa	0.0002	-0.0025	-0.0015	-0.0023	-0.0057	-0.0066
Sugar	0.0034	0.0039	0.0016	0.0155	0.0146	0.0096
Cotton No.2	0.0072	0.0019	-0.0006	0.0270	0.0178	0.0107
Gold	0.0063	-0.0065	-0.0082	-0.0221	-0.0391	-0.0440
Silver	-0.0025	-0.0022	-0.0004	0.0075	0.0092	0.0131
Platinum	0.0221	0.0235	0.0217	0.0210	0.0199	0.0216
Orange Juice	0.0039	0.0036	0.0023	0.0052	0.0057	0.0033
Lumber	0.0105	0.0036	0.0034	-0.0159	-0.0259	-0.0251
Live Cattle	-0.0050	-0.0035	-0.0009	-0.0163	-0.0134	-0.0078
S&P500	-0.0158	-0.0043	-0.0025	0.0136	0.0435	0.0502
10Y Treasury Bond	0.9578	0.9705	0.9770	0.9277	0.9348	0.9429
	Macro Principal Components + Commodity-Specific					
	γ = 0.1	**γ = 0.25**	**γ = 0.5**	**γ = 0.1**	**γ = 0.25**	**γ = 0.5**
Rendimento Mensile Medio	0.0020	0.0000	0.0018	0.0028	0.0011	0.0018
Deviazione Standard Mensile	0.0455	0.0422	0.0358	0.0408	0.0363	0.0296
Sharpe Ratio Annuo	0.1516	0.0023	0.1770	0.2340	0.1039	0.2077
Utilità Media-Varianza	0.0019	-0.0002	0.0015	0.0027	0.0009	0.0016
	Average Portfolio Weights					
Light Crude Oil	0.0014	0.0011	0.0018	0.0012	-0.0030	-0.0024
Corn	-0.0029	-0.0053	-0.0192	0.0360	0.0315	0.0029
Soybeans	0.0091	0.0052	0.0223	-0.0241	-0.0293	0.0076
Wheat	0.0077	0.0090	0.0094	-0.0019	0.0040	-0.0003
Coffee	-0.0024	-0.0024	-0.0040	0.0231	0.0224	0.0162
Cocoa	0.0023	-0.0013	0.0031	-0.0038	-0.0073	0.0050
Sugar	0.0034	0.0012	0.0003	0.0116	0.0057	0.0026
Cotton No.2	0.0014	-0.0037	-0.0064	0.0098	0.0077	0.0044
Gold	-0.0064	-0.0013	-0.0208	0.0260	0.0171	-0.0160
Silver	0.0077	0.0101	0.0209	0.0031	-0.0012	0.0164
Platinum	0.0109	0.0100	0.0087	-0.0086	0.0034	0.0048
Orange Juice	0.0065	0.0047	0.0053	-0.0088	-0.0065	-0.0005
Lumber	0.0048	0.0067	0.0028	-0.0008	-0.0011	-0.0110
Live Cattle	-0.0082	-0.0027	-0.0027	0.0158	0.0174	0.0181
S&P500	-0.0112	-0.0111	-0.0118	0.0384	0.0312	0.0383
10Y Treasury Bond	0.9760	0.9798	0.9904	0.8828	0.9081	0.9140

Table A4 (continued)

Opimal Asset Allocation – Hidden Markov Models

	No Transaction Costs			With Transaction Costs		
	Principal Components + Aggregated Commodity-Specific					
	γ = 0.1	γ = 0.25	γ = 0.5	γ = 0.1	γ = 0.25	γ = 0.5
Rendimento Mensile Medio	0.0028	0.0012	0.0012	0.0024	0.0027	0.0027
Deviazione Standard Mensile	0.0383	0.0287	0.0290	0.0473	0.0277	0.0288
Sharpe Ratio Annuo	0.2496	0.1426	0.1437	0.1727	0.3362	0.3210
Utilità Media-Varianza	0.0027	0.0011	0.0010	0.0022	0.0026	0.0025
	Pesi Medi			**Pesi Medi**		
Light Crude Oil	0.0076	-0.0002	-0.0047	0.0067	0.0025	0.0026
Corn	0.0245	0.0119	0.0074	0.0656	0.0108	0.0013
Soybeans	-0.0302	-0.0139	-0.0087	-0.0267	0.0093	0.0109
Wheat	0.0140	0.0094	0.0101	-0.0217	-0.0019	0.0030
Coffee	-0.0110	-0.0128	-0.0142	-0.0086	-0.0068	-0.0094
Cocoa	0.0112	0.0093	0.0068	0.0389	0.0185	0.0188
Sugar	-0.0080	-0.0076	-0.0057	0.0050	0.0086	0.0106
Cotton No.2	-0.0018	-0.0016	0.0003	0.0075	-0.0118	-0.0052
Gold	0.0394	0.0251	0.0231	-0.0085	-0.0015	0.0011
Silver	0.0099	0.0137	0.0112	0.0312	0.0062	0.0202
Platinum	-0.0084	-0.0043	-0.0023	-0.0149	0.0079	-0.0061
Orange Juice	0.0018	-0.0038	-0.0035	0.0270	0.0008	0.0024
Lumber	0.0037	0.0024	0.0002	-0.0118	-0.0012	-0.0059
Live Cattle	0.0422	0.0362	0.0284	0.0458	0.0263	0.0272
S&P500	-0.0181	-0.0156	-0.0178	0.0459	0.0396	0.0290
10Y Treasury Bond	0.9233	0.9518	0.9693	0.8187	0.8927	0.8995

Table A5

Opimal Asset Allocation – Stepwise Regression Models Based on Macro Principal Components Only

Panel A (No Transaction Costs)

	Forward γ = 0.1	**Backward γ = 0.1**	**Forward γ = 0.25**	**Backward γ = 0.25**	**Forward γ = 0.5**	**Backward γ = 0.5**
Average Realized Return	-0.0005	-0.0009	-0.0006	-0.0007	-0.0003	-0.0004
Realized Standard Deviation	0.0267	0.0295	0.0267	0.0278	0.0252	0.0265
Annualized Sharpe Ratio	-0.0588	-0.0998	-0.0774	-0.0906	-0.0421	-0.0466
Realized Average MV Utility	-0.0005	-0.0009	-0.0007	-0.0008	-0.0005	-0.0005
			Average Portfolio Weights			
Light Crude Oil	0.0435	0.0303	0.0435	0.0357	0.0445	0.0254
Corn	-0.0098	-0.0470	-0.0114	-0.0540	-0.0120	-0.0408
Soybeans	-0.0221	0.0020	-0.0215	0.0015	-0.0223	-0.0007
Wheat	-0.0003	0.0176	0.0010	0.0221	0.0017	0.0166
Coffee	0.0280	0.0258	0.0284	0.0303	0.0296	0.0334
Cocoa	0.0049	0.0268	0.0053	0.0297	0.0057	0.0234
Sugar	0.0168	0.0070	0.0171	0.0059	0.0178	0.0074
Cotton No.2	0.0124	0.0124	0.0130	0.0195	0.0136	0.0213
Gold	-0.0704	-0.0976	-0.0661	-0.1106	-0.0699	-0.1148
Silver	0.0308	0.0405	0.0296	0.0484	0.0309	0.0543
Platinum	-0.0114	-0.0071	-0.0124	-0.0151	-0.0134	-0.0152
Orange Juice	0.0166	0.0217	0.0167	0.0247	0.0173	0.0338
Lumber	0.0320	0.0243	0.0312	0.0289	0.0319	0.0313
Live Cattle	0.0146	0.0217	0.0132	0.0218	0.0136	0.0306
S&P500	0.0499	0.0519	0.0474	0.0349	0.0444	0.0123
10Y Treasury Bond	0.8646	0.8699	0.8650	0.8763	0.8666	0.8816

Panel B (With Transaction Costs)

	Forward γ = 0.1	**Backward γ = 0.1**	**Forward γ = 0.25**	**Backward γ = 0.25**	**Forward γ = 0.5**	**Backward γ = 0.5**
Rendimento Mensile Medio	0.0038	0.0038	0.0032	0.0037	0.0032	0.0041
Deviazione Standard Mensile	0.0353	0.0412	0.0353	0.0410	0.0351	0.0398
Sharpe Ratio Annuo	0.3736	0.3222	0.3187	0.3094	0.3186	0.3526
Utilità Media-Varianza	0.0037	0.0038	0.0031	0.0035	0.0029	0.0037
			Average Portfolio Weights			
Light Crude Oil	0.0295	0.0796	0.0326	0.0907	0.0325	0.0940
Corn	-0.0279	-0.0857	-0.0280	-0.0953	-0.0279	-0.0807
Soybeans	-0.0362	-0.1175	-0.0308	-0.1348	-0.0309	-0.1775
Wheat	-0.0275	0.0348	-0.0284	0.0406	-0.0283	0.0339
Coffee	0.0356	0.1279	0.0349	0.1430	0.0348	0.1836
Cocoa	0.0266	0.1050	0.0230	0.1167	0.0229	0.1327
Sugar	0.0121	0.0078	0.0131	0.0049	0.0131	0.0074
Cotton No.2	0.0279	0.1732	0.0264	0.2022	0.0262	0.2588
Gold	-0.0350	-0.4650	-0.0508	-0.5117	-0.0508	-0.6425
Silver	0.0166	0.2473	0.0243	0.2786	0.0242	0.3585
Platinum	-0.1158	-0.3760	-0.1134	-0.4337	-0.1130	-0.5666
Orange Juice	0.0364	0.1195	0.0382	0.1321	0.0380	0.1716
Lumber	0.0339	0.0320	0.0349	0.0364	0.0347	0.0406
Live Cattle	0.0210	0.0705	0.0227	0.0737	0.0227	0.0982
S&P500	0.2662	0.2032	0.2569	0.2038	0.2565	0.2202
10Y Treasury Bond	0.7365	0.8435	0.7446	0.8531	0.7452	0.8678

Table A6

Opimal Asset Allocation – Stepwise Regression Models Based on Macro Principal Components + Commodity-Specific Factors (Always Included)

Panel A (No Transaction Costs)

	Forward γ = 0.1	Backward γ = 0.1	Forward γ = 0.25	Backward γ = 0.25	Forward γ = 0.5	Backward γ = 0.5
Average Realized Return	0.0021	-0.0029	0.0020	-0.0028	0.0024	-0.0031
Realized Standard Deviation	0.0418	0.0355	0.0412	0.0349	0.0406	0.0348
Annualized Sharpe Ratio	0.1734	-0.2842	0.1700	-0.2808	0.2063	-0.3123
Realized Average MV Utility	0.0020	-0.0030	0.0018	-0.0030	0.0020	-0.0034
	Average Portfolio Weights					
Light Crude Oil	0.0023	0.0082	0.0020	0.0079	0.0024	0.0097
Corn	0.0780	0.0814	0.0765	0.0859	0.0768	0.0763
Soybeans	-0.0397	0.0072	-0.0362	0.0162	-0.0375	0.0227
Wheat	-0.0218	-0.0507	-0.0243	-0.0578	-0.0209	-0.0557
Coffee	0.0133	0.0117	0.0134	0.0109	0.0116	0.0120
Cocoa	0.0108	0.0000	0.0107	0.0045	0.0067	0.0028
Sugar	0.0011	0.0021	0.0040	0.0034	0.0008	0.0047
Cotton No.2	0.0164	-0.0212	0.0180	-0.0250	0.0176	-0.0261
Gold	0.0774	0.0284	0.0593	0.0035	0.0514	-0.0133
Silver	-0.0291	-0.0241	-0.0136	-0.0122	-0.0111	-0.0092
Platinum	-0.0067	0.0593	-0.0243	0.0627	-0.0161	0.0784
Orange Juice	0.0192	0.0274	0.0200	0.0290	0.0161	0.0308
Lumber	-0.0010	0.0061	0.0002	0.0059	0.0015	0.0082
Live Cattle	0.0731	0.0414	0.0838	0.0402	0.0883	0.0410
S&P500	-0.0074	-0.0172	-0.0091	-0.0273	-0.0143	-0.0371
10Y Treasury Bond	0.8142	0.8398	0.8197	0.8520	0.8267	0.8549

Panel B (With Transaction Costs)

	Forward γ = 0.1	Backward γ = 0.1	Forward γ = 0.25	Backward γ = 0.25	Forward γ = 0.5	Backward γ = 0.5
Rendimento Mensile Medio	0.0029	-0.0012	0.0036	-0.0007	0.0022	-0.0001
Deviazione Standard Mensile	0.0576	0.0319	0.0567	0.0309	0.0544	0.0317
Sharpe Ratio Annuo	0.1724	-0.1331	0.2229	-0.0817	0.1408	-0.0147
Utilità Media-Varianza	0.0027	-0.0013	0.0032	-0.0008	0.0015	-0.0004
	Average Portfolio Weights					
Light Crude Oil	-0.0176	-0.0267	-0.0147	-0.0273	-0.0159	-0.0271
Corn	0.1167	0.1027	0.0988	0.1126	0.0921	0.1105
Soybeans	-0.0885	-0.0366	-0.0733	-0.0778	-0.0635	-0.0646
Wheat	-0.0326	-0.0800	-0.0327	-0.0624	-0.0314	-0.0722
Coffee	0.0234	0.0267	0.0310	0.0220	0.0294	0.0248
Cocoa	0.0114	0.0059	0.0113	-0.0057	0.0146	-0.0028
Sugar	-0.0128	0.0012	0.0073	0.0013	0.0052	-0.0014
Cotton No.2	-0.0014	-0.0124	0.0256	-0.0043	0.0187	-0.0098
Gold	0.0819	0.0412	0.0948	0.0821	0.0775	0.0719
Silver	-0.0592	-0.0664	-0.0144	-0.0425	0.0072	-0.0631
Platinum	0.0888	0.0931	-0.0570	0.0438	-0.0739	0.0766
Orange Juice	0.0131	0.0099	0.0129	0.0106	0.0231	0.0140
Lumber	-0.0257	-0.0165	-0.0261	-0.0010	-0.0280	-0.0097
Live Cattle	0.1471	0.0225	0.1282	0.0267	0.1309	0.0152
S&P500	0.0594	0.1271	0.0825	0.1114	0.0839	0.1254
10Y Treasury Bond	0.6961	0.8084	0.7259	0.8105	0.7300	0.8125

Table A7

Opimal Asset Allocation – Stepwise Regression Models Based on Macro Principal Components + Commodity-Specific Factors

Panel A (No Transaction Costs)

	Forward γ = 0.1	Backward γ = 0.1	Forward γ = 0.25	Backward γ = 0.25	Forward γ = 0.5	Backward γ = 0.5
Average Realized Return	-0.0033	-0.0022	-0.0016	-0.0022	0.0000	-0.0020
Realized Standard Deviation	0.0359	0.0324	0.0381	0.0324	0.0327	0.0313
Annualized Sharpe Ratio	-0.3152	-0.2371	-0.1419	-0.2384	-0.0013	-0.2218
Realized Average MV Utility	-0.0033	-0.0023	-0.0017	-0.0024	-0.0003	-0.0022
	Average Portfolio Weights					
Light Crude Oil	0.0091	0.0141	0.0043	0.0111	0.0028	-0.0095
Corn	0.0577	0.0479	0.0554	0.0494	0.0729	0.0170
Soybeans	0.0057	0.0077	-0.0114	-0.0170	-0.0309	0.0431
Wheat	-0.0443	-0.0366	-0.0260	-0.0221	-0.0228	0.0072
Coffee	0.0013	0.0003	0.0088	0.0025	0.0111	-0.0038
Cocoa	0.0153	0.0013	0.0205	0.0047	0.0158	0.0030
Sugar	0.0018	0.0125	0.0047	0.0070	0.0011	-0.0265
Cotton No.2	-0.0145	-0.0001	-0.0111	-0.0079	-0.0021	-0.0675
Gold	0.0821	0.0749	0.0621	0.0440	0.0673	-0.0901
Silver	-0.0324	-0.0099	-0.0345	0.0010	-0.0396	0.0676
Platinum	0.0115	-0.0036	0.0351	0.0106	0.0557	0.0559
Orange Juice	0.0285	0.0198	0.0373	0.0236	0.0235	0.0409
Lumber	0.0000	-0.0035	0.0043	-0.0029	0.0033	0.0225
Live Cattle	0.0286	0.0437	0.0091	0.0596	0.0151	0.1179
S&P500	0.0137	-0.0159	-0.0051	-0.0216	-0.0297	-0.0597
10Y Treasury Bond	0.8359	0.8473	0.8470	0.8580	0.8566	0.8819

Panel B (With Transaction Costs)

	Forward γ = 0.1	Backward γ = 0.1	Forward γ = 0.25	Backward γ = 0.25	Forward γ = 0.5	Backward γ = 0.5
Rendimento Mensile Medio	0.0007	0.0021	0.0005	0.0030	0.0006	0.0016
Deviazione Standard Mensile	0.0209	0.0304	0.0216	0.0283	0.0211	0.0283
Sharpe Ratio Annuo	0.1094	0.2388	0.0848	0.3642	0.0956	0.1938
Utilità Media-Varianza	0.0006	0.0020	0.0005	0.0029	0.0005	0.0014
	Average Portfolio Weights					
Light Crude Oil	0.0229	0.0065	0.0105	0.0070	0.0129	0.0030
Corn	0.1002	0.0651	0.0537	0.0626	0.0470	0.0368
Soybeans	-0.0680	-0.0443	-0.0195	-0.0576	-0.0111	-0.0260
Wheat	-0.0630	-0.0301	-0.0302	-0.0209	-0.0403	-0.0161
Coffee	0.0272	0.0200	0.0202	0.0208	0.0093	0.0159
Cocoa	0.0048	0.0105	0.0079	0.0105	0.0180	0.0133
Sugar	0.0249	0.0150	0.0098	0.0124	0.0082	0.0055
Cotton No.2	0.0146	0.0123	-0.0066	0.0123	-0.0002	-0.0043
Gold	0.0712	0.0613	0.1346	0.0585	0.0635	0.0081
Silver	-0.0025	0.0053	-0.0447	0.0044	-0.0108	0.0191
Platinum	-0.0700	-0.0443	-0.0038	-0.0387	-0.0026	-0.0270
Orange Juice	0.0121	0.0229	0.0080	0.0170	0.0221	0.0169
Lumber	-0.0024	-0.0089	-0.0020	-0.0124	-0.0045	-0.0079
Live Cattle	0.0551	0.0447	0.0350	0.0519	0.0068	0.0662
S&P500	0.0427	0.0912	-0.0107	0.0949	0.0406	0.0995
10Y Treasury Bond	0.9225	0.8584	0.9309	0.8635	0.9346	0.8856

Table A8

Opimal Asset Allocation – Stepwise Regression Models Based on Macro Principal Components + Commodity-Speficic Factors + Aggreage Commodity Specific Factors (Always Included)

Panel A (No Transaction Costs)

	Forward γ = 0.1	Backward γ = 0.1	Forward γ = 0.25	Backward γ = 0.25	Forward γ = 0.5	Backward γ = 0.5
Average Realized Return	-0.0042	-0.0032	-0.0034	-0.0032	-0.0034	-0.0018
Realized Standard Deviation	0.0378	0.0376	0.0366	0.0406	0.0366	0.0374
Annualized Sharpe Ratio	-0.3878	-0.2975	-0.3183	-0.2731	-0.3179	-0.1663
Realized Average MV Utility	-0.0043	-0.0033	-0.0035	-0.0034	-0.0037	-0.0021
	Average Portfolio Weights					
Light Crude Oil	0.0134	-0.0028	0.0091	-0.0110	0.0091	-0.0032
Corn	0.0321	-0.0332	0.0041	-0.0182	0.0041	-0.0137
Soybeans	-0.0202	0.0473	-0.0067	0.0593	-0.0067	0.0354
Wheat	-0.0208	0.0133	-0.0010	0.0115	-0.0010	0.0138
Coffee	0.0153	0.0162	0.0031	0.0113	0.0031	0.0180
Cocoa	-0.0166	0.0014	-0.0114	-0.0195	-0.0114	-0.0098
Sugar	-0.0178	0.0044	-0.0142	0.0051	-0.0142	0.0091
Cotton No.2	-0.0007	0.0236	0.0174	0.0489	0.0174	0.0346
Gold	0.0102	-0.0115	0.0198	0.0101	0.0198	0.0021
Silver	0.0224	0.0293	0.0141	0.0209	0.0141	0.0272
Platinum	0.0030	0.0186	0.0122	0.0163	0.0123	-0.0005
Orange Juice	-0.0166	0.0071	-0.0183	-0.0027	-0.0183	-0.0046
Lumber	-0.0016	0.0067	-0.0126	-0.0013	-0.0125	-0.0025
Live Cattle	0.1806	0.1012	0.1680	0.1401	0.1681	0.1075
S&P500	-0.0134	-0.0658	-0.0321	-0.1190	-0.0323	-0.0736
10Y Treasury Bond	0.8306	0.8443	0.8485	0.8482	0.8486	0.8602

Panel B (With Transaction Costs)

	Forward γ = 0.1	Backward γ = 0.1	Forward γ = 0.25	Backward γ = 0.25	Forward γ = 0.5	Backward γ = 0.5
Rendimento Mensile Medio	0.0017	-0.0002	0.0014	-0.0001	0.0014	0.0004
Deviazione Standard Mensile	0.0287	0.0282	0.0307	0.0282	0.0307	0.0305
Sharpe Ratio Annuo	0.2061	-0.0277	0.1595	-0.0101	0.1531	0.0461
Utilità Media-Varianza	0.0017	-0.0003	0.0013	-0.0002	0.0011	0.0002
	Average Portfolio Weights					
Light Crude Oil	0.0305	0.0210	0.0462	0.0252	0.0436	0.0130
Corn	-0.0460	-0.0443	0.0374	-0.0403	0.0418	-0.0043
Soybeans	-0.0119	0.0295	-0.0524	0.0297	-0.0547	0.0055
Wheat	0.0125	-0.0101	-0.0167	-0.0134	-0.0134	-0.0212
Coffee	0.0122	0.0056	0.0168	0.0075	0.0155	0.0130
Cocoa	0.0762	0.0345	0.0653	0.0286	0.0648	0.0340
Sugar	0.0204	-0.0151	0.0042	-0.0111	0.0010	0.0002
Cotton No.2	0.0087	0.0216	0.0502	0.0306	0.0498	0.0033
Gold	0.0889	-0.0482	0.1503	-0.0470	0.1624	0.0123
Silver	0.0632	0.0444	0.0666	0.0467	0.0605	0.0380
Platinum	-0.1296	-0.0123	-0.1822	-0.0207	-0.1789	-0.0697
Orange Juice	0.0000	0.0333	-0.0099	0.0346	-0.0137	0.0182
Lumber	-0.0341	0.0080	-0.0639	0.0079	-0.0673	0.0008
Live Cattle	-0.2218	0.0062	-0.2534	0.0037	-0.2612	-0.0288
S&P500	0.2668	0.0967	0.2742	0.0860	0.2816	0.1493
10Y Treasury Bond	0.8640	0.8291	0.8675	0.8320	0.8680	0.8365

Figure A1

Recursively Selected Factors Selected by Stepwise Regression Model (Backward Algorithm) for Light Crude Oil during the High Volatility Regime, January 2008 – December 2012

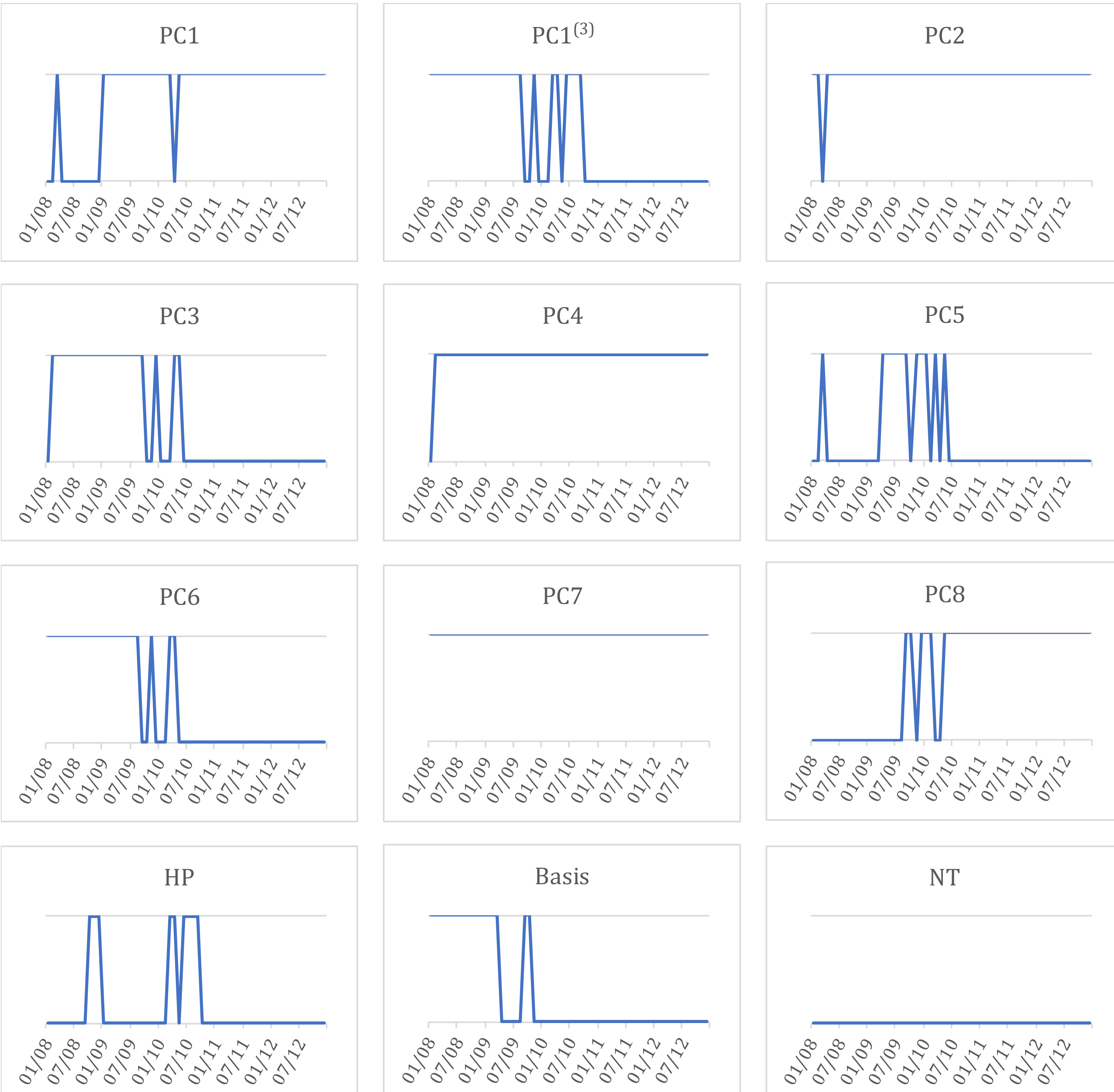

Figure A1 (continued)

Recursively Selected Factors Selected by Stepwise Regression Model (Backward Algorithm) for Light Crude Oil during the High Volatility Regime, January 2008 – December 2012