THESIS DECLARATION

The undersigned

SURNAME	GARGANO
NAME	ANTONIO
PhD Registration Number	1376839

Thesis title:

Essays on Return Predictability and Asset Pricing

PhD in	Finance
Cycle	XXIII
Candidate's tutor	Carlo Ambrogio Favero
Year of discussion	2013

DECLARES

Under his responsibility:

1) that, according to the President's decree of 28.12.2000, No. 445, mendacious declarations, falsifying records and the use of false records are punishable under the penal code and special laws, should any of these hypotheses prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree of 30th April 1999 protocol no. 224/1999, to keep copy of the thesis on deposit at the Biblioteche Nazionali Centrali di Roma e Firenze, where consultation is permitted, unless there is a temporary embargo in order to protect the rights of external bodies and industrial/commercial exploitation of the thesis;

3) that the Servizio Biblioteca Bocconi will file the thesis in its 'Archivio istituzionale ad accesso aperto' and will permit on-line consultation of the complete text (except in cases of a temporary embargo);

4) that in order keep the thesis on file at Biblioteca Bocconi, the University requires that the thesis be delivered by the candidate to Società NORMADEC (acting on behalf of the University) by online procedure the contents of which must be unalterable and that NORMADEC will indicate in each footnote the following information:

- thesis (thesis title) Essays in Return Predictability and Asset Pricing
- by (candidate's surname and first name) Gargano Antonio ;
- discussed at Università Commerciale Luigi Bocconi – Milano in (year of discussion) 2013 ;
- the thesis is protected by the regulations governing copyright (law of 22 April 1941, no. 633 and successive modifications). The exception is the right of Università Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;

5) that the copy of the thesis deposited with NORMADEC by online procedure is identical to those handed in/sent to the Examiners and to any other copy deposited in the University offices on paper or electronic copy and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (law of 22 April 1941, no. 633 and successive integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal and shall be exempt from any requests or claims from third parties;

7[a]) that the PhD thesis is not the result of work included in the regulations governing industrial property, it was not produced as part of projects financed by public or private bodies with restrictions on the diffusion of the results; it is not subject to patent or protection registrations, and therefore not subject to an embargo;

Date _____31/10/2012______

SURNAME	GARGANO	NAME	ANTONIO

Essays on Return Predictability and Asset Pricing

Dissertation in partial fulfillment of the requirements for the academic degree of Doctor of Philosophy in Finance (XXIII cycle).

Antonio Gargano (1376839)
Università Commerciale
Luigi Bocconi

Thesis Committee:
Carlo A. Favero, Università Commerciale L. Bocconi
Gianluca Fusai, Università del Piemonte Orientale
Barbara Alemanni, SDA Bocconi

Acknowledgments

It is a pleasure to thank those who made this thesis possible. I am grateful to my advisors Allan Timmermann and Carlo Favero for their helpful comments, suggestions and especially for having encouraged me during all these years; to my cohautors Daniele Bianchi, Graham Elliott, Barbara Ostdiek, Brad Paye and Allan Timmermann. Barbara Alemanni and Gianluca Fusai have been very nice in accepting to be part of my thesis committee. Part of this thesis has been written while I was visting at University of California in San Diego, I thank Joey Engelberg, Dongmei Li and Chris Parsons for their help and their suggestions, a special thought goes to Jun Liu, Rossen Valkanov and Alberto Rossi. The Thesis has benefited from comments and helpful feedback from Massimo Guidolin, Davide Pettenuzzo, Roberto Steri and Lawrence Schmidt. Many other people have been great friends in these years: among them, special thanks go to Yuanji, Domenico, Roberto, Carlo, Gerry, Alberto and Riccardo. I will always be thankful to my family and to Ilaria.

I dedicate this Thesis to the people that have believed in me and gave me a chance to prove myself.

Contents

Preface

Return predictability plays a central role in many areas of finance research. Asset pricing models aim to explain why returns are predictable and which implications predictability has for our understanding of the markets. Econometricians investigate which variables predict returns and how the accuracy of the forecasts can be improved. From an asset allocation perspective, is of interest if predictability can be uncovered in asset classes other than stocks. Finally, investors cares if predictability can be exploited to build more efficient portfolios. Each paper in my dissertation looks at predictability from one of these perspectives.

The first chapter, **Return Predictability in Recessions: an Asset Pricing Perspective**, develops a consumption-based asset pricing framework to explain business cycle fluctuations in stock return predictability.
The dividend-price ratio predicts aggregate stock market returns with higher precision during recessions than in expansions in a way that is both statistically and economically significant. I recalibrate three popular asset pricing models[1] using data which include the most recent recessions. While

[1] In the specific, habit-persistence (Campbell and Cochrane (1999)), long-run risk (Bansal and Yaron (2004)) and rare disasters (Gourio (2012)).

these models are able to match the unconditional moments of the equity premium, they fail with respect to the business-cycle fluctuations in return predictability.

Instead, I propose a long-run risk model which links volatility and left tail events in consumption and dividend growth. Recessions are, by definition, periods when economic activity contracts and growth rates are more likely to be negative. Moreover, aggregate volatility is countercyclical and peaks during recessions. It is therefore natural to link these two facts in a model that generates lower growth rates during periods of higher uncertainty. I reproduce this dynamic by adding an asymmetric error term to the growth rates. It follows a Gaussian mixture distribution: with probability p the error is drawn from a normal with negative mean, and with probability $1-p$ from a normal with zero mean. This probability is time-varying and positively correlated with the underlying process for volatility. A positive shock to volatility leads, therefore, to an higher probability of observing negative growth rates, as in the data.

The asymmetry in the error structure reproduces the negative skewness and coskewness observed in the growth rates. Such skewness, for which I provide empirical evidence with a quantile-based measure, has an important theoretical implication: in my model investors demand higher risk pre-

mia when they anticipate negative growth rates. The model also incorporates the negative relation between the conditional mean and variance of growth rates because periods of high volatility are more likely to generate negative growth rates. I provide empirical evidence for coskewness through an AR(1)-GARCH(1) specification.
These three features enables the model to generate higher predictability during recessions. In a bivariate system comprising stock returns and the dividend-price ratio as the sole predictor, I show that higher volatility in the predictor is associated with higher predictability. The introduction of skewness and coskewness increase the variance of the dividend price during recessions, which, because of the asymmetry in the shocks, are periods characterized by high volatility *and* low growth rates.

The second part, **Complete Subset Regressions** (with A. Timmermann and G. Elliott), proposes a new method for combining forecasts based on complete subset regressions. For a given set of potential predictor variables we combine forecasts from all possible linear regression models that keep the number of predictors fixed. From a theoretical perspective, we show that subset regression combinations are akin to a complex version of shrinkage. Unlike the ridge estimator and the usual application of Bayesian estimators, it does not impose the same amount of shrinkage on each coefficient. Unlike model selection methods, it also does not assign binary zero-one weights to the OLS coefficients. In an application to predictability of stock returns, we find that combinations of subset regressions can produce more accurate forecasts than conventional approaches based on equal-weighted forecasts, combinations of univariate forecasts, or forecasts generated by methods such as bagging, ridge regression or Bayesian model averaging.

The third chapter, **Predictive Dynamics in Commodity Prices** (with A. Timmermann), investigates predictability in several dimensions of spot commodity prices: returns, volatility and price increases/decreases. We establish out-of-sample predictability by means of variables such as bond spreads, growth in money supply and industrial production. Predictability is strongest for raw industrials and metals indexes and weakest for foods and textiles. Some variables, such as the inflation rate, have little or no predictive power at the monthly horizon, but appear to have stronger predictive power over commodity spot prices at the quarterly and annual horizons. Our results suggest that predictability of commodity returns from macroeconomic variables such as inflation, industrial production and money supply is stronger during economic recessions than during expansions. This finding carries over to models for realized commodity volatility, where economic state variables add predictive power to a simple autoregression mostly during recessions.

The fourth chapter, **High-Dimensional Index Tracking: Cointegration portfolios and Genetic Algorithms** (with D. Bianchi) looks at predictability from a more practitioners oriented perspective. We present a two-steps procedure to solve active index tracking in high-dimensional spaces. The goal is to track the trajectory of vast indexes instead of the ubiquitous returns. Cointegration between index and a subset of constituent stocks is the key point for the stock picking strategy. The binary stock selection problem is efficiently solved through a Genetic Algorithm by using the Augmented Dickey Fuller test as loss function. The index tracking scheme is completed selecting the fraction of wealth invested in each of stocks through non-linear optimization. Transaction costs are taken into account by introducing a penalization term in the target function. This cointegration-based scheme is tested by using several distance and performance measures, against other, well-established, index tracking procedures on three different datasets. We consider several portfolio sizes and rebalancing horizons. We provide empirical evidences of superior risk-tracking performances of the cointegration based scheme.

Chapter I
Return Predictability in Recessions: an Asset Pricing Perspective

1 Introduction

Aggregate stock market returns appear to be highly predictable during recessions and largely unpredictable in expansions. Using a variety of methodologies ranging from standard OLS to more elaborate Kalman filter and Markov switching models, Rapach, Strauss, and Zhou (2011), Henkel, Martin, and Nardari (2011) and Dangl and Halling (2011) show that business cycle troughs are associated with stronger return predictability in the stock market. These findings provide a new set of stylized facts that can be used to test existing asset pricing models.

Any hypothesis about return predictability is a joint hypothesis about the forecasting model and the variables used as predictors. Variables such as consumption growth and the dividend-price ratio, are endogenously generated in consumption based asset pricing models, and so provide the most direct mapping from theory to empirical data. Using the dividend-price ratio as predictor, and consumption growth as the variable determining the state of the economy, I gauge the amount of predictability that asset pricing models are expected to match. I find that the dividend-price ratio predicts returns with an R^2 of 4.48% during recessions and 1.48% during expansions. The difference in predictability across the two states is statistically significant and translates into sizeable economic gains/losses.

I then recalibrate the Campbell and Cochrane (1999) (CC), Bansal and Yaron (2004) (BY) and Gourio (2012) (G) models using data that include the most recent recessions. While these models are able to match the unconditional moments of the equity premium and the "average" degree of predictability, they fail with respect to the state-dependence in the return predictability observed empirically.

To address such shortcoming, I propose a generalized long-run risk model that matches the empirical evidence. Recessions are, by definition, periods when economic activity contracts and growth rates are more likely to be negative. Furthermore, aggregate volatility is countercyclical and peaks during recessions. I show that these facts generate two empirical regularities neglected in the original Bansal and Yaron (2004) model: negative skewness and coskewness[1] in growth rates. The introduction of skewness, in particular, has important asset pricing implications: investors require higher risk premia when they anticipate negative future consumption growth. I reproduce the link between left tail events and volatility by adding an asymmetric error to growth rates. The asymmetry is captured by a Gaussian mixture distribution where the first (second) component has negative (zero) mean and higher variance. Moreover, the probability that the error is drawn from the first component is time-varying and generates fluctuations in the degree of asymmetry. A positive shock to volatility, therefore increases the probability of observing negative growth rates. This, together with the negative skewness and coskewness in growth rates, enhances the model's ability to generate higher predictability during recessions. In a bivariate system compris-

[1] The covariance between the conditional mean and variance.

ing stock returns and the dividend-price ratio as the sole predictor, I show that higher volatility in the predictor is associated with higher return predictability. Intuitively, the introduction of skewness and coskewness increases the variance of the dividend-price during recessions, which, because of the asymmetry in the shocks, are periods characterized by high volatility *and* low growth rates.

To sum up, my paper makes both empirical and theoretical contributions. Empirically, I show that a) the predictive power of the dividend-price ratio is concentrated in recessions; b) the difference in the predictive accuracy across recessions and expansions is economically and statistically significant and c) both the choice of the predictors and the variables determining the state of the economy are crucial, because they lead to qualitatively similar but quantitatively different results and therefore, to potentially false rejections of the models. Turning to the asset pricing models, a) I show that the empirical evidence is not matched by three popular asset pricing approaches and b) I propose an extended long-run risk model that incorporates this feature of the data.

My analysis is closely related to the empirical literature investigating the time-varying predictability of the equity premium.[2] Using recursive model selection techniques, Pesaran and Timmermann (1995) show that the equity premium is forecastable but that the number, identity and accuracy of the relevant predictors change through time. Even with a constant set of predictors, time-variation in the predictability can arise from breaks in the estimated parameters (Paye and Timmermann (2006)) or in the series itself (Lettau and Nieuwerburgh (2008)). More recently, Rapach, Strauss, and Zhou (2011) (RSZ), Henkel, Martin, and Nardari (2011) (HMN) and Dangl and Halling (2011) (DH) show that time-varying predictability is linked to business cycle fluctuations. My contribution is closest to theirs. However, I address the issue from a more theoretical perspective and ask if this finding can be explained by existing asset pricing models. For this purpose, I cannot use the results of Rapach, Strauss,

[2] See the survey of Zhou (2013).

and Zhou (2011) and Dangl and Halling (2011) because their combined forecasts are based on predictors which do not play a role in consumption-based asset pricing models. The results of Henkel, Martin, and Nardari (2011) are not applicable either because the coefficient estimates of the dividend-price ratio can differ substantially in univariate and multivariate models, see Ang and Bekaert (2007).

The paper is also closely related to the theoretical asset pricing literature that focuses on stock return dynamics over the business cycle. Campbell and Cochrane (1999) introduce time-varying habit persistence in a standard power utility setup. Bansal and Yaron (2004) develop the long-run risk framework, where persistent fluctuations in economic uncertainty is modeled by introducing stochastic volatility in the growth rate dynamics. Barro (2006) and Barro (2009) allow for rare but large declines (i.e. rare disasters) in consumption growth. More recently, several extensions to the long-run risk framework have been proposed with both the intent of solving some of the limitations of the original model (see Beeler and Campbell (2012)) and replicating additional stylized facts.[3] Bonomo, Garcia, Meddahi, and Tedongap (2010) introduce Markov Switching dynamics and Disappointment Aversion preferences. Ghosh and Constantinides (2012) use a similar setup while keeping standard Epstein-Zin preferences and removing the stochastic volatility component. Yaron and Drechsler (2011) add jumps by means of Poisson shocks to match the dynamics of the variance premium, while Bollerslev, Tauchen, and Zhou (2009) specify a process for the volatility of volatility in order to theoretically justify the predictive power of the variance risk premium. Finally, Zhou and Zhu (2012) cast the model in continuous-time and allow for both long- and short-run volatility components. My paper contributes to this literature in two ways: it is the first to rationalize business-cycle fluctuations in predictability and to model skewness and coskewness of growth rates.[4]

[3]With respect to the rare-disaster framework, the focus has concentrated on allowing time-variation in the disaster probability (see Gabaix (2012), Gourio (2012) and Watcher (2012)).

[4]From this perspective my approach is closer to Cecchetti, Lam, and

The paper is organized as follows. Section 2 and 3 cast the empirical evidence on time-varying predictability in an asset pricing context. Section 4 recalibrates the Campbell and Cochrane (1999), Bansal and Yaron (2004) and Gourio (2012) models and show that they do not adequately account for time-varying return predictability. Section 5 illustrates the model and how it matches the empirical evidence. Section 6 concludes.

2 Mapping Theory to Data

This section illustrates why dividend-price ratio and consumption growth are the variables that better map theory to data, and the empirical methodology.

Mark (1990), Cecchetti, Lam, and Mark (1993) and Bekaert and Engstrom (2009).

2.1 The Choice of Variables

The predictive models commonly employed in financial economics take the form of a linear regression

$$r_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}, \tag{1}$$

where r_{t+1} are excess returns on a broad stock market portfolio over a risk free rate, x_t is one of the numerous predictors proposed in the literature[5] and ϵ_{t+1} is a serially uncorrelated and unpredictable error term. The series of fitted risk-premia $\hat{r}_{t+1} = \hat{\alpha} + \hat{\beta} x_t$ mechanically inherits the time-series properties of the predictor variables constituting the information set. Therefore, in order to understand the nature of time-varying risk-premia and thus its relation to the business cycle, it is instructive to analyze the time-series behavior of commonly employed state variables. Figure 1 shows the time-series of the term spread (tms), the default spread (dfy), the dividend price-ratio (dp) and the T-bill rate

[5] See Goyal and Welch (2008) for a comprehensive summary.

(tbl) for the period 1947-2011. The gray shaded areas represent NBER recessions.

The term spread and the default spread display a countercyclical pattern, increasing in recessions and decreasing in expansions. This is expected, given that the first is roughly interpreted as the difference between the current and future short term rates. Short-term rates are lowered by the FED's monetary policy in recessions and driving up the term spread measure. The default spread measures instead the so-called flight-to-quality, generally associated with fear in financial markets. The dividend-price ratio and T-bill rate display a near-integrated behavior and exhibit far weaker relation with the business cycle. In general, however, the first increases during recessions while the latter decreases.

In linear regression models such as (1), cyclical variations in the predictor variables mechanically generate cyclical risk-premia. Less obvious, however, is that the equity premium can be predicted with greater precision during recessions, as highlighted by a number of empirical studies.

Rapach, Strauss, and Zhou (2011) emphasize the benefits of combining univariate forecasts in out-of-sample prediction. In their estimates, predictability is concentrated in good and bad states of the economy as defined by real GDP and corporate profit growth. Henkel, Martin, and Nardari (2011) use a Regime Switching Vector Autoregression (RSVAR) to endogenously identify expansion and recession states, where the first (second) are characterized by low (high) risk premia and volatility. They show that in many of the G7 countries, return predictability in recessions is significantly larger than in expansions. Dangl and Halling (2011) allow for model uncertainty and time-varying coefficients using Bayesian Model Averaging and the Kalman Filter. Based on NBER turning points they distinguish four phases of the business cycle: early and late expansions as well as early and late recessions. They find that the highest degree of predictability is concentrated in late recessions.

While their focus is for the most part empirical, these papers hypothesize that their findings can be rationalized by asset pricing models able to generate time-varying risk-premia. There isn't, however, a clear mapping between their results and the theoretical models invoked, both in terms of the state variables chosen and the definition of bad states. Indeed, most of the predictors adopted are not available in a consumption-based asset pricing framework. Not only, to the best of my knowledge, there aren't models that can endogenously generate recessions and expansions as binary states, because what characterizes good and bad states in consumption-based asset pricing models is the marginal utility of consumption.

Consumption Growth as the State Variable. In a standard Lucas (1978) endowment economy, this can be seen from the Euler Equation:

$$1 = E_t[M_{t+1}R_{t+1}], \tag{2}$$

where the stochastic discount factor $M_{t+1} = \beta\frac{u'(C_{t+1})}{u'(C_t)}$ depends on the agent's subjective discount factor β and the intertemporal marginal rate of substitution $\frac{u'(C_{t+1})}{u'(C_t)}$. By assuming log-normal returns this expression implies an expected risk-premium of

$$E_t[r_{t+1} - r_{f,t}] - 0.5V_t(r_{t+1}) = -cov_t(m_{t+1}, r_{t+1}). \tag{3}$$

Under power utility the covariance on the right hand-side of equation (3) is equal to $\gamma\ cov_t(\Delta c_{t+1}, r_{i,t+1})$ where γ is the coefficient of relative risk aversion. Under Epstein-Zin preferences, the expected risk premium becomes

$$\begin{aligned} E_t[r_{t+1} - r_{f,t}] - 0.5V_t(r_{t+1}) =& \frac{\theta}{\psi} cov_t(\Delta c_{t+1}, r_{t+1}) \\ &+ (1-\theta) cov_t(r_{t+1}, r_{\omega,t+1}), \end{aligned} \tag{4}$$

where ψ is the intertemporal elasticity of substitution, $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}$ and $r_{\omega,t+1}$ is the return on the aggregate wealth. In equilibrium, assets whose returns co-vary positively (negatively) with consumption have a high (low) risk-premium. Therefore, consumption is the variable that better maps the empirical results into an asset pricing framework and good, normal and bad periods should be characterized on the basis of variations in consumption growth.

Dividend Price Ratio as Predictor. Of the numerous equity premium predictors proposed in the literature, the dividend price ratio stands out for its theoretical foundations. Starting from the definition of returns, Campbell and Shiller (1987) show

$$d_t - p_t = -\frac{k}{1-\rho} + E_t\left[\sum_{j=0}^{\infty} \rho^j \left(-\Delta d_{t+1+j} + r_{t+1+j}\right)\right] \qquad (5)$$

which implies that the dividend price ratio *must* forecast either future returns, dividend growth or both.[6] Because of this argument, an acid test for asset pricing models is their ability to match the predictability of the equity premium using the dividend-price ratio. For example, for consumption-based

[6]Empirically Cochrane (2008) shows that the dividend-price ratio does not forecast dividend growth.

asset pricing models the Euler equation can be rewritten as

$$\frac{p_t}{d_t} = E_t\left[\sum_{j=1}^{\infty}\beta^j \frac{u'(c_{t+j})}{u'(c_t)}\frac{d_{t+1}}{d_t}\right]. \tag{6}$$

Depending on the specification of the dividend and consumption growth, equation (6) can either be written in closed form or simulated.[7] Therefore, in order to complete the mapping I use the dividend-price ratio as predictor.

2.2 Empirical Implementation

The first empirical exercise proceeds as follows. I first separate expansions and recessions by sorting the consumption growth variable into terciles, with the lowest tercile denoting recession. I then estimate on the full sample the predictive regression:

$$r_{t+1} = \alpha + \beta dp_t + \epsilon_{t+1}, \tag{7}$$

[7] For example, the specifications in Bonomo and Garcia (1996), Cecchetti, Lam, and Mark (1993) allow for closed form expressions while CC, BY rely either on approximations or on simulations.

and compute the R^2 of each regime by using only the residuals associated with each state

$$R_s^2 = 1 - \frac{\frac{1}{T_s}\sum_{t=1}^{T_s}(r_{s,t+1} - \hat{r}_{s,t+1})^2}{\frac{1}{T_s}\sum_{t=1}^{T_s}(r_{s,t+1} - \bar{r}_{s,t+1})^2} \tag{8}$$

where $\hat{r}_{s,t+1}$ and $\bar{r}_{s,t+1}$ denote the forecast/fitted value of the alternative and the prevailing mean models in each of the two states (expansions or recessions) respectively. Information at time $t+1$, i.e. Δc_{t+1} rather than Δc_t, is used to determine the state of the economy because of the focus on the concurrent dynamics of asset prices and the real economy rather than the real-time predictability of the equity premium.[8] Being ordinal in nature, the classification of states based on consumption growth is necessarily sample dependent. I obviate this problem in two ways. First, I always use the largest sample available to define the states. Second, I repeat the analysis using a well established variable for the state of the economy: the NBER recession indicator. The

[8] To a real-time investor forecasting returns at $t+1$ only Δc_t would be available.

latter allows to conduct the study at the monthly frequency as well.

Inoue and Kilian (2005) and Cochrane (2008) argue that in-sample tests have more power in assessing statistical significance of return predictability, while Goyal and Welch (2008) advocate out-of-sample tests as an additional tool to detect misspecifications. In light of this debate, I provide evidence using both approaches. The in-sample significance is assessed with a bootstrap, whereby the following procedure is repeated 10,000 times: (1) resample T pairs of ($\hat{\epsilon}$,$\hat{\eta}$), with replacement, from OLS residuals in the regressions $r_{t+1} = \alpha + \epsilon_{t+1}$ and $x_{t+1} = \mu + Ax_t + \eta_{t+1}$; (2) build up the time series of predictors, x_t, from the unconditional mean $\hat{\mu}(I - \hat{A})^{-1}$ and iterating forward on x_{t+1} using the OLS estimates $\hat{\mu}$, $\hat{A}$ and the resampled values of $\hat{\eta}_{t+1}$; (3) construct the time series of returns, r_t, by adding the resampled values of $\hat{\epsilon}_{t+1}$ to the sample mean (i.e., under the null that returns are not predictable); and (4) use the resulting series x_t and r_t to estimate the regressions by OLS. The bootstrapped p-values associated with the reported R^2 value is the relative frequency with which the reported R^2 is smaller than its counterparts bootstrapped under the null of no predictability.[9] The out-of sample predictability is assessed with the Clark and West (2006) test

$$cw_{t+1} = (\underbrace{r_{t+1} - \bar{r}_{t+1}}_{\bar{e}_{t+1}})^2 - [(\underbrace{r_{t+1} - \hat{r}_{t+1}}_{\hat{e}})^2 - (\hat{r}_{t+1} - \bar{r}_{t+1})^2]. \quad (9)$$

This corrects the difference in out-of-sample predictive accuracy of two models ($\bar{e}^2_{t+1} - \hat{e}^2_{t+1}$) by their forecasts' variance (third term in the equation). By regressing cw on a constant and calculating the t-statistic corresponding to the constant, a p-value for a one-sided (upper-tail) test is obtained with the standard normal distribution.

The in-sample analysis spans the period from 1947 to

[9]The significance of the R^2 value in the expansionary and recessive states is computed by adding a third equation for the consumption growth, $\Delta c_{t+1} = \theta + \iota_{t+1}$; in step (1) I resample T triples of ($\hat{\epsilon}$,$\hat{\eta}$,$\hat{\iota}$), in step (3) I construct the time series of consumption growth, Δc_t, by adding the resampled values of $\hat{\iota}_{t+1}$ to $\hat{\theta}$, I finally select only the residuals that belong to the state s as defined by the bootstrapped consumption growth process.

2010, while the out-of-sample forecasts are obtained from 1964 to 2010, using an expanding window scheme to obtain parameter estimates. I divide the sample into an in-sample portion composed of the first n observations and an out-of-sample portion composed of the last p observations. The initial out-of-sample forecast of the equity premium is given by $\hat{r}_{n+1} = \hat{\alpha} + \hat{\beta}_n' x_n$ where $\hat{\alpha}_n$ and $\hat{\beta}_n$ are the ordinary least squares (OLS) estimates of α and β, and x_n is the vector containing the predictor variables. The next out-of-sample forecast is given $\hat{r}_{n+2} = \hat{\alpha} + \hat{\beta}_{n+1}' x_{n+1}$ where $\hat{\alpha}_{n+1}$ and $\hat{\beta}_{n+1}$ are generated by regressing $\{r_t\}_{t=2}^{n+1}$ on a constant and $\{x_t\}_{t=2}^{n+1}$. Proceeding in this manner to the end of the out-of-sample period, I generate a series of p out-of-sample forecasts of the equity premium based on x_t, $\{\hat{r}\}_{t=n}^{T-1}$.

3 Results

This section presents the results and their sensitivity with respect to the choice of the predictors and the variables determining the state of the economy.

3.1 The Choice of Predictors

Although the dividend-price ratio is the variable that better maps theory to data, it is of interest to analyze the degree of predictability obtained with different specifications. First, because it allows to assess the robustness of the results; second, because it helps to quantify the effect of choosing the “wrong” variables when testing the models. For this purpose, I use three additional model specifications/forecasting methods. The first is inspired by Henkel, Martin, and Nardari (2011) and uses the dividend yield, the short-rate, the term-spread, the default-spread and lagged returns in a linear model. The second combines with equal weights the forecasts of 15 univariate models using the variables from the Goyal and Welch (2008) dataset as in Rapach, Strauss, and Zhou (2011). The third uses Bayesian Model Averaging in line with Dangl and Halling (2011)[10]. I only use models with constant coefficients because I want to isolate the effect of the state variable(s) and aggregator function on the predictability results.

Table 1 displays the summary statistics for the monthly (Panel A) and quarterly (Panel B) predicted values. Quarterly forecasts are also displayed in Figure 2. Due to the estimation error induced by discarding part of the full in-sample information, out-of-sample forecasts are more volatile and less cross-correlated than in-sample fitted values. By the same token, out-of-sample forecasts are less precise both in terms of directional accuracy[11] and mean squared error. When computed in-sample the bias, measured as the aver-

[10] Bayesian Model Averaging is performed according to the MC^3 algorithm of Raftery, Madigan, and Hoeting (1997) as in Avramov (2002). It averages across a pool of N models selected on the basis of their posterior: $\hat{r}_{t+1} = \sum_{i=1}^{N} \hat{r}_{i,t+1}\omega_i$. The weight is $\omega = \frac{p(m_i|X)}{\sum_{i=1}^{N} p(m_i|X)}$ and p is the posterior probability of each model.

[11] Directional accuracy measures the frequency with which the sign of the equity premium is correctly predicted; it is computed as $\frac{I(r)'I(\hat{r})}{T}$, where r and $\hat{r}$ are respectively the vectors of realized and predicted equity premia, and I is an indicator function taking value 1 if its argument is positive.

age difference between the realized and the predicted values, is always zero while it varies more out-of-sample. By construction the variance of the predicted values increases with the number of predictors:[12] the model that only includes the intercept and the multivariate specification display respectively the lowest and the highest standard deviation,[13] the model including dp as the only predictor and the forecast combination approach lie in between, with the latter behaving similarly to the null model because of the the diversification effect obtained by averaging across the univariate forecasts.

Table 2 shows how these features translate into different testable implications for the time-variation of the equity premium predictability. Panel A reports the baseline results

[12] The forecast variance is

$$var(r_{T+1}) = var(x_T\hat{\beta}) = x_T(X'_{1:T-1}X_{1:T-1}\hat{\sigma}^2_\epsilon)x'_T,$$

which is increasing in the column dimension of x_T.

[13] The fitted values of the null model do not display any in-sample variation because they are a $1 \times (T-1)$ vector where each element equals the sample mean of the equity premium. For the same reason they display zero correlation with the fitted values of the other specifications.

for the dividend-price ratio. The in-sample analysis for consumption growth shows that predictability is concentrated in recessions with an R^2 of 4.48% (0.051 p-value), compared to a statistically insignificant R^2 of 1.48% (p-value of 0.122) in expansions. Using NBER recessions and monthly data leads to similar results: predictability is concentrated in recessions with an R^2 of 1.11% (0.108 p-value) rather than expansions that are characterized by an R^2 of 0.61% (0.09 p-value). On the full sample, the R^2 is equal to 2.26% (0.050 p-value) using quarterly data and 0.75% (0.060 p-value) using monthly data, which is expected given that noise increases with the sampling frequency. The out-of-sample evidence is equally clear-cut as predictability prevails in bad times as defined by NBER recessions or consumption and no-predictability seems to be associated with good times. Two facts are worth noting. First, as expected, the out-of-sample R^2 is smaller than the in-sample value, both at the monthly and quarterly frequency. Second, compared to the in-sample results reported above, the differential in predictability across states is much larger out-of-sample. At the quarterly frequency, the out-of-sample R^2 is 4.73% (0.069 p-value) in bad states and it is -1.12% (0.119 p-value) in expansionary states. At the monthly frequency the out-of-sample R^2 is 1.62% (0.086 p-value) in recessions compared to -0.52% (0.137 p-value) in expansions.

In panels B, C and D of Table 2 the above exercise is repeated for the other forecasting methods. Compared to the specification that only includes the dividend-price, these methods deliver qualitatively similar but quantitatively different results. The multivariate specification displays an higher in-sample R^2, mainly because it includes more predictor variables. This leads to more volatile forecasts and translates into a rather poor out-of-sample performance. While still producing higher predictability in recessions, the forecast combination approach delivers positive R^2 in all states because of the stabilizing effect of forecast combination. Bayesian Model Averaging predictions have the best in-sample fit because assign higher weights to the models with the lowest

Schwarz Information Criterion (Schwarz (1978)). Consequently, the selected models share a similar subset of variables. The lack of diversification across information sources translates into a negative out-of-sample R^2, overall. Nevertheless, the approach confirms that predictability is higher during recessions/bad periods.

This pattern is not limited to the four specifications tested in Table 2. Figure 3 displays the in- and out-of-sample distribution of the R^2 of all possible $2^{13} = 8192$ models obtained combining the variables in Goyal and Welch (2008).[14] Full line denote in- while the dotted lines denote out-of-sample R^2, the two upper plots refer to quarterly data (when expansions and recessions are defined according to consumption growth) while the two bottom plots refer to monthly frequency instead. Notably, in all plots the distribution of recessive R^2 is shifted to the right of the one relative to the ex-

[14] Due to the multicollinearity between the log dividend-price (dp), the log earning-price (ep) and the log dividend-earning (de) ratios, $de = dp - ep$, and the long term yield (lty), the term spread (tms) and the t-bill ($tbill$), $tms = lty - tbill$, I only use 13 predictors out of the 15 originally composing the dataset.

pansionary ones. Not surprisingly, the R^2s obtained with the dividend-price ratio, denoted by the red asterisks on each line, are on the left side of the distribution when computed in-sample,[15] on the right one when computed out-of-sample.

3.1.1 Statistical Significance

Even if individual R^2 are statistically different from zero in each state, as tested in Table 2, the actual *difference* between recessive and expansionary R^2 may not be large enough to warrant that asset pricing models replicate it. I address this concern by testing the null that the predictive ability of the alternative model (with respect to the prevailing mean benchmark) during expansions is no worse than in recessions, against the alternative that states the opposite. More

[15] Because outperformed by the multivariate models.

formally, I test

$$
\begin{aligned}
H_0 : \qquad & E[\underbrace{\bar{\epsilon}_0^2 - \hat{\epsilon}_0^2}_{\Delta_0}] \geq E[\underbrace{\bar{\epsilon}_1^2 - \hat{\epsilon}_1^2}_{\Delta_1}] \\
H_1 : \qquad & E[\bar{\epsilon}_0^2 - \hat{\epsilon}_0^2] < E[\bar{\epsilon}_1^2 - \hat{\epsilon}_1^2],
\end{aligned}
$$

where $\bar{\epsilon}$ and $\hat{\epsilon}$ are, respectively, the error of the prevailing mean and of the alternative model, the subscript refers to expansions (0) and recessions (1). I test this hypothesis by using the sample counterparts of Δ_0 and Δ_1: the average difference between the estimated squared errors of the benchmark ($\bar{e}$) and the alternative model ($\hat{e}$) in expansions (i.e. $\hat{\Delta}_0 = \frac{1}{T_0}\sum_{i=1}^{T_0}(\bar{e}_{0,t+i}^2 - \hat{e}_{0,t+i}^2)$) and recessions (i.e. $\hat{\Delta}_1 = \frac{1}{T_1}\sum_{i=1}^{T_1}(\bar{e}_{1,t+i}^2 - \hat{e}_{1,t+i}^2)$) . I test the null hypothesis in four ways: (i) with the unpaired samples, unequal size and variance t-test

$$
t = \frac{\Delta_0 - \Delta_1}{\sqrt{\frac{\sigma^2_{\Delta_0}}{T_0} + \frac{\sigma^2_{\Delta_1}}{T_1}}}, \tag{10}
$$

where σ^2_{Δ} denotes the variance of the error differentials; (ii) the non parametric version of (10), the Mann-Whitney test based on observations' ranks;[16] iii) the monotonicity test (MT) of Patton and Timmermann (2010),[17] based on the relative frequency with which $\Delta_0 - \Delta_1$ is smaller than its 10,000 counterparts bootstrapped under the null of $\Delta_0 = \Delta_1$; iv) the t-stat of the slope coefficient in the following regression

$$\underbrace{\bar{e}^2_{t+1} - \hat{e}^2_{t+1}}_{\Delta_{t+1}} = \alpha + \beta * NBER_{t+1} + \epsilon_{t+1}. \quad (11)$$

Positive values of Δ_{t+1} imply that the squared errors of

[16] The test statistic and the relative p-value are computed as follows: (i) rank all Δ_t from 1 to T ignoring group membership; (ii) select Δ_0 and Δ_1 with their relative rank; (iii) compute $U = min(U_0, U_1)$, where $U_j = R_j - \frac{n_j(n_j+1)}{2}$ and R_j is the sum of the ranks in the j_{th} group and n_j counts the number of observations; (iv) z= $\frac{U - m_U}{\sigma_U}$ follows a standard normal distribution with $m_U = \frac{n_0 n_1}{2}$ and $\sigma_U = \sqrt{\frac{n_0 n_1 (n_0+n_1+1)}{12}}$; (v) given that a left-sided test is used, p-value are computed as $\Phi(z)$, where $\Phi(\circ)$ is the CDF of a standard normal.

[17] The test p-value is computed as follows i) impose the null of equal-predictability across states i.e. compute $\hat{\Delta}_0 = \Delta_0 - \mu(\Delta_0)$ and $\hat{\Delta}_1 = \Delta_1 - \mu(\Delta_1)$; ii) estimate the distribution under the null: using circular bootstrap of Politis and Romano (1994) take B bootstrap samples from $\hat{\Delta}_0$ and $\hat{\Delta}_1$, and compute for each of them $J^b = \mu(\hat{\Delta}^b_0) - \mu(\hat{\Delta}^b_1)$; iii) Compute p-values (left-sided tail) as $p_{val} = \frac{1}{B}\sum_{b=1}^{B} 1[J > J^b]$ where $J = \mu(\Delta_0) - \mu(\Delta_1)$ is based on the data.

the prevailing mean benchmark are higher than the ones of the alternative model. Therefore, the more negative $\hat{\Delta}_0 - \hat{\Delta}_1$ is, the stronger is the rejection of the null. By the same token, a positive estimate of β in equation (11) leads to the rejection of the null because it implies that during recessions the differential between the squared errors increases. Table 3 presents the results of these tests. For the forecast combination approach the difference in predictability between recessions and expansions is strongly significant both in-sample and out-of-sample, at both frequencies. This is due to the low variance of the forecasts/fitted values driving down the denominator of the t-test. By construction the variance of $\hat{\Delta}_{t+1}$ is equal to $var(\bar{e}^2)+var(\hat{e}^2)-cov(\bar{e}^2,\hat{e}^2)$; the first term is the same for all forecasting approaches, the third is null in-sample and negligible out-of-sample,[18] therefore, what drives the results is the variance of the forecasts on which the second term depends. Indeed, there is a strong relation between the predicted values' standard deviations

[18] Because of the low cross-correlations with the null model as displayed in Table 1

in Table 1 and the values of $\sqrt{\sigma^2_{\Delta_0} + \sigma^2_{\Delta_1}}$ reported in column 9 of Table 3. Although the forecasts/fitted values obtained with the multivariate specification and the Bayesian Model Averaging approach are very volatile, the difference in predictability across states (which drives down the numerator of the t-stat, increasing its significance) is large enough to still deliver significance. With respect to the dividend-price ratio, the only scenario where none of the test deliver any significance is for the out-of-sample R^2 at quarterly horizon. As for the other three forecasting approaches, the test that delivers the highest significance is the Whitney-Mann test.

Overall Tables 1 and 2, and Figures 2 and 3 show that the choice of the information set is key because it leads to quantitatively different testable implications. Therefore, from now on, I focus on the dividend price-ratio.

3.1.2 Economic Significance

Statistical significance might not necessarily translate into economic sizable gains/losses. To test whether this is the case, I use an utility-based measure. I consider the value of the predictions from the perspective of a mean-variance investor who chooses portfolio weights to maximize expected utility. Specifically, I assume that the investor optimally allocates wealth to the aggregate stock market given estimates of the first two conditional moments of the return distribution, $E_t[r_{m,t+1}] - r^f_{t+1}$ and $V_t[r_{m,t+1}]$, where $r_{m,t+1}$ is the market return and r^f_{t+1} is the risk-free rate (T-bill rate). Under mean-variance preferences, this gives rise to an optimal allocation to stocks

$$\omega^*_t = \frac{E_t[r_{m,t+1}] - r^f_{t+1}}{\gamma V_t[r_{m,t+1}]}, \tag{12}$$

where γ captures the investor's risk aversion. Marquering and Verbeek (2004) and Fleming, Kirby, and Ostdiek (2001) determine the economic value of a dynamic strategy based on volatility timing. Since my focus is on excess-returns, I keep the volatility specification constant across the models. Following standard methods in the literature on volatility modeling (see Poon and Granger (2003)), I capture time-variation in volatility $V_t[r_{m,t+1}]$, with (i) a GARCH(1,1) specification and (ii) an AR(1) on the log of a realized volatility measure which takes into account positive correlations in daily returns.[19] The investor's ex-post realized utility is

$$u_{t+1} = r_{f,t+1} + \omega_t^*(r_{m,t+1} - r_{f,t+1}) - 0.5\gamma\omega_t^{*2}\sigma_{t+1}^2. \qquad (13)$$

Finally, I compare the investor's average utility under the modeling approach that includes the dividend-price ratio against the corresponding value under the benchmark prevailing mean model. I report results in the form of the certainty equivalent return (CER), i.e., the return which would leave an investor indifferent between using the prevailing mean forecasts versus the forecasts produced by the dividend-price ratio. Negative (positive) values imply that the prevailing mean investor realizes lower (higher) utility than the dividend-price investor. The difference is computed both in expansions and reces-

[19] Specifically, $\sigma_t^2 = \sum_{i=1}^{N_t}(r_{i,t} - \bar{r})[1 + 2\frac{1}{N_t}\sum_{j=1}^{N_t-1}(N_t - j)\hat{\phi}_t^j]$ where N_t is the number of trading days in month t, $r_{i,t}$ the return on day i in month t and $\hat{\phi}_t$ is the first order autocorrelation coefficient estimated using daily returns within month t (see French, Schwert, and Stambaugh (1989)).

sions. Results are reported in Table 4. At quarterly frequency, when consumption growth is used to determine the state of the economy, the prevailing mean investor generally realizes during recessions lower utility than the dividend-price investor. This result is robust to the choice of γ, the model used to forecast volatility and the presence of the short sales constraint. The only exception is when an autoregressive model is used to forecast future volatility and short sales are not allowed: the differentials are negative also in expansions, however they are still smaller than in recessions. As expected, as γ increases, the differentials shrink because the dividend-price investor is more penalized by the higher volatility of the forecasts. Although on a different scale, the results hold at monthly horizon, when NBER is used to categorize the recessive states. Overall, these numbers guarantee the economic significance of the empirical results.

3.2 The Choice of the State Variables

In consumption-based asset pricing models, consumption growth provides the link between the macro-economy and financial markets because, as stated in equation (11), assets should pay a high premium if they perform poorly in bad times. Consumption growth seems to be the most natural choice to assess the ability of a model to match the different degrees of predictability across the various states of the economy. To test how much the choice of the variable determining the state of the economy, affects the results, I repeat the empirical exercise in Table 2 using several macro variables and compare them to consumption growth. Panel A of Table 5 reports the results for growth in investment, industrial production, income, gross domestic product, and unemployment. Also Using these macro variables the dividend-price ratio delivers higher and more statistically significant R^2 in recessions than in expansions (the only exception is unemployment which delivers positive R^2 in recession as well). This indicates that they share comparable time-series characteristics with consumption, possibly because they are all strongly procyclical. The exact amount of predictability however varies. For example, compared to consumption, GDP delivers relatively low in-sample fit in expansions (0.19% against 1.48%), while unemployment generates a relatively high in-sample fit during recessions (7.82% against 4.48%). This is probably due to the fact that some indicators might be leading while other might be coincident business-cycle indicators. To rule out these potential distortions, I use two economic indicators constructed by aggregating multiple time series: the ADS index[20] of Aruoba, Diebold, and Scotti (2009) and the first three common factors of 148 macro series (described in the appendix) which I computed following Lettau and Ng (2009). Once again the results, displayed in panel B, are only qualitatively in line with consumption growth: ADS and the first factor deliver very high and significant recessive R^2, the differential are larger than the one observed when consumption growth is

[20] Available at http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/

used.

From this I conclude that not only the choice of the predictor variable but also the choice of the business cycle indicator is crucial in determining the amount of predictability the models are expected to match in the different states. The results obtained using variables other than consumption growth and dividend-price may not be directly mapped into standard asset pricing models and may lead therefore to false rejections.

4 Habit Persistence, Long-run Risk, Rare Disasters and Return Predictability

This section re-calibrates three asset pricing models (Campbell and Cochrane (1999), Bansal and Yaron (2004) and Gourio (2012)) using data from 1930 to 2010 and shows that neither is able to generate the time-varying predictability observed in the data.[21] These models are chosen a) due

[21] Note that the models are not able to generate the time-varying predictability even under the original parametrization.

to their ability to generate time-varying risk premia (otherwise they could not generate any predictability) b) because they are the seminal paper of its kind (otherwise the choice of which extended version to examine would have been too discretionary). Campbell & Cochrane (1999) and Bansal & Yaron (2004) are the first papers proposing new frameworks generating counter-cyclical risk premia: habit persistence and long run risk respectively. Therefore, they are natural benchmarks to evaluate. The papers originally proposing the rare disaster framework, Barro (2006) and Barro (2009), do not generate time variation in the risk premia because the disaster probability is kept constant. These two papers, therefore, do not provide any insight with respect to return predictability. Of the papers generalizing this framework modeling the *conditional* behavior of the disaster probability, Gabaix (2012) and Gourio (2012), I recalibrate the latter because it generates counter-cyclical risk premia.

Campbell and Cochrane (1999) The CC model assumes identical agents maximizing:

$$E\sum_{t=0}^{\infty}\delta^t\frac{(C_t - X_t)^{1-\gamma} - 1}{1-\gamma} \tag{14}$$

where C_t denotes consumption, X_t the level of habit and δ the subjective discount factor. The surplus consumption ratio is defined as

$$S_t \equiv \frac{C_t - X_t}{C_t} \tag{15}$$

and its log process is assumed to evolve according to:

$$s_{t+1} = (1-\phi)\bar{s} + \phi\, s_t + \lambda(s_t)(c_{t+1} - c_t - g) \tag{16}$$

where ϕ controls the persistence, g is the growth of consumption and $\bar{s}$ is the (log) long-run habit level. Finally, the $\lambda(s_t)$ function determines how the surplus varies as a function of consumption shocks. This function is chosen to generate a constant risk-free rate, a predetermined habit $s_t = \bar{s}$ and a habit that is a non-negative function of s_t.

Consumption growth is assumed to be be a log-normal iid process:

$$\Delta\, c_{t+1} = g + v_{t+1}, \ \ where \ \ v_{t+1} \sim iid \ \mathcal{N}\left(0, \sigma^2\right) \qquad (17)$$

Introducing habit in the standard power utility framework implies that the local curvature of the utility function, and hence the effective risk-aversion, is a function of the γ coefficient and the level of habit S_t and this generates the countercyclical equity premium of the model. It is not clear, however, how predictability varies over the course of the business cycle because of a lack of closed form expressions for it. It is therefore necessary to simulate the model to obtain answer.

The model is simulated using the parametrization reported in Panel A of Table 6. The mean real per capita consumption growth "g" is set to 1.502 and its standard deviation "σ" to 2.282. Note that the first is smaller than the one in the original Campbell and Cochrane (1999) paper, while the second one is larger. This can be attributed to the fact that the main results of the original paper are calibrated on the rather short postwar sample 1947-1995. The long-run real risk-free rate "r^f" is calibrated to 0.446 and the persistence of the habit process "ϕ" is set to 0.90. Finally, the risk-aversion coefficient γ is set to 1.25 (instead of 2) to match the lower Sharpe ratio (0.27) compared to the one of the original paper (0.43). The implied parameters δ, $\bar{S}$, S_{max} are also reported in Panel A of Table 6.[22]

Bansal and Yaron (2004). The BY model is fundamentally different from the CC model in terms of utility function and underlying processes assumed. It assumes Epstein and Zin (1989) utility which implies the conditional Euler equation

$$E_t \left[\delta^{\theta} \; G_{t+1}^{-\frac{\theta}{\psi}} R_{a,t+1}^{-(1-\theta)} R_{i,t+1} \right] = 1, \tag{18}$$

where G_{t+1} is the gross rate of consumption growth, $R_{a,t+1}$ is the gross return on the asset that pays aggregate consumption as its dividend in every period, δ is the discount factor and $\theta \equiv (1-\gamma)/(1-\frac{1}{\psi})$. The risk-aversion coefficient is de-

[22]for their formulas, please refer to equations 8, 9 and 11 in the original paper.

noted by γ and the intertemporal elasticity of substitution by ψ.

Consumption and dividend growth evolve according to

$$g_{t+1} = \mu + x_t + \sigma_t \ \eta_{t+1} \tag{19}$$

$$g_{d,t+1} = \mu_d + \phi x_t + \ \phi_d \ \sigma_t u_{t+1} \tag{20}$$

Both processes share a persistent component and stochastic volatility defined by

$$x_{t+1} = \rho x_t + \phi_e \sigma_t e_{t+1} \tag{21}$$

$$\sigma_{t+1}^2 = \sigma^2 + \nu_1(\sigma_t^2 - \sigma^2) + \sigma_w w_{t+1} \tag{22}$$

and all errors are independent white noise. I calibrate the model by matching the unconditional moments of consumption and dividend growth to the extended data I have available and report the resulting parameter values in Panel B of Table 6. The updated coefficients are very close to those reported in the BY paper because only few years are added to the original dataset.

Gourio (2012) Gourio (2012) adopts a real business cycle setup with Epstein-Zin preferences and Cobb-Douglas output function

$$Y_t = K_t^{\alpha}(z_t N_t)^{(1-\alpha)}$$

where Y_t, K_t, z_t and N_t are respectively output, capital, productivity and worked hours. Economic disasters are introduced through shocks affecting both capital and the permanent and transitory component of productivity

$$\begin{aligned} K_{t+1} &= \Big((1-\delta)K_t + \phi(\frac{I_t}{K_t})K_t\Big)e^{x_{t+1}\xi_{t+1}} \\ z_{p,t+1} &= \mu + z_{p,t} + \epsilon_{t+1} + x_{t+1}\theta_t \\ z_{r,t+1} &= z_{r,t} + (\phi_{t+1} - \theta_{t+1})x_{t+1} \end{aligned} \tag{23}$$

In equations (23), x_t is an indicator equal to one if there is a disaster ongoing at time t, and zero if not. Whenever a disaster happens, the amount capital and productivity are reduced is determined by three random variables, ξ_{t+1}, θ_{t+1} and ϕ_{t+1}.[23] Finally, the probability of entering a disaster state follows a Markov chain which approximates an AR(1) process: $\pi_t = (1-\rho_\pi)\bar{\pi} + \rho_\pi \pi_{t-1} + \epsilon_t^\pi$. The model is simulated using the parametrization reported in Panel C of Table 6. Compared to the original parametrization, I use lower value for the risk aversion (3.2 instead of 3.8) and higher value for the standard deviation of the log of the disaster probability (3.10 instead of 2.80) in order to match the lower mean and the higher volatility of the equity premium in my sample with respect to the one used in the original paper.

To sum up, the three models generate counter-cyclical risk-premia through different channels. In the CC model the curvature of the utility function, and hence the effective degree of risk-aversion, is inversely related to the level of the surplus consumption ratio. The latter increases during good times and decreases in bad ones. Consumption growth, being iid, does not have a direct effect on the cyclicality of

[23] As in the original Barro (2006) model, conditioning on the disaster happening, the size of the disaster is a random variable

the equity premium. Consumption growth plays instead a central role in BY as it is neither independently nor identically distributed, because of the persistent component in the mean, x_t, and the stochastic volatility term, σ_t. This, together with Epstein-Zin preferences with elasticity of intertemporal substitution greater than 1, generates counter-cyclical risk-premia. Gourio (2012) models a production economy, therefore consumption growth is affected only indirectly through the feasibility constraint.[24] Rather than from features of the consumption growth or the utility function, risk premia are affected by the disaster probability.

Models Performance. Panel A of Table 7 reports, in the first column, the sample moments of equity returns. The values of the equity premium (5.18%), risk-free rate (0.45%) and market Sharpe ratio (0.27) are lower than the ones generally reported in the past literature due to the effects of the latest recession. For the same reason, the volatility is higher (19%). The second, third and fourth columns report the re-

[24] Consumption plus investment can not exceed total output, $C_t + I_t = Y_t$

spective moments generated by the CC, BY and G models. They all match the unconditional moments of the updated sample. The BY model generates a slightly higher risk-free rate than the one observed in the data. The same holds for the risk-premium and volatility.[25]

The models are also good at matching predictability in the equity premium at the one-year horizon (Panel B of Table 7). The sample R^2 using yearly data is 3.09% and is matched rather closely by the CC and G model that report respectively an R^2 of 3.82% and 3.50%. The one associated with the BY model is somewhat lower (1.90%), but definitely close to the one estimated empirically. The coefficients are very close as well, -.95 , -0.9 and -0.74 for the CC, G and BY models, compared to -0.71 estimated on the sample. More crucial for the purpose of this paper is that neither model is able to replicate the time-varying predictability uncovered in the data. The first column of Panel C reports the results on actual data. At the quarterly horizon, pre-

[25] Note that the CC model matches the risk-free rate by construction.

dictability is rather higher in recession than expansions. The G and CC models are able to generate only slightly higher predictability in recessions. Predictability in expansionary states is however higher than the one observed in the data. BY performs poorly because it generates almost the same amount of predictability across states. These results suggest that time-varying predictability is something that must be explicitly incorporated in these asset-pricing models. The lack of closed-form expressions makes it hard to fully understand why and how predictability evolves over the business cycle in the CC model. The mechanism is more transparent in the BY model where log-linearized solutions are available. Moreover, BY model is the one performing the worst, generating almost the same degree of predictability across states. Therefore, I present a generalization of the latter that explicitly accounts for time variation in predictability.

5 The Model

This section presents a consumption-based model with a representative agent that provides a rational explanation for business-cycle fluctuations in predictability. The model builds on the long-run risk literature which relies on Epstein and Zin (1989) recursive preferences, persistent shocks and time-varying volatility. In contrast to the original Bansal and Yaron (2004) framework my model have several additional features.

Recessions are, by definition, periods when economic activity contracts and growth rates are more likely to be negative. Moreover, aggregate volatility is countercyclical and peaks during recessions. It is therefore natural to link these two facts in a model that generates lower growth rates during periods of higher uncertainty. I reproduce this dynamic by adding an asymmetric error term to the growth rates which follows a Gaussian mixture distribution: with probability p the error is drawn from a normal with negative mean, and with probability $1-p$ from a normal with zero mean. This probability is time-varying and positively correlated with the underlying process for volatility. A positive shock to volatility leads, therefore, to an higher probability of observing negative growth rates, as in the data. The asymmetry in the error structure introduces an additional features of the data: the negative skewness in consumption growth. This has an important theoretical implication: investors demand higher risk premia when they anticipate more volatile *and* left skewed growth rates. The model also incorporates the negative relation between the conditional mean and the conditional variance of the growth rates because periods of high volatility are more likely to generate negative growth rates. These three features enhance the model ability to generate higher predictability during recessions. In a bivariate system comprising stock returns and the dividend-price ratio as the sole predictor, I show that higher volatility in the predictor is associated with higher predictability. The introduction of skewness and coskewness increases the variance of the dividend-price during recessions, which, because of the asymmetry in the shocks, are periods characterized by high volatility *and* low growth rates.

5.1 Solution Scheme

In a consumption-based framework, the Equity Premium depends on the covariance between the innovations in the stochastic discount factor (m_{t+1}) and in the market return ($r_{m,t+1}$). More formally,

$$E_t[r_{m,t+1} - r_{f,t+1}] - \frac{1}{2}V_t[r_{m,t+1}] = -\,Cov_t\Big(m_{t+1} - E_t[m_{t+1}], \\ r_{m,t+1} - E_t[r_{m,t+1}]\Big). \quad (24)$$

Because the representative agent has Epstein-Zin preferences, the stochastic discount factor in (24) is

$$m_{t+1} = \theta log(\delta) - \frac{\theta}{\psi}g_{t+1} - (1-\theta)r_{a,t+1}. \quad (25)$$

The parameter δ is the time discount factor, $\theta \equiv \frac{1-\gamma}{1-\frac{1}{\psi}}$, with γ being the risk-aversion and ψ the intertemporal elasticity of substitution. Because the first term in (40) is constant, the innovations in the stochastic discount factor are driven by the shocks to consumption growth (g_{t+1}) and to the (log of) return on the aggregate consumption ($r_{a,t+1}$). The data generating process for g_{t+1} is presented in greater detail in the next subsection. Returns are log-linearized following Campbell and Shiller (1989) and are function of the valuation ratios and the growth rates,

$$r_{a,t+1} = k_0 + k_1 z_{t+1} - z_t + g_{t+1}$$
$$r_{m,t+1} = k_{m,0} + k_{m,1} z_{m,t+1} - z_{m,t} + g_{d,t+1}$$

where k_0, k_1, $k_{m,0}$ and $k_{m,1}$ are linearization constants, $g_{d,t+1}$ is the dividend growth, z_{t+1} and $z_{m,t+1}$ are the valuation ratio. Therefore, in order to fully characterize the model, a solution for the valuation ratios of the two assets is needed. I conjecture, and verify, that the solution is affine in the state variables of the model. These results are available in the Appendix A.1.

5.2 Growth dynamics

As in the original BY model, consumption (g_{t+1}) and dividend $(g_{d,t+1})$ growth fluctuate around their means (μ and μ_d respectively) and share an expected (x_{t+1}) and stochastic volatility (σ^2_{t+1}) component both of which are persistent. The degree of persistence in x_{t+1} is regulated by ρ, while the persistence in σ^2_{t+1} depends on ν_1. In my model, growth rates are also affected by an extra error ϵ_{t+1} which follows a mixture distribution of two Gaussians where the probability (p_{t+1}) of the error to be drawn from the first component is time-varying. More formally,

- Growth rates

$$g_{t+1} = \mu + x_t + \sigma_t \eta_{t+1} + \epsilon_{t+1}$$

$$g_{d,t+1} = \mu_d + \phi x_t + \phi_d \sigma_t u_{t+1} + \epsilon_{t+1}$$

with

$$\epsilon_{t+1} \sim \begin{cases} \mathcal{N}_1(\delta, k_\epsilon \sigma_\epsilon^2) & \text{with probability } p_t \\ \mathcal{N}_2(0, \sigma_\epsilon^2) & \text{with probability } 1 - p_t \end{cases}$$

- Latent components

$$\begin{aligned} x_{t+1} &= \rho x_t + \phi_e \sigma_t e_{t+1} \\ p_{t+1} &= p + v_2(p_t - p) + \sigma_t \sigma_\lambda \lambda_{t+1} \\ \sigma_{t+1}^2 &= \sigma^2 + v_1(\sigma_t^2 - \sigma^2) + \sigma_t \sigma_\omega \omega_{t+1} \end{aligned} \tag{26}$$

The latent components in (34) have an intuitive interpretation, they regulate the first three conditional moments of the growth rates: x_t the conditional mean, σ_t^2 the conditional variance and p_t the conditional asymmetry/skewness.

In the original BY model the error covariance matrix is restricted to be an identity matrix. I relax this restriction and set $\sigma_{\lambda,\omega} > 0$, this implies that a positive shock to volatility leads to an higher probability of observing negative growth rates and therefore to a skewed conditional distribution. I also set $\sigma_{e,\omega} < 0$, this is the channel through which I introduce coskewness. I will show that these modifications i) have economic meaning, ii) are empirically supported by the data iii) help to improve the predictability of returns during recessions.

5.3 Skewness

A feature of the data ignored by the BY model is the negative skew in consumption and dividend growth (see Cecchetti, Lam, and Mark (1990), Cecchetti, Lam, and Mark (1993), Barro (2006) and Bekaert and Engstrom (2009)). Following Ghysel, Plazzi, and Valkanov (2012) I use a robust, quantile-based, measure of asymmetry[26] (see also Kim and White (2004)) defined as

$$RA_\theta(g_t) = \frac{[q_\theta(g_t) - q_{50}(g_t)] - [q_{50}(g_t) - q_{100-\theta}(g_t)]}{[q_\theta(g_t) - q_{100-\theta}(g_t)]}$$

[26] I also compute, the sample counterpart of $\frac{E(g_t-\mu_g)^3}{\sigma_g^3}$ but this measure is very sensible to outliers and harder to interpret.

$q_{100-\theta}(g_t)$, $q_{50}(g_t)$ and $q_\theta(g_t)$ are the $1-\theta$, $50th$, and θ unconditional quantiles of the growth rates and quantile θ is defined as $q_\theta(g_t) = F^{-1}(\theta)$ where F is the empirical cumulative distribution function of g_t. This skewness measure captures asymmetry of quantiles $q_{1-\theta}(g_t)$ and $q_\theta(g_t)$ with respect to the median. The normalization in the denominator ensures that the measure is bounded between -1 and 1. When $RA_\theta(g_t)$ is equal to zero the distribution is symmetric, while values diverging to -1 (1) indicate skewness to the left (right). Figure 5 displays the RA statistic computed on a 10-year rolling window for both dividend (bottom graphs) and consumption (top graphs) growth for $\theta = [95 \;\; 75]$.

For most of the sample, consumption growth displays negative skewness. The series obtained using the $95th-5th$ percentiles and the one using the $75th-25th$ percentiles differ for the first fifteen years where the first increases and the latter decreases. However, both series drop around 1984 and increase at the end of the sample. For dividend growth the negative asymmetry is stronger when using the $75th-$

$25th$ percentiles: during the last twenty years the skewness is almost always negative and reach the minimum of -0.3. When using the $95th - 5th$ percentiles the asymmetry is strongly negative at the beginning and at the end of the sample

Neither GARCH nor stochastic volatility dynamics are sufficient to reproduce this aspect of the data because they both rely on symmetric innovations. I introduce negative skewness by adding an extra error term in the growth rates, ϵ_{t+1}, which is negatively skewed and follows a mixture of two normals: ϵ_{t+1} is draw from a normal $\mathcal{N}(\delta_\epsilon < 0, k_\epsilon \sigma_\epsilon^2)$ with probability p, from a normal $\mathcal{N}(0, \sigma_\epsilon^2)$ with probability $1-p$. The effect of these errors can be shut down by setting $\delta_\epsilon = \sigma_\epsilon^2 = 0$. Figure 6 shows how the mixture of two normals where the first component has lower mean and higher variance than the second, can generate a fatter left tail, inducing therefore asymmetry in the distribution. Conditioning on the moments of the two components being constant, what drives the variation in the degree of asymmetry is the probability that the random variable is drawn from the first component. In the top-left graph, where this probability is 15%, the distribution displays a degree of asymmetry of only -0.11. The asymmetry is increasing in p and reaches the maxium of -0.37 in the bottom right plot which is generated under $p = 75\%$. This is the key mechanism of the model: when p_t in (34) increases, the left tail of the growth distribution becomes fatter, negative observations more likely to happen and the conditional distribution negatively skewed. In my model, the probability that the errors are drawn from the distribution with low mean is time-varying and follows an AR(1) process

$$p_{t+1} = p + v_2(p_t - p) + \sigma_t \sigma_\lambda \lambda_{t+1}. \tag{27}$$

The parameter regulating the the covariance between the shocks to the volatility process and (27 is $\sigma_{\omega,\lambda}$. Because it is set to be positive the model implies that the probability of observing growth observations far in the left tail of the distribution increases when the volatility is higher, as seen in recessions. By adding the state variable p_t, the valuation ratio ratio evolves according to the following affine structure

$$dp_{t+1} = B_0 + B_1 x_{t+1} + B_2 \sigma^2_{t+1} + B_3 p_{t+1}, \tag{28}$$

where

$$\begin{aligned}
B_0 &= \frac{K_{m,a} + \sigma^2(1-v_1)\Big(k_{m,1}B_2 - \bar{\theta}A_2\Big)}{(1-k_{m,1})} \\
&+ \frac{p(1-v_2)\Big(k_{m,1}B_3 - \bar{\theta}A_3\Big)}{(1-k_{m,1})} \\
B_1 &= \frac{\phi - \frac{1}{\psi}}{1 - k_{1,m}\rho} \\
B_2 &= \frac{b_m - \sqrt{b_m^2 - 4a_m c_m}}{2a_m} \\
B_3 &= \frac{-\theta\frac{(\psi-1)}{\psi}\delta_\epsilon - 0.5[\delta_\epsilon^2 + \sigma_\epsilon^2(k_\epsilon - 1)](\theta - \frac{\theta}{\psi})^2}{\theta(v_2 k_{m,2} - 1)}
\end{aligned} \tag{29}$$

The terms in (29) are derived in the Appendix A.1.4. x_{t+1} and σ^2_{t+1} are the same persistent and stochastic volatility components of the growth rates in the Bansal & Yaron model, while p_{t+1} is the probability of having negatively skewed innovations. This latter component represents an extra source of variation in the dividend price ratio, and will be crucial to explain the state-dependent predictability pattern observed in the data. Conditioning on $\psi > 1$, B_1 is positive and B_2 is negative as in the BY model. This follows the intuition that higher expected growth rates increase the valuation ratio while the variance has the opposite effect. Compared with the BY model an extra term, B_3, appears. The sign of B_3 depends on δ_ϵ: $\psi > 1$ and $\gamma > 1$ imply that $\theta \equiv \frac{1-\gamma}{1-\frac{1}{\psi}} < 0$, therefore for $\delta_\epsilon < 0$ and $k_\epsilon > 1$, B_3 is negative. This agrees with the intuition that negative (positive) skewness lowers (increases) the dividend-price ratio: in the same way markets dislike volatility they dislike (like) negative (positive) skewness. The asymmetry in the conditional growth distribution also affects the Equity Premium

$$E_t[r_{m,t+1} - r_{f,t+1}] - \frac{1}{2}V_t[r_{m,t+1}] = \lambda_\sigma \sigma_t^2 + \lambda_\pi p_t + \lambda_\epsilon \sigma_\epsilon^2. \quad (30)$$

λ_σ, λ_π, and λ_ϵ in expression (30) are derived in the Appendix A.1.4, and are positive. Equation (30) states that fluctuations in the equity premium depend not only on the stochastic volatility component (σ_t^2) of the growth rates (as in the original BY model) but also on the variable determining the degree of asymmetry (p_t) in the consumption growth distribution. Investors require higher risk premia when they anticipate negative consumption growth. Because λ_π is positive, the higher the likelihood of this event, the larger is the required compensation. Specifically,

$$\lambda_\pi = -(\theta(\frac{\psi - 1}{\psi}) - 1)\Big(\delta_\epsilon^2 + \sigma_\epsilon^2(k_\epsilon - 1)\Big). \qquad (31)$$

Expression (31) is a positive function of δ_ϵ, σ_ϵ^2 and k_ϵ which, conditioning on a given value of p, increase the degree of asymmetry. Interestingly, the equity premium also depends on σ_ϵ^2 which determines how far in the left tail growth rates can be. As for the stochastic volatility process, the channel driving the *conditional* asymmetry can be eliminated by setting $p = \nu_2 = \sigma_\lambda^2 = 0$ in equation (27), by additionally imposing $\sigma_\epsilon^2 = 0$ also the effect of the *unconditional* asymmetry is neutralized, and equation (30) reduces to the one in the original BY model.

5.4 Coskewness

The effect of coskewness on asset prices has been mainly investigated in the CAPM framework (see Smith (2007) and Harvey and Siddique (2000)). I allow $\sigma_{e,\omega}$, the covariance between the shocks to the persistent and the volatility components of the growth rates, to be different from zero. This implies that $Cov(\sigma_t^2, x_t)$ and $Cov_t(\sigma_{t+1}^2, x_{t+1})$ are different from zero, and introduces therefore conditional and unconditional coskewness between the volatility and the persistent component of consumption and dividend growth. To provide empirical support for this parametrization I estimate the following

AR(1)-GARCH(1,1) model

$$g_{t+1} = \mu + \rho g_t + \sqrt{\sigma^2_{t+1}}\epsilon_{t+1}, \tag{32}$$
$$\sigma^2_{t+1} = \omega + p\sigma^2_t + q\epsilon^2_t.$$

I compute the correlation between the fitted values of the mean (the empirical proxy for x_t in the model, i.e. the expected component of growth rates) and the volatility (the empirical proxy for σ^2_t in the model, i.e. the time-varying economic uncertainty) equations. Both series are measured at monthly and quarterly frequency. Dividend growth is computed as $log\left(\frac{P_{t+1}+D_{t+1}}{P_t} - \frac{P_{t+1}}{P_t}\right) - log\left(\frac{P_t+D_t}{P_{t-1}} - \frac{P_t}{P_{t-1}}\right)$ using cum- and ex-dividend returns on the value weighted NYSE-NASDAQ-AMEX index available on CRSP from 1926 to 2011. Consumption growth is computed using real consumption of non durable goods and services available from BEA, it starts in 1959 for the monthly frequency, in 1947 for the quarterly one. Results are reported in Table 8. At the monthly frequency, both series have an intercept significantly different from zero. While dividend growth has a small but statistically significant persistent component ($\hat{\rho} = 0.11$, s.e. 0.039), consumption does not ($\hat{\rho} = 0.04$, s.e. 0.094). Turning to the volatility equation, dividend growth displays both GARCH ($\hat{p} = 0.824$, s.e. 0.013) and ARCH ($\hat{q} = 0.167$, s.e. 0.017) effects, while only the ARCH coefficient is significant for consumption ($\hat{q} = 0.524$, s.e. 0.150). Similar results hold at the quarterly horizon with both series having a small but significant persistent component ($\hat{\rho} = 0.47$ for dividend and $\hat{\rho} = 0.26$ for consumption growth) and significant ARCH and GARCH effects. More relevant for modeling purposes, in both cases the correlations between the fitted values of the two equations are negative and statistically different from zero. This agrees with the intuition that volatility is higher in recessions when growth rates are lower. Figure 4 displays the fitted values from the mean and volatility equations, it provides a clearer and more immediate evidence of the negative correlation between the two series.

5.5 Time-varying Return Predictability

The model is able to generate time-varying predictability, in particular higher R^2 during recessions. The intuition for this result comes from the following reduced-form model

$$r_{t+1} = \alpha + \beta \, dp_t + \iota_{t+1},$$

$$dp_{t+1} = \gamma + \theta \, dp_t + \eta_{t+1},$$

$$\begin{pmatrix} \iota_{t+1} \\ \eta_{t+1} \end{pmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\iota^2 & \sigma_{\iota\eta} \\ \sigma_{\iota\eta} & \sigma_\eta^2 \end{bmatrix} \right)$$

Systems of this form have been extensively used in the literature on asset allocation under return predictability (e.g., Stambaugh (1986), Kandel and Stambaugh (1996), Barberis (2000), Campbell, Chan, and Viceira (2003), and Pettenuzzo and Timmermann (2011)). Under standard distributional assumptions, the in-sample predictive power of dp_t is given by

$$\begin{aligned} R^2(r_{t+1}|dp_t) &= \frac{Var(E(r_{t+1}|dp_t))}{Var(r_{t+1})} \\ &= \frac{Var(\alpha + dp_t\beta)}{Var(\alpha + dp_t\beta + \iota_{t+1})} \\ &= \frac{\beta^2 Var(dp_t)}{\beta^2 Var(dp_t) + \sigma_\iota^2}. \end{aligned}$$

This is decreasing in the noise of the predictive equation σ_ι^2, increasing in the loading β and in the variance of the dividend-price ratio $V(dp_t)$. The latter equals $\sigma_\eta^2/(1-\theta^2)$ and therefore increases both in the persistence θ and the noise σ_η^2. This suggests that, ceteris paribus, periods where the dividend price ratio is more volatile should have higher return predictability. From equation (28) it is straightforward to obtain the conditional volatility of the dividend-price ratio

$$\begin{aligned} V_t(dp_{t+1}) = & \underbrace{B_1^2 V_t(x_{t+1}) + B_2^2 V_t(\sigma_{t+1}^2)}_{Bansal-Yaron} + \\ & \underbrace{B_3^2 V_t(p_{t+1}) + 2B_2 B_3 Cov_t(\sigma_{t+1}^2, p_{t+1})}_{due-to-skewness} + \\ & \underbrace{2B_1 B_2 Cov_t(\sigma_{t+1}^2, x_{t+1})}_{due-to-coskewness} \\ = & B_1^2 \phi_e^2 \sigma_t^2 + B_2^2 \sigma_\omega^2 \sigma_t^2 + \\ & B_3^2 \sigma_\lambda^2 \sigma_t^2 + 2B_2 B_3 \sigma_t^2 \sigma_\omega \sigma_\lambda \sigma_{\lambda,\omega} + 2B_1 B_2 \sigma_t^2 \sigma_\omega \phi_e \sigma_{e,\omega} \end{aligned} \tag{33}$$

The first two components in equation (33), $B_1^2 V_t(x_{t+1})$ and $B_2^2 V_t(\sigma_{t+1}^2)$, are the same as in the original BY model. The third and fourth terms, $B_3^2 V_t(p_{t+1}) + 2B_2 B_3 Cov_t(\sigma_{t+1}^2, p_{t+1})$, are due to the introduction of negative skewness in the growth rates, while the last term, $2B_1 B_2 Cov_t(\sigma_{t+1}^2, x_{t+1})$, is due to the coskewness. Since $B_2 < 0$, $B_3 < 0$ and $Cov_t(\sigma_{t+1}^2, p_{t+1}) > 0$ the fourth term increases the volatility, since $B_1 > 0$, $B_2 < 0$, and $Cov_t(\sigma_{t+1}^2, x_{t+1}) < 0$ the last term has a positive effect as well. Therefore, in this framework an increase in the volatility σ_t^2 has two effects: (i) it leads to an higher probability of observing growth rates in the left tail of the distribution, because $\sigma_{\omega,\lambda} > 0$ (ii) it leads to an increase in the volatility of the dividend-price ratio as stated in equation (33). This generates more predictability in recessions, exactly when volatility is higher *and* consumption growth is lower. The original BY model can't replicate this pattern because the first channel is shut down, and periods of higher volatility do not correspond to periods of lower consumption growth. Indeed, in the original BY model, even if the volatility is higher, growth rates might be in the right tail of the distribution if the two shocks η_{t+1} and u_{t+1} are positive.

Table 9 displays the values chosen for the calibration. The values for μ, μ_d, ρ and ϕ are set in order to match the unconditional mean and the first order autocorrelation of the growth rates. Similarly, the values for σ^2, ν_1, ϕ_d, ϕ_e and σ_ω are chosen in order to match the unconditional variance of the dividend and consumption growth. The parameters determining the process for p_t are critical due to the fact that being p_t a probability it must be bounded between 0 and 1. For this purpose I set the unconditional value, p, to 0.45 while ν_2 and σ_λ are set in order to p_t be bounded between 0 and 1. The values for δ_ϵ, k_ϵ and σ^2_ϵ are chosen to reproduce the same magnitude of negative asymmetry observed in the data (i.e. to reproduce a minimum RA measure of -0.3). In setting the risk aversion and the discount factor I follow standard guidelines in the literature. As in the original Bansal and Yaron (2004) model, the intertemporal rate of substitution ψ is set to be larger than one in order to have positive values for A_1 and B_1. As shown in Table 10 the model matches very closely the state-dependent pattern in return predictability observed in the data: it generates an R^2 of 1.72% in expansions (compared to 1.48% observed in the data) and an R^2 of 4.87% in recessions (compared to 4.48%); while still matching the standard moments of the Equity premium.

6 Conclusions

Aggregate tock market returns are more predictable in recessions than in expansions. I address this issue from a theoretical perspective and provide a consumption-based explanation. First, I use variables related to consumption-based asset pricing models to determine the amount of predictability a model is expected to generate in different states of the economy. Second, I test if asset pricing frameworks relying on habit-persistence (Campbell and Cochrane (1999)), long-run risk (Bansal and Yaron (2004)) and rare disasters (Gourio (2012)) are able to reproduce the empirical evidence. Because the patterns they generate differ from what is observed, I propose a generalized long-run risk model which links left tail events and volatility and introduces negative skewness and coskewness in the growth rates. These modifications are empirically supported by the data and generate business-cycle fluctuations in predictability by increasing the variance of the dividend price ratio (and so delivering more predictability) during recessions .

This paper is the first to rationalize time variation in return predictability within a consumption-based asset pricing framework and it opens several future perspectives. First, it is of interest whether other mechanisms such as behavioral biases, market frictions or limits to arbitrage could generate the same pattern. A second path of investigation is whether skewness and/or higher moments of growth rates can rationalize other stylized facts observed in the data. It is also of interest to analyze which are the implications of time-varying predictability in an asset allocation exercise. I leave for future research.

References

Ang, A., and G. Bekaert, 2007, “Stock return predictability: is it there?,” *Review of Financial Studies*, 20, 651–707.

Aruoba, S. B., F. X. Diebold, and C. Scotti, 2009, “Real-time measurement of business conditions,” *Journal of Business and Economic Statistics*, 27, 417–427.

Avramov, D., 2002, “Stock return predictability and model uncertainty,” *Journal of Financial Economics*, 64, 423–458.

Bansal, R., and A. Yaron, 2004, “Risks for the long run: a potential resolution of asset pricing puzzles,” *Journal of Finance*, 59, 1481–1509.

Barberis, N., 2000, “Investing for the long run when returns are predictable,” *Journal of Finance*, 55, 225–264.

Barro, R. J., 2006, “Rare Disasters and Asset Markets in the twentieth century,” *The Quaterly Journal of Economics*, 121, 823–866.

——— , 2009, “Rare Disasters, Asset Prices, and Welfare Costs,” *The American Economic Review*, 99, 243–264.

Beeler, J., and J. Y. Campbell, 2012, “The long-run risks model and aggregate asset prices: an empirical assessment,” *Critical Finance*, 1.

Bekaert, G., and E. Engstrom, 2009, “Asset Return Dynamics under Bad Environment Good Environment Fundamentals,” *Working Paper*.

Bollerslev, T., G. Tauchen, and H. Zhou, 2009, “Expected stock returns and variance risk premia,” *Review of Financial Studies*, 22, 4464–4492.

Bonomo, M., and R. Garcia, 1996, “Consumption and equilibrium asset pricing: An empirical assessment,” *Journal of Empirical Finance*, 3, 239–265.

Bonomo, M., R. Garcia, N. Meddahi, and R. Tedongap, 2010, “Generalized Disappointment Aversion, Long-run

Volatility Risk, and Asset Prices," *Review of Financial Studies*, 24, 82–122.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira, 2003, "A multivariate model of strategic asset allocation," *Journal of Financial Economics*, 67, 41–80.

Campbell, J. Y., and J. H. Cochrane, 1999, "By force of habit: a consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107, 205–251.

Campbell, J. Y., and R. J. Shiller, 1987, "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

Cecchetti, S. G., P. S. Lam, and N. C. Mark, 1990, "Mean Reversion in Equilibrium Asset Prices," *The American Economic Review*, 80, 298–418.

Cecchetti, S. G., P. S. Lam, and N. C. Mark, 1993, "The equity premium and the risk-free rate: matching the moments," *Journal of Monetary Economics*, 31, 21–45.

Clark, T. E., and K. D. West, 2006, “Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis,” *Journal of Econometrics*, 135, 155–186.

Cochrane, J. H., 2008, “The dog that did not bark: a defense of return predictability,” *Review of Financial Studies*, 21, 1533–1575.

Dangl, T., and M. Halling, 2011, “Predictive regressions with time-varying coefficients,” *Journal of Financial Economis (forthcoming)*.

Epstein, L. G., and S. E. Zin, 1989, “Substitution, risk aversion and temporal behavior of consumption and asset returns: a thoeretical framework,” *Econometrica*, 57, 937–969.

Fleming, J., C. Kirby, and B. Ostdiek, 2001, “The Economic Value of Volatility Timing,” *Journal of Finance*, 56, 329–352.

French, K., G. Schwert, and R. Stambaugh, 1989, “1989,” *Journal of Financial Economics*, 25, 23–49.

Gabaix, X., 2012, “Variable Rare Disasters: an Exactly Solved Framework for ten Puzzles in Macro-finance,” *The Quarterly Journal of Economics*, 127, 645–700.

Ghosh, A., and G. Constantinides, 2012, “The Predictability of returns with regime shifts in consumption and dividend growth,” *Working Paper*.

Ghysel, E., A. J. Plazzi, and R. Valkanov, 2012, “Conditional Skewness of Stock Market Returns in Developed and Emerging Markets and Its Economic Fundamentals,” *Working Paper*.

Gourio, F., 2012, “Disaster RIsk and Business Cycles,” *The American Economic Review*, forthcoming.

Goyal, A., and H. Welch, 2008, “A comprehensive look at the empirical performance of equity premium prediction,” *Review of Financial Studies*, 18, 101–145.

Harvey, C., and A. Siddique, 2000, “Conditional skewness in asset pricing models tests,” *Journal of Finance*, 65, 1263–1295.

Henkel, S. J., J. S. Martin, and F. Nardari, 2011, “Time-varying shor-horizon predictability,” *Journal of Financial Economics*, 99, 560–580.

Inoue, A., and L. Kilian, 2005, “In-sample and out-of-sample tests of predictability: which one should we use?,” *Econometric Reviews*, 23, 371–402.

Kandel, S., and R. F. Stambaugh, 1996, “On the predictability of stock returns: an asset-allocation perspective,” *Journal of Finance*, 51, 285–424.

Kim, T.-H., and H. White, 2004, “On more robust estimation of skewness and kurtosis,” *Finance Research Letters*, 1, 56–73.

Lettau, S., and S. Ng, 2009, “Macro Factors in Bond Risk Premia,” *Review of Financial Studies*, 22, 5027–5067.

Lettau, S., and S. V. Nieuwerburgh, 2008, “Reconciling the Return Predictability Evidence,” *Review of Financial Studies*, 21, 1607–1652.

Lucas, R. E., 1978, “Asset prices in an exchange economy,” *Econometrica*, 46, 1429–1443.

Marquering, W., and M. Verbeek, 2004, “The Economic Value of Predicting Stock Index Returns and Volatility,” *Journal of financial and quantitative analysis*, 39, 407–429.

Paye, B. S., and A. Timmermann, 2006, “On the instability of predictive relations,” *Journal of Empirical Finance*, 13, 274–315.

Pesaran, H., and A. Timmermann, 1995, “Predictability of stock returns, robustness and economic significance,” *Journal of Finance*, 50(4), 1201–1227.

Pettenuzzo, D., and A. Timmermann, 2011, “Predictability of stock returns and asset allocation under structural breaks," *Journal of Econometrics*, 164, 60–78.

Politis, D. N., and J. P. Romano, 1994, "The Stationary Bootstrap," *Journal of American Statistical Association*, 89, 1303–1313.

Poon, S.-H., and C. W. Granger, 2003, "Forecasting Volatility in Financial Markets: a Review," *Journal of Economic Literature*, 41, 478–539.

Raftery, A., D. Madigan, and J. Hoeting, 1997, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 92(437), 230–247.

Rapach, D., J. Strauss, and G. Zhou, 2011, "Out-of-sample equity prediction: Combination forecasts and links to the real economy," *Review of Financial Studies*, 23(2), 821–862.

Schwarz, G., 1978, "Estimating the dimension of a model," *Annals of Statistics*, 6(2), 461–464.

Smith, D. R., 2007, “Conditional Coskewness and Asset Pricing,” *Journal of Empirical Finance*, 14, 91–119.

Watcher, J., 2012, “Can time-varying risk of rare disasters explain aggregate stock market volatility?,” *Journal of Finance (forthcoming)*.

Yaron, A., and I. Drechsler, 2011, “What’s Vol Got to Do With It?,” *Review of Financial Studies*, 24, 2–45.

Zhou, D. R. . G., 2013, *Handbook of Economic Forecasting*Elsevier, chap. Forecasting Stock Returns, pp. 1–78.

Zhou, G., and Y. Zhu, 2012, “A long-run risks model with long- and short-run volatilities: explaining predictability and volatility risk premium,” *Working Paper*.

Panel A: Monthly

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)				
	Bias (%)		St.dev (%)		Sign (%)		MSE (%)		Correlation				
	In	Out	In	Out	In	Out	In	Out					
dp	0.000	0.041	0.367	0.425	57.366	48.188	0.178	0.199		0.620	0.715	0.607	0.000
multi	0.000	0.077	0.613	0.952	57.236	53.442	0.175	0.202	-0.139		0.843	0.642	0.000
fc	0.000	-0.171	0.098	0.137	58.670	55.797	0.177	0.197	-0.064	0.552		0.914	0.000
bma	0.000	0.099	0.461	0.842	59.322	53.260	0.173	0.202	-0.243	0.755	0.772		0.000
null	0.000	-0.293	0.000	0.125	58.670	56.159	0.179	0.200	0.018	-0.016	0.223	-0.002	

Panel B: Quarterly

	Bias (%)		St.dev (%)		Sign (%)		MSE (%)		Correlation				
	In	Out	In	Out	In	Out	In	Out					
dp	0.000	0.149	1.191	1.397	60.392	50.001	0.611	0.707		0.664	0.696	0.449	0.000
multi	0.000	0.191	1.915	2.790	61.568	51.087	0.589	0.751	-0.038		0.928	0.722	0.000
fc	0.000	-0.541	0.348	0.458	63.137	59.239	0.611	0.691	-0.008	0.673		0.835	0.000
bma	0.000	0.199	1.143	2.939	63.529	51.630	0.586	0.766	-0.136	0.572	0.799		0.000
null	0.000	-0.907	0.000	0.389	63.137	59.782	0.625	0.712	0.020	0.036	0.036	-0.165	

Table 1: **Summary Statistics:** This Table displays the summary statistics for the monthly (Panel A) and quarterly (Panel B) predicted values. Columns 1, 3, 5 and 7 refer to the in-sample fitted values, columns 2, 4, 6 and 8 to the out-of-sample forecasts. Bias is the bias of the predicted values, $\frac{\iota' r - \iota' \hat{r}}{N}$; St.dev. is the standard deviation, $\sqrt{\frac{\hat{r}'\hat{r}}{N}}$; Sign is the directional accuracy, $\frac{I(r)' I(\hat{r})}{N}$; MSE is the Mean Squared Error, $\frac{(r-\hat{r})'(r-\hat{r})}{N}$; r and $\hat{r}$ are the vectors of the realized and predicted returns, I is an indicator function taking value one if its argument is positive, and ι is an $N \times 1$ unit vector. Cross-correlations between fitted values (out-of-sample forecasts) are displayed in the upper- (lower) triangle part of each correlation matrix. *dp* refers to the univariate regression including only the dividend-price ratio, *multi* to the multivariate regression including the variables used in Henkel, Martin and Nardari (2011), *fc* to the forecast combination approach of Rapach, Strauss and Zhou (2010), *bma* to a Bayesian Model Averaging scheme similar in spirit to Dangl and Halling (2012), *null* to the model containing only an intercept. The in-sample window spans from 1947 to 2010, the out-of-sample period starts in 1964 and ends in 2010.

Panel A: Dividend-Price Ratio

	In-Sample R^2			**Out-of-Sample R^2**		
	All	Expansion	Recession	All	Expansion	Recession
Δc	2.26	1.48	4.48	0.71	-1.12	4.73
	(0.050)	(0.122)	(0.051)	(0.039)	(0.119)	(0.069)
NBER	0.75	0.61	1.11	0.15	-0.52	1.62
	(0.060)	(0.091)	(0.108)	(0.054)	(0.137)	(0.086)

Panel B: Multivariate

	In-Sample R^2			**Out-of-Sample R^2**		
	All	Expansion	Recession	All	Expansion	Recession
Δc	5.84	3.95	7.97	-5.50	-14.17	1.83
	(0.004)	(0.046)	(0.035)	(0.098)	(0.341)	(0.069)
NBER	2.10	1.38	3.92	-1.40	-3.50	3.27
	(0.001)	(0.039)	(0.005)	(0.041)	(0.166)	(0.071)

Panel C: Forecast Combination

	In-Sample R^2			**Out-of-Sample R^2**		
	All	Expansion	Recession	All	Expansion	Recession
Δc	2.24	1.80	2.74	2.97	1.63	4.11
	(0.013)	(0.023)	(0.015)	(0.001)	(0.062)	(0.004)
NBER	0.86	0.64	1.42	1.20	0.81	2.07
	(0.052)	(0.083)	(0.061)	(0.000)	(0.007)	(0.015)

Panel D: Bayesian Model Averaging

	In-Sample R^2			**Out-of-Sample R^2**		
	All	Expansion	Recession	All	Expansion	Recession
Δc	6.23	3.91	8.85	-7.54	-16.98	0.44
	(0.375)	(0.711)	(0.329)	(0.088)	(0.224)	(0.127)
NBER	3.21	1.84	6.71	-0.88	-3.15	4.16
	(0.002)	(0.075)	(0.000)	(0.009)	(0.009)	(0.118)

Table 2: **Equity Premium Predictability Across Different States of the Economy.** This Table displays in- and out-of-sample R^2 (in percentage) in expansions and recessions as defined by (the $30th$ percentile of) consumption growth as well as NBER recession index. The results across all periods are also reported. Panel A uses the dividend-price ratio as the only predictor, Panel B uses the specification of Henkel, Martin and Nardari (2011) which includes the *dividend yield*, the *short-rate*, the *term-spread* the *default-spread* and the *lagged excess return*, Panel C uses the forecast combination approach of Rapach, Strauss and Zhou (2010), Panel D uses a Bayesian Model Averaging approach similar in spirit to Dangl and Halling (2012). In all panels coefficients are estimated with Ordinary Least Squares. In-sample p-values (reported in brackets) relative to the null hypothesis of no predictability are computed using bootstrap, out-of-sample p-values relative to the null hypothesis of equal predicting ability are computed using the MSPE-adjusted statistic of Clark and West (2008). The in-sample analysis spans from 1947 to 2010 while the out-of-sample results are based on the 1964-2010 window. The results using consumption data are computed at the quarterly frequency, the ones based on NBER recessions are computed at the monthly frequency instead.

Panel A: Defining States by NBER Indicator

	Panel A.1: In-Sample				**Panel A.2: Out-of-Sample**			
	dividend price	multivariate	forecast combination	bayesian model averaging	dividend price	multivariate	forecast combination	bayesian model averaging
R^2_{Exp}	0.612	1.381	0.643	1.842	-0.523	-3.497	0.809	-3.155
R^2_{Rec}	1.113	3.920	1.422	6.713	1.624	3.274	2.071	4.169
$\frac{\mu_{\Delta_0}-\mu_{\Delta_1}}{\sqrt{\sigma^2_{\Delta_0}+\sigma^2_{\Delta_1}}}$	-0.041 / 0.052	-0.103 / 0.088	-0.035 / 0.015	-0.194 / 0.092	-0.109 / 0.076	-0.191 / 0.217	-0.072 / 0.041	-0.137 / 0.213
MT	0.151	0.068*	0.002***	0.008***	0.049**	0.192	0.043**	0.261
T-test	-1.024	-1.474*	-2.819***	-2.530***	-1.619*	-0.842	-1.666**	-0.611
Whitney	-1.342*	-2.903***	-3.796***	-3.317***	-2.160**	-1.800**	-2.903***	-1.911**
T-stat	1.234	1.948**	3.724***	3.936***	1.795**	1.515*	2.979***	1.137

Panel B: Defining States by Consumption Growth

	Panel B.1: In-Sample				**Panel B.2: Out-of-Sample**			
	dividend price	multivariate	forecast combination	bayesian model averaging	dividend price	multivariate	forecast combination	bayesian model averaging
R^2_{Exp}	1.477	3.948	1.800	3.906	-1.122	-14.171	1.628	-16.984
R^2_{Rec}	4.484	7.968	2.738	8.852	4.729	1.825	4.106	0.441
$\frac{\mu_{\Delta_0}-\mu_{\Delta_1}}{\sqrt{\sigma^2_{\Delta_0}+\sigma^2_{\Delta_1}}}$	-0.278 / 0.238	-0.601 / 0.435	-0.185 / 0.089	-0.690 / 0.322	-0.593 / 0.419	-0.895 / 0.840	-0.454 / 0.186	-0.848 / 0.856
MT	0.116	0.085*	0.026**	0.028**	0.150	0.190	0.032**	0.211
T-test	-1.181	-1.341*	-2.002**	-2.052**	-1.035	-0.874	-1.912**	-0.800
Whitney	-1.698**	-1.214	-2.012**	-2.363***	-0.692	-0.600	-1.915**	-1.189
T-stat	0.952	1.511*	2.279**	2.383***	1.126	0.995	2.426***	0.944

Table 3: **Tests of Equal Predictability in Expansions and Recessions** This Table displays the test statistics relative to the null hypothesis of predictability during expansions being no lower than in recessions: H_0 : $R^2_{Exp} \geq R^2_{Rec}$ against H_1 : $R^2_{Exp} < R^2_{Rec}$ as defined by NBER recession index (Panel A) and (the $30th$ percentile of) consumption growth (Panel B). The first two rows display the R^2 in expansions and recessions, μ_{Δ_i} and $\sigma^2_{\Delta_i}$ are the mean and variance of the difference between the squared errors of the benchmark and the alternative model, expansions (recessions) are denoted by the 1 (0) subscript. The last four rows display respectively the fraction of times the realized $\mu_{\Delta_0} - \mu_{\Delta_1}$ is smaller than its bootstrapped counterparts in the Monotonicity Test of Patton and Timmermann (2010), the unpaired samples, unequal size and variance t-test, the Mann-Whitney test, and the t-stat of the slope coefficient of the regression of $(r_{t+1} - \bar{r}_{t+1})^2 - (r_{t+1} - \hat{r}_{t+1})^2$ on the NBER index and a constant. In each sub-panel, columns 1, 2, 3 and 4 refer respectively to the model containing only the dividend-price ratio, to the model including the predictors used in Henkel, Martin and Nardari (2011), to the forecast combination approach of Rapach, Strauss and Zhou (2010), and to a Bayesian Model Averaging approach similar in spirit to Dangl and Halling (2012). Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level

		NBER				Consumption Growth			
		Panel A: AR(1) on log(VOL)							
		γ				γ			
		1	3	5	10	1	3	5	10
Short	Expansion	0.0660	0.0220	0.0132	0.0066	1.352	0.451	0.270	0.135
	Recession	-1.4185	-0.4728	-0.2837	-0.1418	-9.062	-3.021	-1.812	-0.906
No Short	Expansion	-0.3774	-0.1258	-0.0755	-0.0377	-0.444	-0.148	-0.089	-0.044
	Recession	-1.3110	-0.4370	-0.2622	-0.1311	-8.421	-2.807	-1.684	-0.842
		Panel B: GARCH(1,1)							
		γ				γ			
		1	3	5	10	1	3	5	10
Short	Expansion	0.2544	0.0848	0.0509	0.0254	1.575	0.525	0.315	0.158
	Recession	-1.3549	-0.4516	-0.2710	-0.1355	-6.603	-2.201	-1.321	-0.660
No Short	Expansion	-0.1304	-0.0435	-0.0261	-0.0130	0.279	0.093	0.056	0.028
	Recession	-1.3153	-0.4384	-0.2631	-0.1315	-6.144	-2.048	-1.229	-0.614

Table 4: **Economic Significance:** This Table presents the out-of-sample average utility differentials between two mean-variance investors who optimally allocate their wealth into the aggregate market proportionally to their expectations about future variance, $V_t[r_{m,t+1}]$, excess returns, $E_t[r_{m,t+1} - r_{f,t+1}]$, and their risk aversion, γ: $\omega_t^* = \frac{E_t[r_{m,t+1} - r_{f,t+1}]}{\gamma V_t[r_{m,t+1}]}$. The forecasting methods to predict future variance are the AR(1) on log volatility in Panel A, and the GARCH(1,1) in Panel B. The reported quantity is computed as $\Delta u = \frac{1}{N}\sum_{i=1}^{N}(r_{f,t+1} + \bar{\omega}_t^*(r_{ex,t+1}) - 0.5\gamma\bar{\omega}_t^{*2}VOL_{t+1}) - (r_{f,t+1} + \hat{\omega}_t^*(r_{ex,t+1}) - 0.5\gamma\hat{\omega}_t^{*2}VOL_{t+1})$, where N is the number of observations in the recessive/expansionary state, $\bar{\omega}^*$ is the optimal weight of the prevailing mean investor, and $\hat{\omega}^*$ the one of the dividend-price investor. At the monthly frequency, expansions and recessions are defined by the NBER index, while at the quarterly frequency they are defined by terciles of consumption growth. If short sales are not allowed, weights are restricted to be zero when negative.

Panel A: Macro Variables

	In-Sample		Out-of-Sample	
	Expansion	Recession	Expansion	Recession
Consumption	1.48	4.48	-1.12	4.73
	(0.122)	(0.051)	(0.119)	(0.069)
Investment	1.02	5.36	-1.89	5.51
	(0.261)	(0.021)	(0.156)	(0.032)
Production	0.72	4.90	-1.86	4.31
	(0.261)	(0.033)	(0.147)	(0.048)
Income	0.68	4.98	-3.77	6.84
	(0.288)	(0.038)	(0.266)	(0.008)
GDP	0.19	6.77	-4.05	7.97
	(0.534)	(0.012)	(0.311)	(0.002)
Unempl	1.44	7.82	0.35	2.64
	(0.146)	(0.009)	(0.071)	(0.164)

Panel B: Aggregate Indicators

	In-Sample		Out-of-Sample	
	Expansion	Recession	Expansion	Recession
ADS	-2.38	5.99	-4.34	7.64
	(1.000)	(0.049)	(0.322)	(0.005)
PC(1)	-2.47	7.17	-3.04	7.27
	(1.000)	(0.019)	(0.245)	(0.011)
PC(2)	0.00	4.77	0.69	0.80
	(0.665)	(0.071)	(0.060)	(0.197)
PC(3)	-0.68	5.23	0.22	2.25
	(1.000)	(0.058)	(0.068)	(0.171)

Table 5: **Equity Premium Predictability Across States of the Economy:** This Table displays in- and out-of-sample R^2 in expansions and recessions as defined by (the $30th$ percentile of) several macro variables. *Consumption* is the real consumption growth of non-durables goods and services (FRED mnemonic PCEC), *Investment* is the growth in real fixed private investments (FPIC), *Production* is the growth in real Industrial Production (INDPRO), *GDP* is the growth in the real Gross Domestic Product (GDP), *Unemployment* is the unemployment rate (UNRATE), *ADS* is the high-frequency business cycle indicator of Arouba, Diebold and Scotti (2009), *PC(1)*,*PC(2)* and *PC(3)* are the first three factors obtained from 148 macro time series (described in the appendix) following the methodology of Lettau and Ng (2009). In-sample p-values (reported in brackets) relative to the null hypothesis of no predictability are computed using bootstrap, out-of-sample p-values relative to the null hypothesis of equal predicting ability are computed using the MSPE-adjusted statistic of Clark and West (2008). The in-sample analysis spans from 1947 to 2010 while the out-of-sample results are based on the 1964-2010 window. In all panels predictive regression including dividend-price ratio as the only predictor is used.

Panel A: Habit Persistence [Campbell and Cochrane (1999)]

Parameter	**Variable**	**Value**	**CC(1999)**
Mean Consumption Growth (%)*	g	1.50	1.89
Standard Deviation of Consumption Growth (%)*	σ	2.28	1.50
Log Risk-Free Rate (%)*	r^f	0.45	0.94
Persistence Coefficient*	ϕ	0.90	0.87
Utility Curvature	γ	1.25	2.00

Panel B: Long-Run Risk [Bansal and Yaron (2004)]

Parameter	**Variable**	**Value**	**BY(2004)**
Consumption Growth Process:			
Unconditional Mean (%)	μ	0.150	0.150
Dividend Growth Process:			
Unconditional Mean (%)	μ_d	0.150	0.150
Leverage on Expected Consumption Growth	ϕ	3.5	3
Leverage on Consumption Growth Volatility	ϕ_d	4.67	4.5
Persistent Component:			
Persistency	ρ	0.984	0.979
Leverage on Consumption Growth Volatility	ϕ_e	0.039	0.044
Stochastic Volatility:			
Unconditional Mean	σ^2	0.0067^2	0.0078^2
Persistence coefficient	ν_1	0.987	0.987
Error Variance	σ_ω	$0.23 * 10^{-5}$	$0.23 * 10^{-5}$
Epstein-Zin Utility Parameters:			
Risk-Aversion	γ	10	10
Intertemporal Elasticity of Substitution	ψ	2.0	1.5
Discount Factor	δ	0.998	0.998

Panel C: Rare Disasters [Gourio (2012)]

Parameter	**Variable**	**Value**	**Gourio(2012)**
Capital share	α	.34	.34
Depreciation rate	δ	0.02	0.02
Leisure preference	ν	2.33	2.33
Discount factor	β	.999	.999
Adjustment cost	η	1.7	1.7
Trend growth of TFP	μ	.005	.005
Std. dev. of TFP shock	σ_ϵ	.01	.01
Elasticity of substitution	$1/\psi$	2	2
Risk aversion	γ	3.2	3.8
Mean permanent shock	μ_θ	-0.007	-0.007
Std. dev. permanent shock	σ_θ	0.092	.092
Mean transitory shock	μ_ϕ	-0.055	-0.055
Std. dev. transitory shock	σ_ϕ	0.041	.041
Persistence of productivity	ρ_z	.71	.71
Persistence of disaster state	q	.914	.914
Persistence of log(p)	ρ_p	.90	.90
Std. dev. of log(p)	$\sigma_p/\sqrt{1-\rho^2}$	3.10	2.80
Average prob. of disaster		.72	.72

Table 6: **Calibrated Values** This Table reports the calibrated and assumed parameter values for the Campbell and Cochrane (1999) (Panel A), Bansal and Yaron (2004) (Panel B) and Gourio (2012) (Panel C) models. Left column displays the values of the calibration using data from 1930 to 2010, the right column the ones used in the original paper. For Campbell and Cochrane (1999), the parameter values have been annualized, as in the original paper.

Panel A. Moments of the Equity Premium

Statistic	Data	CC(1999)	BY(2004)	G(2012)
$\mu(r_f)$	0.45%	0.45%	1.50%	0.68%
$E(r-r^f)$	5.18%	5.30%	5.98%	6.15%
$\sigma(r-r^f)$	18.97%	18.27%	19.50%	17.31%
$E(r-r^f)/\sigma(r-r^f)$	0.27	0.29	0.31	0.35

Panel B. One-Year Predictability

Statistic	Data	CC(1999)	BY(2004)	G(2012)
Coefficient ($10\times$)	-0.71	-0.95	-0.74	-0.90
R^2	3.09%	3.82%	1.90%	3.5%

Panel C. Quarterly Time-Varying Predictability (R^2)

State	Data	CC(1999)	BY(2004)	G(2012)
Expansion	1.48%	3.15 %	1.94%	2.99%
Recession	4.48%	3.94 %	2.11%	3.88%

Table 7: **Equity Premium Moments and Predictability.** This Table presents realized and model-generated moments for the equity premium. The theoretical models under consideration are the Campbell and Cochrane (1999), the Bansal and Yaron (2004), and Gourio (2012) all calibrated using data from 1930 to 2010. Panel A reports the results for the risk-free rate r^f, the equity premium $E(r-r^f)$, its volatility $\sigma(r-r^f)$ and its Sharpe ratio $E(r-r^f)/\sigma(r-r^f)$. Panel B reports the results for the one-year predictive regression $r_{t+1} = \alpha + \beta dp_t + \epsilon_{t+1}$. Panel C reports the in-sample R^2 in expansions and recessions as defined by the $30th$ percentile of the consumption growth distribution generated by the models.

Panel A: Monthly Frequency

	Panel A.1: Dividend Growth		**Panel A.2:** Consumption Growth	
	Mean Equation			
	Estimate	(s.e.)	Estimate	(s.e.)
μ	0.0042	0.0003	0.0023	0.0003
ρ	0.1131	0.0390	0.0430	0.0941
	Volatility Equation			
	Estimate	(s.e.)	Estimate	(s.e.)
$\omega(\times 1000)$	0.0091	0.0011	0.0076	0.0020
p	0.8240	0.0133	0.0072	0.1368
q	0.1674	0.0174	0.5243	0.1509
corr	-0.11***		-0.04	

Panel B: Quarterly Frequency

	Panel B.1: Dividend Growth		**Panel B.2:** Consumption Growth	
	Mean Equation			
	Estimate	(s.e.)	Estimate	(s.e.)
μ	0.0062	0.0013	0.0052	0.0006
ρ	0.4748	0.055	0.2695	0.0705
	Volatility Equation			
	Estimate	(s.e.)	Estimate	(s.e.)
$\omega(\times 1000)$	0.0203	0.0082	0.0036	0.0017
p	0.7977	0.0368	0.6788	0.0911
q	0.2020	0.0437	0.224	0.077
corr	-0.13**		-0.06*	

Table 8: **AR(1)-GARCH(1,1) estimates:** This Table displays the estimates of the AR(1)-GARCH(1,1) $g_{t+1} = \mu + \rho g_t + \sqrt{\sigma^2_{t+1}}\epsilon_{t+1}; \sigma^2_{t+1} = \omega + p\sigma^2_t + q\epsilon^2_t$ for both dividend (Panel A) and consumption (Panel B) growth. Both series are measured at the monthly and quarterly frequency, data for dividend growth span from 1929 to 2011 while data for consumption start in 1947 for the quarterly frequency, in 1959 for the monthly one. Dividend growth is computed as $log(\frac{P_{t+1}+D_{t+1}}{P_t} - \frac{P_{t+1}}{P_t}) - log(\frac{P_t+D_t}{P_{t-1}} - \frac{P_t}{P_{t-1}})$ using cum- and ex-dividend returns on the value weighted NYSE-NASDAQ-AMEX index available on CRSP, consumption growth is computed using real consumption of non durable goods and services available from BEA.

Unconditional Mean of Consumption Growth	μ	0.150	Unconditional Probability	p	0.45
Unconditional Mean of Dividend Growth	μ_d	0.150	Persistency of Probability	ν_2	0.975
Leverage on Expected Component	ϕ	3.57	Error Variance of Probability	σ^2_λ	0.12^2
Leverage on Volatility Component	ϕ_d	4.70	First Mixture mean	δ_ϵ	-0.0072
Persistency of x_t	ρ	0.979	Mixture Variance	σ^2_ϵ	0.01^2
Leverage on x_t Volatility	ϕ_e	0.041	Leverage on First Mixture Variance	k_ϵ	1.5
Unconditional Mean of Stochastic Volatility	σ^2	0.0069^2	Risk-Aversion	γ	10
Persistency of Stochastic Volatility	ν_1	0.987	Intertemporal Elasticity of Substitution	ψ	2.0
Error Variance of Stochastic Volatility	σ_ω	$0.22 * 10^{-5}$	Discount Factor	δ	0.998

Table 9: **Calibrated Values** This Table reports the calibrated for the extended long-run risk model

Panel A. Moments of the Equity Premium

Statistic	**Data**	Model
$\mu(r_f)$	0.45%	1.60%
$E(r - r^f)$	5.18%	5.75%
$\sigma(r - r^f)$	18.97%	19.60%
$E(r - r^f)/\sigma(r - r^f)$	0.27	0.29

Panel B. Quarterly Time-Varying Predictability (R^2)

State	**Data**	Model
Expansion	1.48%	1.72%
Recession	4.48%	4.87%

Table 10: **Model Performance** This Table presents realized and model-generated moments for the equity premium. Panel A reports the results for the risk-free rate r^f, the equity premium $E(r - r^f)$, its volatility $\sigma(r - r^f)$ and its Sharpe ratio $E(r - r^f)/\sigma(r - r^f)$. Panel B reports the in-sample R^2 in expansions and recessions as defined by the $30th$ percentile of the consumption growth distribution generated by the model.

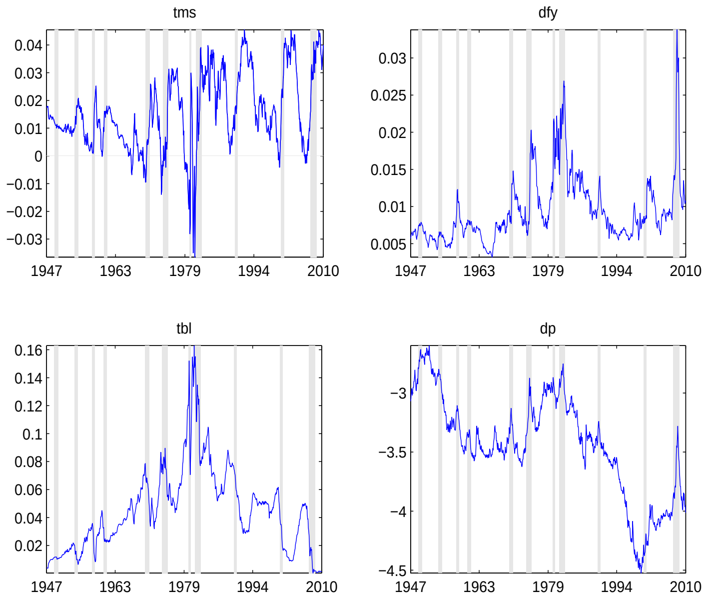


Figure 1: **Business cycle and state variables.** These plots show the time series behavior of four commonly employed predictor variables: the term spread (tms), the default yield (dfy), the t-bill rate (tbl) and the dividend-price ratio (dp). The series are plotted at the monthly frequency over the time-period 1947-2010. The gray shaded ares represent NBER recessions.

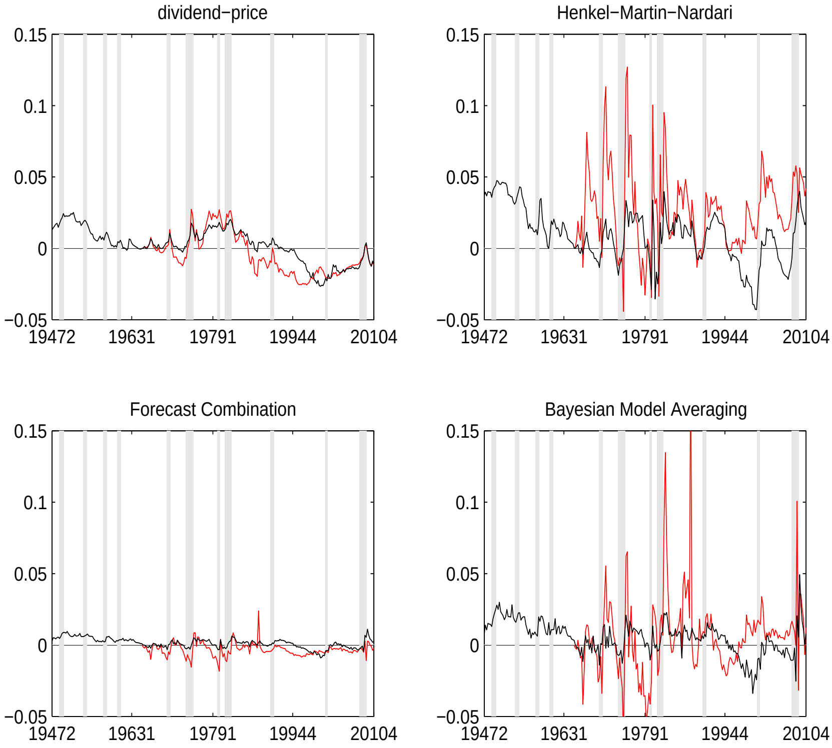


Figure 2: **Forecasts.** This Figure displays the in-sample fitted values (black line) and out-of-sample forecasts (red line) for four different approaches: univariate regression including dividend-price ratio (upper-left), multivariate regression (upper-right) including the variables used in Henkel, Martin and Nardari (2011), forecast combination (lower-left) as in Rapach, Strauss and Zhou (2010), and Bayesian Model Averaging (lower-right) similar to Dangl and Halling (2012). The gray shaded areas represent NBER recessions.

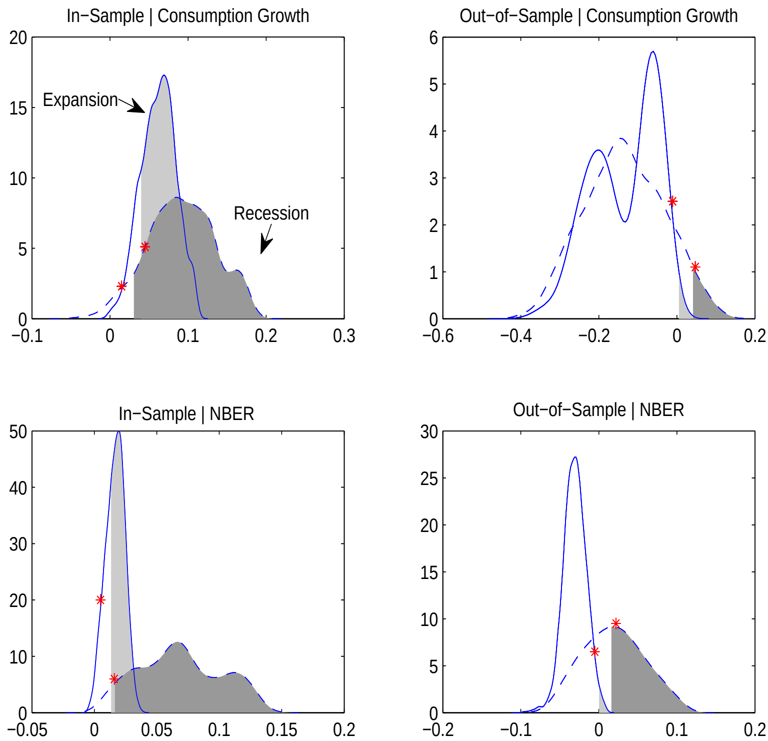


Figure 3: **$\mathbf{R^2}$'s distribution.** This Figure displays the distribution of in- and out-of-sample R^2 in expansions (full line) and recessions (dotted line) obtained with all possible $2^{K=13} = 8192$ models combining 13 of the 15 variables in Goyal & Welch (2008). The log dividend-earning and term spread are excluded to avoid multicollinearity. Red asterisks represent the R^2 of the dividend-price ratio. Grey shaded areas underneath the lines denote the fraction of R^2 statistically different from zero. In-sample p-values relative to the null hypothesis of no predictability are computed using bootstrap, out-of-sample p-values relative to the null hypothesis of equal predicting ability are computed using the MSPE-adjusted statistic of Clark and West (2008). In the top graphs the $30th$ percentile of consumption growth is used to define expansions and recessions while in the bottom ones NBER recession index is used instead.

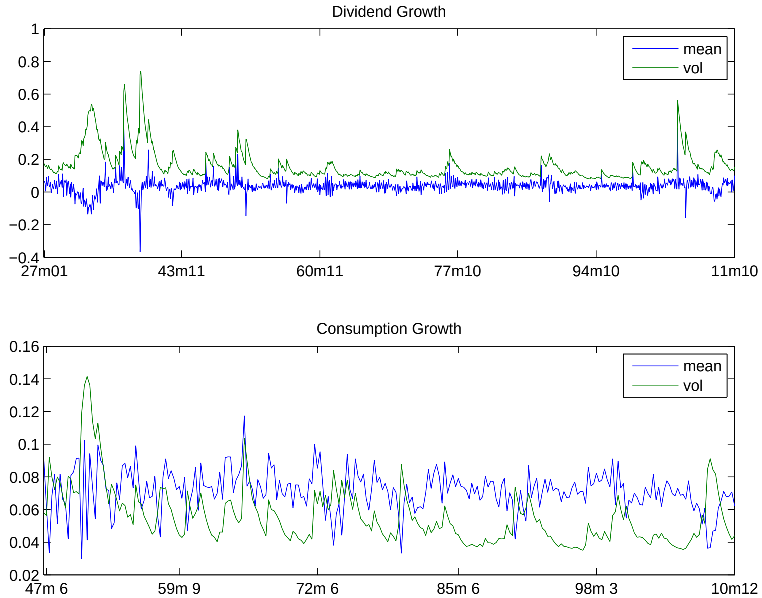


Figure 4: **Coskewness:** This Figure displays the fitted persistent and volatility components of growth rates from the following AR(1)-GARCH(1,1) $g_{t+1} = \mu + \rho g_t + \sqrt{\sigma^2_{t+1}}\epsilon_{t+1}$; $\sigma^2_{t+1} = \omega + p\sigma^2_t + q\epsilon^2_t$ for both dividend and consumption growth. Dividend growth is measured at the monthly frequency, while consumption growth is measured at the quarterly frequency. Dividend growth is computed as $log(\frac{P_{t+1}+D_{t+1}}{P_t} - \frac{P_{t+1}}{P_t}) - log(\frac{P_t+D_t}{P_{t-1}} - \frac{P_t}{P_{t-1}})$ using cum- and ex-dividend returns on the value weighted NYSE-NASDAQ-AMEX index available on CRSP, consumption growth is computed using real consumption of non durable goods and services available from BEA.

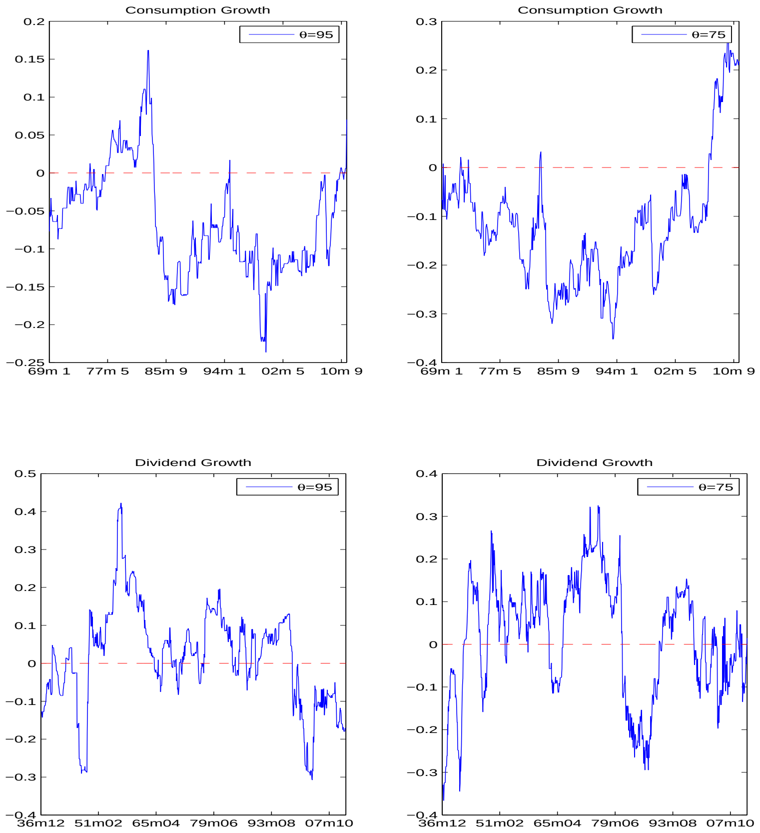


Figure 5: **Skewness:** The two upper (bottom) graphs display Hinkley (1975) robust coefficient of asymmetry, $RA_\theta(g_t) = \frac{[q_\theta(g_t)-q_{50}(g_t)]-[q_{50}(g_t)-q_{100-\theta}(g_t)]}{[q_\theta(g_t)-q_{100-\theta}(g_t)]}$, for consumption (dividend) growth using a 10-year rolling window for the $95th$ percentile ($\theta = 95$, left graphs) and the $75th$ percentile ($\theta = 75$, right graphs). Dividend growth is computed as $log(\frac{P_{t+1}+D_{t+1}}{P_t} - \frac{P_{t+1}}{P_t}) - log(\frac{P_t+D_t}{P_{t-1}} - \frac{P_t}{P_{t-1}})$ using cum- and ex-dividend returns on the value weighted NYSE-NASDAQ-AMEX index available on CRSP. Consumption growth is computed using real consumption of non durable goods and services available from FRED.

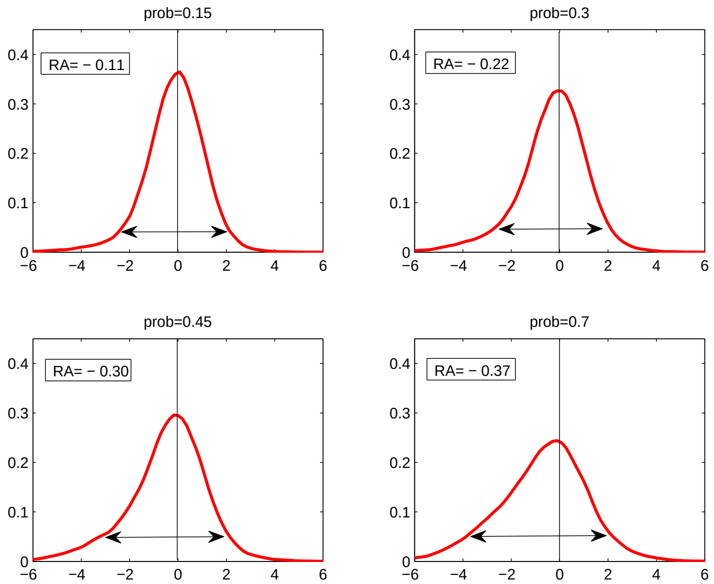


Figure 6: **Mixture of Gaussians:** This Figure shows how the mixture of two normals can be used to generate a fatter left tail, and induce negative asymmetry in the distribution. The probability density function of the mixture is plotted for four different values of p, i.e. the probability that the random variable is drawn from the negative-mean component. The Hinkley (1975) robust coefficient of asymmetry is displayed in the upper-left corner of each plot.

1 Appendix: Model Derivation

This Appendix derives the generalized Bansal & Yaron (2004) model which, by linking left tail events in growth rates to volatility, generates (negative) skewness in consumption and dividend growth, and coskewness between their persistent and volatility components. Sections 1.1 and 1.2 describe the model's dynamics and their conditional moments, Sections 1.3, 1.4 and 1.5 derive the risk premia for the return on aggregate wealth and the market, and the risk free rate.

1.1 Model Dynamics

The main dynamics of the model evolve according to

$$\begin{aligned} g_{t+1} &= \mu + x_t + \sigma_t \eta_{t+1} + \epsilon_{t+1} \\ g_{d,t+1} &= \mu_d + \phi x_t + \phi_d \sigma_t u_{t+1} + \epsilon_{t+1} \\ x_{t+1} &= \rho x_t + \phi_e \sigma_t e_{t+1} \\ p_{t+1} &= p + v_2(p_t - p) + \sigma_t \sigma_\lambda \lambda_{t+1} \\ \sigma^2_{t+1} &= \sigma^2 + v_1(\sigma^2_t - \sigma^2) + \sigma_t \sigma_\omega \omega_{t+1} \end{aligned} \tag{34}$$

with

$$\begin{bmatrix} \eta_{t+1} \\ u_{t+1} \\ e_{t+1} \\ \omega_{t+1} \\ \lambda_{t+1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{\eta,u} & \sigma_{\eta,e} & \sigma_{\eta,\omega} & \sigma_{\eta,\lambda} \\ & 1 & \sigma_{u,e} & \sigma_{u,\omega} & \sigma_{u,\lambda} \\ & & 1 & \sigma_{e,\omega} & \sigma_{e,\lambda} \\ & & & 1 & \sigma_{\omega,\lambda} \\ & & & & 1 \end{bmatrix} \right)$$

and

$$\epsilon_{t+1} \sim \begin{cases} \mathcal{N}(\delta_\epsilon, k_\epsilon \sigma_\epsilon^2) & \text{with probability } p_t \\ \mathcal{N}(0, \sigma_\epsilon^2) & \text{with probability } 1 - p_t \end{cases}$$

g_{t+1} and g_t denote, respectively, consumption and dividend growth; they are both affected by (i) x_{t+1}, the persistent component; (ii) σ^2_{t+1}, the stochastic volatility component; and (iii) p_{t+1}, the probability that the error ϵ_{t+1} is drawn from a normal distribution with mean δ_ϵ. Following the Campbell and Shiller (1989) log-linearization the return of the asset on aggregate consumption (r_a) evolves according to

$$r_{a,t+1} = k_0 + k_1 z_{t+1} - z_t + g_{t+1} \tag{35}$$

while the return on the market (r_m) is

$$r_{m,t+1} = k_{m,0} + k_{m,1} z_{m,t+1} - z_{m,t} + g_{d,t+1} \tag{36}$$

where k_0, k_1, $k_{m,0}$ and $k_{m,1}$ are linearization constants. For their valuation ratio (price over consumption and price over dividend, respectively) I conjecture the following affine forms

$$z_{t+1} = A_0 + A_1 x_{t+1} + A_2 \sigma_{t+1}^2 + A_3 p_{t+1}$$
$$z_{m,t+1} = B_0 + B_1 x_{t+1} + B_2 \sigma_{t+1}^2 + B_3 p_{t+1} \quad (37)$$

Given that consumption and dividend growth, i.e. g_{t+1} and $g_{d,t+1}$, are exogenous, I first solve for the A and B coefficients in (37) on which z_{t+1} and $z_{m,t+1}$ depend. This in turn leads to $r_{a,t+1}$, $r_{m,t+1}$ (equations (35) and (36)) and their conditional moments, from which I obtain the stochastic discount factor (i.e. $m_{t+1} = \theta log(\delta) - \frac{\theta}{\psi} g_{t+1} - (1-\theta) r_{a,t+1}$) and the risk premia of the two assets.

1.2 Conditional Moments

Using standard properties of the conditional expectation and conditional variance[27]

$$
\begin{aligned}
E_t(g_{t+1}) &= \mu + x_t + \delta_\epsilon p_t \\
E_t(g_{d,t+1}) &= \mu_d + \phi x_t + \delta_\epsilon p_t \\
E_t(x_{t+1}) &= \rho x_t \\
E_t(p_{t+1}) &= p + \upsilon_2(p_t - p) \\
E_t(\sigma^2_{t+1}) &= \sigma^2 + \upsilon_1(\sigma^2_t - \sigma^2) \\
V_t(g_{t+1}) &= \sigma^2_t + p_t(\delta^2_\epsilon + \sigma^2_\epsilon(k_\epsilon - 1)) + \sigma^2_\epsilon \qquad (38)\\
V_t(g_{d,t+1}) &= \phi^2_d\sigma^2_t + p_t(\delta^2_\epsilon + \sigma^2_\epsilon(k_\epsilon - 1)) + \sigma^2_\epsilon \\
V_t(x_{t+1}) &= \phi^2_e\sigma^2_t \\
V_t(p_{t+1}) &= \sigma^2_\lambda\sigma^2_t \\
V_t(\sigma^2_{t+1}) &= \sigma^2_\omega\sigma^2_t
\end{aligned}
$$

[27] First: $E_t[ax_{t+1}] = aE_t[x_{t+1}]$, Second: $E_t[x_t] = x_t$, Third: $V_t[ax_{t+1}] = a^2V_t[x_{t+1}]$, Fourth: $V_t[x_t] = 0$

Using standard properties of the conditional covariance[28]

$$C_t(g_{t+1}, g_{d,t+1}) = \sigma_t^2 \phi_d \sigma_{\eta,u} + p_t(\delta_\epsilon^2 + \sigma_\epsilon^2(k_\epsilon - 1)) + \sigma_\epsilon^2$$
$$C_t(g_{t+1}, x_{t+1}) = C_t(\sigma_t \eta_{t+1} + \epsilon_{t+1}, \phi_e \sigma_t e_{t+1}) = \sigma_t^2 \phi_e \sigma_{\eta,e}$$
$$C_t(g_{t+1}, p_{t+1}) = C_t(\sigma_t \eta_{t+1} + \epsilon_{t+1}, \sigma_t \sigma_\lambda \lambda_{t+1}) = \sigma_t^2 \sigma_\lambda \sigma_{\eta,\lambda}$$
$$C_t(g_{t+1}, \sigma_{t+1}^2) = C_t(\sigma_t \eta_{t+1} + \epsilon_{t+1}, \sigma_t \sigma_\omega \omega_{t+1}) = \sigma_t^2 \sigma_\omega \sigma_{\eta,\omega}$$

$$C_t(g_{d,t+1}, x_{t+1}) = C_t(\phi_d \sigma_t u_{t+1} + \epsilon_{t+1}, \phi_e \sigma_t e_{t+1}) = \sigma_t^2 \phi_d \phi_e \sigma_{u,e}$$
$$C_t(g_{d,t+1}, p_{t+1}) = C_t(\phi_d \sigma_t u_{t+1} + \epsilon_{t+1}, \sigma_t \sigma_\lambda \lambda_{t+1}) = \sigma_t^2 \sigma_\lambda \phi_d \sigma_{u,\lambda}$$
$$C_t(g_{d,t+1}, \sigma_{t+1}^2) = C_t(\phi_d \sigma_t u_{t+1} + \epsilon_{t+1}, \sigma_t \sigma_\omega \omega_{t+1}) = \sigma_t^2 \sigma_\omega \phi_d \sigma_{u,\omega}$$
$$C_t(x_{t+1}, p_{t+1}) = C_t(\phi_e \sigma_t e_{t+1}, \sigma_t \sigma_\lambda \lambda_{t+1}) = \sigma_t^2 \sigma_\lambda \phi_e \sigma_{e,\lambda} \qquad (39)$$
$$C_t(x_{t+1}, \sigma_{t+1}^2) = C_t(\phi_e \sigma_t e_{t+1}, \sigma_t \sigma_\omega \omega_{t+1}) = \sigma_t^2 \sigma_\omega \phi_e \sigma_{e,\omega}$$
$$C_t(\sigma_{t+1}^2, p_{t+1}) = C_t(\sigma_t \sigma_\omega \omega_{t+1}, \sigma_t \sigma_\lambda \lambda_{t+1}) = \sigma_t^2 \sigma_\omega \sigma_\lambda \sigma_{\lambda,\omega}$$

The covariances for $r_{a,t}$ and $r_{m,t}$ are less trivial. The only terms in the expressions (35) and (36) that matter when

[28] First: $C_t[ax_{t+1}, by_{t+1}] = abC_t[x_{t+1}, y_{t+1}]$, Second: $C_t[x_{t+1}, k] = 0$, Third: $C_t[ax_{t+1}, y_t] = 0$

computing conditional covariances are

$$r_{a,t+1} = k_1 A_1 x_{t+1} + k_1 A_2 \sigma^2_{t+1} + k_1 A_3 p_{t+1} + g_{t+1}$$
$$r_{m,t+1} = k_{m,1} B_1 x_{t+1} + k_{m,1} B_2 \sigma^2_{t+1} + k_{m,1} B_3 p_{t+1} + g_{d,t+1}$$

1.3 Return on aggregate consumption, R_a

To solve the model for R_a I need to solve its (log) Euler equation. To do so I need the stochastic discount factor that I recover from the first order condition

$$E_t[\underbrace{\delta^\theta G_{t+1}^{-\theta/\psi} R_{a,t+1}^{-(1-\theta)}}_{M_{t+1}} R_{i,t}] = 1$$

from which

$$ln(M_{t+1}) = m_{t+1} = \theta log(\delta) - \frac{\theta}{\psi} g_{t+1} - (1-\theta) r_{a,t+1} \quad (40)$$

The Euler equation for R_a is therefore

$$1 = E_t[M_{t+1} R_{a,t+1}] = E_t[e^{ln(M_{t+1})} e^{ln(R_{a,t+1})}] \quad (41)$$

taking the log and assuming lognormality[29] equation (41) becomes

$$
\begin{aligned}
0 =& E_t[log(M_{t+1}R_{a,t+1})] + 0.5V_t[log(M_{t+1}R_{a,t+1})] \\
=& E_t[m_{t+1}] + E_t[r_{a,t+1}] + 0.5[V_t(m_{t+1}) + V_t(r_{a,t+1}) + \\
& 2C_t(m_{t+1}, r_{a,t+1})]
\end{aligned}
\tag{42}
$$

Using the expressions in (38) and (39) it is trivial to derive $\mathbf{E_t}[\mathbf{m_{t+1}}]$, $\mathbf{E_t}[\mathbf{r_{a,t+1}}]$, $\mathbf{V_t}[\mathbf{m_{t+1}}]$, $\mathbf{V_t}[\mathbf{r_{a,t+1}}]$ and $\mathbf{C_t}[\mathbf{r_{a,t+1}}, \mathbf{m_{t+1}}]$. Plugging these terms in equation (42), and collecting all the x_t terms I obtain

$$0 = -\frac{1}{\psi}x_t + (k_1 A_1 \rho x_t - A_1 x_t + x_t)$$

which leads to

$$A_1 = \frac{1 - \frac{1}{\psi}}{1 - k_1\rho}. \tag{43}$$

[29] If $log(x) \sim N(\mu, \sigma)$ then $x \sim logN\left(e^{\mu+0.5\sigma^2}, (e^{\sigma^2} - 1)(e^{2\mu+\sigma^2})\right)$. From which $E[x] = e^{\mu+0.5\sigma^2}$ $\Rightarrow$log(E[x])=$\mu + 0.5\sigma^2 = E[log(x)] + 0.5Var[log(x)]]$

It is the same as in the BY model and it is positive for $\psi > 1$.

Collecting all the p_t terms I obtain

$$0 = A_3(v_2k_1 - 1)\theta + \delta_\epsilon\theta\frac{(\psi - 1)}{\psi} + \frac{1}{2}\bar{\delta}^2(\theta - \frac{\theta}{\psi})^2$$

which leads to

$$A_3 = \frac{-\theta\delta_\epsilon\frac{\psi-1}{\psi} - \frac{1}{2}\bar{\delta}^2(\theta - \frac{\theta}{\psi})^2}{\theta(v_2k_1 - 1)}$$

where $\bar{\delta}^2 \equiv \delta_\epsilon^2 + \sigma_\epsilon^2(k_\epsilon - 1)$; $\psi > 1$ and $\gamma > 1$ imply that $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}} < 0$, therefore for $\delta_\epsilon < 0$ and $k_\epsilon > 1$, A_3 is negative.

Collecting all the σ_t^2 terms I obtain

$$0 = \sigma_t^2\Big(F + \frac{1}{2}[G + H\sigma_{\eta,e} + I\sigma_\omega^2 + K\sigma_{e,\omega} + \Lambda\sigma_{\eta,\omega} + M\sigma_\lambda^2 + N\sigma_{\omega,\lambda} + O\sigma_{e,\lambda} + P\sigma_{\eta,\lambda}]\Big)$$

with

$$
\begin{aligned}
F &= \theta(k_1 v_1 - 1)A_2 \\
G &= (\theta - \frac{\theta}{\psi})^2 + \theta^2(k_1 A_1 \phi_e)^2 \\
H &= k_1 A_1 \phi_e 2\theta^2 \frac{(\psi - 1)}{\psi} \\
\Lambda &= k_1 A_2 \sigma_\omega 2\theta^2 \frac{(\psi - 1)}{\psi} \\
P &= k_1 A_3 \sigma_\lambda 2\theta^2 \frac{(\psi - 1)}{\psi} \\
I &= k_1^2 A_2^2 \theta^2 \\
M &= k_1^2 A_3^2 \theta^2 \\
K &= 2k_1^2 A_1 A_2 \sigma_\omega \phi_e \theta^2 \\
N &= 2k_1^2 A_2 A_3 \sigma_\omega \sigma_\lambda \theta^2 \\
O &= 2k_1^2 A_1 A_3 \sigma_\lambda \phi_e \theta^2
\end{aligned}
$$

which is a quadratic equation of the form $x^2a + bx + c = 0$ in A_2 where

$$
\begin{aligned}
a =& k_1^2\theta^2\sigma_\omega^2 \\
b =& k_1\sigma_\omega\theta^2\Big(\alpha_1\sigma_{e,\omega} + \alpha_3\sigma_{\omega,\lambda} + \bar{\psi}\sigma_{\eta,\omega}\Big) + \theta(k_1v_1 - 1) \\
c =& \frac{1}{2}\Big((\theta - \frac{\theta}{\psi})^2 + (\theta\alpha_1)^2 + (\theta\alpha_3)^2\Big) \\
& + \theta^2\Big(\bar{\psi}(\alpha_1\sigma_{\eta,e} + \alpha_3\sigma_{\eta,\lambda}) + \alpha_1\alpha_3\sigma_{e,\lambda}\Big)
\end{aligned}
$$

with $\bar{\psi} = \frac{\psi-1}{\psi}$, $\alpha_1 = k_1A_1\phi_e$ and $\alpha_3 = k_1A_3\sigma_\lambda$. I take the negative root

$$A_2 = \frac{b - \sqrt{b^2 - 4ac}}{2a} \tag{44}$$

Collecting all the constant terms I obtain

$$
\begin{aligned}
0 =& \theta log(\delta) - \frac{\theta}{\psi}\mu + \theta k_0 - \theta A_0(1 - k_1) + \theta k_1 A_2\sigma^2(1 - v_1) + \\
& \theta\mu + A_3k_1(p - v_2p)\theta + 0.5\sigma_\epsilon^2\theta^2(1 - \frac{1}{\psi})^2
\end{aligned}
$$

which leads to

$$A_0 = \frac{K_a + A_2 k_1 \sigma^2 (1 - v_1) + A_3 k_1 p (1 - v_2)}{(1 - k_1)} \tag{45}$$

with $K_a = log(\delta) - \frac{\mu}{\psi} + k_0 + \mu + 0.5\sigma_\epsilon^2 \theta (1 - \frac{1}{\psi})^2$ The premium for the return on the aggregate wealth is obtained using the covariance between the innovations in the SDF ($m_{t+1} - E_t[m_{t+1}]$) and in the return ($r_{a,t+1} - E_t[r_{a,t+1}]$)

$$\begin{aligned} E_t[r_{a,t+1} - r_{f,t+1}] - \frac{1}{2} V_t[r_{a,t+1}] &= -Cov_t\Big(m_{t+1} - E_t[m_{t+1}], \\ &\qquad r_{a,t+1} - E_t[r_{a,t+1}]\Big) \\ &= \lambda_\sigma \sigma_t^2 + \lambda_\pi p_t + \lambda_1 \sigma_\epsilon^2 \end{aligned}$$

where

$$\lambda_\pi = \lambda_1 \bar{\delta}^2$$

$$\lambda_\sigma = \sum_{i=1}^{6} \lambda_i$$

$$\lambda_1 = -(\theta(\frac{\psi - 1}{\psi}) - 1)$$

$$\lambda_2 = -(\theta - 1)k_1^2 A_1^2 \phi_e^2$$

$$\lambda_3 = -(\theta - 1)k_1^2 A_2^2 \sigma_\omega^2$$

$$\lambda_4 = -(\theta - 1)k_1^2 A_3^2 \sigma_\lambda^2$$

$$\lambda_5 = -2(\theta - 1)k_1^2 \phi_e \sigma_\omega A_1 A_2 \sigma_{e,\omega}$$

$$\lambda_6 = -2(\theta - 1)k_1^2 \sigma_\omega \sigma_\lambda A_2 A_3 \sigma_{\lambda,\omega}$$

All λ are positive. Because $\theta < 0$ and $\psi > 1$, λ_1 is positive. λ_2, λ_3, λ_4 are positive because $-(\theta - 1) > 0$. λ_5 is positive because $A_1 * A_2$ and $\sigma_{e,\omega}$ are negative while $-(\theta - 1)$ is positive. λ_6 is positive because $A_2 * A_3$ and $\sigma_{\lambda,\omega}$ are positive.

1.4 Return on market, R_m

To solve for R_m I start from its Euler equation

$$\begin{aligned} 0 =& E_t[m_{t+1}] + E_t[r_{m,t+1}] + 0.5[V_t(m_{t+1}) + V_t(r_{m,t+1}) + \\ & 2C_t(m_{t+1}, r_{m,t+1})], \end{aligned} \quad (46)$$

Using the expressions in (38) it is trivial to derive $\mathbf{E_t}[\mathbf{m_{t+1}}]$, $\mathbf{V_t}[\mathbf{m_{t+1}}]$, $\mathbf{E_t}[\mathbf{r_{m,t+1}}]$, $\mathbf{V_t}[\mathbf{r_{m,t+1}}]$, and $\mathbf{C_t}[\mathbf{r_{m,t+1}}, \mathbf{m_{t+1}}]$. Plugging these terms in equation (46) and collecting all the x_t terms I obtain

$$0 = -\frac{x_t}{\psi} + k_{m,1} B_1 \rho x_t - B_1 x_t + \phi x_t$$

which, after substituting the value of A_1, leads to

$$B_1 = \frac{\phi - \frac{1}{\psi}}{1 - k_{1,m}\rho} \quad (47)$$

Collecting all the p_t terms I obtain

$$0 =\delta_\epsilon\theta\frac{(\psi-1)}{\psi} - (1-\theta)A_3(v_2k_1 - 1) + B_3(k_{m,1}v_2 - 1)+$$
$$0.5\bar{\delta}^2(\theta - \frac{\theta}{\psi})^2$$

which, after substituting the value of A_3, leads to

$$B_3 = \frac{-\theta\delta_\epsilon\frac{(\psi-1)}{\psi} - 0.5\bar{\delta}^2(\theta - \frac{\theta}{\psi})^2}{\theta(v_2k_{m,1} - 1)} \tag{48}$$

where $\bar{\delta}^2 \equiv \delta_\epsilon^2 + \sigma_\epsilon^2(k_\epsilon - 1)$; $\psi > 1$ and $\gamma > 1$ imply that $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}} < 0$, therefore for $\delta_\epsilon < 0$ and $k_\epsilon > 1$, B_3 is negative. Collecting all the σ_t^2 terms I obtain

$$0 = \sigma_t^2\Big(F_m + \frac{1}{2}[H_m + I_m\sigma_\omega^2 + K_m\sigma_{e,\omega} + \Lambda_m\sigma_{\eta,e}+$$
$$M_m\sigma_{\eta,\omega} + N_m\sigma_{u,e} + O_m\sigma_{u,\omega} + \Pi_m\sigma_{\eta,u}+$$
$$Q_m\sigma_\lambda^2 + R_m\sigma_{\lambda,\omega} + S_m\sigma_{e,\lambda} + T_m\sigma_{\eta,\lambda} + U_m\sigma_{u,\lambda}]\Big)$$

with

$$
\begin{aligned}
F_m =&(\theta-1)(k_1A_2v_1-A_2)+(k_{m,1}B_2v_1-B_2)\\
H_m =&(\frac{\theta}{\psi}+(1-\theta))^2+(\bar{\theta}A_1\phi_e)^2+(k_{m,1}B_1\phi_e)^2-\\
&2\bar{\theta}k_{1,m}A_1B_1\phi_e^2+\phi_d^2\\
I_m =&[\bar{\theta}^2A_2^2+k_{m,1}^2B_2^2-2(1-\theta)k_1k_{m,1}A_2B_2]\\
Q_m =&[(1-\theta)^2k_1^2A_3^2+k_{m,1}^2B_3^2-2(1-\theta)k_{m,1}k_1A_3B_3]\\
K_m =&\sigma_\omega\phi_e2[\bar{\theta}^2A_1A_2+k_{m,1}^2B_1B_2-\bar{\theta}k_{m,1}A_1B_2-\bar{\theta}k_{m,1}A_2B_1]\\
R_m =&\sigma_\omega\sigma_\lambda2[\bar{\theta}^2A_3A_2+k_{m,1}^2B_3B_2-\bar{\theta}k_{m,1}A_2B_3-\bar{\theta}k_{m,1}A_3B_2]\\
S_m =&\sigma_\lambda\phi_e2[\bar{\theta}^2A_1A_3+k_{m,1}^2B_1B_3-\bar{\theta}k_{m,1}A_1B_3-\bar{\theta}k_{m,1}A_3B_1]\\
\Lambda_m =&\phi_e2[(1-\theta)^2k_1A_1+\frac{\theta}{\psi}\bar{\theta}A_1-\frac{\theta}{\psi}k_{m,1}B_1-(1-\theta)k_{m,1}B_1]\\
M_m =&\sigma_\omega2[(1-\theta)^2k_1A_2+\frac{\theta}{\psi}\bar{\theta}A_2-\frac{\theta}{\psi}k_{m,1}B_2-(1-\theta)k_{m,1}B_2]\\
T_m =&\sigma_\lambda2[(1-\theta)^2k_1A_3+\frac{\theta}{\psi}\bar{\theta}A_3-\frac{\theta}{\psi}k_{m,1}B_3-(1-\theta)k_{m,1}B_3]\\
N_m =&\phi_d\phi_e[2k_{m,1}B_1-2(1-\theta)k_1A_1]\\
O_m =&\phi_d\sigma_\omega[2k_{m,1}B_2-2(1-\theta)k_1A_2]\\
U_m =&\sigma_\lambda\phi_d[2k_{m,1}B_3-2(1-\theta)k_1A_3]\\
\Pi_m =&\phi_d[-2\frac{\theta}{\psi}-2(1-\theta)]
\end{aligned}
$$

which leads to a quadratic equation $ax^2 + bx + c = 0$ in B_2 where

$$
\begin{aligned}
a_m =& k_{m,1}^2 \sigma_\omega^2 \\
b_m =& k_{m,1}\Big(v_1 + \sigma_\omega \xi_m - \bar{\theta} A_2 \sigma_\omega^2\Big) - 1 \\
c_m =& \frac{1}{2}\Big(H_m + \Lambda_m \sigma_{\eta,e} + N_m \sigma_{u,e} + \Pi_m \sigma_{\eta,u} + Q_m \sigma_\lambda^2 + \\
& S_m \sigma_{e,\lambda} + T_m \sigma_{\eta,\lambda} + U_m \sigma_{u,\lambda}\Big) + \\
& \bar{\theta} A_2 \Big(\frac{1}{k_1} - v_1 - \sigma_\omega \xi_m + 0.5 \bar{\theta} A_2 \sigma_\omega^2\Big)
\end{aligned}
$$

with $\bar{\theta} = (1-\theta)k_1$ and $\xi_m = \phi_e \sigma_{e,\omega}(k_{m,1} B_1 - \bar{\theta} A_1) + \sigma_\lambda \sigma_{\omega,\lambda}(k_{m,1} B_3 - \bar{\theta} A_3) + \sigma_{\eta,\omega}(\theta \bar{\psi} - 1) + \phi_d \sigma_{u,\omega}$. I take the negative root:

$$
B_2 = \frac{b_m - \sqrt{b_m^2 - 4 a_m c_m}}{2 a_m} \tag{49}
$$

Collecting all the constant terms I obtain

$$
\begin{aligned}
0 =& K_{m,a} + B_0(k_{m,1} - 1) + \sigma^2(1 - v_1)\Big(k_{m,1} B_2 - (1-\theta) k_1 A_2\Big) + \\
& p(1 - v_2)\Big(k_{m,1} B_3 - (1-\theta) k_1 A_3\Big)
\end{aligned}
$$

which leads to

$$B_0 = \frac{K_{m,a} + \sigma^2(1-v_1)\Big(k_{m,1}B_2 - \bar{\theta}A_2\Big)}{(1-k_{m,1})} + \frac{p(1-v_2)\Big(k_{m,1}B_3 - \bar{\theta}A_3\Big)}{(1-k_{m,1})} \tag{50}$$

with $K_{m,a} = \theta log(\delta) - \frac{\theta}{\psi}\mu + k_{m,0} + \mu_d - (1-\theta)(k_0 + A_0(k_1 - 1) + \mu) + 0.5\sigma_\epsilon^2(\theta - \frac{\theta}{\psi})^2$. The premium for the market return is obtained using the covariance between the innovations in the SDF ($m_{t+1} - E_t[m_{t+1}]$) and in the return ($r_{m,t+1} - E_t[r_{m,t+1}]$)

$$\begin{aligned} E_t[r_{m,t+1} - r_{f,t+1}] - \frac{1}{2}V_t[r_{m,t+1}] &= -Cov_t\Big(m_{t+1} - E_t[m_{t+1}], \\ &\qquad r_{m,t+1} - E_t[r_{m,t+1}]\Big) \\ &= \lambda_\sigma \sigma_t^2 + \lambda_\pi p_t + \lambda_1 \sigma_\epsilon^2 \end{aligned} \tag{51}$$

where

$$\lambda_\pi = \lambda_1 \bar{\delta}^2$$

$$\lambda_\sigma = \sum_{i=1}^{5} \lambda_i$$

$$\lambda_1 = -(\theta(\frac{\psi - 1}{\psi}) - 1)$$

$$\lambda_2 = -(\theta - 1) k_1 k_m A_1 B_1 \phi_e^2$$

$$\lambda_3 = -(\theta - 1) k_1 k_m A_2 B_2 \sigma_\omega^2$$

$$\lambda_4 = -(\theta - 1) k_1 k_m A_3 B_3 \sigma_\lambda^2$$

$$\lambda_5 = -2(\theta - 1) k_1 k_m A_2 B_1 \phi_e \sigma_\omega \sigma_{e,\omega}$$

$$\lambda_6 = -2(\theta - 1) k_1 k_m \sigma_\omega \sigma_\lambda A_3 B_2 \sigma_{\lambda,\omega}$$

θ, A_3, B_3, A_2, B_2 and $\sigma_{e,\omega}$ are negative while A_1, B_1 and $\sigma_{\lambda,\omega}$ are positive. Therefore, all λ have positive sign. Fluctuations in the equity premium not only depend on the underline economic uncertainty but also on the parameters regulating the degree of asymmetry.

2 Appendix: Macro Data Description

Table 11: **Macro Variables:**. This table presents, for each macro variable, its mnemonic (column 2), a brief description of the series (column 3), the source (column 4) and the transformation applied to the series (column 5). In the transformation column, lv denotes the level of the series, and $\Delta^2 lv$ denotes the first difference of the series, ln denotes logarithm, Δln and $\Delta^2 ln$ denote the first and second difference of the logarithm. The data span the period from 1959 to 2010. In the source column, GFD stands for Global Financial Database and BEA for Bureau of Economic Research

N.	Mnemonic	Description	Source	Transf.
1	mtq	SALES - MANUFACTURING and TRADE, CHAINED 1996 DOLLARS (BCI)	Global Insight	Δln
2	ips11	INDUSTRIAL PRODUCTION INDEX - PRODUCTS, TOTAL	Global Insight	Δln
3	ips299	INDUSTRIAL PRODUCTION INDEX - FINAL PRODUCTS	Global Insight	Δln
4	ips12	INDUSTRIAL PRODUCTION INDEX - CONSUMER GOODS	Global Insight	Δln
5	ips13	INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS	Global Insight	Δln
6	ips18	INDUSTRIAL PRODUCTION INDEX - NONDURABLE CONSUMER GOODS	Global Insight	Δln
7	ips25	INDUSTRIAL PRODUCTION INDEX - BUSINESS EQUIPMENT	Global Insight	Δln
8	ips97	INDUSTRIAL PRODUCTION INDEX - NON-ENERGY MATERIALS	Global Insight	Δln
9	ips32	INDUSTRIAL PRODUCTION INDEX - MATERIALS	Global Insight	Δln
10	ips13	INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS	Global Insight	Δln
11	ips38	INDUSTRIAL PRODUCTION INDEX - NONDURABLE GOODS MATERIALS	Global Insight	Δln
12	ips43	INDUSTRIAL PRODUCTION INDEX - MANUFACTURING (SIC)	Global Insight	Δln
13	ipn95	INDUSTRIAL PRODUCTION INDEX - DURABLE MANUFACTURING	Global Insight	Δln
14	IPSGMFN	INDUSTRIAL OUTPUT Industrial Production Index Nondurable manufacturing (NAICS)	Global Insight	Δln
15	ips67	INDUSTRIAL PRODUCTION INDEX - MINING	Global Insight	Δln
16	ips307	INDUSTRIAL PRODUCTION INDEX - RESIDENTIAL UTILITIES	Global Insight	Δln
17	utl10	CAPACITY UTILIZATION - TOTAL INDEX	Global Insight	Δlv
18	utl11	CAPACITY UTILIZATION - MANUFACTURING (SIC)	Global Insight	Δlv
19	utl13	CAPACITY UTILIZATION - DURABLE MANUFACTURING (NAICS)	Global Insight	Δlv
20	utl25	CAPACITY UTILIZATION - NONDURABLE MANUFACTURING (NAICS)	Global Insight	Δlv
21	utl35	CAPACITY UTILIZATION - MINING NAICS=21	Global Insight	Δlv
22	utl36	CAPACITY UTILIZATION - ELECTRIC AND GAS UTILITIES	Global Insight	Δlv
23	yp	PERS INCOME,BILLIONS OF $,SAAR-US	Global Insight	Δln
24	a0m051	PERS INCOME LESS TRSF PMT (AR BIL. CHAIN 2000 $),SA-US	Global Insight	Δln
25	pi031	Personal Consumption Expenditures, Billions of Dollars , SAAR	Global Insight	Δln
26	pi032	Personal Consumption Expenditures - Durable Goods, Billions of Dollars , SAAR	Global Insight	Δln
27	pi033	Personal Consumption Expenditures - Nondurable Goods, Billions of Dollars , SAAR	Global Insight	Δln
28	pi034	Personal Consumption Expenditures - Services, Billions of Dollars , SAAR	Global Insight	Δln
29	n.a.	Motor vehicles and parts: Personal Consumption Expenditures by Type of Product	Global Insight	Δln
30	mtq	SALES - MANUFACTURING & TRADE, CHAINED 1996 DOLLARS (BCI)	Global Insight	Δln
31	ips10	INDUSTRIAL PRODUCTION INDEX - TOTAL INDEX	Global Insight	Δln
32	CES0000000001	ALL EMPL: TOT NFARM,	Global Insight	Δln
33	lhu26	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)	Global Insight	Δln
34	CES003	EMPLOYEES, NONFARM - GOODS-PRODUCING	Global Insight	Δln
35	lhu15	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)	Global Insight	Δln
36	CES0500000001	ALL EMPL: TOT PRIV,	Global Insight	Δln
37	CES011	EMPLOYEES, NONFARM - CONSTRUCTION	Global Insight	Δln
38	lhelx	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF	Global Insight	Δlv
39	lhu5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)	Global Insight	Δln
40	lhu14	UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,SA)	Global Insight	Δln
41	CES046	EMPLOYEES, NONFARM - SERVICE-PROVIDING	Global Insight	Δln
42	CES048	EMPLOYEES, NONFARM - TRADE, TRANSPORT, UTILITIES	Global Insight	Δln
43	CES049	EMPLOYEES, NONFARM - WHOLESALE TRADE	Global Insight	Δln
44	CES088	EMPLOYEES, NONFARM - FINANCIAL ACTIVITIES	Global Insight	Δln
45	CES0700000001	EMPLOYMENT Establishment Data All Employees: Service-Providing	Global Insight	Δln
46	CES140	EMPLOYEES, NONFARM - GOVERNMENT	Global Insight	Δln
47	HPEAP	Average weekly hours of production and nonsupervisory employees. Total private units: Hours, SA	Global Insight	lv
48	CES3000000007	AVG WKLY HR,PROD WORKERS: MFG,	Global Insight	lv
49	CES3000000009	AVG WEEKLY OT,PROD WORKERS: MFG,SA-US	Global Insight	lv
50	lhu680	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA)	Global Insight	Δlv

N.	Mnemonic	Description	Source	Transf.
51	lhur	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%,SA)	Global Insight	Δlv
52	CES033	EMPLOYEES, NONFARM - NONDURABLE GOODS	Global Insight	Δln
53	CES015	EMPLOYEES, NONFARM - MFG	Global Insight	Δln
54	n.a.	Index of help-wanted advertising in newspapers (1967=100; sa)	Conference Board	Δlv
55	CES017	EMPLOYEES, NONFARM - DURABLE GOODS	Global Insight	Δln
56	lhem	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)	Global Insight	Δln
57	lhnag	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)	Global Insight	Δln
58	CES006	EMPLOYEES, NONFARM - MINING	Global Insight	Δln
59	hssou	HOUSING STARTS:SOUTH (THOUS.U.)S.A.	Global Insight	log
60	n.a.	Value of construction put in place: Total construction, SA	Census	Δlv
61	n.a.	Value of construction put in place: Total private construction, SA	Census	Δlv
62	n.a.	Value of construction put in place: Total public construction, SA	Census	Δlv
63	condo9	CONSTR CONTRACTS,COML&IND BLDG (BCD),SA-US	Global Insight	log
64	n.a.	Houses For Sale by Region and Months' Supply at Current Sales Rate, for sale at end of period, SA	Census	log
65	n.a.	Houses For Sale by Region and Months' Supply at Current Sales Rate, months' supply at current sales rate, SA	Census	log
66	n.a.	Houses sold during period, NSA	Census	log
67	hsbr	HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)	Global Insight	log
68	hswst	HOUSING STARTS:WEST (THOUS.U.)S.A.	Global Insight	log
69	hmob	MOBILE HOMES: MANUFACTURERS' SHIPMENTS (THOUS.OF UNITS,SAAR)	Global Insight	log
70	hsmw	HOUSING STARTS:MIDWEST(THOUS.U.)S.A.	Global Insight	log
71	hsne	HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.	Global Insight	log
72	hsfr	HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)(THOUS.,SA)	Global Insight	log
73	n.a.	Real Manufacturing and Trade Inventories Seasonally Adjusted End of Period	BEA	Δln
74	n.a.	Real Manufacturing and Trade Inventories Seasonally Adjusted End of Period, business	BEA	Δln
75	n.a.	Real Manufacturing and Trade Inventories Seasonally Adjusted End of Period,business durable	BEA	Δln
76	n.a.	Real Manufacturing and Trade Inventories Seasonally Adjusted End of Period, business nondurable	BEA	Δln
77	n.a.	Real Manufacturing and Trade Inventories Seasonally Adjusted End of Period, wholesalers	BEA	Δln
78	n.a.	Real Manufacturing and Trade Inventories Seasonally Adjusted End of Period, retail	BEA	Δln
79	pmi	PURCHASING MANAGERS' INDEX (SA)	Global Insight	lv
80	pmp	NAPM PRODUCTION INDEX (PERCENT)	Global Insight	lv
81	pmno	NAPM NEW ORDERS INDEX (PERCENT)	Global Insight	lv
82	pmdel	NAPM VENDOR DELIVERIES INDEX (PERCENT)	Global Insight	lv
83	pmnv	NAPM INVENTORIES INDEX (PERCENT)	Global Insight	lv
84	pmemp	NAPM EMPLOYMENT INDEX (PERCENT)	Global Insight	lv
85	pmcp	NAPM COMMODITY PRICES INDEX (PERCENT)	Global Insight	lv
86	mocmq	NEW ORDERS (NET) - CONSUMER GOODS & MATERIALS, 1996 DOLLARS (BCI)	Global Insight	Δln
87	msondq	NEW ORDERS, NONDEFENSE CAPITAL GOODS, IN 1996 DOLLARS (BCI)	Global Insight	Δln
88	n.a.	Manufacturing new orders, total, SA	Census	Δln
89	n.a.	Manufacturing new orders, durable goods, SA	Census	Δln
90	n.a.	Manufacturing unfilled orders, total, SA	Census	Δln
91	n.a.	Manufacturing unfilled orders, durable goods, SA	Census	Δln
92	fygt10	INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)	Global Insight	Δlv
93	A0M101	COML&IND LOANS OUTST,SA-US	Global Insight	Δln
94	n.a.	Dow Jones common stock price index: composite	GFD	Δln
95	n.a.	Dow Jones common stock price index: industrial	GFD	Δln
96	n.a.	Dow Jones common stock price index: transportation	GFD	Δln
97	n.a.	Dow Jones common stock price index: utility	GFD	Δln
98	n.a.	NYSE Financial index	GFD	Δln
99	fspcom	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE	Global Insight	Δln
100	fspin	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS	Global Insight	Δln
101	n.a.	S&ps composite common stock: dividend yield	GFD	Δlv
102	n.a.	S&ps composite common stock: price- earnings ratio (%, nsa)	GFD	Δln
103	fyff	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE)	Global Insight	Δlv
104	fygm3	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO	Global Insight	Δlv
105	fygm6	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO	Global Insight	Δlv
106	fygt1	INTEREST RATE: U.S.TREASURY CONST MATURITIES, 1-Y	Global Insight	Δlv
107	fygt5	INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.	Global Insight	Δlv
108	MNY2	M2 - MONEY SUPPLY - M1 + SAVINGS DEPOSITS, SMALL TIME DEPOSITS	Global Insight	$\Delta^2 ln$
109	fyaaac	BOND YIELD: MOODY'S AAA CORPORATE	Global Insight	Δlv
110	fybaac	BOND YIELD: MOODY'S BAA CORPORATE	Global Insight	Δlv
111	sfygm3	Fygm3-fyff (AC)	Global Insight	lv
112	sfygm6	Fygm6-fyff (AC)	Global Insight	lv
113	sfygt1	Fygt1-fyff (AC)	Global Insight	lv
114	sfygt5	Fygt5-fyff (AC)	Global Insight	lv
115	fygt10-fyff	Fygt10-fyff (AC)	Global Insight	lv
116	sfyaaac	Fyaaac-fyff (AC)	Global Insight	lv
117	sfybaac	Fybaac-fyff (AC)	Global Insight	lv
118	exruk	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)	Global Insight	Δln
119	exrcan	FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)	Global Insight	Δln
120	fm1	MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)	Global Insight	$\Delta^2 ln$

N.	Mnemonic	Description	Source	Transf.
121	fm2	MONEY STOCK:M2	Global Insight	$\Delta^2 ln$
122	n.a.	Monetary services index: All assets. S.A.	FRED	$\Delta^2 ln$
123	fmfba	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES	Global Insight	$\Delta^2 ln$
124	fmrra	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS	Global Insight	$\Delta^2 ln$
125	CES274	AVG HRLY EARNINGS, PROD WRKRS, NONFARM - TOTAL PRIVATE	Global Insight	Δlv
126	CES277	AVG HRLY EARNINGS, PROD WRKRS, NONFARM - CONSTRUCTION	Global Insight	Δlv
127	CES278	AVG HRLY EARNINGS, PROD WRKRS, NONFARM - MFG	Global Insight	Δlv
128	CES283	AVG HRLY EARNINGS, PROD WRKRS, NONFARM - TRADE, TRANSPORT, UTILITIES	Global Insight	Δlv
129	n.a.	AVG HRLY EARNINGS, PROD WRKRS, NONFARM - WHOLESALE + RETAIL TRADE	Global Insight	Δlv
130	CES289	AVG HRLY EARNINGS, PROD WRKRS, NONFARM - FINANCIAL ACTIVITIES	Global Insight	Δlv
131	pwfsa	PRODUCER PRICE INDEX: FINISHED GOODS	Global Insight	$\Delta^2 ln$
132	pwfcsa	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS	Global Insight	$\Delta^2 ln$
133	pwimsa	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS	Global Insight	$\Delta^2 ln$
134	pwcmsa	PRODUCER PRICE INDEX:CRUDE MATERIALS	Global Insight	$\Delta^2 ln$
135	pwfxsa	PRODUCER PRICE INDEX: FINISHED GOODS,EXCL. FOODS	Global Insight	$\Delta^2 ln$
136	punew	CPI-U: ALL ITEMS	Global Insight	$\Delta^2 ln$
137	pu81	CPI-U: FOOD & BEVERAGES	Global Insight	$\Delta^2 ln$
138	puh	CPI-U: HOUSING	Global Insight	$\Delta^2 ln$
139	pu83	CPI-U: APPAREL & UPKEEP	Global Insight	$\Delta^2 ln$
140	pu84	CPI-U: TRANSPORTATION	Global Insight	$\Delta^2 ln$
141	pu85	CPI-U: MEDICAL CARE	Global Insight	$\Delta^2 ln$
142	pu882	CPI-U: NONDURABLES	Global Insight	$\Delta^2 ln$
143	puc	CPI-U: COMMODITIES	Global Insight	$\Delta^2 ln$
144	pucd	CPI-U: DURABLES	Global Insight	$\Delta^2 ln$
145	pus	CPI-U: SERVICES	Global Insight	$\Delta^2 ln$
146	puxf	CPI-U: ALL ITEMS LESS FOOD	Global Insight	$\Delta^2 ln$
147	puxhs	CPI-U: ALL ITEMS LESS SHELTER	Global Insight	$\Delta^2 ln$
148	puxm	CPI-U: ALL ITEMS LESS MIDICAL CARE	Global Insight	$\Delta^2 ln$

Chapter II
Complete Subset Regressions

1 Introduction

Methods for controlling estimation error in forecasting problems involving small sample sizes and many potential predictor variables has been the subject of much recent research.[1] One lesson learned from this literature is that a strategy of including all possible variables is too profligate; given the relatively short data samples typically available to estimate the parameters of economic forecasting models, it is important to limit the number of parameters that have to be estimated or in other ways reduce the effect of parameter estimation error. This has led to the preponderance of forecast methods such as shrinkage or ridge regression (Hoerl and Kennard (1970)), model averaging (Bates and Granger (1969), Raftery, Madigan, and Hoeting (1997)), and bagging (Breiman (1996)) which accomplish this in different ways.

This paper proposes a new method for combining forecasts based on complete subset regressions. For a given

[1] See, e.g., Stock and Watson (2006) for a review of the literature.

set of potential predictor variables we combine forecasts from all possible linear regression models that keep the number of predictors fixed. For example, with K possible predictors, there are K unique univariate models and $n_{k,K} = K!/((K-k)!k!)$ different $k-$variate models for $k \leq K$. We refer to the set of models corresponding to a given value of k as a complete subset and propose to use equal-weighted combinations of the forecasts from all models within these subsets indexed by k. Moreover, we show that an optimal value of k can be determined from the covariance matrix of the potential regressors and so lends itself to be selected recursively in time.

Special cases of subset regression combinations have appeared in the empirical literature. For example, Rapach, Strauss and Zhou (2010) consider equal-weighted combinations of all possible univariate equity premium models and find that they produce better forecasts of stock returns than a simple no-predictability model. This corresponds to setting $k = 1$ in our context. Papers such as Aiolfi and Favero

(2003) consider equal-weighted combinations of forecasts of stock returns from all possible 2^K models. While not directly nested by our approach, this can nevertheless be obtained from a combination of the individual subset regression forecasts.

From a theoretical perspective, we show that subset regression combinations are akin to a complex version of shrinkage which, in general, does not reduce to shrinking the OLS estimates coefficient by coefficient. Rather, the adjustment to the coefficients depends on all least squares estimates and is a function of both k, the number of variables included in the model, and K, the total number of potential predictors. Only in the special case where the covariance matrix of the predictors is orthonormal, does subset regression reduce to ridge regression or, equivalently, to a Bayes estimator with a specific prior distribution. For this special case we derive the exact degree of shrinkage implied by different values of k and thus formalize how k, the number of parameters in the conditional mean equation, is equivalent to other measures of model complexity that have previously been proposed in the literature.

One attraction of the proposed method is that, unlike the ridge estimator and the usual application of Bayesian estimators, it does not impose the same amount of shrinkage on each coefficient. Unlike model selection methods, it also does not assign binary zero-one weights to the OLS coefficients. In this regard, a more similar method is Bagging which applies differential shrinkage weights to each coefficient.

We also show that the weights implied by subset regression reflects omitted variable bias in a way that can be useful for forecasting. This holds particularly in situations with strongly positively correlated regressors since the subset regression estimates account for the omitted predictors.

To illustrate the subset regression approach empirically we consider, like many previous studies, predictability of U.S. stock returns. In particular, following Rapach et al. (2010), we study quarterly data on U.S. stock returns in an application that has 12 potential predictor variables and so generates subset regressions with $k = 1$, 2, ..., 12 predictor variables. We find that subset regression combinations that use $k = 2, 3,$ or 4 predictors produce the lowest out-of-sample mean squared error (MSE) values. Moreover, these subset models generate superior predictive accuracy relative to the equal-weighted average computed across all possible models, a benchmark that is well-known to be difficult to beat, see Clemen (1989). We also find that the value of k in the subset regression approach can be chosen recursively (in pseudo "real time") in such a manner that the approach produces forecasts with lower out-of-sample MSE-values than those produced by recursive versions of Bayesian model averaging, ridge regression, Lasso, or Bagging.

The outline of the paper is as follows. Section 2 introduces the subset regression approach and characterizes its theoretical properties. Section 3 presents a Monte Carlo simulation study, Section 4 conducts the empirical analysis of US stock returns, while Section 5 concludes.

2 Theoretical Results

This section presents the setup for the analysis and derives theoretical results for the proposed complete subset regression method.

2.1 Setup

Suppose we are interested in predicting the univariate variable y_{T+1} using a regression model based on K predictors $x_T \in \mathbb{R}^K$, and a history of data, $\{y_{t+1}, x_t\}_{t=0}^{T-1}$. Let $E[x_t x_t'] = \Sigma_X$ for all t and, without loss of generality, assume that $E[x_t] = 0$ for all t. To focus on regressions that include only a subset of the predictors, define β to be a $K \times 1$ vector with slope coefficients in the rows representing included regressors and zeros in the rows of the excluded variables. Let β_0 be the pseudo true value for β, i.e., the population value of the projection of y on X, where $y = (y_1, ..., y_T)$ is a $T \times 1$ vector and $X = (x_0', x_1', ..., x_{T-1}')'$ stacks the x observations into a $T \times K$ matrix. Denote the generalized inverse of a matrix A

by A^{-}. Let S_i be a $K \times K$ matrix with zeros everywhere except for ones in the diagonal cells corresponding to included variables, zeros for the excluded variables, so that if the $[j, j]$ element of S_i is one, the jth regressor is included, while if this element is zero, the jth regressor is excluded. Sums over i are sums over all permutations of S_i.

We propose an estimation method that uses equal-weighted combinations of forecasts based on all possible models that include a particular subset of the predictor variables. Each subset is defined by the set of regression models that include a fixed (given) number of regressors, $k \leq K$. Specifically, we run the 'short' regression of y_t on a particular subset of the regressors, then average the results across all k dimensional subsets of the regressors to provide an estimator, $\hat{\beta}$, for forecasting, where $k \leq K$. With K regressors in the full model and k regressors chosen for each of the short models, there will be $n_{k,K} = K!/(k!(K-k)!)$ subset regressions to average over. In turn, each regressor gets included a total of $n_{k-1,K-1}$ times.

As an illustration, consider the univariate case, $k = 1$, which has $n_{1,K} = K!/(1!(K-1)! = K$ short regressions, each with a single variable. Here all elements of $\hat{\beta}_i$ are zero except for the least squares estimate of y_t on x_{it} in the i^{th} row. The equal-weighted combination of forecasts from the individual models is then

$$\hat{y}_{T+1} = \frac{1}{K} \sum_{i=1}^{K} x'_T \hat{\beta}_i. \tag{1}$$

Following common practice, our analysis assumes quadratic or mean square error (MSE) loss. For any estimator, we have

$$\begin{aligned}
&E\left[\left(y_{T+1} - \hat{\beta}_T x_T\right)^2\right] \\
&= E\left[\left(y_{T+1} - \beta'_0 x_T + (\beta_0 - \hat{\beta}_T)' x_T\right)^2\right] \\
&= E\left[\left(\varepsilon_{T+1} + (\beta_0 - \hat{\beta}_T)' x_T\right)^2\right] \\
&= \sigma_\varepsilon^2 \left(1 + T^{-1} \sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_T - \beta_0)' x_T x'_T (\hat{\beta}_T - \beta_0)\right]\right)
\end{aligned}$$

Here ε_{T+1} is the residual from the population projection of

y_{T+1} on x_T. We concentrate on the last term since the first term does not depend on $\hat{\beta}$. Hence, we are interested in examining $\sigma_\varepsilon^{-2} E\left[(\hat{\beta}_T - \beta)' x_T x_T' (\hat{\beta}_T - \beta)\right]$.

2.2 Complete Subset Regressions

Subset regression coefficients can be computed as averages over least squares estimates of the subset regressions. When the covariates are correlated, the individual regressions will be affected by omitted variable bias. However, as we next show, the subset regression estimators are themselves a weighted average of the full regression OLS estimator:

Theorem 1 *Assume that as the sample size gets large $\hat{\beta}_{OLS} \to^p \beta_0$ for some β_0 and $T^{-1}X'X \to^p \Sigma_X$. Then, for fixed K, the estimator for the complete subset regression, $\hat{\beta}_{k,K}$, can be written as*

$$\hat{\beta}_{k,K} = \Lambda_{k,K} \hat{\beta}_{OLS} + o_p(1),$$

where

$$\Lambda_{k,K} \equiv \frac{1}{n_{k,K}} \sum_{i=1}^{n_{k,K}} \left(S_i' \Sigma_X S_i\right)^{-} \left(S_i' \Sigma_X\right).$$

A proof of this result is contained in the Appendix.

This result on the relationship between $\hat{\beta}_{k,K}$ and the OLS estimator makes use of high level assumptions that hold under very general conditions on the data; see White (2000, chapter 3) for a set of sufficient conditions. For example, the result allows $\{X_t\}$ to be dependent, mixing with a sufficiently small mixing coefficient, and even allows $E[X_t' X_t]$ to be heterogenous over time, in which case Σ_X is the average variance covariance matrix, although, for simplicity, we assume that Σ_X is constant over time.

In general, $\Lambda_{k,K}$ is not diagonal and hence the coefficients $\hat{\beta}_{k,K}$ are not (approximately) simple OLS coefficient-by-coefficient shrinkages. Rather, subset regression coefficients are functions of all the OLS coefficients in the regression. Insight into how the method works as a shrinkage estimator can be gained from the special case when the covariates are orthonormal.[2] In this case, $\hat{\beta}_{k,K} = \lambda_{k,K}\hat{\beta}_{OLS}$, where $\lambda_{k,K} = 1 - (n_{k,K-1}/n_{k,K})$ is a scalar and so subset regression is numerically equal to ridge regression.[3]

To see this, note that for this special case $\hat{\beta}_{OLS} = X'y$ while each of the subset regression estimates can be written $\hat{\beta}_i = S_i X'y$ where S_i is a $K \times K$ diagonal vector with ones (zeros) on the diagonal for each included (excluded) regressor, and zeros off the diagonal. The complete subset regression estimator is then given by

$$\begin{aligned}\hat{\beta}_{k,K} &= \frac{1}{n_{k,K}}\sum_{i=1}^{n_{k,K}}\hat{\beta}_i \\ &= \frac{1}{n_{k,K}}\sum_{i=1}^{n_{k,K}} S_i X'y \\ &= \left(\frac{1}{n_{k,K}}\sum_{i=1}^{n_{k,K}} S_i\right)\hat{\beta}_{OLS}.\end{aligned}$$

[2]We refer to subset regressions as similar to shrinkage although for some configurations of the variance covariance matrix of the predictors and some OLS estimates, subset regression will not actually shrink the coefficient estimates.

[3]Equivalently, this case corresponds to a Bayes estimator under normality with prior $N(\mu, \gamma_{k,K}^{-1}\sigma_\varepsilon^2))$, $\hat{\beta} = (X'X + \gamma_{k,K}I)^{-1}(X'y + \gamma_{k,K}\mu)$, prior mean $\mu = 0$, and $\gamma_{k,K} = (1 - \lambda_{k,K})/\lambda_{k,K}$. If the assumption on the regressors is weakened to $\Sigma_X = I_K$, the same result holds asymptocially.

The result now follows by noting that the elements of $\sum_{i=1}^{n_{k,K}} S_i$ are zero for the off-diagonals terms, and equal the number of times the regressor is included in the subset regressions for the diagonal terms. In turn the diagonal terms equal $n_{k,K}$ minus the number of times a regressor is excluded, which gives the result, noting that the solution is the same for each diagonal.

Several points follow from this result. First, the amount of shrinkage implied by $\lambda_{k,K}$ is a function of both k and K. As an illustration, Figure 1 plots $\lambda_{k,K}$ as a function of k for the orthonormal case. Higher curves represent smaller values of K, where $K = \{10, 15, 20\}$. For any value of K, $\lambda_{k,K}$ is a linear function of k that increases to one. In fact, setting $k = K$, corresponds to simply running OLS with all variables included. Further, as K increases, the slope of the $\lambda_{k,K}$ line gets reduced, so the amount of shrinkage is decreasing for any k, the larger is K, the total number of potential predictors. Essentially, the smaller is k relative to K, the greater the amount of shrinkage.

Second, in general $\Lambda_{k,K}$ reduces to the ridge estimator, either approximately or exactly, only when the regressors are uncorrelated. When this does not hold, subset regression coefficients will not be simple regressor-by-regressor shrinkages of the OLS estimates, and instead depend on the full covariance matrix of all regressors. Specifically, $\Lambda_{k,K}$ is not diagonal and each element of $\hat{\beta}$ is approximately a weighted sum of all of the elements in $\hat{\beta}_{OLS}$. The weights depend not only on $\{k, K\}$ but on all elements in Σ_X. For example, if $K = 3$ and $k = 1$, we have

$$\Lambda_{1,3} = \frac{1}{3} \begin{pmatrix} 1 & \frac{\Sigma_{12}}{\Sigma_{11}} & \frac{\Sigma_{13}}{\Sigma_{11}} \\ \frac{\Sigma_{12}}{\Sigma_{22}} & 1 & \frac{\Sigma_{23}}{\Sigma_{22}} \\ \frac{\Sigma_{13}}{\Sigma_{33}} & \frac{\Sigma_{23}}{\Sigma_{33}} & 1 \end{pmatrix}. \quad (2)$$

Each row of $\Lambda_{1,3}$ is the result of including a particular subset regression in the average. For example, the first row gives the first element of $\hat{\beta}_{1,3}$ as a weighted sum of the OLS regressors $\hat{\beta}_{OLS}$. Apart from the division by $1/3$, the own coefficient is given a relative weight of one while the remaining coefficients are those we expect from omitted variable bias formulas. Clearly the effect of dividing by $n_{1,3} = 3$ is to shrink all coefficients, including the own coefficient, towards zero.

For $k > 1$, each regressor gets included more often in the regressions. This increases their effect on $\Lambda_{k,K}$ through a higher inclusion frequency, but decreases their effect through the omitted variable bias. Since the direct effect is larger than the omitted variable bias, an increased k generally reduces the amount of shrinkage. Of course, in the limit as $k = K$, there is no shrinkage and the method is identical to OLS.

2.3 Risk

We next examine the risk of the subset regression estimator. In common with all biased methods, for values of β_0 far from zero, the risk is large and so it is appropriate not to shrink coefficients towards zero. Shrinkage methods only add value when β_0 is near zero. To capture such a situation, we assume that β_0 is local to zero. Specifically, we assume that $\beta_0 = T^{-1/2}\sigma_\varepsilon b$ for some fixed vector b.

Under general, dependent data generating processes, the risk is difficult to derive. However, if we restrict the setup to i.i.d. data $\{y_{t+1}, x_t\}$, we get the following result:

Theorem 2 *Assume that the data* $\{y_{t+1}, x_t\}$ *are i.i.d.,* $E[(\hat{\beta} - \beta_0)^2 | x_{T+1}] = E[\hat{\beta} - \beta_0]^2$*, and* $T^{-1/2}(\hat{\beta}_{OLS} - \beta) \rightarrow^d N(0, \Sigma_X^{-1})$*. Then, in large samples,*

$$\sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_T - \beta)'\Sigma_X(\hat{\beta}_T - \beta)\right] \approx \sum_{j=1}^{K} \zeta_j + b'(\Lambda_{k,K} - I)'\Sigma_X(\Lambda_{k,K} - I)b, \quad (2)$$

where ζ_j are the eigenvalues of $\Lambda'_{k,K}\Sigma_X\Lambda_{k,K}\Sigma_X^{-1}$.

The expected loss depends on many aspects of the problem. First, it is a function of the variance covariance matrix through both Σ_X and $\Lambda_{k,K}$. Second, it depends on the dimension of the problem, K, and of the subset regression, k.

Third, it is a function of the elements of b. Different trade-offs can be explored by varying these parameters. Some will be very attractive when compared to OLS, while others might not be. As in the simple orthogonal case, the larger are the elements of b, the worse the complete subset regression methods will perform.

For different choices of $\{K, k, \Sigma_X, b\}$, we can compute the expected loss frontier as a function of k. If $\Sigma_X = I$, so the regressors are mutually orthogonal, (2) reduces to

$$\sigma_{\varepsilon}^{-2} E\left[(\hat{\beta}_T - \beta)'\Sigma_X(\hat{\beta}_T - \beta)\right] = \lambda_{k,K}^2 K + (1 - \lambda_{k,K})^2 b'b, \quad (3)$$

which depends on $\{K, k, b'b\}$. For fixed values of $b'b$ and K, as k increases, $\lambda_{k,K}$ gets bigger and the increase in the first term in (3) is offset by the decrease in the second term in this equation. The extent of this offset depends on the relative sizes of K and $b'b$. As an illustration of this, the left window in Figure 2 plots the value of the expected loss (3) as a function of k, for $K = 10$ and $b'b = (1, 3, 4)$. Each line corresponds to a separate value of $b'b$ with larger intercept on the x axis, the greater is $b'b$. Setting $k = K = 10$ yields OLS loss, so all lines converge at that point. A variety of shapes are possible. If $b'b$ is quite small, so that the regressors are not that useful for forecasting, then a large amount of shrinkage, and hence a small value of k, works best. Conversely, if $b'b$ is large, bigger values of k become optimal.

In practice, different choices of k can be motivated by theoretical considerations. As always with shrinkage estimation, the smaller b is expected to be, the more useful it is to apply strong shrinkage. As we discuss above, the amount of shrinkage tends to be greater, the smaller one chooses k. Since $\{k, K\}$ are known and Σ_X can be estimated by $T^{-1}X'X$, (2) can be used to produce curves such as those in the left window of Figure 2 but relevant for the application at hand. One can then choose k as the point at which expected loss is lowest given reasonable choices for b. As an illustration of this point, the right window of Figure 2 uses data from the application in the empirical section to estimate

Σ_X and shows expected loss curves for $b'b = 1$, 2, or 3. Although the expected loss curve varies quite a bit across different values of $b'b$, an interior optimal value$-$corresponding to a minimal expected loss$-$around $k = 2$, 3, or 4 is obtained in all three cases.

2.3.1 Comparison with OLS and Ridge

It is informative to compare the risk for subset regressions to that of models estimated by OLS. In some cases, this comparison can be done analytically. For example, this can be done for general K when Σ_X has ones on the diagonal and ρ elsewhere and $k = 1$, corresponding to combinations of univariate models. First, note that the risk for OLS regression is K while for this case the risk of the subset regression method reduces to

$$\frac{1}{K}\left(1+(K-1)\rho^2\right)+(\rho-1)^2\left(\frac{K-1}{K}\right)^2(K+K(K-1)\rho)\,. \tag{4}$$

Hence, subset regressions produce lower risk than OLS for any (K, ρ) pair for which

$$\frac{1}{K}\left(1+(K-1)\rho^2\right)+(\rho-1)^2\left(\frac{K-1}{K}\right)^2(K+K(K-1)\rho)<K.$$

For small values of K this holds for nearly all possible correlations. To illustrate this, Figure 3 plots the ratio of the subset regression MSE to the OLS MSE as a function of ρ, the correlation between the predictors, and k, the number of predictors included. The figure assumes that $T = 100$. Whenever the plotted value falls below one, the subset regression approach dominates OLS regression in the sense that it produces lower risk. For any $K \leq 6,$ subset regression always (for any ρ for which Σ_X is positive definite) has a lower risk than OLS based on the complete set of regressors. For $K > 6$, we find that there is a small region with small values of ρ and $k = 1$ for which the reverse is true, but otherwise subset regression continues to perform better than OLS.

The figure thus illustrates that a simple equal-weighted average of univariate forecasts can produce better forecasts than the conventional multivariate model that includes all predictors even in situations where the univariate models are misspecified due to omitted variable bias.

Figure 4 uses heat maps to compare the expected loss of the subset regressions to that of the Ridge regression approach for different values of the limit of the shrinkage parameter, γ/T. The figure assumes that there are $K = 8$ predictor variables, sets $b = 1$, a vector of ones, and lets Σ_X have ones on the diagonal and ρ on all off-diagonal cells. The correlation between predictor variables, ρ, varies along the horizontal axis, while the shrinkage parameter, γ, varies along the vertical axis. We use colors to indicate the value for $\min(0, MSE^{ridge} - MSE^{subset})$, with dark red indicating that $MSE^{ridge} > MSE^{subset}$, while, conversely, yellow and blue indicate areas where $MSE^{ridge} < MSE^{subset}$. Each window corresponds to a different value of k. Suppose that, moving along the vertical axis corresponding to a particular value of ρ, there is no red color. This shows that, for this particular value of ρ, ridge regressions always produce a lower expected loss than the corresponding subset regression. Conversely, if, for a given value of ρ, the area is red for all values of γ, subset regressions dominate all ridge regressions, regardless of the chosen shrinkage parameter.

Figure 4 shows that when $k = 1$, ridge regressions mostly produce lower MSE-values than subset regressions for $\rho < 0.6$. Conversely, for $\rho > 0.85$, this univariate subset regression uniformly dominate all ridge results. If $k = 2,$ subset regressions uniformly dominate when $\rho > 0.6$, while if $k = 4$, subset regressions always dominate when $\rho < 0.5$. This means that, for $k = 2, 3,$ or 4, one of the subset regressions will always dominate the best ridge regression as they produce the lowest MSE loss.

3 Monte Carlo Simulation

To better understand how the subset combination approach works, we first consider a Monte Carlo simulation experiment that allows us to study both the absolute forecast performance of the subset regression approach as well as its performance relative to alternative methods.

3.1 Simulation setup

Our Monte Carlo design assumes a simple linear regression model:

$$Y_{t+1} = \sum_{k=1}^{K} \beta_k X_{kt} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2). \tag{5}$$

We assume a sample size of $T = 100$ observations and consider one-step-ahead forecasts of Y_{T+1}. The covariance matrix of the $X-$variables $\Sigma_X = Cov(X_1, ..., X_K)$ takes the simple form

$$\Sigma_X = \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \ddots & & \vdots \\ & & \ddots & & \\ \vdots & & \ddots & 1 & \rho \\ \rho & & \dots & \rho & 1 \end{pmatrix},$$

where $\rho \in \{0, 0.25, 0.5, 0.75, 0.95\}$. Small values of ρ correspond to small values of the predictive R^2, while the R^2 increases as ρ is raised. Data are assumed to be i.i.d., and we include up to $K = 8$ predictors. Two designs are considered for the regression parameter: $b = 1_K$ and $b = (1\ 1\ 1\ 1\ 0\ 0\ 0\ 0)$. In the first experiment, all predictors are relevant and matter equally; in the second experiment only the first four predictors matter to the outcome.

3.2 Comparison with other approaches

We are interested not only in how well the subset combination approach performs in absolute terms, but also in how it compares with other approaches. Many alternative ways to combine or shrink forecasts from different models have been considered in the literature. Among the most prominent ones are Bayesian model averaging (Raftery, Madigan and Hoeting, 1997), Bagging (Breiman, 1996), ridge regression (Hoerl and Kennard, 1970), and the Lasso (Tibshirani, 1996). Given the availability of these alternatives, it becomes important to compare the subset regression combination approach to such methods. We briefly discuss each of the methods and explain how we implement them.

3.2.1 Bayesian model averaging

Our implementation of Bayesian model averaging is based on the Raftery, Madigan and Hoeting (1997) algorithm. This depends on two hyperparameters regulating the prior variance of the regressor coefficients, ν and λ, which, following the authors' discussion, are set to 2.58 and 0.28, respectively. A third parameter, ϕ, determines the number of models with zero weight. The BMA predictions are obtained by weighting each model's forecast by its posterior probability:

$$\hat{y}^{BMA}_{T+1|T} = \sum_{j=1}^{2^K} \omega_j \hat{y}_{T+1|T,j}, \tag{6}$$

$$\omega_j \propto p(X_{t+1:T}|m_j)p(m_j)$$

where $p(X_{t+1:T}|m_j)$ and $p(m_j)$ denote the likelihood and the prior, respectively, of model j. The approach determines ω_j numerically from the frequency the Markov chain visits the jth model. Based on standard diagnostic tests (e.g., Koop (2003)) we use a chain of 2,500 draws, discarding the first 500 points.

3.2.2 Bagging

Our implementation of Bagging is based on 1,000 bootstrapped samples of the original data arranged in the $\{y_{t+1:T}, X_{t:T-1}\}$ tuple. We preserve the autocorrelation structure of the predictors by applying the circular block bootstrap of Politis and Romano (1992) with block size chosen optimally according to Politis and White (2004).[4] Contemporaneous dependence across observations is preserved by using the same blocks for all variables. For each bootstrapped sample, an estimate of β, $\hat{\beta}^b$, is obtained and forecasts are computed as

$$\hat{y}^b_{T+1|T} = (x'_T S_T)\hat{\beta}^b. \tag{7}$$

Here S_T is the stochastic selection matrix whose (i,i) elements equal the indicator function $I(|t_i| > c)$. A predictor is added only if its t-statistic is significant at the threshold implied by c. The larger the value of c, the higher the threshold and so the more parsimonious the final model will be. To control for this effect, we follow Inoue and Kilian and consider different values of c. The final Bagging forecasts are obtained by averaging across the bootstrap draws

$$\hat{y}^{BAGG}_{T+1|T} = \frac{1}{B}\sum_{b=1}^{B} \hat{y}^b_{T+1|T}. \tag{8}$$

[4]For robustness, we also implemented the parametric bootstrap, but found that the results are not sensitive to this choice.

3.2.3 Ridge regression

The only parameter that has to be chosen under the ridge approach is γ which regulates the amount of shrinkage imposed on the regression coefficients. Given a value of γ, the forecasts are obtained by

$$\hat{y}^{RIDGE}_{T+1|T} = x'_T \hat{\beta}_\gamma, \tag{9}$$

where

$$\hat{\beta}_\gamma = argmin \ \left(\sum_{t=1}^{T-1} (y_{t+1} - x'_t \beta)^2 + \gamma \sum_{j=1}^{K} \beta_j^2 \right) \tag{10}$$

Note that, by construction, as $\gamma \to \infty$, $\hat{y}^{RIDGE}_{T+1} \to \frac{1}{T-1} \sum_{j=2}^{T} y_j$, so the ridge forecast simply converges to the sample mean.

3.2.4 Lasso

Least absolute shrinkage and selection operator, LASSO (Tibshirani 1996), retains the features of both model selection and ridge regression: it shrinks some coefficients and sets others to zero. Lasso forecasts are computed as

$$\hat{y}^{LASSO}_{T+1|T} = x'_T \hat{\beta}_\tau, \tag{11}$$

where

$$\hat{\beta}_\psi = argmin \sum_{t=1}^{T-1} (y_{t+1} - x'_t \beta)^2, \tag{12}$$
$$s.t. \sum_{j=1}^{K} |\beta_j| \leq \psi.$$

Here the parameter ψ controls for the amount of shrinkage. For sufficiently large values of ψ the constraint is not binding and the LASSO estimator reduces to OLS. Given the absolute value operator $|.|$, the constraint is not linear and a closed form solution is not available. $\hat{\beta}_\psi$ is therefore computed following the algorithm described in section 6 of Tibshirani (1996). Because the forecasts depend on ψ, we consider a grid of values for ψ.

3.3 Simulation results

Table 1 shows results from the simulation experiment, using 25,000 simulations. We report performance in terms of the R^2-value, which is inversely related to the MSE-value, but conveys the same message and is slightly easier to interpret. First consider the performance of the subset regression approach when $\beta = 1_K$ (left panel). Since the R^2 is positive for the (infeasible) model that uses the correct parameter values, negative R^2-values show that parameter estimation error dominates whatever evidence of predictability the model identifies. This case only occurs for the subset regressions when $\rho = 0$ and $k = 8$, corresponding to the "kitchen sink" approach that includes all predictors and so does not average across multiple models. For small values of ρ the best subset regressions use three or four predictors. As ρ increases, the number of variables included in the best-performing subset regressions tends to decrease and the best performance is obtained for $k = 1$ or $k = 2$. In general, the difference between the best and worst subset combination (usually the kitchen sink, $k = 8$) tends to be greater, the smaller the value of ρ. This is likely to reflect the greater importance of estimation error in situations where the predictive signal is weaker, parameter estimation error matters more and affects the larger models (large k) more than the smaller models (small k).

The ridge regression results most closely resemble those from the subset regressions. Compared with subset regression, ridge regression performs quite well, although, consistent with Figure 4, the best subset regression produces better performance than the best ridge regression in all cases. In turn, the best subset and ridge regressions generally perform better than the best lasso, bagging and BMA approaches.

Interestingly, however, both the ridge and BMA approaches produce more homogenous performance across different values of their design parameters, γ, ϕ, with the worst performance bettering that of the worst subset regression which typically sets $k = 8$.

Similar conclusions emerge when we set $\beta = (1\ 1\ 1\ 1\ 0$

$0\ 0\ 0)'$, the results for which are shown in the right panel of Table 1. This case represents a setup with a smaller degree of predictability over the outcome variable, and so lower R^2-values are obtained. Unsurprisingly, for this case the best subset regressions use a smaller value of k than in the previous case where all predictors had an effect on the outcome. The subset regressions that include relatively few predictors, e.g., $k = 2$, 3, or 4, continue to perform particularly well, whereas performance clearly deteriorates for the models that include more predictors.

4 Empirical Application: Stock Return Predictions

To illustrate the complete subset regression approach to forecast combination and to compare its performance against that of alternative approaches, this section provides an empirical application to US stock returns. This application is well suited for our analysis in part because predictability of stock returns has been the subject of an extensive literature in finance, recently summarized by Rapach and Zhou (2012), in part because there is a great deal of uncertainty about which, if any, predictors help forecast stock returns. Clearly this is a case where estimation error matters a great deal, see, e.g., the discussion in Goyal and Welch (2008).

Specifically, we investigate if there is any improvement in the subset regression forecasts that combine k-variate models for $k \geq 2$ relative to using a simple equal-weighted combination of univariate models ($k = 1$), as proposed in Rapach et al. (2010), or relative to other combination schemes such as those described in the previous section.

4.1 Data

Data are taken from Goyal and Welch (2008), updated to 2010, and are recorded at the quarterly horizon over the period 1947Q1 - 2010Q4. The list of predictors comprises

12 variables for a total of $2^{12} = 4096$ possible models.[5]

The 12 variables are the *Dividend Price Ratio (dp),* the difference between the log of the 12-month moving sum of dividends and the log of the S&P 500 index; *Dividend Yield (dy),* the difference between the log of the 12-month moving sum of dividends and the lagged log S&P 500 index; *Earnings Price Ratio (ep),* the difference between the log of the 12-month moving sum of earnings and the log S&P 500 index; *Book to Market (bm),* the ratio of the book value to market value for the Dow Jones Industrial Average; *Net Equity Expansion (ntis),* the ratio of the 12-month moving sum of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks; *Treasure Bill (tbl),* the 3-Month Treasury Bill (secondary market) rate; *Long Term Rate of Returns (ltr),* the long-term rate of return

[5] Data are available at http://www.hec.unil.ch/agoyal/. Variable definitions and data sources are described in more detail in Goyal and Welch (2008). To avoid multicollinearity when estimating some of the multivariate models, we exclude the log dividend earnings ratio and the long term yield. By construction, the log dividend earnings ratio is equal to the difference between the log dividend price ratio and the log earnings price ratio, while the long term yield is equal to the sum of the term spread and the Treasury Bill rate.

on US Bonds; *Term Spread (tms),* the difference between the long term yield on government bonds and the Treasury Bill rate; *Default Yield Spread (dfy),* the difference between yields on AAA and BAA-rated bonds; *Default Return Spread (dfr),* the difference between long-term corporate bond and long-term government bond returns; *Inflation (infl),* the (log) growth of the Consumer Price Index (All Urban Consumers); and *Investment to Capital Ratio (ik),* the ratio of aggregate investments to aggregate capital for the whole economy.

The equity premium, our dependent variable, is the difference between the continuously compounded return on the S&P 500 index (including dividends) and the 3-month Treasury Bill rate. As in Rapach et al. (2010) and Goyal and Welch (2008), we adopt a recursively expanding estimation scheme. The initial estimation sample goes from 1947Q1 to 1964Q4, yielding a first forecast for 1965Q1, while the last forecast is for 2010Q4. Each quarter parameters are (re)estimated using all available information up to that point. This pseudo out-of-sample forecasting exercise simulates the practice of a real time forecaster. As in the theoretical analysis, forecasts are generated from the following predictive regression

$$r_{2:t+1} = \alpha + (X_{1:t}S)\beta + \epsilon_{2:t+1}, \quad (13)$$

where $r_{2:t+1}$ is the equity premium variable defined above, $X_{1:t}$ is the full regressor matrix, $\epsilon_{2:t+1}$ is a vector of error terms, α and β are unknown parameters estimated by OLS, and S is a diagonal selector matrix whose unity elements determine which variables get included in the model. For example, the “kitchen sink” model containing all predictors is obtained by setting $S = I_{12}$, while the constant ‘null’ model is obtained by setting S equal to a 12×12 matrix of zeros. Following the analysis in Section 2, our focus is on the combination of k-variate models, more specifically

$$\hat{r}^k_{t+1} = \frac{1}{n_{k,K}} \sum_{j=1}^{n_{k,K}} (\hat{\alpha}_j + x'_t S_j \hat{\beta}_j) \quad s.t. \quad tr(S_j) = k, \quad (14)$$

where $tr(\circ)$ is the trace operator.

4.2 Bias-variance trade-off

Figure 5 plots time-series of out-of-sample forecasts of returns for the different $k-$variate subset regression combinations. The forecasts display similar patterns except that as k increases, the variance of the combined forecasts also increases. The least volatile forecasts are generated by the constant model ($k = 0$), while the most volatile forecasts arise when we use the model that contains all regressors $(k = K = 12)$. Neither of these cases perform any forecast combination. As we shall subsequently see, forecasts from the best k-variate combinations are in turn more volatile than those from combinations of univariate models but less volatile than those from the other k-variate combinations. The extent to which volatility of the forecast reduces or enhances forecast performance depends, of course, on how strongly this variation is correlated with the outcome$-$a point we further address below.

Figure 6 provides insight into the relation between the variance and bias of the forecasts. Along the $x-$axis, the upper left window lists the number of predictors included in each model, k, while the y-axis lists the time-series variance associated with a given model. Thus, for example, for $k = 1$ the circles show the variance for each of the 12 univariate forecasting models, while for $k = 2$, the circles show the forecast variance for each of the 66 bivariate models. The upper left graph shows that the variance of the forecast is increasing in the number of variables included in the forecast models. To see why, define $x_t^S = x_t S$ and $X_{1:T}^S = X_{1:T} S$, and note that

$$var(\hat{r}_{t+1}) = var(\hat{\alpha} + x_t^S \hat{\beta}) = [\iota' \iota + x_t^S (X_{1:T}' X_{1:T})^{-1} x_t^{S\prime}] \hat{\sigma}_\epsilon, \quad (15)$$

which is increasing in $\hat{\sigma}_\epsilon$ and in the column dimension of ι', $x_t^{S\prime}$ and X^S. Therefore, the larger the dimensional of the pooled models, the higher the forecast variance.

The upper right window in Figure 6 shows gains from pooling the models due to the reduction in the (squared) bias. Specifically, the combination of the three-variate models has the lowest bias. The constant model produces the most (upward) biased forecasts. At the other end of the spectrum, the “kitchen sink” model with all variables included generates the most biased forecasts because of its occasional extreme negative forecasts (see Figure 5). Except for the models based on dp, dy and ep, the individual univariate models generate a large bias.

Putting together the forecast variance and bias results, the bottom window of Figure 6 establishes a (squared) bias-variance trade-off that resembles the well-known mean-variance efficient frontier known from modern portfolio theory in finance. Specifically, the (squared) bias is largest for models with either very few or very many predictors, while the variance increases monotonically in k.

4.3 Performance of subset regressions

To gain insights into the forecast performance of the various models, Figure 7 plots the out-of-sample R^2 (top window) and the MSE-value (bottom window) for the individual $k-$variate forecasting models along with those for the subset regression combinations.[6] The lower $x-$axis shows the number of predictors included in each model, while the upper $x-$axis in the top window lists the total number of $k-$variate models, i.e., $n_{k,12}$. For $1 \leq k \leq 6$, the $k-$variate combinations generate lower MSE values than the equal-weighted average forecast computed across all 4,096 models, a benchmark frequently used in the forecast combination literature. They also perform better than the constant equity premium model ($k = 0$), a benchmark considered difficult to beat in the finance literature, see Goyal and Welch

[6]The out-of-sample R^2-value is computed as

$$R^2 = 1 - \frac{\sum_{\tau=R}^{T}(r_{\tau+1} - \hat{r}_{\tau+1|\tau})^2}{\sum_{\tau=R}^{T}(r_{\tau+1} - \hat{r}^{bmk}_{\tau+1|\tau})^2}.$$

(2008).

Interestingly, the two and three-variate combinations generate out-of-sample R^2-values that are 1% higher than the univariate combination approach used by Rapach et al. (2010), while the first six k-variate combinations produce better performance than the combination of all models, i.e., the “thick” forecast modelling approach described in Aiolfi & Favero (2003). This may not seem like such a large difference but, as emphasized by Campbell and Thompson (2008), even small differences in out-of-sample R^2 can translate into economically large gains in investor utility.

Figure 6 showed that the forecast results do not depend simply on the number of pooled forecasts. For example, there are 66 two-variate as well as ten-variate models, but the corresponding equal-weighted combinations produce very different outcomes. This is not surprising given that the worst two-variate model is better than the best ten-variate model. To control for the mere effect of the number of models included in the combination, we also combine models that are randomly selected across different values of k. Figure 8 plots the out-of-sample MSE and R^2-values as a function of the number of models in the combined forecast. Less than 100 models, i.e. about 2% of the total, need to be pooled in order to approximate the behavior of the forecasts obtained by combining all models.[7] This finding is not surprising given that about 60% of the models contain five, six or seven predictors so that the combinations get dominated by forecast models with five, six and seven variables included.[8] In fact, when models are randomly selected, the probability of picking a 6-variate model is about 0.225 against 0.002 for the univariate or eleven-variate models. Indeed the combinations of the six-variate models has very similar performance to the total combination.

The benefit of subset combination is evident from three observations. First, the k-variate subset combinations have similar, if not better (for k =1, 2, 3, 10 and 11), perfor-

[7]This finding becomes very relevant in situations where it is infeasible to estimate all 2^K models, e.g., when $K > 20$, since the number of models is exponentially related to the number of predictors.

[8]This fraction is given by $\binom{12}{5} + \binom{12}{6} + \binom{12}{7}/2^{12}$.

mance as the single best k-variate model, the identity of which, however, is difficult to establish ex-ante. Second, for $k \leq 10$ the $k-$variate combinations produce better results than models selected by recursively applying information criteria such as the AIC or the BIC. This happens despite the fact that these subset combinations contain, on average, the same or a larger number of predictors.[9] Third, while some univariate models, the ones containing dp, dy, dfr, and ik, produce better results than the equal-weighted combination of all models, in contrast no single predictor model does better than the three best-performing k-variate subset combinations.

4.4 Performance comparisons

Table 2 presents out-of-sample R^2-values. First consider the univariate models shown in Panel A. Only five of the twelve variables generate positive out-of-sample R^2-values,

[9] On average, the BIC and AIC criteria select 2.73 and 4.88 predictors, respectively.

the highest such value being 2.28% for the investment-capital ratio. Panel B shows that all subset regressions with $k \leq 6$ generate positive out-of-sample R^2-values, the largest values occurring for $k = 2$ or $k = 3$ which lead to an R^2 around 4%. As k grows larger, the out-of-sample forecasting performance quickly deteriorates with values below -10% when $k = 11$ or $k = 12$.

Turning to the alternative approaches described earlier, Panel C shows that the Lasso forecasts are only capable of producing small positive R^2-values for $\psi \leq 3$ and generate large negative R^2-values for the largest values of ψ. Panel D shows that the ridge regressions generate large negative R^2-values when the shrinkage parameter, γ, is small, corresponding to the inclusion of many predictors. Better performance is reached for higher values of γ, but even the best value of γ only leads to an R^2 of 2.8%. The bagging approach (panel E) suffers from similar deficiencies when c is small, leading to large prediction models, but improves for values of c around two at which point an R^2 of 1.7% is reached. Finally, Bayesian model averaging performs quite poorly in this case, producing negative R^2-values with exception of the cases where $\phi = 0$ or $\phi = 0.01$.

To compare model performance more formally, we use the test proposed by Clark and West (2007), treating the simple prevailing mean forecast as our benchmark. This test re-centers the difference in mean squared forecast errors to account for the higher variability associated with forecasts from larger models. The test results show that three of the univariate models (corresponding to dp, dy, and ik) produce better forecasting performance than the benchmark at the 5% significance level. For the bagging and BMA methods, forecasting performance superior to the benchmark is obtained only when $c = 1.96$ and when ϕ is selected to be very small ($\phi \leq 0.05$), respectively. The ridge regressions produce significantly improved forecasts for $\gamma \geq 20$, while the subset regressions do so for all but the largest models, i.e., as long as $k \leq 9$. Notably, the rejections are much stronger for many of the subset regressions, with $p-$values below 1%

as long as $k \leq 5$. Similar results are obtained when the encompassing test of Harvey, Leybourne, and Newbold (1998) is adopted.

4.4.1 Recursive selection of hyperparameters

Our results so far show that the choice of hyperparameter can matter a great deal for the performance of many of the combination approaches. It is therefore important to establish whether such hyperparameters can be chosen recursively, in "real time" so as to deliver good forecasting performance. To this end, we conduct an experiment that, at each point in time, uses the data up to this point (but not thereafter) to select the value of the hyper parameter which would have given the best performance. Figure 9 shows the recursively chosen values for the hyperparameters. The subset regression approach always chooses $k = 2$ or $k = 3$, with $k = 2$ being chosen almost exclusively from 1990 onwards. The value for γ chosen under the ridge approach fluctuates between 100 and 200. The critical value, c, in the bagging approach fluctuates between 1.2 and 2.2, while ϕ fluctuates between zero and 100 under the BMA approach.

Table 3 shows the resulting forecast performance numbers from this exercise. First consider the results under an expanding window, reported in Panel A. The univariate regression approach is very poor by this measure, as is the BMA approach, both generating negative R^2-values. Bagging produces an R^2 of 0.3%, while the Ridge approach generates an R^2-value around 0.7%. The best approach, however, is the subset regression method which generates an R^2-value of 1.5%. Using the Clark-West $p-$values, the subset, ridge, bagging and BMA forecasts all improve on the prevailing mean forecast at the 10% significance level.

5 Conclusion

We propose a new forecast combination approach that averages forecasts across complete subset regressions with the same number of predictor variables and thus the same degree of model complexity. In many forecasting situations the trade-off between model complexity and model fit is such that subset combinations perform well for a relatively small number of included predictors. Moreover, we find that subset regression combinations often can do better than the simple equal-weighted combinations which include all models, small and large, and hence do not penalize sufficiently for including variables with weak predictive power. In many cases subset regression combinations amount to a form of shrinkage, but one that is more general than the conventional variable-by-variable shrinkage implied by ridge regression.

Empirically in an analysis of U.S. stock returns, we find that the subset regression approach appears to perform quite well when compared to competing approaches such as ridge regression, bagging, Lasso or Bayesian Model Averaging.

6 Appendix

This appendix provides details of the technical results in the paper.

6.1 Proof of Theorem 1

Proof. The proof follows from aggregating over the finite number $n_{k,K}$ of subset regression estimators

$$\hat{\beta}_i = (S_i' X' X S_i)^{-} (S_i' X' y) = (S_i' \Sigma_X S_i)^{-} (S_i' \Sigma_X) \hat{\beta}_{OLS} + o_p(1)$$

. First, note that

$$\begin{aligned}
\hat{\beta}_i &= (S_i' X' X S_i)^{-} (S_i' X' y) \\
&= (S_i' X' X S_i)^{-} (S_i' X' X) \hat{\beta}_{OLS} \\
&= (S_i' \Sigma_X S_i)^{-} (S_i' \Sigma_X) \hat{\beta}_{OLS} \\
&= \left[(S_i' X' X S_i)^{-} (S_i' X' X) - (S_i' \Sigma_X S_i)^{-} (S_i' \Sigma_X) \right] \hat{\beta}_{OLS}.
\end{aligned}$$

Since $\hat{\beta}_{OLS} \to^p \beta$ and $T^{-1}X'X \to \Sigma_X$, we have

$$\begin{aligned} &(S_i'X'XS_i)^{-}(S_i'X'X) - (S_i'\Sigma_X S_i)^{-}(S_i'\Sigma_X) \\ &= (S_i'T^{-1}X'XS_i)^{-}(S_i'\Sigma_X) - (S_i'\Sigma_X S_i)^{-}(S_i'\Sigma_X) + o_p(1) \quad (16) \\ &= \left[(S_i'T^{-1}X'XS_i)^{-} - (S_i'\Sigma_X S_i)^{-}\right](S_i'\Sigma_X) + o_p(1). \end{aligned}$$

$S_i'T^{-1}X'XS_i$ can be rearranged so that the upper $k \times k$ block is $T^{-1}X^{*\prime}X^*$ where X^* contains the k regressors included in the i^{th} regression. Since $T^{-1}X'X \to^p \Sigma_X$, then $T^{-1}X^{*\prime}X^* \to^p \Sigma_X^*$ (which is the variance covariance matrix of the included regressors) by the definition of convergence in probability for matrices. Rearranging the term $(S_i'T^{-1}X'XS_i)^{-} - (S_i'\Sigma_X S_i)^{-}$ in this way yields an upper $k \times k$ block that is $o_p(1)$ with the remaining blocks equal to zero. The final regressor is a sum over these individual regressors, yielding the result.

rule0.5em0.5em

6.2 Proof of Theorem 2

Proof. From the results of Theorem 1, we have

$$
\begin{aligned}
&\sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_T - \beta_0)' x_T x_T' (\hat{\beta}_T - \beta_0)\right] \\
=&\sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_T - \beta_0)' \Sigma_X (\hat{\beta}_T - \beta_0)\right] \\
&+ \sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_{T,ols} - \beta_0)' \Lambda' (x_T x_T' - \Sigma_X) \Lambda (\hat{\beta}_{T,ols} - \beta_0)\right] \\
&= \sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_T - \beta_0)' \Sigma_X (\hat{\beta}_T - \beta_0)\right] + o_p(1),
\end{aligned}
$$

where the second term is zero by the LIE as we assume $E[(\hat{\beta}_{OLS} - \beta_0)^2 | x_T] = E[(\hat{\beta}_{OLS} - \beta_0)^2]$ and $E[x_T x_T' - \Sigma_X] = 0$.

Now

$$
\begin{aligned}
&T^{1/2} \sigma_\varepsilon^{-1} \Sigma_X^{1/2} (\hat{\beta}_T - \beta_0) \\
&= T^{1/2} \sigma_\varepsilon^{-1} \Lambda (\hat{\beta}_{T,OLS} - \beta_0) + T^{1/2} \sigma_\varepsilon^{-1} (\Lambda - I) \beta_0 + o_p(1) \\
&= T^{1/2} \sigma_\varepsilon^{-1} \Lambda (\hat{\beta}_{T,OLS} - \beta_0) + (\Lambda - I) b + o_p(1),
\end{aligned}
$$

and so

$$\begin{aligned}
&\sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_T-\beta_0)'\Sigma_X(\hat{\beta}_T-\beta_0)\right]\\
=&\sigma_\varepsilon^{-2} E\left[T(\hat{\beta}_{T,OLS}-\beta_0)'\Lambda'\Sigma_X\Lambda(\hat{\beta}_{T,OLS}-\beta_0)\right]\\
&+b'(\Lambda-I)'\Sigma_X(\Lambda-I)b\\
&+2b'(\Lambda-I)'\Sigma_X\Lambda\left(\sigma_\varepsilon^{-1}\left[ET^{1/2}(\hat{\beta}_{T,OLS}-\beta)\right]\right)+o_p(1).
\end{aligned} \tag{17}$$

Since $T^{1/2}(\hat{\beta}_{T,OLS}-\beta) \to^d N(0,\Sigma_X)$, the third term is zero in large enough samples and $\sigma_\varepsilon^{-2}T(\hat{\beta}_{T,OLS}-\beta_0)'\Lambda'\Sigma_X\Lambda(\hat{\beta}_{T,OLS}-\beta_0) \to^d Z'\Lambda'\Sigma_X\Lambda Z$ with $Z \sim N(0,\Sigma_X)$ and $E\left[Z'\Lambda'\Sigma_X\Lambda Z\right] = \sum_{j=1}^K \zeta_j$.

rule0.5em0.5em

References

[1] Aiolfi, M. and C. A. Favero, 2003, Model uncertainty: thick modelling and the predictability of stock returns. Journal of Forecasting 24, 233-254.

[2] Bates, J.M., Granger, C.W.J. (1969). The combination

of forecasts. Operations Research Quarterly 20, 451–468.

[3] Breiman, L., 1996, Bagging predictors. Machine Learning 36, 105-139.

[4] Campbell, J.Y., and Thompson, 2008, Predicting the equity premium out of sample: can anything beat the historical average? Review of Financial Studies 21, 1201-2355

[5] Clark, T.E., and K.D. West, 2007, Approximately normal estimator for equal predictive accuracy in nested models. Journal of Econometrics 127, 291-311.

[6] Clemen, R.T., 1989, Combining forecasts: A review and annotated bibliography. International Journal of Forecasting 5, 559-581.

[7] Goyal, A., and I. Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, Review of Financial Studies 21, 1455-1508

[8] Harvey, D.I., S.J. Leybourne, and P. Newbold 1998, Tests for forecast encompassing, Journal of Business and Economic Statistics 16, 254-259.

[9] Hoerl, A.E., and R.W. Kennard, 1970, Ridge regression: Biased estimation for Nonorthogonal Problems, Technometrics 12, 55-67.

[10] Inoue, A., and L. Killian, 2008, How useful is bagging in forecasting economic time series? A case study of

US consumer price inflation. Journal of the American Statistical Association 103, 511-522.

[11] Koop, G., 2003, Bayesian Econometrics, NewYork: John Wiley.

[12] Politis, D., and J.P. Romano, 1992, A circular block-resampling procedure for stationary data, in Exploring the limits of bootstrap, New York: John Wiley, 263-270

[13] Politis, D., and H. White, 2004, Automatic block-length selection for the dependent bootstrap, Econometric Reviews 23, 53-70.

[14] Raftery, A., D. Madigan, and J. Hoeting, 1997, Bayesian model averaging for linear regression models. Journal of the American Statistical Association 97, 179-191.

[15] Rapach, D.E., J.K. Strauss, and G. Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821-862.

[16] Stock, J., and M.W. Watson, 2006, Forecasting with many predictors. Pages 515-554 in Elliott, G., C.W.J. Granger, and A. Timmermann (eds.) Handbook of Economic Forecasting. North Holland

[17] Tibshirani, R., 1996, Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B 58, 267-288.

[18] White, H., 2000, Asymptotic Theory for Econometricians, revised edition. Academic Press, San Diego.

Figure 1: Degree of shrinkage as a function of k (the number of included predictors) and K (the total number of all predictors) assuming a diagonal covariance matrix.

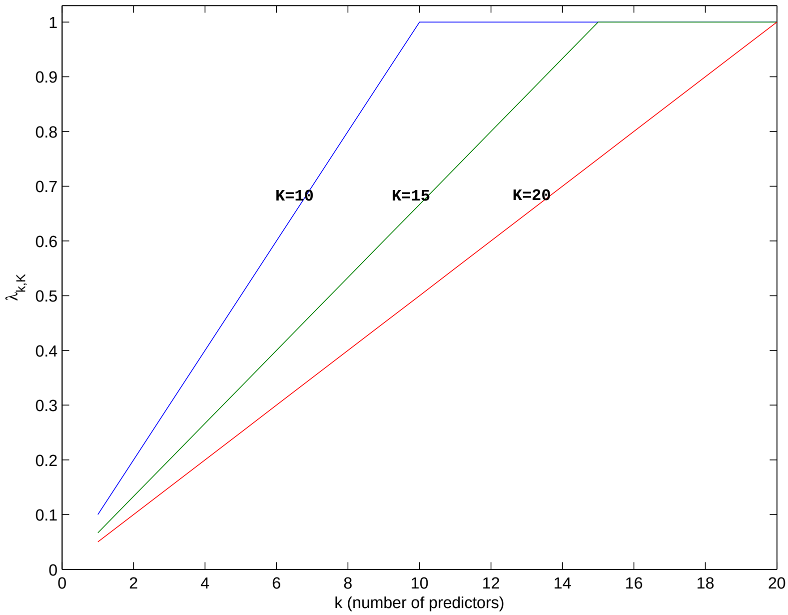


Figure 2: Expected loss for different values of the local-to-zero parameters (b) for $\Sigma = I$ (left) and $\Sigma = \hat{\Sigma}_X$(right)

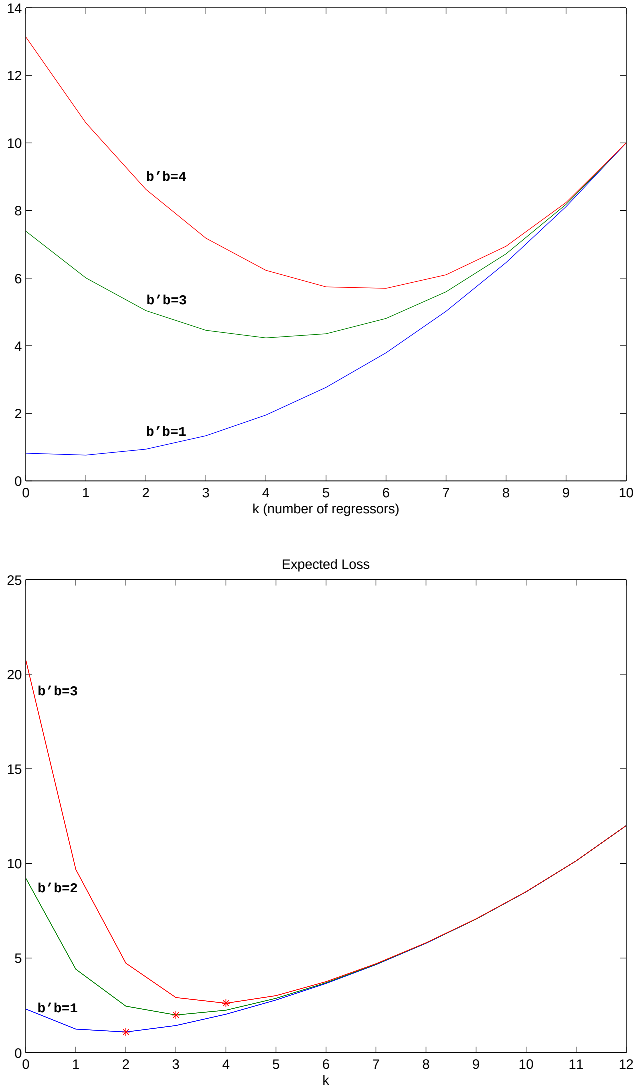


Figure 3: Relative performance of OLS versus subset regression. The figure shows the MSE loss under subset regression relative to OLS (i.e. MSE_{subset}/MSE_{ols}) as a function of ρ, the correlation between predictors, k, the number of included predictors, and K, the total number of predictors. Values below unity show that the subset regression risk is lower than the OLS risk, whereas values above unity indicate that OLS is better.

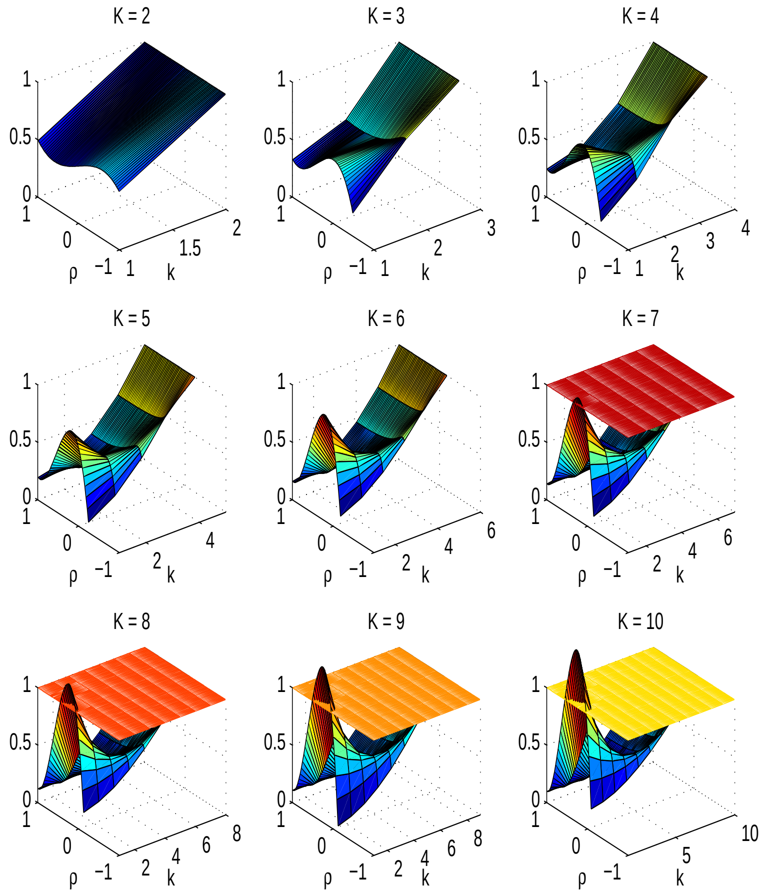


Figure 4: Relative performance of ridge versus subset regression. This figure shows min(0,MSE_{ridge}-MSE_{subset}) as a function of ρ, the correlation between the predictor variables on the x-axis, and γ, the shrinkage parameter used by the ridge approach on the y-axis. Dark red color shows areas where the subset regression produces a lower MSE than the ridge approach, while yellow and blue colors indicate areas where the subset approach produces the highest MSE values. Each box corresponds to a different value of k, the number of predictors included in the forecast model. The graph assumes that b is a vector of ones and K=8.

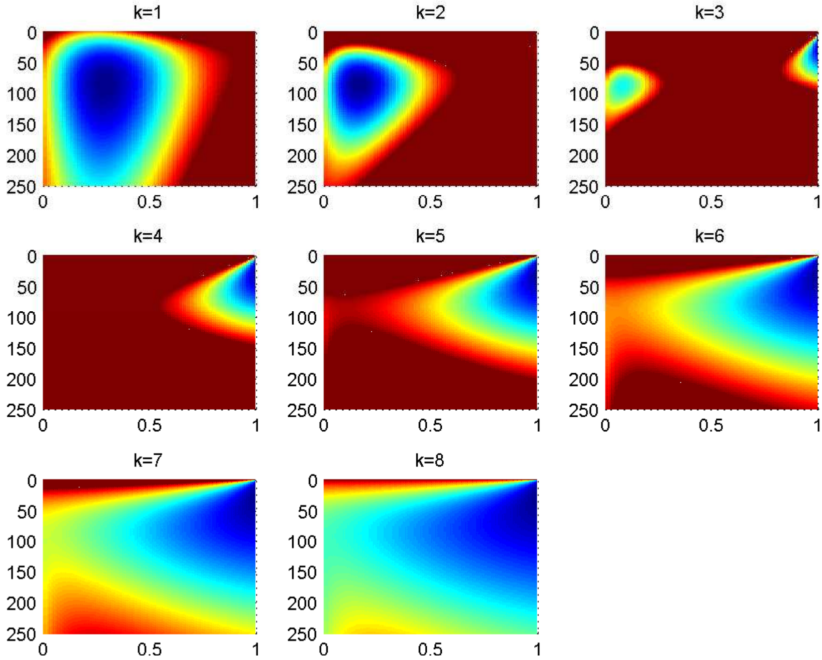


Figure 5: Out-of-sample forecasts of monthly stock returns for different k-variate subset combinations

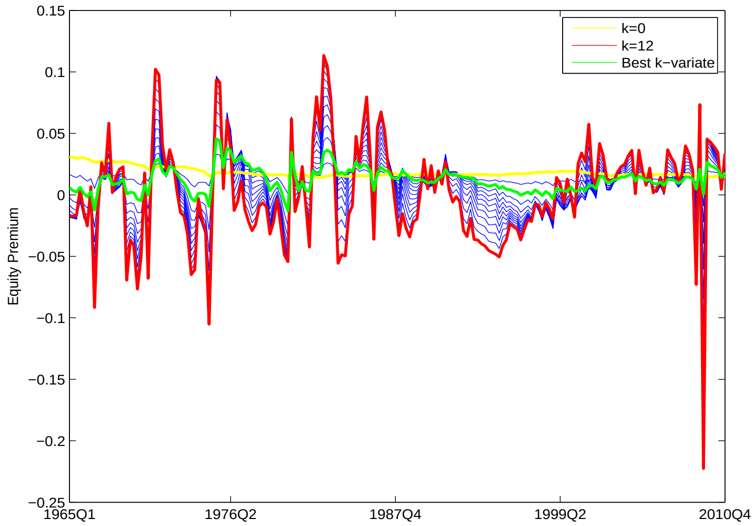


Figure 6: Bias-Variance trade-off. Each blue circle represents a single regression model, grouped according to the number of predictors the model contains. Green circles represent average values computed across all models with a given number of predictors, k, i.e., for a given subset. The horizontal green line shows the average performance computed across all 4096 models while the red dotted line refers to the performance of the equal-weighted forecast combination based on all models. The red line tracks the combination of the k-variate models. The best and worst univariate models are displayed as a text string; AIC and BIC refer to the models recursively selected by these information criteria. The bottom figure displays the scatter plot of the squared bias against the variance for each of the k-variate subset combinations (with k denoted in red) as well as for the individual univariate models.

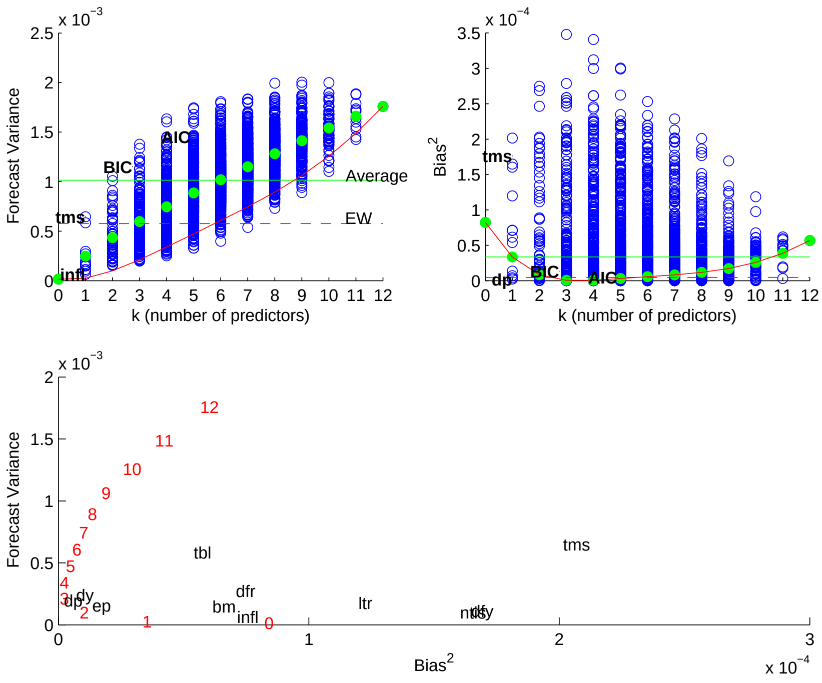


Figure 7: Out-of-sample forecast performance. Each blue circle represents a single regression model, grouped according to the number of predictors the model contains. For a given value of k, the number of possible k-variate models, $\binom{12}{k} = \frac{12!}{k!(12-k)!}$, is reported on the upper x-axes at the top of the diagram. Green circles represent average values computed across all models with a given number of predictors, k. The horizontal green line shows the average performance computed across all 4096 models while the red dotted line refers to the performance of the equal-weighted forecast combination based on all models. The red line tracks the subset combination of the k-variate models. The best and worst univariate models are displayed as text strings above $k = 1$; AIC and BIC refer to the models recursively selected by these information criteria.

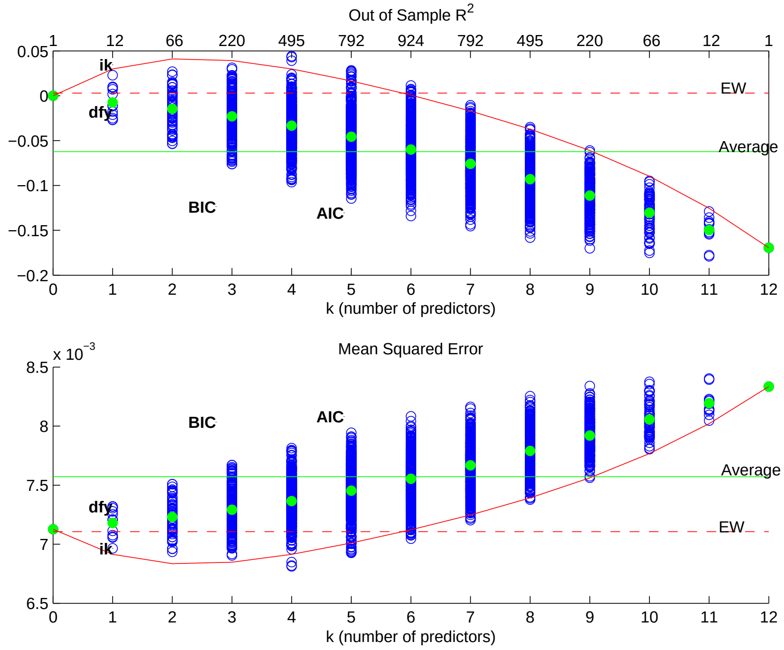


Figure 8: Performance of pools of randomly selected models. At each point in time, n models are randomly selected (without replacement), their forecasts pooled, and the forecast performance recorded. This procedure is repeated 1,000 times. The red line tracks the median value across these trials. For comparison, the green line shows the performance of the combination of all 4,096 models.

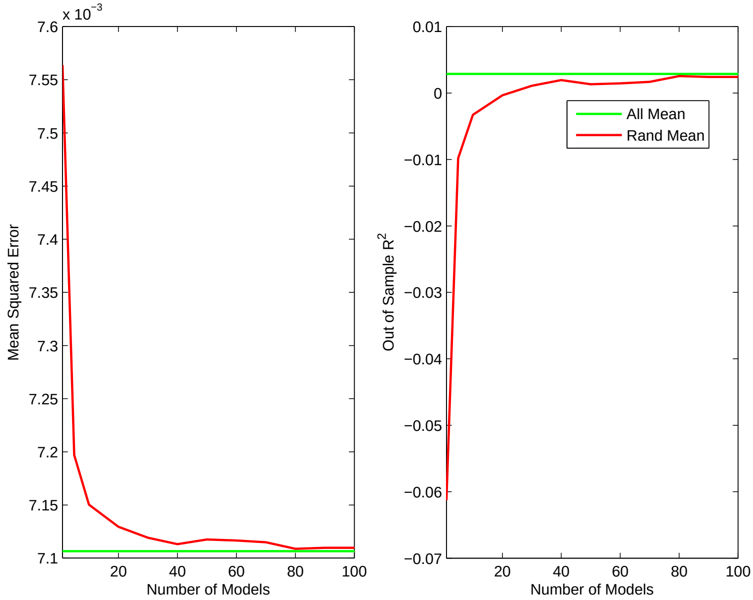


Figure 9: Recursive choice of parameter values. For each approach the graphs shows the value of the design parameter that, at each point in time, gets selected based on its recursive performance evaluated up to that point in time. Parameter selection is performed from the discrete grid of values described in the text of the paper.

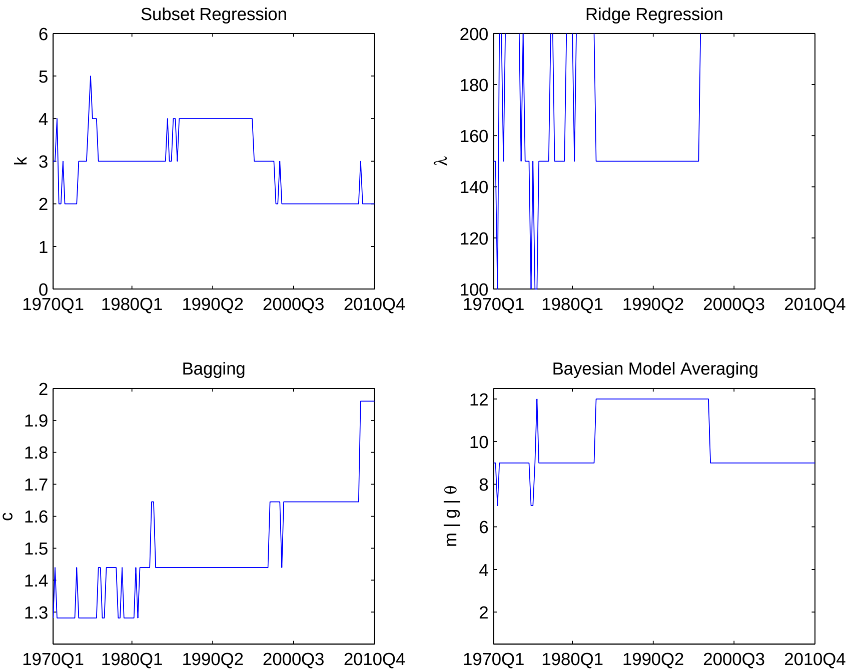


Table 1: Monte Carlo Simulation Results. This table reports the R^2 from a linear prediction model $y_{t+1} = x_t'\beta + \epsilon_{t+1}$, with X containing eight predictors. The X-variables and ϵ are assumed to be normally distributed and i.i.d., while $\beta_i = \frac{\sigma_\epsilon}{\sqrt{T}}$. The covariance matrix of the predictor variables has ones on the diagonal and ρ in all off-diagonal cells, so ρ controls the degree of correlation among the predictors. All forecasting methods only use information up to time T to produce predictions $\hat{y}_{T+1}^{(j)}$, where j refers to the simulation number. The prevailing mean forecast is $\bar{y}_{T+1}^{(j)} = \frac{1}{T}\sum_{t=1}^{T} y_t^{(j)}$. The reported out of sample R^2 are computed as $R^2 = 1 - \frac{\sum_{j=1}^{25.000}(y_{T+1}^{(j)} - \hat{y}_{T+1}^{(j)})^2}{\sum_{j=1}^{25.000}(y_{T+1}^{(j)} - \bar{y}_{T+1}^{(j)})^2}$ The results are based on 25,000 simulations and a sample size of 100 observations.

	b=[1 1 1 1 1 1 1 1]								b=[1 1 1 1 0 0 0 0]							
	Subset Regression															
	R^2								R^2							
k	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
ρ																
0	1.613	2.737	3.378	**3.535**	3.196	2.340	0.935	-1.063	0.827	1.266	1.317	0.975	0.227	-0.949	-2.582	-4.714
0.25	10.093	13.824	15.163	**15.354**	14.866	13.890	12.503	10.724	2.922	3.913	**4.066**	3.698	2.910	1.729	0.144	-1.875
0.5	21.285	24.443	**25.032**	**24.797**	24.121	23.110	21.788	20.144	6.527	**7.361**	**7.266**	6.721	5.811	4.543	2.897	0.838
0.75	31.438	**32.414**	**32.278**	31.781	31.023	30.016	28.750	27.200	10.272	**10.439**	10.089	9.417	8.435	7.126	5.464	3.410
0.95	**37.171**	**37.143**	36.893	36.448	35.802	34.944	33.855	32.511	**12.744**	12.678	12.355	11.764	10.890	9.715	8.216	6.358
	Ridge Regression															
	R^2								R^2							
γ	30	60	90	120	150	180	210	240	30	60	90	120	150	180	210	240
ρ																
0	2.280	3.260	3.523	3.523	3.420	3.277	3.124	2.971	-1.027	0.327	0.911	1.176	1.293	1.334	1.335	1.316
0.25	14.115	15.047	15.196	15.018	14.687	14.285	13.853	13.414	2.011	3.268	3.762	3.951	3.998	3.972	3.907	3.822
0.5	23.938	24.721	24.796	24.591	24.247	23.828	23.368	22.888	5.596	6.730	7.115	7.236	7.238	7.179	7.086	6.973
0.75	31.732	32.192	32.149	31.924	31.606	31.230	30.817	30.380	9.382	10.107	10.286	10.308	10.260	10.176	10.069	9.948
0.95	37.121	37.030	36.823	36.551	36.231	35.875	35.492	35.088	12.698	12.729	12.685	12.613	12.521	12.414	12.296	12.170
	Lasso															
	R^2								R^2							
ψ	1	15	30	45	60	75	90	100	1	15	30	45	60	75	90	100
ρ																
0	-0.000	1.963	1.828	0.726	-0.129	-0.566	-0.766	-0.831	0.000	0.899	-0.597	-2.274	-3.322	-3.828	-4.050	-4.121
0.25	-0.000	10.209	13.100	12.562	11.604	10.995	10.703	10.609	-0.000	3.107	2.220	0.473	-0.755	-1.367	-1.634	-1.717
0.5	0.000	18.190	22.198	22.017	20.912	20.082	19.695	19.566	0.000	5.702	5.452	3.469	1.959	1.177	0.855	0.751
0.75	0.000	26.133	30.329	30.340	29.150	28.135	27.676	27.543	-0.000	8.731	9.156	7.197	5.195	4.157	3.738	3.620
0.95	0.000	32.307	36.020	36.093	34.588	33.020	32.338	32.154	-0.000	11.248	12.085	10.751	8.171	6.721	6.184	6.043
	Bagging															
	R^2								R^2							
c	0.38	1.28	1.64	2.24	2.80	3.29	3.89	5.32	0.38	1.28	1.64	2.24	2.80	3.29	3.89	5.32
ρ																
0	-1.142	-0.007	0.730	1.560	1.629	1.270	0.776	0.108	-4.755	-2.949	-1.797	-0.281	0.385	0.471	0.346	0.062
0.25	10.771	11.526	11.740	10.985	8.525	5.848	3.049	0.358	-1.803	-0.061	1.019	2.168	2.062	1.523	0.841	0.107
0.5	20.152	20.996	21.256	19.663	14.859	9.736	4.812	0.487	0.865	2.914	4.155	5.016	4.045	2.661	1.320	0.137
0.75	27.175	28.391	28.774	26.229	18.824	11.661	5.264	0.433	3.411	5.703	7.056	7.474	5.483	3.400	1.515	0.132
0.95	32.567	34.175	34.477	29.855	19.877	11.523	4.814	0.336	6.422	8.918	10.232	9.857	6.740	3.948	1.644	0.116
	Bayesian Model Averaging															
	R^2								R^2							
ϕ	0	0.01	0.05	0.1	0.3	1	5	100	0	0.01	0.05	0.1	0.3	1	5	100
ρ																
0	3.328	3.334	2.962	1.920	1.344	1.351	1.270	0.177	0.804	0.812	0.512	-0.303	-0.620	-0.088	0.286	-0.023
0.25	14.900	14.798	13.966	13.042	12.447	11.953	10.899	7.838	3.525	3.443	3.028	2.275	1.966	2.426	2.598	1.405
0.5	24.515	24.372	23.177	22.524	22.150	21.859	21.106	18.189	6.620	6.518	5.837	5.342	4.973	5.636	5.730	4.267
0.75	31.639	31.463	30.102	29.974	29.912	30.055	29.817	27.610	9.352	9.241	8.616	8.437	8.119	8.536	8.867	8.011
0.95	36.388	36.199	35.178	35.339	35.751	35.814	36.276	35.955	11.694	11.544	11.072	11.172	11.304	11.189	11.908	11.277

Table 2: Out-of-sample forecast performance for U.S. stock returns. Panel A displays the out-of-sample forecast performance for the univariate models, Panel B for the Subset Regression, Panel C for the combination of random models (averaged across 1000 draws), Panel D for Ridge Regression, Panel E for Bagging and Panel F for Bayesian Model Averaging. The p-values associated with the out-of-sample R^2 are based on the one-sided test of Clark & West (2007), and the encompassing test of Harvey, Leybourne & Newbold (1998). All forecasts of quarterly stock returns are computed recursively and cover the period 1965Q1-2010Q4. Except for the p-values, all entries are in percentages.

Out Of Sample from 1965Q1 to 2010Q4

Panel A: Univariate

variable	MSE	oosR^2	p-val$_{CW}$	p-val$_{HLN}$
dp	0.708	0.708	0.039	0.041
dy	0.706	0.986	0.030	0.031
ep	0.720	-1.066	0.297	0.298
bm	0.725	-1.767	0.427	0.428
ntis	0.728	-2.115	0.630	0.629
tbl	0.731	-2.502	0.046	0.048
ltr	0.721	-1.150	0.305	0.306
tms	0.732	-2.672	0.056	0.058
dfy	0.732	-2.699	0.717	0.716
dfr	0.706	0.906	0.110	0.112
infl	0.711	0.192	0.307	0.308
ik	0.696	2.281	0.010	0.011

Panel B: Subset Regression

k	MSE	oosR^2	p-val$_{CW}$	p-val$_{HLN}$
0	0.713	0.000		
1	0.691	2.991	0.002	0.002
2	0.684	4.097	0.004	0.004
3	0.685	3.923	0.006	0.007
4	0.691	2.985	0.009	0.010
5	0.701	1.643	0.014	0.015
6	0.712	0.073	0.020	0.021
7	0.725	-1.696	0.027	0.028
8	0.739	-3.716	0.035	0.037
9	0.756	-6.096	0.046	0.047
10	0.777	-8.979	0.058	0.059
11	0.802	-12.535	0.072	0.074
12	0.833	-16.948	0.090	0.092

Panel C: Lasso

ψ	MSE	oosR^2	p-val$_{CW}$	p-val$_{HLN}$
1	0.713	0.000	0.255	0.256
2	0.712	0.055	0.392	0.393
3	0.712	0.073	0.370	0.371
4	0.714	-0.195	0.450	0.450
5	0.717	-0.662	0.502	0.502
10	0.724	-1.565	0.352	0.353
20	0.733	-2.829	0.140	0.142
50	0.782	-9.721	0.056	0.058
100	0.831	-16.616	0.088	0.090

Panel D: Ridge Regression

γ	MSE	oosR^2	p-val$_{CW}$	p-val$_{HLN}$
0.5	0.824	-15.630	0.084	0.086
1	0.817	-14.671	0.080	0.082
2	0.807	-13.268	0.074	0.076
3	0.800	-12.227	0.070	0.072
4	0.794	-11.389	0.068	0.070
5	0.789	-10.684	0.065	0.067
10	0.771	-8.185	0.057	0.059
20	0.750	-5.289	0.047	0.049
50	0.722	-1.314	0.032	0.034
100	0.704	1.203	0.024	0.025
150	0.697	2.266	0.020	0.021
200	0.693	2.793	0.017	0.019

Panel E: Bagging

c	MSE	oosR^2	p-val$_{CW}$	p-val$_{HLN}$
0.3853	0.754	-5.754	0.156	0.158
0.6745	0.745	-4.557	0.137	0.139
1.2816	0.717	-0.613	0.082	0.084
1.4395	0.710	0.379	0.065	0.067
1.6449	0.705	1.137	0.055	0.057
1.96	0.700	1.725	0.044	0.045
2.2414	0.703	1.368	0.059	0.061
2.5758	0.706	0.979	0.080	0.082
2.807	0.706	0.945	0.069	0.071
3.0233	0.709	0.579	0.116	0.118
3.2905	0.710	0.418	0.131	0.133
3.4808	0.711	0.222	0.218	0.219
3.8906	0.712	0.105	0.268	0.269
4.4172	0.712	0.072	0.223	0.224
5.3267	0.713	0.021	0.349	0.350

Panel F: Bayesian Model Averaging

ϕ	MSE	oosR^2	p-val$_{CW}$	p-val$_{HLN}$
0	0.713	0.016	0.022	0.023
0.01	0.708	0.726	0.016	0.017
0.05	0.737	-3.474	0.049	0.051
0.1	0.754	-5.817	0.077	0.079
0.2	0.763	-7.095	0.099	0.101
0.3	0.768	-7.799	0.121	0.123
0.4	0.770	-8.034	0.144	0.146
0.5	0.768	-7.796	0.160	0.162
1	0.762	-6.961	0.193	0.195
2	0.754	-5.822	0.237	0.238
5	0.749	-5.054	0.337	0.338
100	0.722	-1.363	0.445	0.445

Table 3: Out-of-sample forecast performance with recursively selected model parameters. This table displays the out-of-sample forecast performance when the model parameters are chosen recursively in a pseudo out-of-sample experiment with an expanding estimation window. p-values are based on Clark & West (2007). The forecast evaluation period is 1970Q1-2010Q4. Except for the p-values, all entries are in percentages.

	MSE	$oosR^2$	p-val$_{CW}$	p-val$_{HLN}$
Univariate	0.826	-9.805	0.740	0.739
Subset	0.741	1.515	0.074	0.076
Ridge	0.747	0.704	0.076	0.079
Bagging	0.750	0.328	0.075	0.096
BMA	0.761	-1.093	0.080	0.089
Lasso	0.769	-2.137	0.407	0.408

Chapter III
Predictive Dynamics in Commodity Prices

1 Introduction

Commodity markets have gained significant investor interest in recent years. According to the Investment Company Institute, total net assets of commodity exchange traded funds grew from $1bn in 2004 to more than $100bn in 2010.[1] Commodity markets, particularly those for precious metals, have also been proposed as a vehicle for hedging investors' exposure to inflation risk. This has featured prominently recently due to central bank implementation of quantitative easing policies combined with increased uncertainty about future inflation rates. Increases in commodity prices, notably crude oil, have also been linked to economic recessions and deterioration in growth prospects.[2]

With few exceptions, however, little is known about the extent to which commodity prices are predictable and how they co-vary with economic state variables. Bessembinder

[1] 2011 Investment Company Fact Book, 51st Edition.

[2] Hamilton (2011) notes that ten of eleven postwar recessions were preceded by sharp increases in the price of crude petroleum.

and Chan (1992) find that T-bill yields, the dividend yield and the junk bond premium have limited predictive power over movements in agricultural, metals and currency futures prices. Hong and Yogo (2011) find evidence of limited in-sample predictability of commodity spot and futures returns. While predictability of commodity prices has thus been studied by previous authors, several unexamined questions remain.

First, previous studies were concerned with in-sample predictability of commodity prices. However, as pointed out by studies of stock market return predictability, in-sample predictability is not synonymous with the ability to predict returns out-of-sample given the historically available sample information, see, e.g., Pesaran and Timmermann (1995) and Goyal and Welch (2008). Specifically, there is no guarantee that in-sample return predictability could have been used in real time by investors to produce more accurate forecasts of commodity prices than a simple constant return benchmark model.

Second, the commodity price literature has mostly considered predictor variables identified in the literature on predictability of stock and bond returns. Broader measures of macroeconomic risk have not been examined to the same extent. This is important since such variables could well provide good measures of either production or storage costs or, alternatively, time-varying risk premia, both of which could induce predictability in commodity spot price changes.

Third, the literature on predictability of commodity prices has not considered the extent to which such return predictability varies over the economic cycle. This is an important shortcoming given the evidence in Rapach, Strauss and Zhou (2010) and Henkel, Martin, and Nardari (2011) that predictability of stock returns is largely confined to economic recessions. One interpretation of this finding is that expected returns vary more during economic recessions than during expansions. Clearly it is of interest to see if a similar finding carries over to commodity markets for which the state of the economy can be expected to play an important role.

Fourth, much of the work on predictability of commodity price movements has focused on futures prices, while spot prices have received less attention. Spot prices are of separate interest, however, as they affect producer costs and, in turn, price inflation. Moreover, spot and futures prices can be expected to be affected by similar risk premium variations. For example, Acharya et al. (2011) propose a model in which producers' hedging demand induces a common component in spot and futures prices. Speculators are assumed to be liquidity constrained and so producers' hedging demand affect optimal inventory holdings and equilibrium spot prices. In their model, expected spot prices reflect a common risk term as well as inventory stock-out and supply effects. Empirically, Acharya et al. (2011) find mild evidence of predictability of petroleum spot returns from fundamental hedging demand variables as well as from the term spread.

Our paper makes several contributions. First, we explore out-of-sample return predictability for a range of commodity spot price indexes over the 20-year period 1991-2010. In so doing, we consider a wider set of predictors, including macro variables measuring the state of the economy such as inflation, money supply growth, growth in industrial production and changes in the unemployment rate. We separately consider predictability at the monthly, quarterly, and annual horizons and in different economic states.

We find that out-of-sample return predictability varies considerably across different horizons. Specifically, there is modest evidence of out-of-sample predictability of monthly movements in metals and raw industrials commodity spot price indexes as well as for the aggregate commodity spot price index. Specifically, individual predictor variables such as the T-bill rate, the default return spread (the return difference between long-term corporate and government bonds), and money supply growth appear able to predict commodity prices at the monthly frequency. At longer horizons, the evidence on out-of-sample commodity price predictability strengthens considerably. For example, at the quarterly horizon, variables such as the T-bill rate, investment-capital ratio, money supply growth and also the rate of inflation have some predictive power over changes in raw industrials and metals spot prices. At the annual horizon, the evidence is even stronger with a host of similar predictor variables apparently capable of predicting movements in commodity prices. Interestingly, the estimated coefficient on inflation is negative, suggesting that current inflation is negatively related to future commodity price movements.

Second, we find that the only variable capable of consistently predicting commodity spot price movements at both the monthly, quarterly, and annual horizons is the growth in the narrow money supply, M1. Moreover, this variable proved successful at predicting the recent surge in commodity prices following the global financial crisis and the ensuing expansionary monetary policy.

Third, return predictability is notably stronger for the raw industrials and metals indexes and weaker for foods, fats-oils, livestock, and textile indexes.

Fourth, whereas there is little evidence of predictability of commodity prices during expansion periods, there is stronger evidence that some macrovariables predict commodity price movements during recessions. For example, this holds for inflation which fails to predicts commodity price movements in expansions, whereas its predictive power over commodity returns is far stronger during recessions. Similar findings hold for growth in industrial production and money supply growth.

Fifth, and finally, we consider predictability of the realized (log-) commodity volatility. Few, if any, state variables appear capable of improving upon the out-of-sample predictive accuracy of an AR(1) model for commodity volatility. Interestingly, however, during economic recessions several variables, most notably the macroeconomic variables (growth in industrial production, money supply growth, and changes in the unemployment rate), produce better out-of-sample forecasts of monthly commodity market volatility when added to the AR(1) model. The evidence is weaker at the quarterly and annual horizons, although at the quarterly horizon the inflation rate produces notably better out-of-sample forecasts of commodity market volatility in recessions when added to the AR(1) model. We also find that the variables that are capable of predicting increasing commodity prices are different from those predicting price declines. Specifically, whereas lagged volatility, lagged returns, and the lagged money supply growth proved capable of predicting the magnitude of increases in commodity prices, the inflation rate plays a much more prominent role when it comes to predicting the magnitude of declines in commodity prices, decreasing inflation being linked to lower expected commodity prices.

The outline of the paper is as follows. Section 2 introduces the data. Section 3 presents empirical results for univariate models used to capture predictability of movements in commodity spot prices associated with individual predictor variables. Section 4 explores predictability from multivariate predictability models. Section 5 analyses predictability of commodity price volatility and separately considers price decreases versus increases. Finally, Section 6 concludes.

2 Data

This section describes the data sources for the commodity prices and predictor variables and provides a brief characterization of our data.

2.1 Commodity prices

Commodity spot prices are measured by the Reuters Jeffries indexes compiled by the Commodity Research Bureau. These are computed as an unweighted geometric mean of the individual commodity prices relative to their base periods which reduces the impact of extreme movements in individual commodity prices in the index. We use end-of-month prices measured at close, denominated in US dollars. When available, the spot price is based on the listed exchange price for a commodity of standard quality but bid or ask prices are used if a spot price is not readily available.

The sample period is 1947m1-2010m12.[3]

The data comprises an aggregate spot market index (ticker: CMCRBSPD) that is based on 22 individual commodities. This broad index is split into two major indexes, namely raw industrials (CMCRBIND, including burlap, copper scrap, cotton, hides, lead scrap, print cloth, rosin, rubber, steel scrap, tallow, tin, wool tops, and zinc), and foodstuffs (CMCRBFOD, including butter, cocoa beans, corn, cottonseed oil, hogs, lard, steers, sugar, and wheat). In turn, these indexes are subdivided into metals (CMCRBMED, including copper scrap, lead scrap, steel scrap, tin, and zinc), textiles and fibers (CMCRBTXD, including burlap, cotton, print cloth, and wool tops), fats and oils (CMCRBFAD, including butter, cottonseed oil, lard, and tallow), and livestock and products (CMCRBLID, including hides, hogs, lard, steers, and tallow).

[3] For further detail, see http://www.crbtrader.com/crbindex/spot.

2.2 Predictor variables

As predictors we consider a set of 11 state variables. The first seven variables are from the literature on stock return predictability and was previously used by Goyal and Welch (2008). Specifically, the *Dividend Price Ratio (dp),* is measured as the difference between the log of the 12-month moving sum of dividends and the log of the S&P 500 index; *Treasure Bill (tbl),* is the 3-Month Treasury Bill (secondary market) rate; *Long Term Rate of Returns (ltr)* is the long-term rate of returns on US Bonds; *Term Spread (tms)* is the difference between the long term yield on government bonds and the Treasury Bill rate; *Default Return Spread (dfr)* is the difference between long-term corporate bond and long-term government bond returns; *Inflation (infl)* is the (log) growth of the Consumer Price Index (All Urban Consumers); *Investment to Capital Ratio (ik)* is the ratio of aggregate investments to aggregate capital for the whole economy. These series have been constructed by Goyal & Welch (2008) and are available on the authors' web site.

To measure the broad state of the economy, we consider a range of macroeconomic variables. *Industrial Production (ΔIND)* is the monthly growth in Industrial Production as reported by the Federal Reserve Bank of St. Louis (FRED mnemonic: INDPRO). Quarterly and annual series are obtained averaging monthly values over each quarter and year. For example, letting $IND_{Y2:M2}$ and $IND_{Y2:Q2}$ denote industrial production during the second month and second quarter of the second year in the sample, monthly, quarterly and annual growth rates are computed as follows

$$\begin{aligned} \Delta I_{Y2:M2} &= ln(I_{Y2:M2}) - ln(I_{Y2:M1}) \\ \Delta I_{Y2:Q2} &= ln\left(\sum_{j=4}^{6} I_{Y2:Mj}\right) - ln\left(\sum_{i=1}^{3} I_{Y2:Mi}\right) \\ \Delta I_{Y2} &= ln\left(\sum_{j=1}^{12} I_{Y2:Mj}\right) - ln\left(\sum_{i=1}^{12} I_{Y1:Mi}\right) \end{aligned} \quad (1)$$

Unemployment (ΔUN), is the change in the monthly unemployment rate (FRED mnemonic: UNRATE); quarterly and annual series are obtained averaging monthly values over each quarter and year. Monthly, quarterly and annual growth rates are computed as for Industrial production. *Money Stock ($\Delta M1$),* is the year-on-year growth in the monthly M1 money stock (FRED mnemonic: M1SL), with quarterly and annual series again obtained by averaging monthly values over each quarter and year:

$$\begin{aligned} \Delta M1_{Y2:M2} &= ln(M1_{Y2:M2}) - ln(M1_{Y1:M2}) \\ \Delta M1_{Y2:Q2} &= ln\left(\sum_{j=1}^{3} M1_{Y2:Mj}\right) - ln\left(\sum_{i=1}^{3} M1_{Y1:Mi}\right) \\ \Delta M1_{Y2} &= ln\left(\sum_{j=1}^{12} M1_{Y2:Mj}\right) - ln\left(\sum_{i=1}^{12} M1_{Y1:Mi}\right) \end{aligned} \quad (2)$$

In addition to these variables, we construct a realized commodity price volatility measure. *Commodity volatility (cvol),* is the square root of the sum of squared daily returns on the Dow Jones-AIG Commodity Index available from Global Financial Data (mnemonic: DJCD) over the months, quarters and years according to the adopted frequency:[4]

$$cvol_{Y1:M2} = \sqrt{\sum_{t \in Y1:M2} r_t^2} \quad cvol_{Y1} = \sqrt{\sum_{t \in Y1} r_t^2}. \tag{3}$$

2.3 Data Characteristics

Figure 1 presents plots of the nominal commodity spot prices for the seven indexes. Many of the indexes underwent sharp increases during 1973 following the concurrent spike in oil prices. This was followed by more stable nominal prices until 2006, at which point prices rose sharply until mid-2008, only to decline dramatically (with exception of textiles) during the global financial crisis. Between March 2009 and the end of our sample (2010), commodity prices recovered sharply.

Figure 2 shows the associated monthly commodity returns. Percentage price changes from holding a commodity from the end of period t to the end of period $t+1$ is computed as $r_{t+1} = (P_{t+1} - P_t)/P_t$, where P_t and P_{t+1} are the as-

[4]We use the Dow Jones-AIG index, rather than the Reuters/Jeffries-CRB index, because the latter does not have complete daily return records going back to 1947.

sociated commodity prices. Periods of high volatility clearly accompanied the episodes with large adjustments in price levels. In addition to the high volatility during the global financial crisis, commodity markets also saw high volatility in the late 40s/early 50s and again around the oil price hikes in the early seventies.

Table 1 reports descriptive statistics for the commodity spot price changes. For comparison, we also use returns data on a stock market portfolio (based on the value-weighted CRSP index) and on the 10-year Treasury bond. To facilitate our subsequent analysis of monthly, quarterly, and annual price movements, we present statistics for all three frequencies. All commodity indexes earned positive nominal mean returns over the period, ranging from 0.18% per month for textiles to 0.43% per month for metals. These values are dominated by the mean returns on both stocks and T-bonds, however, at 0.98% and 0.48%, respectively.

Volatility varied a great deal across commodities, being lowest for industrials (2.84% per month) which was less than half the level observed for fat and oils (6.61%). All commodity returns were more volatile than the bond returns, while three indexes (fat and oils, livestock, and metals) were more volatile than the stock return series. Interestingly, while stock returns are left-skewed, all but one of the commodity series (metals) are right-skewed, suggesting that large increases in commodity prices are more common than large declines. Moreover, the kurtosis of commodity returns, a measure frequently used to gauge how “fat-tailed” returns are, exceeds that of both stock and bond returns.

While stock and bond returns are not serially correlated, three of the commodity spot return series (industrials, metals, and the broad index) are quite persistent with a first order autocorrelation around 0.3 at the monthly horizon. This serial correlation is only mildly reduced at the quarterly horizon, but disappears in the annual data. Since trades in spot markets can be associated with storage costs and risk premia may also be time-varying, serial correlation in spot market returns is clearly not proof of arbitrage opportunities.[5] Deaton and Laroque (1992, 1996) consider a model where speculators' trades induce serial correlation in commodity price levels, although they also find that speculation cannot explain the observed degree of serial correlation in commodity prices.

An analysis of cross-correlations among commodity returns shows that fats and oils, foods, and livestock prices are strongly correlated, while in turn industrials and metals are also strongly correlated. Textile prices tend to have the weakest correlation with other commodity price indexes.

3 Empirical results

Following studies on stock return predictability such as Goyal and Welch (2008), Campbell and Thompson (2008), and Rapach, Strauss and Zhou (2010), we first consider simple

[5]We also examined serial correlation in a range of futures indexes and found, as expected, that such serial correlation is absent in the corresponding futures returns.

univariate prediction models of commodity price changes. These have the advantage of revealing the marginal predictive power of individual predictor variables.

We specify the univariate return regressions as follows:

$$r_{t+1:t+h} \equiv \frac{P_{t+h} - P_t}{P_t} = \beta_{0h} + \beta_{1h} x_t + \varepsilon_{t+1:t+h}, \quad (4)$$

where $r_{t+1:t+h}$ is the cumulated return between the end of period t and the end of period $t+h$, h is the horizon (equal to one, three, and twelve, for the monthly, quarterly, and annual regressions, respectively), and x_t is the lagged predictor variable.

3.1 In-sample return predictability

Pairing each of the commodity price series with each of the individual predictor variables, Table 2 reports in-sample estimates of slope coefficients obtained from equation (4). Panel A reports results for the monthly regressions, while Panels B and C show results for the corresponding quarterly and annual regressions. At the monthly horizon, variables such as the dividend-price ratio and the T-bill rate, which have been identified as predictors of stock and bond returns, fail to be significant for commodity price changes. Conversely, the long term return (ltr) has a negative and significant coefficient for the industrials, metals and broad commodity price index, while this variable generates positive slopes for stock and bond returns. This is similar to the observation by Hong and Yogo (2011) that the yield spread has the opposite sign for stocks and commodity futures returns. Exposure to this variable through a long position in stocks or bonds can therefore be partially hedged by simultaneously taking a long positions in commodities. The coefficient of the default return spread (dfr) is positive and significant for industrials, livestock, and metals, but is insignificant for stocks and the other commodity indexes.[6]

[6]To evaluate statistical significance, we compute bootstrapped $p-$values repeating the following procedure 5,000 times: (i) resample T pairs of $(\hat{\varepsilon}, \hat{\eta})$, with replacement, from OLS residuals in regressions $r_{t+1} = \alpha + \varepsilon_{t+1}$ and $x_{t+1} = \mu + \rho x_t + \eta_{t+1}$; (ii) build up time series of predictors, x_t, from the unconditional mean $\hat{\mu}/(1-\hat{\rho})$ and iterate forward on the x_{t+1} equation using the OLS estimates $\hat{\mu}, \hat{\rho}$ and the resampled

Turning to the macroeconomic state variables, growth in industrial production and growth in the money supply are positively and significantly linked to the subsequent month's price changes in industrials, metals, textiles, and the broad commodity index. In contrast, changes in the unemployment rate are negatively correlated with subsequent metals and industrials price changes. These findings suggest that evidence of increased economic activity are positively correlated with subsequent commodity spot price movements. Unsurprisingly, given the earlier findings of a strong autoregressive component in many of the indexes, the lagged return is significant for most of the commodity price series, though not for stocks.

The evidence on predictability of commodity price move-

---

values of $\hat{\eta}_{t+1}$; (iii) construct time series of returns, r_t, by adding the resampled values of $\hat{\varepsilon}_{t+1}$ to the sample mean (under the null that returns are not predictable); (iv) use the resulting series x_t and r_t to estimate return regressions by OLS; (v) leave out the last $T-N+1$ observations to produce out-of-sample forecasts. The boostrapped p-values associated with the reported βs is the relative frequency with which the (absolute value) of the bootstrapped t-statistics in point (iv) exceed the actual value. The boostrapped p-values associated with the out-of-sample R^2 is the relative frequency with which the bootstrapped values in point (v) exceed the value recorded in the actual data.

ments varies substantially across different horizons. For example, whereas the inflation rate turned out to be insignificant for all commodity price indexes at the monthly horizon, in sharp contrast, at the quarterly horizon this predictor is significant at the 5% level for fats and oils, industrials, livestock, and the broad index, and it is significant at the 10% level for metals. In all cases the slope coefficient is negative. In contrast, the slope coefficients on the growth in money supply, $\Delta M1$, continue to be positive and highly significant for all commodity indexes except for fats and oils, and foods.

Predictability of commodity returns is strongest at the annual horizon, particularly for industrials, metals, and the broad commodity index for which the majority of predictor variables turn out to be significant. Macroeconomic state variables such as inflation, growth in industrial production, growth in the money supply and changes in the unemployment rate generate significant slope coefficients at the annual horizon for these indexes. In contrast, the lagged return is no longer significant at the annual horizon.

We conclude from these results that return predictability varies a great deal across different horizons. Variables such as the inflation rate are insignificant at the monthly horizon but become significant at the quarterly and annual horizons, whereas growth in industrial production is significant in the monthly and annual regressions, but not in the quarterly ones. Only growth in the money supply seems capable of predicting commodity returns across all three horizons. Return predictability is also stronger for industrials and metals and weakest for fats-oils, foods, and textiles.

3.2 Out-of-sample return predictability

Measures of in-sample return predictability such as those reported in Table 2 are not true ex-ante measures of expected returns since they reflect data from the full sample which of course would not have been available to investors in real time. To address this issue, it is common to report out-of-sample predictability measures using recursively estimated parameter values to generate forecasts. For example, setting $z_t = (1\ x_t)'$ and using data from $\tau = 1, ..., t$, least squares parameter estimates can be obtained at time t and used to generate a forecast of r_{t+1}, $\hat{r}_{t+1|t} = \hat{\beta}_t' z_t$. The following period, $t+1$, data from $\tau = 1, ..., t+1$ can be used to obtain an estimate, $\hat{\beta}$, generate a forecast, $\hat{r}_{t+2|t+1} = \hat{\beta}_{t+1}' z_{t+1}$, and so forth. This procedure continues until the end of the sample and ensures that look-ahead bias is absent from the coefficient estimates used to compute the forecasts. In our analysis we reserve data up to 1990:12 to estimate the model parameters and use the remaining 20 years of data, 1991:01-2010:12, for out-of-sample forecast evaluation.

For each of the univariate models Table 3 reports the out-of-sample R^2-value, measured relative to the value obtained from the benchmark model that only includes a constant and so sets $\beta_{1h} = 0$ in equation (4):

$$R^2 = 1 - \frac{\sum_{t=R}^{T-1} (r_{t+1} - \hat{r}_{t+1|t})^2}{\sum_{t=R}^{T-1} (r_{t+1} - \hat{r}_{t+1|t}^{bmk})^2}, \tag{5}$$

where $R = 1990m12$, and $T = 2010m12$. First consider the monthly results in Panel A. Many R^2-values are negative as a result of the effect of parameter estimation error which reduces the precision of the forecast, see, e.g., the discussion in Clark and West (2007) and Inou and Kilian (2008). However, for some of the predictor variables$-$notably, the T-bill rate, the term spread, the default return spread, inflation, growth in industrial production, and money supply growth$-$we find positive out-of-sample R^2-values for three or more of the commodity price series. Excluding lagged returns, the highest values, 4.3% and 4.5%, are obtained for the industrial raw materials and metals returns when the default return spread is used as the predictor. Once again, predictability appears strongest for industrials, metals, and the broad commodity price index, and notably weaker for fats-oils, foods, and livestock.

Overall, however, the single best predictor variable is the one-month lagged return which generates out-of-sample R^2-values of 5.2% (commodity price index), 7.2% (metals), and 9.1% (industrials). Interestingly, this predictor also generates a large negative R^2-value for textiles (-7.3%).

We evaluate the statistical significance of the out-of-sample predictability results using the test statistic proposed by Clark and West (2007). This test statistic measures the difference between the out-of-sample MSE-value of a given forecast versus that of the benchmark constant return model, but corrects for the higher variability of the forecasts from the univariate models that include an additional predictor variable by basing inference on the adjusted mean-squared error:

$$\Delta MSE^{adj} = P^{-1}\sum_{t=R}^{T-1}\bar{e}^2_{t+1|t} - P^{-1}\sum_{t=R}^{T-1}\hat{e}^2_{t+1|t} + P^{-1}\sum_{t=R}^{T-1}(\bar{r}_{t+1|t} - \hat{r}_{t+1|t})^2. \quad (6)$$

Here $\bar{e}^2_{t+1|t}$ is the squared forecast error from the prevailing mean model, $\hat{e}^2_{t+1|t}$ is the squared forecast error from the univariate forecasting model, while $\bar{r}_{t+1|t}$ is the prevailing mean forecast and $\hat{r}_{t+1|t}$ is the forecast from the univariate model that nests the prevailing mean model. $P = T - R$ is the size of the forecast evaluation sample. Positive values of this measure suggest that the benchmark is associated with larger forecast errors and so the univariate prediction model dominates. Notice that the final term in (6) corrects for the typically higher variability associated with the forecasts generated by the larger (univariate) model, relative to the prevailing mean forecast.

The results are very much in line with the out-of-sample R^2-values and show that the forecasts based on the T-bill rate, the default return spread, and money supply growth are significant at the 10% level (or less) for the industrials and metals commodity price indexes. Finally, the forecasts based on the lagged return generate highly significant, positive R^2-values for industrials, metals, and the broad commodity price index.

At the quarterly horizon (Panel B), out-of-sample return predictability grows stronger. In fact, for the univariate models based on the T-bill rate, inflation, and the money supply growth we find positive, and in many cases statistically significant R^2-values, for the majority of the commodity series. Note that predictor variables such as the inflation rate work far better at the quarterly than at the monthly horizon, so, once again, it is not clear that the best prediction model is identical across different horizons.

The tendency for the return predictability results to strengthen when going from the monthly to the quarterly regressions carries over to the annual results where out-of-sample R^2-values in the range 10-20% are found for the models based on the T-bill rate or the term spread and the macroeconomic predictors (growth in industrial production, money supply growth, and changes in the unemployment rate). In sharp contrast with the earlier results, the lagged return does not generate positive out-of-sample R^2-values at the annual horizon.

Conversely, some of the R^2-values become more negative at the quarterly and, particularly, annual horizons. This is to be expected: our forecasts are based on non-overlapping observations which means that there are far fewer data points on which to estimate the annual models than the monthly models. In turn this results in larger estimation errors and so explains the large negative out-of-sample R^2-values.[7]

One way to inspect how return predictability evolves over time is by examining the cumulated sum of squared error differential between the benchmark model and a candidate prediction model proposed by Goyal and Welch (2008):

$$\Delta SSE_t = \sum_{\tau=1}^{t} e_\tau^2(Bmk) - \sum_{\tau=1}^{t} e_\tau^2(Model). \qquad (7)$$

Positive values of this measure indicate that the candidate forecasting model has produced more accurate forecasts than the benchmark model up to that point in time. Periods associated with an increase in ΔSSE suggest that the particular forecasting model produced a lower MSE-value than the benchmark, while conversely declines in ΔSSE suggest that the forecasts were less precise than those based on the benchmark. Hence plots of ΔSSE provide a useful diagnos-

[7] Note that this is not just an issue for the commodity return predictions but also hold for the stock return forecasts for which every single out-of-sample R^2-value is negative at the annual horizon.

tic that helps identify periods of (relative) out- or underperformance. Figure 3 provides such plots for the raw industrials (left windows) and metals (right windows) indexes based on the univariate return prediction model that uses money supply growth as the predictor variable. At the monthly and quarterly horizons, the forecasts underperform in the early sample up to around 1993, before steadily outperforming up to 2002. This is followed by a period of underperformance from 2005-2008, before superior performance returns between 2008 and 2010. Compared with the monthly and quarterly models, the superior performance of the annual forecasts against the prevailing mean model evolves more steadily, as can be seen from the two lower diagrams.[8]

Given the strong performance of the monthly, quarterly, and annual forecasts based on the simple AR(1) model, it is natural to ask if any of the financial and macroeconomic pre-

[8]Further inspection of the out-of-sample forecasts for the monthly and quarterly models suggest that the ability of individual predictor variables to improve on the forecasts from the simple constant return model is not particularly stable over time. The predictor variables that are best able to generate stable outperformance over the benchmark is the T-bill rate, the default return spread (dfr) and money supply growth.

dictor variables help improve the precision of the forecasts, over and above the lagged return. To address this point, we consider a bivariate regression model that includes the lagged return and a single predictor variable:

$$r_{t+1:t+h} = \beta_{0h} + \beta_{1h} r_{t-h+1:t} + \beta_{2h} x_t + \varepsilon_{t+1:t+h}. \qquad (8)$$

The benchmark model is now the AR(1) specification which is obtained by setting $\beta_{2h} = 0$ in equation (8). Table 4 shows the marginal R^2-values, i.e., the change in the out-of-sample R^2 of the forecasting model in equation (8) compared with the AR(1) specification. At the monthly horizon the results reveal little evidence that any predictor adds to the predictive performance of the AR(1) model. At the quarterly horizon, the inflation rate in particular, but also the investment-capital ratio and money supply growth help significantly improve on the AR(1) forecast of returns on the industrials and metals indexes. At the annual horizon, the results are largely unchanged relative to the model that used the simpler prevailing mean specification as the benchmark, and most predictor variables add value over the AR(1) benchmark, particularly for industrials, metals, and the broad commodity index.

3.3 Forecasts of Levels of Commodity Prices

So far we have focused on modeling percentage changes (i.e., returns) on the commodity price indexes. An alternative is to directly predict the price level as a function of its past value and the same list of predictors considered thus far. Although statistical tests suggest that there is a unit root in all of the commodity price indexes, this procedure is still of interest given that it allows us to relax this assumption. Table 5 reports out-of-sample R^2-values for this exercise, computed in percentage terms for comparison with Table 3. The reuslts are quite comparable to those in Table 3. At the monthly horizon, there is only modest evidence of predictability with the default rate and growth in industrial production producing significant results for the industrials, metals and broad commodity indexes. Inflation works well as a predictor at the quarterly horizon, while a broader set of macroeconomic variables in addition to the term spread generate significant R^2-values at the annual horizon.

3.4 Forecasting performance in recessions and expansions

Studies such as Rapach, Strauss and Zhou (2010) and Henkel, Martin, and Nardari (2011) find that predictability of stock returns is stronger during slow growth or recessionary states of the economy. Since many of our predictor variables, particularly the macroeconomic ones, are related to the economic cycle, we next explore if there is state-dependence in the strength of the predictive evidence. To this end, Table 6 compares the out-of-sample R^2-values in recessions, as defined by the NBER, versus expansions. Specifically, we evaluate the statistical significance of differences in predictive power in recessions relative to expansions using regressions of the squared error return difference

$$(r_{t+1:t+h} - \bar{r}_{t+1:t+h|t})^2 - (r_{t+1:t+h} - \hat{r}_{t+1:t+h|t})^2 =$$
$$\alpha + \beta NBER_{t+1} + \varepsilon_{t+1:t+h},$$

where $(r_{t+1:t+h} - \bar{r}_{t+1:t+h|t})^2$ is the squared forecast error of the constant (prevailing mean) benchmark, $(r_{t+1:t+h} - \hat{r}_{t+1:t+h|t})^2$ is the squared forecast error for the univariate prediction model, and $NBER_{t+1}$ is a recession indicator which is unity during recessions and zero during expansions. Positive and significant values of β suggest that the univariate prediction model is more accurate, relative to the benchmark, during recessions than during expansions. Note that by considering forecasting performance relative to the benchmark, we control for the fact that commodity price volatility may be higher during recessions than during expansions.

Table 6 shows that there is little evidence of commodity price predictability during expansions. In contrast, predictability is significantly stronger during recessions. The strongest evidence to this effect is found for the inflation rate. This variable shows no predictive power during expansions, but has strong predictive power$-$with out-of-sample R^2-values up to 3.2% for industrials at the monthly frequency and an R^2-value above 10% for most of the commodity indexes at the quarterly frequency. Similarly, industrial production growth, and growth in money supply show evidence of having significantly stronger predictive power during recessions than during expansions, as does the lagged return.

These findings suggest that predictability of commodity prices is highly state dependent. For example, inflation *does* predict commodity prices, but only in recession states. This suggests perhaps the need for developing models that account for such dependencies, one example being the regime switching models recently reviewed by Ang and Timmermann (2011).

4 Multivariate Regressions

So far we have analyzed the effect of individual predictor variables. It is natural, however, to inquire what happens if multivariate information is used. To this end we study three strategies. First, we use the Akaike (AIC) and Bayesian (BIC) information criteria to select which variables to include among the full set of predictor variables. These criteria trade off model parsimony against fit, with the BIC most heavily penalizing additional included variables. Again we implement the variable selection recursively, at each point in time considering all possible 2^N combinations of predictor variables.

Second, we consider shrinkage methods such as ridge regression and subset combinations which are designed to reduce the effect of parameter estimation error on the forecasts. Ridge regression requires selecting a parameter λ which regulates the amount of shrinkage imposed on the regression coefficients:

$$\hat{\beta}_{\lambda t} = \arg\min_{\lambda} \left(\sum_{\tau=1}^{t} (r_\tau - z'_{\tau-h}\beta_{\lambda t})^2 + \lambda \sum_{j=1}^{K} \beta^2_{\lambda tj} \right). \quad (9)$$

Given a value of λ, and a vector of predictors $z_t = (1\ x'_t)'$, the forecasts are obtained as

$$\hat{r}^{RIDGE}_{t+h|t} = z'_t \hat{\beta}_{\lambda t}. \quad (10)$$

By construction, as $\lambda \to \infty$, $\hat{r}^{RIDGE}_{t+h|t} \to \frac{1}{t-1}\sum_{j=2}^{t} r_j$, so the ridge forecast simply converges to the sample mean. Following Inoue and Kilian (2008), we consider a range of shrinkage values λ.

The subset regression approach, recently introduced by Elliott, Gargano, and Timmermann (2012), uses equal-weighted combinations of forecasts based on all possible models that include a particular subset of the predictor variables. Suppose the set of potential predictor variables includes K different predictors. In our case $K = 11$ or $K = 12$ depending on the horizon. Each subset is defined by the set of regression models that includes a fixed (given) number of regressors, $k \leq K$. Specifically, we run the 'short' regression of r_{t+1} on a particular subset of the regressors, then average the results across all $k \leq K$ dimensional subsets of the regressors to provide an estimator, $\hat{\beta}$, for forecasting. With K regressors in the full model and k regressors chosen for each of the short models there are $K!/(k!(K-k)!)$ subset regressions to average over. For example, the univariate case ($k = 1$) has K such short regressions, each with a single variable. The equal-weighted combination of the forecasts from the individual models is then

$$\hat{r}_{t+1|t} = \frac{1}{K} \sum_{i=1}^{K} x'_{ti} \hat{\beta}_{it}. \tag{11}$$

This strategy was used by Rapach, Strauss and Zhou (2010).

4.1 Empirical Results

First consider the information criteria. Since we apply these criteria recursively, they provide interesting insights into which variables get selected at different points in time. For each of the three frequencies, Figure 4 uses the raw industrials index to present this information for the variables under consideration. At each point in time this graph shows which variables get selected by the AIC (left windows) or BIC (right windows) to predict spot returns.

At the monthly frequency the lagged return, money supply growth, industrial production growth, and the T-bill rate get included by the AIC in almost all periods, whereas the inflation rate, long-term return, and the default return spread get included at certain contiguous blocks of time. In contrast the unemployment rate, commodity volatility, the term spread and the dividend-price ratio never or rarely get selected. The BIC is known to penalize inclusion of additional parameters more heavily than the AIC and so only includes industrial production and lagged returns throughout the sample, whereas money supply growth gets included towards the end of the sample.

At the quarterly frequency, the AIC includes almost all variables all of the time except for the investment-capital ratio and the dividend yield, which never get selected, and the unemployment rate which rarely gets selected. Similarly, the BIC also includes more variables than at the monthly frequency, with the lagged return, long term returns, industrial production, the T-bill rate, and money supply growth featuring prominently. The quarterly BIC results suggest that the preferred model does not remain invariant through time, with industrial production and long term returns getting selected up to 1998, and money supply growth and the T-bill rate selected most periods after 2001.

At the annual frequency, both the AIC and BIC select the unemployment rate, money supply growth and inflation as predictors most periods. In addition, the AIC selects the term spread and commodity volatility during the last three years of the sample.

Figure 5 shows results for the metals price index. While there are many similarities with the results for the raw industrials index, there are also important differences. For example, the long term return now gets selected by the BIC while the commodity volatility gets selected by the AIC at the monthly frequency. At the quarterly horizon, the lagged return gets selected less frequently for metals than it did for raw industrials, while industrial production growth remains important in the models selected by BIC. At the annual frequency the AIC chooses more predictors for metals than it did for raw industrials, and the models selected by the BIC are less stable with both the T-bill rate and commodity price volatility now getting selected in some periods.

Table 7 reports the out-of-sample forecasting performance of the models selected by the AIC or the BIC. For industrials, metals, and the broad commodity index, the monthly out-of-sample R^2-values, reported in Panel A1, are positive and lie in the range 5-11%. Small, positive R^2-values are also obtained for foods and livestock under the models selected by AIC. For the other cases, most notably fats-oils and textiles, negative R^2-values are obtained. At the quarterly frequency (Panel B1), only the AIC manages to generate positive R^2-values for industrials, metals and broad commodity returns. In contrast, at the annual frequency (Panel C1) both the AIC and the BIC produce positive and, in some cases, very large R^2-values.

Figure 7 plots the annual out-of-sample forecasts of raw industrials and metals returns generated by the complete subset regressions over the period 1991-2010. Each line corresponds to a different value, k, tracking the number of included variables in the prediction model. The fewer variables get included in the models, the smoother the averaged forecast tends to be. The figure illustrates that although the forecasting models clearly missed the magnitude of the decline in commodity prices in 2008, they did a better job at predicting the subsequent bounceback in 2009 and 2010.

The predictive performance of the ridge and subset regressions is reported in Table 7. At the monthly frequency, the ridge regressions generate positive out-of-sample R^2-values around 10% (industrials), 8% (metals) and 4-5% (broad index), and small, positive R^2-values for livestock returns.

In contrast, negative R^2-values are obtained for fats-oils, foods, and textiles. Results are very similar for the subset regressions that include a suitable number of predictor variables. At the quarterly frequency, similar results are obtained, although R^2-values tend to be somewhat higher than at the monthly horizon for industrials, metals and the broad commodity index. Performance is further boosted at the annual frequency, where out-of-sample R^2-values in the range 20-35% is seen for the broad commodity index and some of the disaggregate indexes. Notice the contrast to the negative out-of-sample R^2-values for the annual stock return predictions.

The simple equal-weighted average of all possible univariate forecasts is shown as the first line ($k = 1$) under the subset regressions. Rapach, Strauss, and Zhou (2010) found that this method provided good out-of-sample forecasts for stock returns. At the monthly horizon the out-of-sample R^2 for metals and industrials is around 3% under this approach. This rises to around 4% for industrials at the quarterly horizon and grows further to 10% for industrials and 6% for metals at the annual horizon. In fact this strategy is dominated by combining forecasts from models with many more predictor variables. Including on the order of 5-8 predictor variables can in many cases double or triple the value of the out-of-sample R^2 compared with the equal-weighted combination of univariate forecasts. This is related to the fact that the best models include relatively many predictors.

5 Forecasting commodity price volatility, increases, and decreases

Our analysis has so far focused on predictability in the mean of commodity returns. However, it is clearly of interest to explore whether the volatility of commodity returns is predictable through time and to what extent such predictability might vary with the state of the economy. While we are unaware of studies that have addressed this question for commodity prices, a large literature has found that stock market volatility follows a pronounced counter-cyclical pattern (Schwert (1989)). Interestingly, there is relatively weak evidence that macroeconomic state variables contain information useful for predicting stock market volatility. Engle, Ghysels and Sohn (2007) find some evidence that inflation volatility helps predict the volatility of stock returns. However, the volatility of interest rate spreads and growth in industrial production, GDP or the monetary base fail to consistently predict future volatility, with evidence being particularly weak in the post-WWII sample. This is consistent with findings in Paye (2010) and Ghysels, Santa-Clara and Valkanov (2006).

Figure 7 shows a plot of the logarithm of the realized commodity volatility series constructed using equation (3). The series displays low frequency movements, trending downwards from 1947 until 1963, before increasing up to the late seventies, slowly drifting down until the early nineties, and then trending up until the end of the sample.

Following Paye (2010) and others, we model the logarithm of the realized commodity variance (i.e., the square of the realized commodity volatility measure in (3)) as the basis for our analysis. Realized commodity variance is highly skewed and fat-tailed, whereas the logged value is much closer to normality. This makes inference easier and dampens the impact of outliers. Unsurprisingly, data analysis confirmed that commodity volatility (or its logged value) is highly persistent, so we include a first-order autoregressive component in our models. Specifically, we explore forecasting models of the form

$$\log(cvol^2_{t+1}) = \beta_0 + \beta_1 \log(cvol^2_t) + \beta_2 x_t + u_{t+1}. \qquad (12)$$

Table 8 presents empirical results from estimating (12). The estimate of β_1 in the univariate regression is close to 0.8 at all three horizons and highly significant, which is in line with work on stock return volatility. β_2 tracks the predictive content of the state variables after controlling for serial correlation in commodity volatility. The coefficients of the dividend-price ratio and term spread are significant at both the monthly and quarterly frequencies, while the inflation rate and growth in money supply are significant at the monthly frequency and the investment-capital ratio is significant at the quarterly frequency.

Turning to the out-of-sample predictive performance, we report the incremental change in the out-of-sample R^2-value, relative to that from an AR(1) model obtained by setting $\beta_2 = 0$ in equation (12). The evidence is very weak when it comes to establishing that the economic covariates improves upon the predictive power of the AR(1) model. In fact, only the term spread variable at the monthly horizon and the default return spread at the annual horizon seems able to marginally improve on the AR(1) model. This is consistent with earlier findings for stock market volatility such as those reported by Paye (2010).

Although few, if any, state variables appear capable of improving upon the out-of-sample predictive accuracy of the AR(1) model for (log-) commodity volatility, the story is quite different when it comes to separately assessing the predictive performance in expansions versus recessions, as judged by the NBER recession indicator. The last two columns in Table 8 show that during economic recessions several variables, most notably the macroeconomic variables (growth in industrial production, money supply growth, and changes in the unemployment rate), produce better out-of-sample forecasts of monthly commodity market volatility when added to the AR(1) model. The evidence is weaker at the quarterly and annual horizons, although at the quarterly horizon the inflation rate produces notably better out-of-sample forecasts of commodity market volatility in recessions when added to the AR(1) model.

5.1 Predictability of commodity price increases and decreases

Hamilton (2003, 2011) suggests that large increases in oil prices can have a particularly negative effect on economic growth. Specifically, he proposes using

$$\max(0, p_t - \max(p_{t-1}, ..., p_{t-12}))$$

where p_t is the oil price, as a predictor of economic growth. Given the interest in predicting increases in oil prices, we next explore whether increases in commodity prices more broadly defined can be predicted. We initially simplify the analysis and consider monthly price increases, defined as $\max(0, r_{t+1:t+h})$.

First note that if $X \sim N(\mu, \sigma^2)$, from the moments of a truncated normal distribution, we have

$$\begin{aligned} E[\max(0,X)] &= E[\max(0,X)|X \geq 0]p(X \geq 0) \\ &= E[X|X \geq 0]p(X \geq 0) \\ &= E[X|X \geq 0]\left(1 - \Phi(\frac{-\mu}{\sigma})\right) \\ &= \left(\mu + \frac{\sigma\phi(-\mu/\sigma)}{1 - \Phi(-\mu/\sigma)}\right)\left(1 - \Phi(\frac{-\mu}{\sigma})\right) \\ &= \mu\left(1 - \Phi(\frac{-\mu}{\sigma})\right) + \sigma\phi(\frac{-\mu}{\sigma}). \end{aligned} \tag{13}$$

Hence the expected value of $\max(0, X)$ depends on both the mean and the volatility of commodity returns. This suggests including both the lagged return and the lagged volatility in our benchmark model and then add the individual predictor variables:

$$\max(0, r_{t+1:t+h}) = \beta_{0h} + \beta_{1h} r_{t-h+1:t} + \\ \beta_{2h} \sigma_{t-h+1:t} + \beta_{3h} x_t + \varepsilon_{t+1:t+h}.$$

Table 9 reports estimates from this regression applied to the monthly (panel A), quarterly (panel B), and annual (panel C) data. We show both full-sample slope estimates (based on the period 1947-2010) as well as out-of-sample R^2-values computed for the 20-year period 1991-2010. First consider the monthly coefficient estimates. Unsurprisingly, given (13), the lagged volatility is highly significant across all commodity indexes, as are lagged commodity returns. In addition, the inflation rate is positive and significant for foods and livestock, whereas money supply growth and industrial production are significant for three of the commodity indexes, including industrials and metals. For most other cases, the individual predictors are insignificant. The out-of-sample R^2 estimates show similar results with money supply growth notably continuing to help improve predictive accuracy for four of the commodity indexes.

At the quarterly frequency the lagged volatility continues to be highly significant in-sample, while the results for the AR(1) coefficient are somewhat weaker. Money supply growth and inflation continue to be significant, however, for raw industrials and metals. These results carry over to the out-of-sample R^2-values where money supply growth now improve the predictive accuracy by more than 2% for four of the commodity indexes.

Figure 8 plots the quarterly out-of-sample forecasts of $\max(0, r_{t+1})$ for the raw industrials and metals indexes using the prediction model that includes the lagged return, lagged volatility, and money supply growth as predictors. Clearly the forecasts are far from perfect, but they increased notably during the rebound in commodity prices that began in March 2009.

Finally, at the annual frequency, we find that both the coefficient estimates and out-of-sample R^2-values are significant for a broader range of the individual predictor variables (notably inflation, the investment-capital ratio and the change in the unemployment rate) although, in line with the earlier results, both lagged commodity volatility and lagged returns play far less of a role compared with the monthly and quarterly results.[9]

We also explore whether $\min(0, r_{t+1})$ is separately predictable by means of the same list of economic state variables. At the monthly frequency, the strongest evidence comes from growth in industrial production which appears capable of significantly increasing the out-of-sample R^2-value when added to the autoregressive and volatility terms for raw

[9] We also considered predictability of the variable in Hamilton (2003), $\max(0, p_t - \max(p_{t-1}, ..., p_{t-12}))$. For four of the six commodity indexes we found that the inflation rate increased the out-of-sample R^2-value by about 1-2% when added to a lag and the conditional volatility. For the two remaining indexes, industrials and metals, growth in industrial production significantly increased the out-of-sample R^2-value.

industrials, metals, and textiles. At the quarterly frequency, the results are very strong for the inflation rate which significantly raises the out-of-sample R^2-value for all commodity indexes, in some cases by more than 5%. This strong predictive performance from the inflation rate largely carries over to the annual frequency.

Comparing the results for $\max(0, r_{t+1})$ to those for $\min(0, r_{t+1})$, different state variables appear able to predict increasing versus decreasing commodity prices. Whereas money supply growth, lagged volatility, and the lagged return possess predictive power over increases in commodity prices, inflation and, to some extent, growth in industrial production, are far better predictors of declines in commodity prices.

6 Conclusion

Using spot price data on a sample of commodity indexes over the period 1947-2010, we examine the predictability of commodity spot price changes at the monthly, quarterly, and annual horizons. We establish out-of-sample return predictability by means of variables such as the default return spread, growth in money supply, and the T-bill rate. Some variables, such as the inflation rate, have little or no predictive power at the monthly horizon, but appear to have stronger predictive power over commodity spot price changes at the quarterly and annual horizons. At the annual horizon, a wide set of macroeconomic variables such as the growth in industrial production, money supply growth, and the change in the unemployment rate possess predictive power over returns. In addition, our results suggest that predictability of commodity spot price changes is stronger during economic recessions than during expansions and that different variables help predict commodity price increases versus decreases. This is important in light of research linking commodity price increases to economic recessions.

While a large literature has focused on establishing in-sample return predictability for futures and forward prices, our study is one of the first to empirically examine the behavior of the underlying spot prices in an out-of-sample context. Our results suggest that far from following a random walk, spot prices contain a sizeable predictive component which could prove helpful when pricing futures contracts.

References

[1] Acharya, V.V., L.A. Lochstoer and T. Ramadorai, 2010, Limits to Arbitrage and Hedging: Evidence from Commodity Markets.

[2] Ang, A., and A. Timmermann, 2011, Regime changes and financial markets. CEPR discussion paper 8480.

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

[4] Campbell, J.Y. and S.B. Thompson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Review of Financial Studies 21, 1509-1532.

[5] Clark, T.E., and K.D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291-311.

[6] Deaton, A., and G. Laroque, 1992, On the Behavior of commodity prices. Review of Economic Studies 59, 1-23.

[7] Deaton, A., and G. Laroque, 1996, Competitive Storage and Commodity Price Dynamics. Journal of Political Economy 104, 896-923.

[8] Elliott, G., A. Gargano, and A. Timmermann, 2012, Complete Subset Regressions. Manuscript, UCSD.

[9] Engle, R.F., E. Ghysels and B. Sohn, 2007, On the Economic Sources of Stock Market Volatility. Manuscript, New York University and University of North Carolina.

[10] Ghysels, E., P. Santa-Clara and R. Valkanov, 2006, Predicting Volatility: How to get the Most out of Return Data Sampled at Different Frequencies. Journal of Econometrics 131, 59-95.

[11] Gorton, G., and K.G. Rouwenhorst, 2006, Facts and Fantasies about Commodity Futures. Financial Analysts Journal 62, 47-68.

[12] Goyal, A., and I. Welch, 2008, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455-1508.

[13] Hamilton, James D., 2003, What is an oil shock? Journal of Econometrics 113, 363–398.

[14] Hamilton, J.D., 2011, Nonlinearities and the Macroeconomic Effects of Oil Prices. Macroeconomic Dynamics 15, 364-378.

[15] Henkel, S.J., J.S. Martin, and F. Nardari, 2011, Time-varying short-horizon predictability. Journal of Financial Economics 99, 560-580.

[16] Hong, H., and M. Yogo, 2009, Digging into commodities. Manuscript, Princeton University.

[17] Hong, H., and M. Yogo, 2011, What does futures market interest tell us about the macroeconomy and asset prices? Forthcoming in Journal of Financial Economics.

[18] Inoue, A., and L. Killian, 2008, How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. Journal of the American Statistical Association 103, 511-522.

[19] Paye, B., 2010, Which Variables Forecast Aggregate Stock Market Volatility? Working Paper, Rice University.

[20] Pesaran, M.H. and A. Timmermann, 1995, Predictability of Stock Returns: Robustness and Economic Significance. Journal of Finance 50(4), 1201-1228.

[21] Rapach, D.E., J.K. Strauss, and G. Zhou, 2010, Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. Review of Financial Studies 23, 821-862.

Figure 1: **Commodity prices.** This figure plots monthly values of the Reuters/Jeffries-CRB spot price indexes compiled by the Commodity Research Bureau. Prices are measured in nominal US dollar terms. The indexes are based on 22 individual commodities including raw industrials (burlap, copper scrap, cotton, hides, lead scrap, print cloth, rosin, rubber, steel scrap, tallow, tin, wool tops, and zinc) and foodstuffs (beans, butter, cocoa, corn, cottonseed oil, hogs, lard, steers, sugar, and wheat). The sample period is 1947-2010.

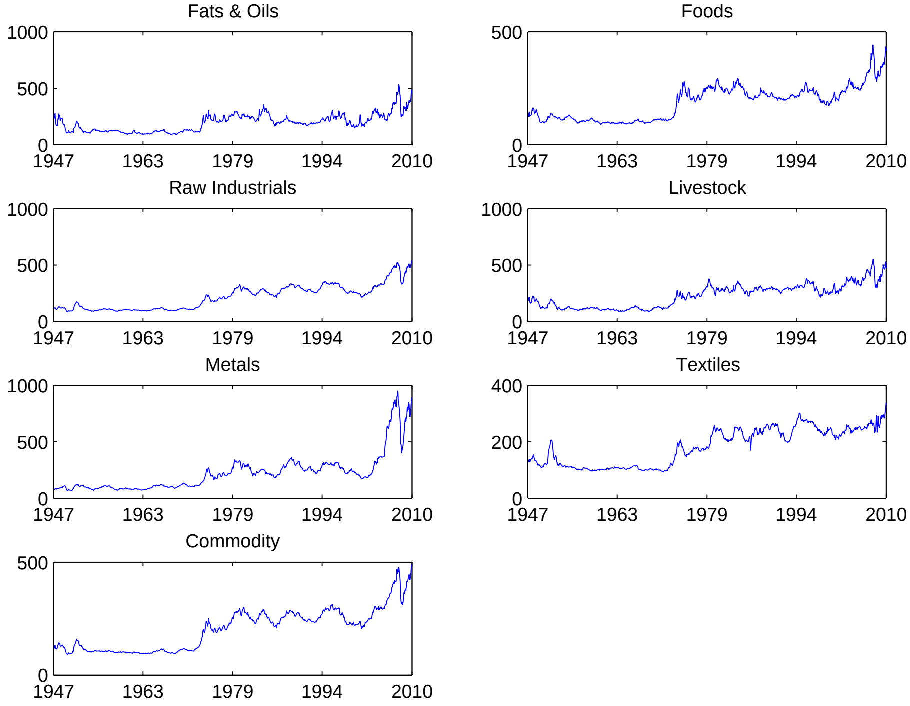


Figure 2: **Commodity returns.** This figure plots monthly returns on the Reuters/Jeffries-CRB spot price indexes compiled by the Commodity Research Bureau. Prices are measured in nominal US dollar terms.

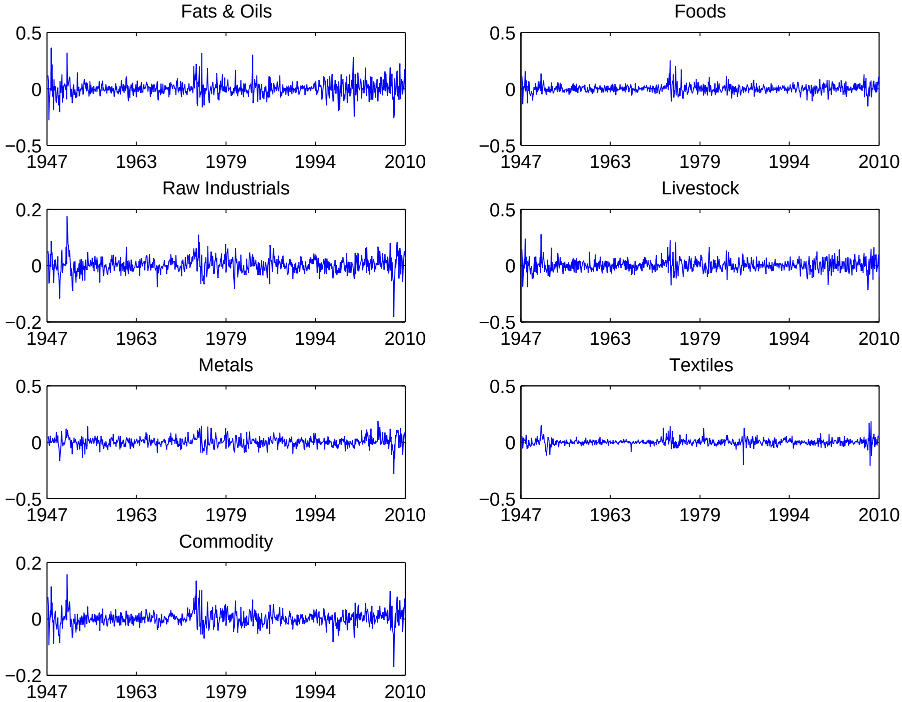


Figure 3: **Cumulated sum of squared forecast error differences.** The figures plot the sum of squared forecast error differences between the benchmark constant mean model and a prediction model that includes a constant and the lagged money supply as the predictor variable. Positive and rising values suggest that the time-varying predictor model outperforms the constant benchmark, while negative and declining values suggest the opposite. The top row uses monthly returns; the middle row uses quarterly returns, while the bottom row uses annual returns, all over the period 1991-2010. All forecasts are generated recursively, using an expanding window of data going back to 1947.

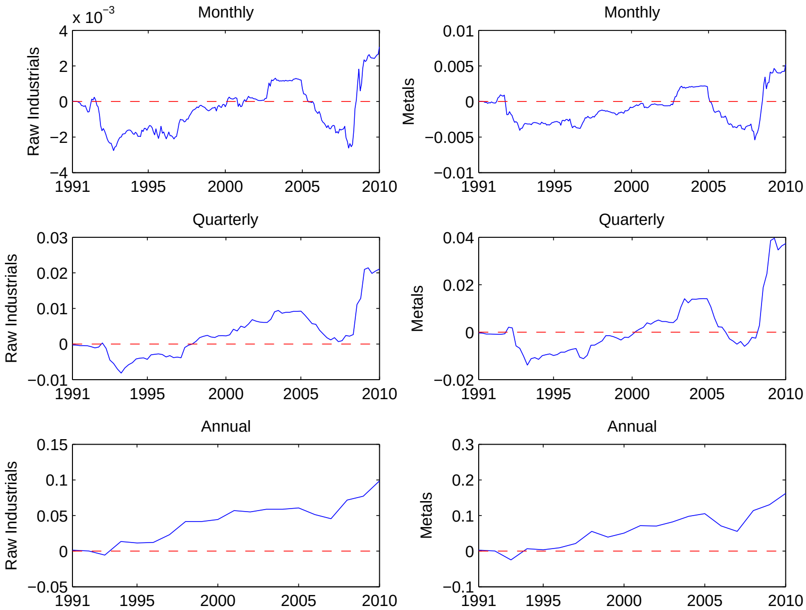


Figure 4: **Variable selection plots: Raw industrials index.** The plots mark which variables are selected at a given point in time by the Akaike (AIC) or Bayes (BIC) information criterion using asterisks to indicate inclusion. The dependent variable is the return on the raw industrials commodity price index, while the predictor variables are selected from the list indicated on the vertical axis. Estimation and variable selection is conducted recursively over the out-of-sample period 1991-2010. The top row uses monthly returns; the middle row uses quarterly returns, while the bottom row uses annual returns, all over the period 1991-2010.

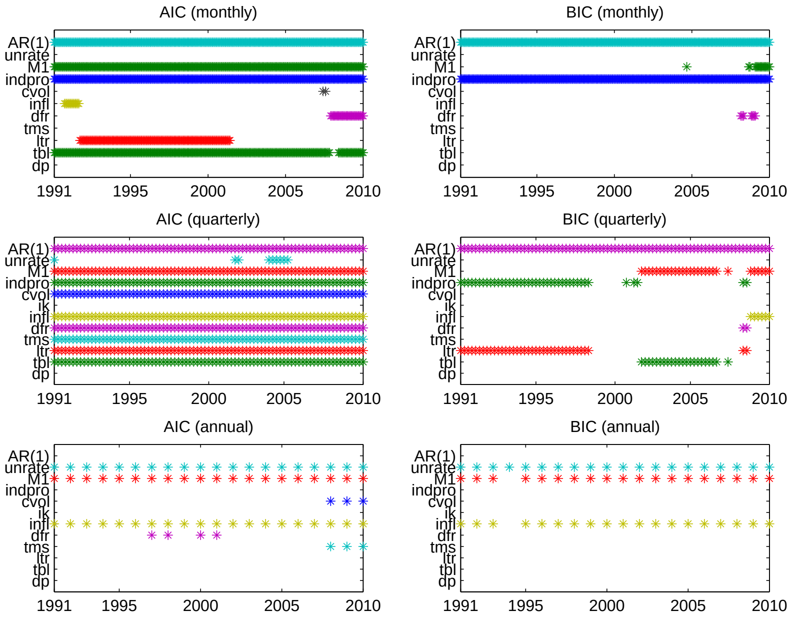


Figure 5: **Variable selection plots: Metals index.** The plots mark which variables are selected at a given point in time by the Akaike (AIC) or Bayes (BIC) information criterion using asterisks to indicate inclusion. The dependent variable is the return on the metals commodity price index, while the predictor variables are selected from the list indicated on the vertical axis. Estimation and variable selection is conducted recursively over the out-of-sample period 1991-2010. The top row uses monthly returns; the middle row uses quarterly returns, while the bottom row uses annual returns, all over the period 1991-2010.

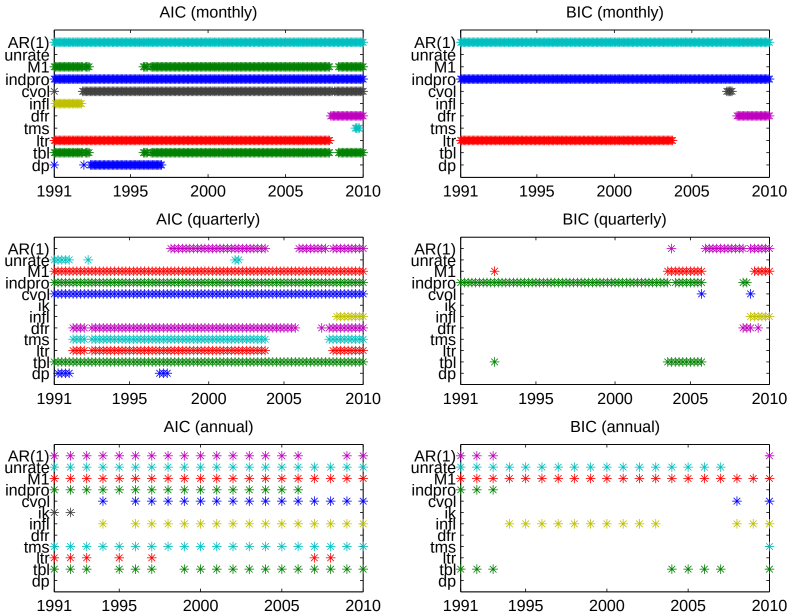
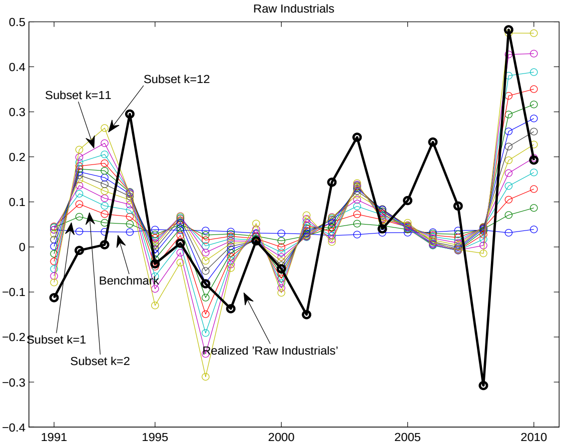


Figure 6: **Forecasts from complete subset regressions.** The figure plots the annual forecasts for the complete subset regressions that combine forecasts from all possible models with k=1, k=2,..., k= 12 predictor variables. The thick black line tracks the actual (realized) return on the corresponding commodity spot price index.



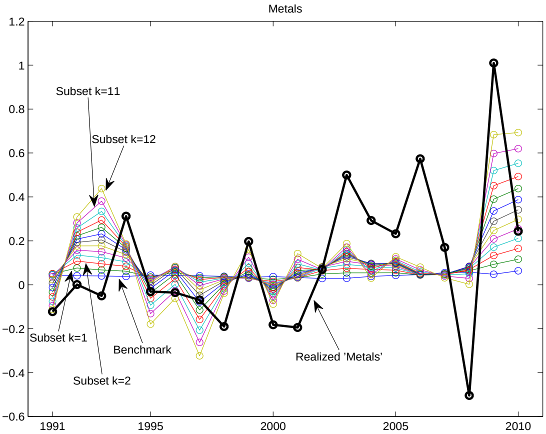


Figure 7: **Commodity variance.** This figure plots monthly values of the logarithm of the realized commodity price variance computed as the sum of squared daily returns of the Dow Jones-AIG Commodity Index over the month. The sample period is 1947-2010.

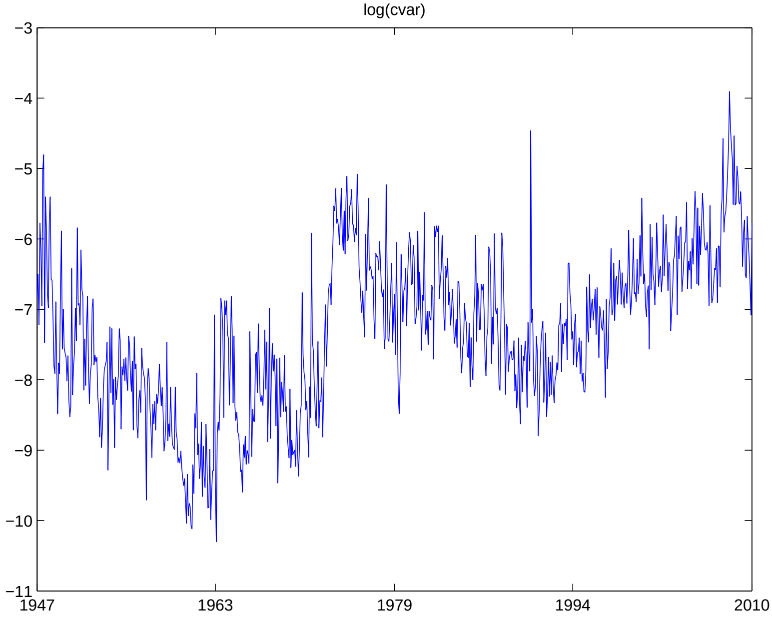
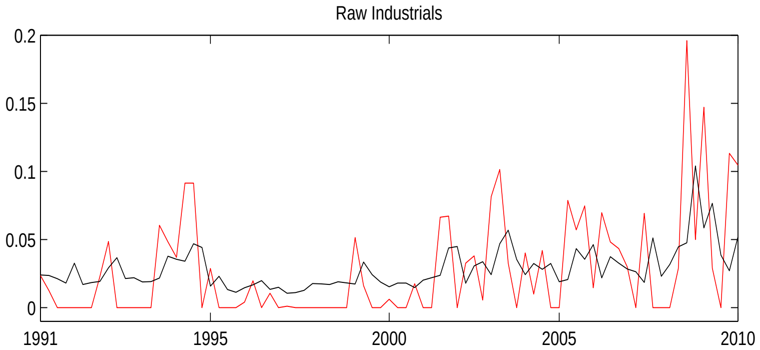


Figure 8: **Forecasts and actual values of max(ret,0) on raw industrials and metals spot price indexes.** The plots show the actual and the predicted value of the max between zero and the returns on the raw industrials and metals Commodity Research Bureau price indexes at quarterly frequency. Forecasts are generated out-of-sample and use the lagged growth in the money supply, lagged volatility and AR(1) as the predictor variable. The out-of-sample period is 1991-2010.



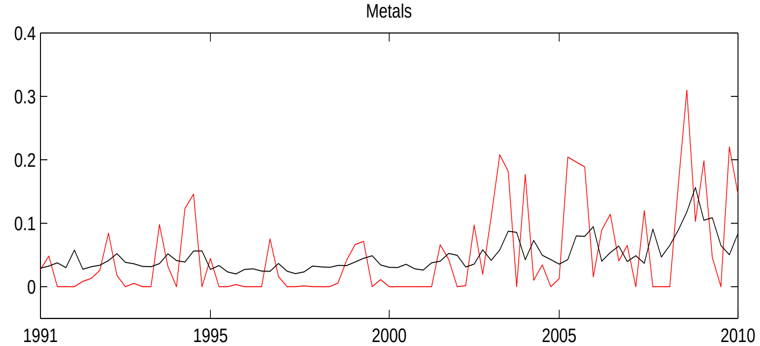


Table 1: **Summary statistics for commodity returns.** This table reports mean, standard deviation, coefficient of skew, coefficient of kurtosis, and the first-order autocorrelation (AR(1)) for commodity returns at the monthly (Panel A), quarterly (Panel B), and annual (Panel C) horizons over the sample period 1947-2010. Commodity prices use the Reuters/Jeffries CRB Commodity Research Bureau spot price indexes and are measured at the end of the month. The last two columns show the comparable values for stocks (tracked by the value-weighted CRSP index) and 10-year T-bonds. Panel D shows correlations between monthly return series above the diagonal and correlations between annual returns below the diagonal.

Panel A: Monthly

	Fats & Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
mean (%)	0.310	0.236	0.254	0.277	0.431	0.179	0.226	0.975	0.484
std (%)	6.610	3.778	2.840	5.289	4.329	3.161	2.669	4.208	2.084
skew	0.552	0.759	0.044	0.269	-0.186	0.280	0.267	-0.411	0.509
kurt	7.324	7.894	7.716	5.671	6.615	12.027	8.412	4.659	5.048
AR(1)	0.089	0.100	0.364	0.098	0.299	0.129	0.280	0.039	0.073

Panel B: Quarterly

	Fats & Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
mean (%)	0.841	0.643	0.813	0.754	1.432	0.524	0.676	2.979	1.466
std (%)	11.200	6.464	6.459	8.987	9.466	5.892	5.476	7.804	3.972
skew	0.268	0.255	0.806	0.160	0.030	1.229	0.241	-0.574	0.934
kurt	5.041	4.775	9.859	4.636	4.736	11.482	6.676	4.051	4.414
AR(1)	0.034	0.088	0.299	0.060	0.220	0.157	0.255	0.102	0.019

Panel C: Annual

	Fats & Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
mean (%)	3.148	2.559	3.936	2.938	6.928	2.284	3.006	12.218	5.964
std (%)	22.439	14.637	18.861	17.862	26.463	15.193	14.977	17.553	8.942
skew	1.395	1.192	1.449	0.548	1.017	1.401	1.503	-0.346	0.972
kurt	6.448	6.536	6.530	3.315	4.830	6.621	6.622	2.948	4.378
AR(1)	-0.076	0.136	-0.128	-0.120	-0.128	-0.075	-0.008	-0.049	-0.095

Panel D: Correlation matrix

	Fats & Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
Fats & Oils	-	0.758	0.507	0.783	0.195	0.290	0.751	0.013	-0.063
Foods	0.845	-	0.397	0.705	0.225	0.278	0.812	0.053	-0.062
Raw Industrials	0.765	0.610	-	0.553	0.781	0.555	0.842	0.130	-0.205
Livestock	0.768	0.665	0.783	-	0.238	0.258	0.751	0.046	-0.112
Metals	0.640	0.536	0.899	0.635	-	0.203	0.606	0.123	-0.162
Textiles	0.685	0.573	0.844	0.692	0.593	-	0.500	0.053	-0.150
Commodity	0.871	0.821	0.952	0.819	0.851	0.826	-	0.104	-0.155
Stock	-0.147	-0.274	0.169	-0.043	0.189	0.085	0.017	-	0.120
Bond	-0.176	-0.116	-0.364	-0.261	-0.466	-0.173	-0.309	-0.015	-

Table 2: **Univariate regression coefficient estimates.** This table reports slope coefficients estimated by OLS using commodity returns as the dependent variable and a constant and the (single) variable listed in the row as predictor. All regressions use non-overlapping returns data over the period 1947-2010. The predictor variables are the dividend-price ratio (dp), the 3-month T-bill rate (tbl), the long term return (ltr), the term spread (tms), the default return spread (dfr), inflation (infl), the investment-capital ratio (ik), commodity price volatility (cvol), growth in industrial production (ΔIND), money supply growth ($\Delta M1$), the change in the unemployment rate (ΔUN), and the one-period lagged return (AR(1)). P-values are computed by bootstrap generating data under the null $r_{t+1} = \alpha + \epsilon_{t+1}$. Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Panel A: Monthly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.007	-0.003	-0.003	-0.005	-0.006*	-0.001	-0.003	0.009**	0.000
tbl	-0.014	-0.014	-0.062*	-0.034	-0.116**	-0.025	-0.041	-0.033	0.096***
ltr	-0.034	0.008	-0.119***	-0.069	-0.226***	-0.040	-0.071**	0.153***	0.060**
tms	0.111	0.013	0.167**	0.194	0.209*	0.078	0.102	0.140	0.096*
dfr	0.134	0.066	0.325***	0.435***	0.559***	0.204**	0.220***	0.109	-0.003
infl	1.065*	0.458	0.522**	0.551	0.517	0.012	0.493**	-0.656*	-0.262
cvol	0.020	0.023	0.056	-0.044	0.190**	0.108	0.029	-0.119	0.050
ΔIN	0.267	0.204	0.580***	0.267	0.756***	0.340***	0.421***	0.079	-0.127*
$\Delta M1$	0.054	0.034	0.094***	0.078*	0.108***	0.091***	0.067***	-0.036	0.011
ΔUN	0.009	0.011	-0.073***	-0.036	-0.114***	-0.015	-0.041*	0.025	0.032*
AR(1)	0.088**	0.099***	0.363***	0.097***	0.299***	0.129***	0.278***	0.039	0.072**

Panel B: Quarterly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.017	-0.011	-0.007	-0.012	-0.016	-0.002	-0.009	0.030**	0.000
tbl	-0.178	-0.062	-0.245*	-0.137	-0.408**	-0.095	-0.166	-0.080	0.319***
ltr	-0.078	0.057	0.004	-0.021	-0.033	0.115	0.021	0.137	0.010
tms	0.367	0.039	0.479*	0.424	0.588	0.247	0.301	0.447	0.250
dfr	0.333	0.220	0.476**	0.484*	0.783***	0.033	0.373**	0.542**	-0.072
infl	-1.744**	-0.557	-1.222***	-1.386**	-1.452**	-0.600	-0.949***	-0.479	0.335
ik	-1.490	0.136	-2.525**	-1.192	-3.616**	-1.295	-1.375	-3.089**	0.725
ΔIN	0.538	0.332*	0.645***	0.373	0.665**	0.381**	0.512***	-0.185	-0.075
$\Delta M1$	0.185	0.141	0.326***	0.284**	0.417***	0.256***	0.246***	-0.107	0.032
ΔUN	-0.105	-0.073	-0.102*	-0.053	-0.114	-0.065	-0.089*	0.117*	0.024
AR(1)	0.033	0.087	0.297***	0.058	0.220***	0.157**	0.252***	0.102	0.018

Panel C: Annual

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.046	-0.043	0.001	-0.029	-0.046	0.009	-0.018	0.145***	-0.006
tbl	-1.042	-0.651	-1.524*	-0.921	-2.696**	-0.657	-1.095*	-0.253	1.159***
ltr	0.169	-0.063	0.431*	0.305	0.651**	0.228	0.220	0.205	-0.088
tms	1.690	1.226	3.091*	2.278	4.650**	1.830	2.292*	1.086	1.257
dfr	0.387	0.514	-0.317	0.109	-0.732	-0.099	0.081	-0.358	0.339
infl	-1.790*	-0.863	-1.864**	-1.408*	-2.672**	-1.022	-1.384**	-0.103	0.719*
ik	-13.388*	-3.163	-15.009**	-11.075*	-19.331**	-7.836	-9.646*	-10.079	3.203
cvol	0.638	0.447	0.737	0.616	1.655***	0.251	0.575	-0.076	0.020
ΔIN	-1.019*	-0.062	-1.363***	-1.022**	-1.330**	-0.982***	-0.791**	-0.367	-0.140
$\Delta M1$	1.028	0.926*	1.211*	1.055*	1.682*	0.867*	1.091**	-0.514	0.288
ΔUN	0.287**	0.035	0.392***	0.272**	0.384**	0.269***	0.232**	0.125	-0.004
AR(1)	-0.075	0.135	-0.128	-0.119	-0.128	-0.073	-0.008	-0.049	-0.094

Table 3: **Out-of-sample R^2 values for the univariate prediction models.** This table reports out-of-sample R^2-values (in percent) for univariate return prediction models that include a constant and the predictor variable listed in each row. The forecast evaluation period is 1991-2010. All forecasts are updated recursively, using an expanding estimation window. Returns are based on the Reuters/Jeffries CRB spot price indexes. Statistical significance is measured by bootstrap generating data under the null $r_{t+1} = \alpha + \epsilon_{t+1}$. Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Panel A: Monthly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.388	-0.308	-2.747	-0.830	-1.066	-1.279	-1.663	-1.555	-1.463
tbl	-0.151	-0.099	0.874**	-0.022	0.997**	-0.047	0.415*	-0.712	-0.840
ltr	-1.074	-0.783	-0.972	-0.751	-0.119	-1.361	-2.050	-0.588	0.323
tms	-0.082	-0.234	0.607**	0.070	0.297	-0.083	0.224	-2.289	-0.076
dfr	-0.548	-0.799	4.325***	2.094***	4.500***	0.333*	1.536***	-0.648	-0.599
infl	0.609**	0.438*	-0.172	0.259	-0.587	-0.942	0.486*	-1.704	0.464*
cvol	-0.673	-0.632	-0.739	-0.910	-0.373	0.217	-0.933	0.110	-1.553
ΔIN	-0.193	0.314*	3.060***	-0.321	1.996***	0.142	2.122***	-0.494	-2.072
$\Delta M1$	-0.551	-0.664	1.593***	-0.324	0.931**	0.793**	-0.140	-0.184	-1.141
ΔUN	-0.089	-0.047	0.062	-0.134	-0.162	-0.133	-0.256	-0.858	-0.022
AR(1)	0.853**	0.313	9.139***	1.018**	7.189***	-7.282	5.279***	0.040	0.128

Panel B: Quarterly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-1.573	-1.132	-4.562	-2.701	-2.141	-6.014	-3.833	-3.541	-3.768
tbl	0.212	-0.195	2.515**	0.212	2.510**	0.876*	1.574*	-1.368	-1.224
ltr	-0.334	-0.111	-2.529	-1.369	-0.914	-6.988	-1.709	-3.550	-0.377
tms	0.014	-0.766	1.058*	0.173	0.460	0.528	0.535	-5.024	0.098
dfr	-2.776	-1.727	4.077***	-0.580	3.862***	-1.223	1.813**	0.313	-1.586
infl	5.137***	1.464*	7.673***	5.328***	3.821***	1.322*	6.391***	0.223	1.442*
ik	-0.195	-3.188	5.819***	-0.239	3.870***	2.662**	1.781*	-1.923	-1.709
ΔIN	-0.380	0.837	-2.855	-0.594	-2.399	-1.257	-0.644	-2.290	-1.216
$\Delta M1$	-0.864	-1.067	6.853***	0.645	3.941***	5.378***	3.255**	-0.811	-2.759
ΔUN	-0.313	0.629	-1.097	-0.713	-0.440	-0.422	-0.006	-2.217	-1.105
AR(1)	-0.273	-0.082	10.340***	-0.302	5.927***	4.549***	6.026***	-0.296	-0.482

Panel C: Annual

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-9.272	-16.877	-4.761	-11.394	-3.771	-6.876	-9.357	-17.316	-7.524
tbl	4.837*	5.625*	13.460**	6.174*	12.870**	6.611*	13.002**	-3.425	1.097
ltr	-3.341	-2.811	-4.448	-0.418	1.839	-25.565	-10.076	-4.343	-1.346
tms	1.943	3.986	11.943**	7.615**	8.875**	9.130**	12.686**	-6.553	-6.047
dfr	-9.357	-6.847	-18.734	-13.906	-17.182	-17.513	-18.585	-19.775	5.167*
infl	6.227*	4.518*	9.746**	8.099**	8.704**	4.511*	9.551**	-0.436	2.545
ik	12.448**	-1.014	17.682***	9.849**	10.003**	9.811**	16.468***	-0.468	-6.388
cvol	2.758	6.873*	5.054*	6.016*	10.272**	-3.245	8.355**	-3.850	-11.143
ΔIN	16.136***	-15.030	18.684***	20.216***	8.286**	12.599**	19.132***	-5.594	-1.423
$\Delta M1$	6.194*	13.426**	15.490**	11.818**	7.347**	22.638***	20.624***	-4.546	-15.396
ΔUN	14.810***	-3.003	14.249**	17.173***	6.953*	10.287**	15.921***	-2.193	-0.730
AR(1)	-4.480	1.442	-2.620	0.117	-7.016	-1.632	-2.758	-4.833	-0.119

Table 4: **Improvement in predictive accuracy relative to the first-order autoregressive return model.** This table reports the marginal improvement in the out-of-sample R^2-value (in percent) of a bivariate return prediction model that includes a constant, the lagged return, and the predictor variable listed in each row, measured relative to the R^2-value of a model that only includes a constant and the lagged commodity return. For example, an R^2-value of 1% means that adding a particular predictor improves on the R^2-value of the pure autoregressive model. The forecast evaluation period is 1991-2010. All forecasts are updated recursively, using an expanding estimation window. Returns are based on the Reuters/Jeffries CRB spot price indexes. Statistical significance is measured by means of the Clark-West (2006) test for out-of-sample predictive accuracy, using the first-order autoregressive model as the benchmark. Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Panel A: Monthly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.466	-0.352	-1.314	-0.837	-0.638	-0.740	-1.100	-1.353	-1.337
tbl	-0.147	-0.110	0.477	-0.042	0.570	-0.075	0.217	-0.616	-0.619
ltr	-0.891	-0.427	-0.517	-0.567	-0.569	-0.579	-0.910	-0.739	0.014
tms	-0.104	-0.237	0.184	0.026	0.013	-0.140	0.034	-2.051	0.087
dfr	-0.630	-0.811	1.018	1.705*	2.415*	-0.376	0.156	-0.706	-1.027
infl	0.331	0.167	-0.286	-0.053	-0.733	-0.321	-0.074	-1.428	0.290
cvol	-0.640	-0.574	-0.569	-0.904	0.018	0.170	-0.842	0.054	-1.524
ΔIND	-0.201	0.309	2.030	-0.301	1.403	0.806	1.591	-0.425	-1.590
$\Delta M1$	-0.516	-0.609	0.911*	-0.297	0.520	0.905	-0.090	-0.168	-0.980
ΔUN	-0.072	-0.007	-0.086	-0.155	-0.135	-0.188	-0.192	-0.919	0.010

Panel B: Quarterly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-1.711	-1.301	-3.058	-2.687	-2.037	-4.833	-3.069	-4.225	-3.709
tbl	0.213	-0.180	2.281	0.193	2.055	1.158	1.365	-1.101	-1.081
ltr	-0.487	0.054	-4.413	-1.671	-1.297	-8.163	-2.375	-2.698	-1.076
tms	-0.026	-0.738	0.118	0.023	0.033	0.526	0.046	-4.235	0.176
dfr	-2.852	-1.495	-0.634	-0.911	1.221	-1.562	0.123	-0.626	-1.917
infl	5.174**	1.411**	8.417*	5.287**	4.098*	1.512	6.439**	0.154	1.411
ik	-0.218	-2.753	4.564**	-0.217	3.245**	2.352*	1.876**	-1.423	-1.680
cvol	-1.759	-1.494	-1.061	-1.576	-0.286	-0.326	-1.401	-0.696	-4.408
ΔIND	-0.550	0.669	-3.861	-0.589	-3.217	-1.833	-1.580	-1.934	-1.152
$\Delta M1$	-0.926	-0.848	5.263**	0.668	3.054*	4.350*	3.059**	-0.790	-2.668
ΔUN	-0.399	0.509	-1.174	-0.661	-0.810	-0.654	-0.442	-1.712	-1.023

Panel C: Annual

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-9.167	-17.053	-5.464	-12.601	-3.384	-7.490	-9.307	-15.806	-7.379
tbl	4.575	6.143	14.143**	5.946*	14.735**	5.901*	13.479**	-4.171	5.262*
ltr	-2.462	-5.064	-4.517	-2.119	3.523	-24.930	-8.431	-5.632	-5.080
tms	1.832	4.214	13.405**	9.460*	10.884**	8.884*	13.158**	-7.565	-0.622
dfr	-7.000	-9.572	-16.830	-11.121	-15.192	-16.430	-17.004	-16.652	2.976
infl	7.421*	6.833	9.826**	7.272**	9.639**	4.531	11.242**	-1.297	1.390
ik	12.478**	2.003	19.813***	9.396*	11.856***	10.271*	16.792***	-5.628	-7.120
cvol	5.197	0.518	10.128*	8.179	17.691*	-1.775	9.006	-3.052	-9.772
ΔIND	14.670*	-10.211	20.325**	19.075*	10.379**	13.506*	20.884**	-10.121	0.778
$\Delta M1$	6.236	12.148*	20.838***	17.099**	9.826**	26.474***	21.837***	-6.284	-13.542
ΔUN	15.004*	-0.823	16.354*	17.648*	9.369*	10.863	16.890*	-6.774	-0.747

Table 5: **Out-of-sample R^2 values for the univariate prediction models (levels).** This table reports out-of-sample R^2-values (in percent) for bivariate prediction models that include a constant, the lagged price and the predictor variable listed in each row: $P_{t+1} = \alpha + \gamma P_t + \beta x_t + \epsilon_{t+1}$. The forecast evaluation period is 1991-2010. All forecasts are updated recursively, using an expanding estimation window. Statistical significance is measured by bootstrap generating data under the null $P_{t+1} = \alpha + \beta P_t + \epsilon_{t+1}$. Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Panel A: Monthly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity
dp	0.333	0.222	-0.650	0.204	-0.088	-0.463	-0.320
tbl	-1.438	-0.715	0.380*	-0.443	-0.108	-0.416	0.195
ltr	-0.699	-0.491	-0.121	-0.493	0.088	-0.984	-0.966
tms	0.116	-0.091	0.072	0.316	-0.258	0.080	-0.097
dfr	-0.620	-0.879	3.351***	1.906***	1.945***	-0.102	1.378***
infl	0.144	0.002	-0.255	-0.035	-0.415	-0.680	-0.037
cvol	-0.968	-0.596	-0.781	-0.857	-0.630	0.607**	-0.746
ΔIN	-0.032	0.374*	2.057***	-0.151	1.171***	0.175	1.652***
$\Delta M1$	-1.119	-1.003	0.438*	-0.729	0.018	0.151	-0.707
ΔUN	-0.051	-0.038	0.031	-0.108	-0.023	-0.111	-0.117

Panel B: Quarterly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity
dp	0.678	0.541	-1.145	0.206	-0.102	-2.251	-0.645
tbl	-2.253	-2.216	0.442	-1.075	-0.664	-0.237	-0.169
ltr	-0.532	0.038	-1.614	-0.765	-0.725	-4.641	-1.036
tms	0.551	-0.117	0.319	0.839	-0.548	1.618*	0.100
dfr	-3.553	-2.449	3.093**	-1.796	2.705**	-1.117	0.400
infl	3.386**	1.381*	4.862***	4.244***	2.947**	0.947*	4.690***
ik	0.078	-1.077	4.214***	0.613	1.778*	2.261*	2.254**
cvol	-1.000	-1.157	-1.422	-0.989	-1.377	0.240	-1.352
ΔIN	0.309	0.955*	-0.028	-0.094	-0.299	-0.245	0.799
$\Delta M1$	-1.899	-2.316	2.436**	-1.003	0.623	0.510	0.052
ΔUN	-0.150	0.322	-0.456	-0.510	-0.262	-0.305	-0.023

Panel C: Annual

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity
dp	-0.740	-0.017	-1.825	-3.746	-0.368	-3.934	-1.955
tbl	-3.618	6.749*	4.891	1.902	-3.072	-0.389	7.223*
ltr	-8.116	-5.050	-5.427	-5.463	-2.883	-23.458	-9.066
tms	4.402*	5.092*	10.213**	10.194**	3.758	18.142***	9.515**
dfr	-11.912	-7.130	-21.366	-16.469	-19.566	-21.049	-18.642
infl	4.155	8.094*	4.731	3.891	0.961	4.900*	6.514*
ik	8.404**	-0.257	10.454**	8.655**	3.383	7.910**	10.402**
cvol	25.992***	26.095***	21.060***	22.481***	18.958**	13.823**	23.896***
ΔIN	15.894***	-3.733	12.979**	17.355***	4.268*	20.799***	11.315**
$\Delta M1$	-1.571	-5.394	9.801**	7.699**	2.744	17.091***	7.541*
ΔUN	10.307**	-2.428	7.783**	11.781**	1.040	12.810**	7.577**

Table 6: **Forecasting performance in recessions versus expansions.** This table compares the out-of-sample R^2 values of monthly and quarterly return prediction models in expansions versus recessions, defined by the NBER recession indicator. The forecast evaluation period is 1991-2010. All forecasts are updated recursively, using an expanding estimation window. Returns are based on the Reuters/Jeffries CRB spot price indexes. Statistical significance measures whether the average squared forecast error of a given model, measured relative to the constant return benchmark, is significantly different in recessions versus expansions. Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Panel A: Monthly

A.1 Expansions

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.372	-0.256	-3.986	-0.999	-1.083	-2.762	-2.391	-3.364	-1.738
tbl	-0.161	-0.094	2.037	0.077	1.973	0.235	1.255	-0.472	-0.919
ltr	-0.881	-0.882	-0.753	0.044	-1.547	-0.628	-2.434	-1.005	1.046
tms	0.061	-0.363	2.358	0.640	1.125	0.309	1.390	-2.496	0.010
dfr	0.015	-0.021	1.654	1.163	0.527	-0.461	0.813	-0.191	-0.397
infl	0.513	-0.021	-2.538	0.005	-1.820	-0.899	-1.258	0.209	0.283
cvol	-0.030	0.235	1.666	-1.019	3.466	0.850	1.249	-0.631	-1.302
ΔIN	-0.247	-0.013	-0.268	-0.593	-0.456	2.086	-0.417	-0.383	-0.669
$\Delta M1$	-0.515	-0.418	0.987	-0.028	-0.040	1.138	-0.145	0.197	-1.108
ΔUN	-0.043	0.028	0.021	-0.294	0.052	0.230	-0.587	-0.640	-0.142
AR(1)	-0.530	0.038	3.417	-0.837	3.659	-2.611	3.295	-0.562	0.409

A.2 Recessions

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-0.420	-0.403	-0.994	-0.462	-1.038	-0.115	-0.832	3.112**	-0.414
tbl	-0.131	-0.109	-0.771	-0.240	-0.692	-0.268	-0.543	-1.333	-0.539
ltr	-1.453	-0.602	-1.283	-2.489	2.352	-1.936	-1.611	0.488	-2.426
tms	-0.366	-0.000	-1.870	-1.173	-1.137	-0.392	-1.105	-1.755	-0.409
dfr	-1.658	-2.211	8.106**	4.128*	11.379***	0.957	2.361	-1.827	-1.364
infl	0.797	1.273***	3.175***	0.813*	1.547***	-0.975	2.478***	-6.643	1.154
cvol	-1.938	-2.206	-4.144	-0.673	-7.023	-0.278	-3.423	2.027*	-2.509
ΔIN	-0.088	0.909*	7.771**	0.273	6.244**	-1.383	5.021**	-0.780	-7.411
$\Delta M1$	-0.621	-1.112	2.452*	-0.969	2.613**	0.523	-0.134	-1.169	-1.269
ΔUN	-0.180	-0.186	0.121	0.214	-0.535	-0.419	0.122	-1.421	0.435
AR(1)	3.577**	0.811	17.239***	5.073***	13.301***	-10.948	7.544*	1.598***	-0.938

Panel B: Quarterly

B.1 Expansions

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-1.103	-0.616	-6.539	-2.532	-1.615	-6.735	-4.612	-8.107	-4.429
tbl	0.800	-0.067	5.858	0.930	4.713	1.131	4.154	-0.908	-1.773
ltr	-0.821	-0.156	-3.985	-2.025	-0.836	-5.351	-2.621	-2.020	-0.127
tms	0.709	-1.271	4.014	1.448	1.848	0.762	2.697	-5.091	1.016
dfr	-0.459	-0.634	4.569	3.210	1.863	-0.734	2.277	-1.555	-0.007
infl	1.822	0.523	-0.314	1.604	-0.530	-1.805	0.940	-0.428	-0.354
ik	-0.306	-4.638	7.304	-0.306	3.870	2.330	2.044	-5.468	-2.528
cvol	1.440	2.774	6.896	2.089	7.453	0.497	7.090	-0.801	-2.279
ΔIN	-1.454	-0.732	-3.114	-1.656	-1.922	1.007	-3.232	-1.366	-0.723
$\Delta M1$	-1.215	-1.409	4.155	0.169	0.773	3.618	1.190	1.483	-2.192
ΔUN	-0.857	-0.333	-0.132	-1.065	0.872	0.897	-0.487	-1.148	-0.392
AR(1)	-0.316	-0.268	11.092	-0.375	5.271	5.153	5.176	-2.334	-0.053

B.2 Recessions

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
dp	-2.842	-2.353	-2.187	-3.016	-2.938	-0.796	-2.854	6.204*	-0.770
tbl	-1.372	-0.496	-1.501	-1.126	-0.831	-0.972	-1.669	-2.351	1.265
ltr	0.982	-0.006	-0.780	-0.147	-1.033	-18.836	-0.561	-6.814	-1.508
tms	-1.860	0.424*	-2.494	-2.203	-1.643	-1.168	-2.183	-4.879	-4.069
dfr	-9.034	-4.309	3.486	-7.650	6.894	-4.761	1.229	4.304	-8.747
infl	14.089***	3.686***	17.271***	12.271***	10.422***	23.973***	13.244***	1.614**	9.597***
ik	0.103	0.239	4.035*	-0.114	3.871**	5.067	1.451	5.644**	2.003
cvol	-9.654	-11.558	-11.083	-8.513	-9.537	-6.362	-12.316	0.991	-14.303
ΔIN	2.519	4.546***	-2.544	1.385	-3.123	-17.661	2.610	-4.263	-3.454
$\Delta M1$	0.082	-0.257	10.095***	1.531	8.746***	18.120	5.851**	-5.708	-5.336
ΔUN	1.155	2.903**	-2.257	-0.057	-2.431	-9.979	0.597	-4.498	-4.339
AR(1)	-0.156	0.358	9.437	-0.165	6.921	0.181	7.095	4.053**	-2.430

Table 7: **Multivariate out-of-sample prediction results:** This table reports the out-of-sample R^2-value for a range of multivariate model selection and estimation methods. AIC and BIC are the Akaike and Bayes Information Criteria which select the prediction model using penalized likelihood criteria, at each point searching across all possible combinations of predictor variables. Ridge regression includes all predictor variables in the forecasting model but shrinks, trough λ, the least squares coefficient estimate towards zero. Subset regression computes an equal-weighted average of forecasts considering all possible models with k predictor variables included. The set of predictor variables is identical to that listed in Table 2. All estimation and model selection is conducted recursively, using an expanding estimation window and 1991-2010 as the out-of-sample forecast evaluation period. Returns are based on the Reuters/Jeffries CRB spot price indexes. Statistical significance is measured by means of the Clark-West (2006) test for out-of-sample predictive accuracy, using the prevailing mean model, which only includes a constant, as the benchmark. Stars indicate statistical significance: ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Panel A: Monthly

A.1 Model Selection

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
AIC	-1.342	0.313	10.614	0.298	7.478	-9.348	5.403	-5.799	0.305*
BIC	-0.975	-0.618	9.417	-2.265	8.679	-11.329	6.403	-6.253	0.698

A.2 Ridge Regression

λ	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
0.5	-2.476	-1.472	10.449***	0.244	8.430***	-8.107	4.430***	-5.249	0.242
1	-2.471	-1.469	10.456***	0.247	8.435***	-8.095	4.436***	-5.240	0.245
2	-2.461	-1.461	10.469***	0.252	8.443***	-8.072	4.448***	-5.221	0.250
3	-2.452	-1.455	10.481***	0.256	8.451***	-8.049	4.459***	-5.202	0.255
4	-2.442	-1.448	10.493***	0.261	8.460***	-8.026	4.471***	-5.183	0.260
5	-2.433	-1.441	10.505***	0.266	8.468***	-8.003	4.482***	-5.164	0.264
10	-2.387	-1.408	10.563***	0.288	8.507***	-7.892	4.536***	-5.073	0.280
20	-2.301	-1.349	10.663***	0.329	8.577***	-7.682	4.633***	-4.901	0.291
50	-2.076	-1.203	10.885***	0.431	8.739***	-7.134	4.854**	-4.449	0.235
100	-1.781	-1.029	11.085***	0.552	8.899***	-6.399	5.082**	-3.861	0.060
150	-1.555	-0.903	11.163***	0.637	8.974***	-5.808	5.210**	-3.412	-0.102
200	-1.376	-0.805	11.170***	0.700	8.993***	-5.316	5.280**	-3.057	-0.230
1000	-0.419	-0.272	9.242***	0.862	7.504***	-2.107	4.605**	-1.153	-0.512

A.3 Subset Regression

k	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
1	-0.033	-0.039	3.336***	0.342	2.658***	-0.234	1.637**	-0.228	-0.095
2	-0.110	-0.102	5.800***	0.570	4.649***	-0.636	2.848**	-0.490	-0.233
3	-0.228	-0.187	7.596***	0.706	6.118***	-1.167	3.722**	-0.788	-0.362
4	-0.384	-0.291	8.882***	0.768	7.179***	-1.801	4.331**	-1.126	-0.447
5	-0.577	-0.412	9.775***	0.773	7.919***	-2.517	4.728**	-1.507	-0.467
6	-0.807	-0.551	10.363***	0.734	8.407***	-3.302	4.955**	-1.934	-0.414
7	-1.073	-0.705	10.711***	0.664	8.693***	-4.148	5.045**	-2.417	-0.292
8	-1.374	-0.876	10.866***	0.572	8.816***	-5.052	5.022**	-2.968	-0.120
9	-1.710	-1.061	10.860***	0.467	8.801***	-6.012	4.902**	-3.607	0.065
10	-2.079	-1.261	10.715***	0.356	8.668***	-7.032	4.700**	-4.359	0.211
11	-2.481	-1.476	10.443***	0.242	8.426***	-8.119	4.424***	-5.259	0.238

Panel B: Quarterly

B.1 Model Selection

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
AIC	-5.140	-11.730	12.763	-5.412	5.974	-5.186	10.330	-14.699	-0.182*
BIC	0.111	0.000	-11.205	-3.519	-10.697	-6.364	-5.151	-12.589	1.923

B.2 Ridge Regression

λ	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
0.5	-8.134	-5.527	10.947***	-6.451	8.772***	-13.050	7.736**	-10.083	-10.078
1	-8.052	-5.450	11.103***	-6.352	8.858***	-12.730	7.866**	-10.013	-9.871
2	-7.894	-5.300	11.402***	-6.160	9.021***	-12.119	8.114**	-9.874	-9.484
3	-7.741	-5.157	11.681***	-5.977	9.174***	-11.541	8.347**	-9.740	-9.130
4	-7.594	-5.021	11.944***	-5.801	9.316***	-10.994	8.565**	-9.609	-8.804
5	-7.451	-4.891	12.191***	-5.633	9.448***	-10.476	8.770**	-9.482	-8.504
10	-6.809	-4.319	13.224***	-4.892	9.993***	-8.248	9.626**	-8.896	-7.304
20	-5.788	-3.466	14.579***	-3.771	10.672***	-5.074	10.741**	-7.918	-5.848
50	-3.892	-2.090	16.071***	-1.872	11.284***	-0.338	11.923**	-5.926	-4.214
100	-2.340	-1.180	16.079***	-0.500	11.034***	2.395	11.847**	-4.084	-3.447
150	-1.533	-0.780	15.460***	0.138	10.513***	3.343	11.293**	-3.028	-3.065
200	-1.043	-0.559	14.745***	0.490	9.975***	3.711*	10.689**	-2.342	-2.778
1000	0.195	-0.014	7.886***	0.901	5.226**	2.593*	5.483**	-0.195	-1.076

B.3 Subset Regression

k	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
1	0.209	-0.060	4.283***	0.527	2.779**	1.351*	2.846**	-0.042	-0.517
2	0.245	-0.170	7.585***	0.841	4.953**	2.327*	5.098**	-0.316	-1.046
3	0.118	-0.329	10.088***	0.963	6.626**	2.957	6.862**	-0.752	-1.546
4	-0.164	-0.533	11.957***	0.914	7.897**	3.269	8.238**	-1.288	-1.992
5	-0.592	-0.781	13.340***	0.714	8.856***	3.277	9.314**	-1.890	-2.374
6	-1.161	-1.075	14.353***	0.377	9.577***	2.968	10.156**	-2.550	-2.704
7	-1.870	-1.426	15.056***	-0.101	10.111***	2.285	10.792**	-3.287	-3.029
8	-2.727	-1.856	15.446***	-0.747	10.472***	1.115	11.197**	-4.148	-3.444
9	-3.752	-2.408	15.436***	-1.611	10.628***	-0.722	11.284**	-5.193	-4.103
10	-4.975	-3.147	14.852***	-2.779	10.492***	-3.493	10.892**	-6.493	-5.239
11	-6.443	-4.169	13.425***	-4.371	9.916***	-7.555	9.776**	-8.123	-7.166
12	-8.217	-5.607	10.785***	-6.553	8.682***	-13.379	7.601**	-10.155	-10.294

Panel C: Annual

C.1 Model Selection

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
AIC	6.565	22.597	22.595	29.393	26.251	6.807	38.418	-57.660	40.517
BIC	0.000	0.000	18.264	11.804	2.111	10.287	30.521	-40.457	40.018

C.2 Ridge Regression

λ	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
0.5	-2.635	17.729*	23.439**	32.078**	30.197**	9.492**	26.831***	-54.765	13.828***
1	1.039	18.977*	26.831**	32.633**	31.651**	11.380**	29.876***	-53.011	18.686***
2	6.121*	20.249*	31.093**	33.060**	33.229**	13.921**	33.526***	-49.949	24.331***
3	9.386*	20.622*	33.543**	33.013**	33.839**	15.460**	35.449***	-47.336	27.087***
4	11.580*	20.539*	35.021**	32.739**	33.950**	16.410**	36.464***	-45.056	28.326***
5	13.093*	20.214*	35.923**	32.352**	33.789**	16.995**	36.957***	-43.036	28.705***
10	16.044*	17.683	36.867**	30.122**	31.723**	17.606**	36.457***	-35.438	26.091***
20	16.084*	13.805	35.009**	26.794**	27.745*	16.598*	33.305**	-26.406	19.150**
50	13.763*	9.516	30.113**	22.155**	21.564*	14.865*	27.688**	-14.684	9.843*
100	11.437*	7.168	25.239**	18.368**	16.841*	13.597*	22.896**	-8.070	5.356
150	9.876*	5.890	21.850**	15.822**	14.056*	12.451*	19.690**	-5.426	3.669
200	8.696*	5.013	19.264**	13.899**	12.113*	11.384*	17.278**	-4.040	2.783
1000	2.958*	1.481	6.585**	4.672**	3.837*	4.381*	5.787**	-0.738	0.548

C.3 Subset Regression

k	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
1	4.884*	2.004	10.618**	7.775**	6.159**	6.535*	9.271**	-1.532	0.645
2	8.238*	3.638	18.005**	13.289**	11.026**	10.327*	16.090**	-3.611	2.040
3	10.451*	4.978	23.126**	17.117**	14.929**	12.464*	21.024**	-6.182	4.289
4	11.885*	6.218	26.866**	19.863**	18.205**	13.911*	24.734**	-9.189	7.469*
5	12.813*	7.614	29.847**	22.060**	21.133**	15.297*	27.783**	-12.608	11.581**
6	13.360*	9.391	32.356**	24.098**	23.882*	16.833*	30.510**	-16.456	16.455**
7	13.468*	11.643	34.347**	26.175**	26.491*	18.349**	32.961***	-20.802	21.609**
8	12.877*	14.244	35.503**	28.283**	28.855*	19.403**	34.880***	-25.777	26.133**
9	11.137	16.790*	35.304**	30.225**	30.721*	19.412**	35.726***	-31.572	28.659***
10	7.648	18.588*	33.057**	31.657**	31.659*	17.756**	34.699***	-38.440	27.458***
11	1.695	18.672*	27.868**	32.130**	31.033**	13.814**	30.716***	-46.691	20.631***
12	-7.550	15.776*	18.477**	31.108**	27.972**	6.947**	22.252***	-56.712	6.265***

Table 8: **Predictability of realized commodity variance:** This table shows results for the logarithm of the realized commodity price variance computed as the sum of squared daily returns over the month (Panel A), the quarter (Panel B) or the year (Panel C). The first column shows full-sample OLS estimates of the slope coefficient, β_2, on the lagged covariates from a model $log(cvol^2_{t+1}) = \beta_0 + \beta_1 * log(cvol^2_t) + \beta_2 * x_t + \epsilon_{t+1}$. The second column shows the out-of-sample R^2 value computed over the sample 1991-2010. For the AR(1) model, the out-of-sample R^2 value is measured relative to a constant volatility benchmark. The third and fourth columns report out-of-sample R^2-values separately for recessions and expansions. Stars indicate statistical significance using the Clark-West (2007) statistic. ***: significant at the 1% level; **: significant at the 5% level; *: significant at the 10% level.

Panel A: Monthly

	β	$OoSR^2$	$OoSR^2_{Expan}$	$OoSR^2_{Recess}$
dp	-0.123***	-1.449	-3.115	1.825*
tbl	0.633	-4.198	-5.203	-2.222
ltr	-0.964	0.125	0.462	-0.536
tms	4.403***	1.332**	0.691	2.592**
dfr	-1.174	-0.450	-0.521	-0.310
infl	10.870*	-5.838	-5.605	-6.297
ΔIN	-2.384	0.277	-0.250	1.314***
$\Delta M1$	1.158**	-3.897	-6.989	2.178**
ΔUN	0.250	-0.034	-0.235	0.362***
AR(1)	0.811***	72.695***		

Panel B: Quarterly

	β	$OoSR^2$	$OoSR^2_{Expan}$	$OoSR^2_{Recess}$
dp	-0.146**	1.411	0.060	2.843*
tbl	0.604	-4.189	-8.602	0.484
ltr	0.264	-0.324	-0.038	-0.626
tms	3.613*	0.540	0.537	0.543
dfr	-2.364	0.704	1.351	0.020
infl	5.106	-4.424	-17.504	9.426***
ik	15.188**	-0.996	-2.510	0.605
ΔIN	-0.758	-0.145	-0.108	-0.183
$\Delta M1$	0.874	-5.241	-11.893	1.802*
ΔUN	0.191	-0.275	-0.446	-0.093
AR(1)	0.844***	74.740***		

Panel C: Annual

	β	$OoSR^2$	$OoSR^2_{Expan}$	$OoSR^2_{Recess}$
dp	-0.278***	2.252	-	-
tbl	0.978	-5.514	-	-
ltr	1.278**	5.088	-	-
tms	1.178	-1.479	-	-
dfr	-3.654***	0.285**	-	-
infl	1.555	-4.151	-	-
ik	39.200**	2.763	-	-
ΔIN	0.388	-0.533	-	-
$\Delta M1$	0.340	-17.576	-	-
ΔUN	-0.051	-1.095	-	-
AR(1)	0.828***	67.659***		

Table 9: **Predictability of min(0,r_{t+1})/max(0,r_{t+1}).** This table shows results from regressions using min(0,r_{t+1})/ max(0,r_{t+1}) as the dependent variables where returns are computed at monthly (Panel A), quarterly (Panel B) and annual (Panel C) frequency. Slope coefficients are estimated by OLS using min(0,r_{t+1})/max(0,r_{t+1}) as the dependent variables and a constant, the lagged values of volatility, the AR(1) and the (single) variable listed in the row as predictors. For AR(1) and cvol slopes are computed from univariate regressions. Stars indicate statistical significance using Newey-West standard errors (with one lag). Out-of-sample R^2 values are computed over the sample 1991-2010 and measured relative to a benchmark model that includes a constant, lagged cvol and lagged returns. For the AR(1) and the cvol models, the out-of-sample R^2 value is measured relative to a constant volatility benchmark. Stars indicate statistical significance using the Clark-West (2007) statistic. ***: significant at the 1% level; **: significant at the 5% level; *: significant at the 10% level

Panel A: Monthly

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
	min(0,r_{t+1})									**max(0,r_{t+1})**								
	Slopes									**Slopes**								
dp	-0.000	-0.003*	-0.003**	-0.004	-0.004**	-0.003**	-0.003**	0.003	0.000	-0.006	0.000	0.001	-0.001	0.001	0.002*	0.000	0.005**	0.000
tbl	0.061	-0.005	-0.004	0.030	-0.041	0.013	-0.006	-0.026	-0.028*	-0.086	-0.008	-0.044**	-0.066	-0.054*	-0.035	-0.027	-0.003	0.113***
ltr	0.024	0.021	-0.013	0.011	-0.097**	2.739	0.004	0.048	-0.003	-0.037	-0.006	-0.035	-0.059	-0.068**	-0.007	-0.034	0.112***	-0.078
tms	0.038	0.028	0.073*	0.128*	0.083	-0.021	0.055	0.093	0.027	0.059	-0.024	0.038	0.066	0.035	0.055	0.018	0.081	0.061
dfr	0.194	0.081	0.122	0.253*	0.299*	-0.000	0.115	0.119	0.034	-0.108	-0.027	0.092*	0.153	0.165**	0.142	0.042	-0.033	0.042
infl	0.363	-0.308	0.059	-0.172	0.202	-0.096	-0.132	-0.231	-0.212*	0.316	0.626**	0.044	0.561*	-0.057	-0.114	0.262	-0.359	-0.037
ΔIN	0.031	0.070	0.171*	0.130	0.433***	0.158***	0.129	0.097	0.045	0.200	0.131	0.240***	0.105	0.233**	0.161	0.195***	-0.064	-0.116*
$\Delta M1$	0.069**	0.024	0.032**	0.054**	0.029	-0.007	0.025**	-0.003	-0.028***	-0.027	0.005	0.034**	0.023	0.044*	0.076***	0.026*	-0.022	0.030**
ΔUN	0.013	0.007	-0.006	-0.011	-0.045**	-0.017	0.002	-0.000	-0.008	0.007	0.008	-0.021	-0.016	-0.027	0.008	-0.018	0.038	0.028**
cvol	-0.402***	-0.223***	-0.146**	-0.295**	-0.158	-0.113**	-0.172***	-0.192**	-0.031	0.422***	0.247***	0.202***	0.250***	0.349***	0.221***	0.202***	0.072	0.082*
AR(1)	0.218***	0.159***	0.340***	0.153***	0.261***	0.123*	0.318***	0.144**	0.034	0.092**	0.097**	0.349***	0.060*	0.265***	0.247***	0.263***	-0.036	0.178***
	OsS R^2									**OsS R^2**								
dp	-1.890	0.455	-1.996	-1.696	-0.945	-0.835	-0.646	-1.455	-0.993	0.458	-0.510	0.130	-0.387	-0.527	0.737*	-0.321	-2.749	-0.567
tbl	-0.988	-0.419	-0.383	-0.750	-0.438	-0.331	-0.384	-0.819	-4.629	0.114	-0.261	0.542	0.507	0.469	0.229	0.040	-0.397	-5.446
ltr	-0.479	-0.287	-0.082	-0.402	1.095	-0.153	-0.240	-0.345	-0.372	-0.884	-0.616	-1.615	-0.448	-1.601	-1.092	-1.878	-0.816	-4.906
tms	-0.134	-0.051	-0.261	0.166	-0.495	-0.216	0.052	-0.994	-1.189	-0.140	-0.171	0.017	-0.138	-0.157	0.098	-0.162	-1.924	0.592
dfr	0.106	-0.675	0.086	1.405	3.539**	-1.262	0.151	-0.790	-0.516	-0.392	-0.398	0.059	0.692	0.365	0.250	-0.541	-0.351	-1.682
infl	-0.131	0.025	-0.733	-1.362	-1.007	-0.076	-0.079	-1.890	-1.625	0.003	2.123**	-0.396	0.340	-0.529	-0.558	0.456	-1.717	-1.156
ΔIN	-0.271	-0.019	2.028*	-0.302	4.385*	1.494***	1.047	-0.019	0.455**	-0.120	0.202	1.115*	-0.246	0.089	-0.139	0.679	-0.391	-1.788
$\Delta M1$	-0.032	-1.160	-1.559	-1.275	-0.669	-0.777	-1.667	-0.359	0.328	-0.087	-0.232	1.584**	-0.139	0.873**	3.911**	0.494	-0.125	-1.887
ΔUN	-0.080	-0.080	-0.147	-0.045	-0.195	0.195	-0.071	-0.473	0.267*	-0.100	-0.091	0.086	-0.294	0.053	-0.176	-0.484	-0.327	1.116**
cvol	3.174*	0.433	0.626	2.030	0.837	-0.440	0.976	3.324**	0.389	3.113***	0.608**	4.054***	3.245***	5.904***	4.491**	2.484***	-0.555	-0.611
AR(1)	6.695**	3.597**	9.547**	2.681	2.821**	-3.834	12.978**	2.416*	-0.127	0.159	-0.237	8.405***	0.245	7.190***	-2.015	0.914**	-0.500	0.167

	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond	Fats-Oils	Foods	Industrials	Livestock	Metals	Textiles	Commodity	Stock	Bond
Panel B: Quarterly																		
	min(0,r_{t+1})									**max(0,r_{t+1})**								
	Slopes									**Slopes**								
dp	-0.011	-0.008*	-0.009*	-0.010	-0.011	-0.012**	-0.009**	0.010*	0.001	-0.002	-0.000	0.005	0.001	0.001	0.010	0.003	0.018**	0.000
tbl	0.042	-0.017	-0.056	0.006	-0.120	-0.033	-0.037	-0.050	0.008	-0.231	-0.047	-0.177**	-0.163	-0.258**	-0.077	-0.124*	-0.014	0.300***
ltr	-0.017	0.042	0.088	0.020	0.091	0.112***	0.064	0.008	-0.021	-0.054	0.024	0.028	-0.023	0.002	0.051	0.033	0.120	-0.205
tms	0.168	-0.035	0.191	0.212	0.095	0.172	0.059	0.176	0.007	0.096	-0.006	0.160	0.098	0.231	0.083	0.145	0.339	0.182
dfr	0.453	0.225	0.267	0.381	0.442	0.006	0.223	0.335**	0.027	-0.135	-0.000	0.053	0.084	0.184	-0.050	0.066	0.128	-0.054
infl	-0.677*	-0.568**	-0.673***	-0.813**	-0.685*	-0.498*	-0.601***	-0.310	-0.045	-1.229***	-0.074	-0.722*	-0.689*	-1.022*	-0.225	-0.469*	-0.077	0.313
ik	-1.206	-0.194	-0.882*	-1.085	-1.048	-0.261	-0.608	-1.350**	-0.057	-0.329	0.236	-1.463**	-0.128	-2.339**	-0.971	-0.740	-1.408	0.744
ΔIN	0.160	0.091	0.168*	0.215	0.416**	-0.015	0.140*	0.122	0.079	0.475	0.293**	0.380	0.218	0.260	0.329	0.318	-0.387**	-0.095
$\Delta M1$	0.191**	0.126**	0.104**	0.163**	0.124	0.045	0.099**	-0.021	-0.039	-0.037	-0.009	0.145*	0.089	0.190**	0.167**	0.093	-0.055	0.057
ΔUN	-0.019	-0.014	-0.023	-0.036	-0.080	0.018	-0.014	-0.021	-0.013	-0.094	-0.062**	-0.038	-0.017	-0.012	-0.063	-0.046	0.151***	0.024
cvol	-0.262	-0.156	-0.105	-0.116	-0.234	-0.036	-0.136	-0.236*	-0.016	0.481***	0.326**	0.344***	0.299**	0.696***	0.143*	0.332***	0.025	0.077
AR(1)	0.083	0.120	0.185**	0.132*	0.164	0.143	0.205**	0.103	-0.005	0.072	0.138	0.312***	0.028	0.292***	0.223***	0.303***	0.130**	0.121*
	OsS R^2									**OsS R^2**								
dp	-6.114	-1.505	-10.085	-6.857	-4.299	-6.916**	-5.980	-3.133	-1.630	-1.678	-2.217	0.429	-2.203	-1.903	6.473***	-0.958	-4.675	-2.215
tbl	-1.267	-1.367	-0.567	-0.943	-1.292	-0.367	-0.556	-2.559	-2.170	-0.580	-1.203	0.847	0.009	2.311	-2.156	-0.086	-1.089	-1.613
ltr	-0.660	0.284	-1.767	-0.729	-0.121	-17.149	0.279	-1.325	-1.788	-0.495	-0.602	-3.853	-1.543	-1.557	-1.097	-3.719	-0.592	-5.668
tms	-0.013	-1.711	-1.081	0.366	-0.870	-0.361	-0.212	-1.578	-1.168	-0.535	-0.650	-0.065	-0.546	-0.084	-0.287	-0.709	-2.677	1.049
dfr	1.176	1.841	3.532*	2.363*	4.272*	-0.920	1.942	2.295	-1.141	-1.185	-2.160	-3.370	-2.745	-1.986	-1.930	-3.975	-2.991	-2.264
infl	2.437**	5.135**	4.979**	5.106***	3.188**	3.816**	5.253**	-0.690	-17.137	3.753*	-0.880	3.750	2.116	3.484	-2.478	2.567	-1.624	2.301*
ik	0.414	-3.225	1.076	1.273*	0.392	-0.259	0.791*	-0.595	-2.025	-0.676	-1.033	4.770*	-1.250	4.560**	3.132*	1.215	-2.635	-1.692
ΔIN	0.207	0.018	1.137	0.884	1.629	-0.119	0.636	0.392	0.943	-0.099	2.391**	-4.679	-0.510	-2.344	-0.545	0.355	0.188	-0.583
$\Delta M1$	1.358**	0.925*	-0.443	1.740**	-0.097	-4.535	-0.105	-1.501	0.553	-0.776	-0.817	5.638**	-0.628	2.909*	10.457***	2.095*	-1.374	-1.852
ΔUN	-0.205	-0.157	0.421	0.164	1.302*	-0.519	0.100	0.079	0.841	-0.204	1.494**	-2.194	-0.690	-0.557	-1.186	-0.104	1.542	-0.406
cvol	0.341	-2.455	-2.464	-2.199	-0.795	-6.734	-2.648	4.077*	-1.447	2.827*	0.202	5.687*	2.346	14.883***	-7.115	5.436**	-2.976	-4.311
AR(1)	0.315	1.302	2.260	2.261*	0.938	0.369	2.318	1.405	-0.227	-2.018	-3.441	8.505***	-1.002	12.145***	6.725**	-1.125**	-4.003	-3.766
Panel C: Annual																		
	min(0,r_{t+1})									**max(0,r_{t+1})**								
	Slopes									**Slopes**								
dp	-0.020	-0.028	-0.024	-0.020	-0.012	-0.020	-0.020	0.045**	0.002	-0.002	-0.010	0.046	0.009	-0.006	0.035	0.017	0.103***	-0.007
tbl	0.168	-0.186	-0.227	0.212	-0.302	-0.182	-0.274	-0.002	0.140**	-1.396**	-0.539*	-1.333**	-1.222***	-2.547***	-0.489	-0.873**	-0.273	1.042**
ltr	0.070	-0.015	0.153*	0.134**	0.128	0.055	0.082	0.057	-0.013	0.051	-0.032	0.253**	0.149	0.314	0.176	0.152	0.167	-0.306
tms	0.245	0.789*	1.332***	0.773**	1.580**	0.866*	0.901***	0.607	0.087	0.828	0.337	1.185	0.965	1.610	0.895	1.024*	0.647	1.437**
dfr	0.036	0.168	0.259	0.100	0.713**	0.069	0.197	0.050	0.098*	0.434	0.334*	-0.461	0.065	-0.975*	-0.161	-0.106	-0.387	0.238
infl	-0.476	-0.415	-0.612**	-0.426	-0.517*	-0.524*	-0.602***	-0.114	0.133	-1.919**	-1.050**	-1.732**	-1.455***	-3.014***	-0.756	-1.301**	-0.021	0.619
ik	-3.414	-0.558	-4.416**	-2.871	-5.076*	-2.925	-2.881**	-4.990**	-0.508	-10.663	-3.517	-10.684	-8.352*	-14.394**	-4.984	-7.077	-5.817	3.672
ΔIN	-0.434	0.071	-0.526***	-0.388	-0.237	-0.463**	-0.301*	-0.192	-0.040	-0.474	-0.101	-0.750	-0.548	-0.336	-0.657	-0.472	-0.368	-0.139
$\Delta M1$	0.571**	0.392**	0.646**	0.706***	1.017**	0.305*	0.444**	0.223	-0.004	0.215	0.382	0.484	0.205	0.510	0.541	0.499	-0.829*	0.314
ΔUN	0.082	-0.028	0.120***	0.064	0.078**	0.093*	0.069**	0.032	0.004	0.176	0.060	0.245*	0.184	0.189	0.182	0.151	0.138*	-0.007
cvol	0.027	-0.038	0.045	0.207*	0.093	-0.024	-2.695	-0.141	0.011	0.610*	0.485**	0.692*	0.409	1.562**	0.275	0.575**	0.064	0.008
AR(1)	0.172	0.107	0.003	0.103	-0.177**	0.001	0.109	0.088	-0.083	-0.041	0.225	-0.010	-0.027	-0.003	0.041	0.082	-0.101	-0.048
	OsS R^2									**OsS R^2**								
dp	-48.857	-13.819	-31.370	-36.562	-35.060	-27.281	-19.780	-2.122	-12.968	-6.347	-14.913	3.639*	-4.988	-3.418	6.862*	-1.336	-33.059	-6.488
tbl	-6.296	-0.864	0.867	-7.684	-0.247	0.340	4.220	-3.574	4.188	16.983**	5.743	17.810*	20.657*	26.438*	3.614	18.300*	-12.522	3.970*
ltr	-1.295	-1.912	-16.252	-4.993	-8.618	-25.619	-10.944	-0.471	-3.096	-6.609	-13.152	-1.383	-0.544	0.975	-18.004	-10.777	-5.698	3.608**
tms	-0.425	11.290**	10.562***	2.050*	7.282***	3.891	6.616***	-2.420	-2.579	-0.820	-6.447	-0.565	1.570	-0.146	3.558	-0.601	-6.078	2.470*
dfr	-9.033	-21.277	-7.007	-4.673	6.994**	-8.677	-14.603	-14.672	0.251	-3.514	-2.431	-4.859	-16.829	0.479	-14.925	-13.421	-17.083	2.345
infl	-0.090	16.469**	8.840**	2.771	-0.310	17.550**	14.000**	-3.621	1.826	11.532**	9.548*	7.269	17.379*	28.913*	-8.381	11.652	-6.701	1.050
ik	1.318	-10.865	7.555**	-1.491	5.573**	5.148	5.308**	7.970**	-4.973	18.854**	5.127	19.043**	15.821*	12.243**	11.100**	21.885***	-14.883	-3.891
ΔIN	3.983*	-10.518	2.242	2.550	2.028*	0.182	1.420	-0.956	0.028	5.721	-1.713	-1.444	10.766	-0.946	8.405**	6.963*	-10.026	1.116
$\Delta M1$	-4.479	-14.732	14.493**	12.940*	13.589**	-2.819	10.007**	-0.663	-3.769	-1.983	-3.680	1.876	-1.241	-2.890	16.142***	3.601	-12.366	-10.727
ΔUN	4.533*	-9.188	2.855	1.761	2.247	-3.655	2.473	0.032	-0.804	11.428*	2.859	3.540*	17.536*	-0.141	12.482**	10.156*	-8.617	-0.951
cvol	-4.219	-28.885	-3.872	3.709	-2.283	-17.104	-7.754	-2.558	-16.408	10.186	20.571*	11.134*	1.983	15.942*	0.317	20.747*	-5.751	-7.345
AR(1)	-4.868	2.817	-3.499	-6.863	0.489	-1.318	-6.199	-5.009	1.052	-5.033	5.998	-1.706	-1.847	-3.696	0.038	1.656	-6.257	-2.494

Chapter IV
High-Dimensional Partial Index Tracking: Cointegration portfolios and Genetic Algorithms

1 Introduction

The risk-adjusted scarse profitability of traditional funds has been extensively empirically investigated. (5) and (23) among the others suggested that active portfolio management, on average, does not significantly outperform the market when transaction costs are considered. As opposed to traditional fund management, Index Tracking aims to match the risk-return profile of a benchmark index, allowing the investor to get round the perceived drawbacks of more aggressive vehicles, and to focus on a specific asset, industry or geographical sector. This alternative investment vehicle started to be popular in the 1990s when a handful of investment banks began offering these products to small investors. Since then, ETFs and tracker funds have seen a sharpe increase in volume, which is almost doubled during 2000s and still increasing. Several Index Tracking schemes have been proposed in the literature. The simplest index replication strategy is to invest in all of the assets composing the index.

This is called *full replication*. Despite, theoretically, sounds the most correct way, full replication is not only cumbersome but also rather costly. Indeed, trading and monitoring costs hamper this approach especially in high-dimensional indexes, since the index composition is rather time-varying recquiring frequent rebalancing. Another relevant approach is represented by *synthetic replication* towards equity derivatives like future contracts. The latter, usually have singularly less transaction costs. However, rolling contracts to dynamically track the underlying index is rather risky, making equity derivatives on average less attractive vehicles. Finally, an investor might consider *partial replication* of a benchmark index. Partial Index Tracking is the core business of ETFs and Tracker Funds, and the framework we focus on this paper. Managers pursuing a partial index replication strategy encounters two main problems, (1) select the optimal set of stocks to create the tracker index and, (2) optimally quantify the amount of wealth to invest in each of the stocks selected.

In the present paper we propose a computationally tractable solution to design near-optimal Partial Index Tracking strategies. The aim to track the trajectory of a benchmark index instead the usual ubiquitous returns. We develop a two-step procedure where the key point is cointegration between the index and constituents stocks. As pointed out in the seminal paper by (3), correlation-based index tracking schemes produce relatively unstable portfolios out-of-sample. (2), (3) and (11) proposed the use cointegration to capture long-run equilibria between tracking funds and the benchmark. Their methodology, however, considered the stock picking framework as a black-box, leaving the portfolio selection to the managers skills. This is a limiting approach especially in high-dimensional settings, where the manager is forced to select a large amount of stocks, and stock picking based on fundamental arguments is likely prohibitive.

Yet, (3) and (11) proposed a portfolio allocation scheme based on the coefficient of univariate linear models, through positive transformations[1]. In this paper we consider consis-

[1] In (3) the vector of weights is obtained through the OLS regression coefficient of a standard linear model, without considering short-sales

tently the portfolio weights selection in the spirit of (14), (10) and (15).

We extend the reference literature in several directions by, (1) developing a stock picking algorithm to extract the strongest cointegration relationships through a Genetic Algorithm. We structure (2) the objective function in the allocation step based on the benchmark and tracking portfolio trajectories consistently with the rationale of cointegration[2]. Finally (3) we provide some further insight on the joint effect of cointegration and trajectory-based allocations, disentangling their benefit as opposed to correlation/returns based index tracking schemes.

We find overall that optimally-cointegrated tracking portfolios provide superior performances in terms of tracking-risk minimization, reductions of transaction costs and portfolio

limitations. A similar approach is implemented in (11) where the portfolio weights are obtained through a positive normalization of reduced rank regressions coefficients by an exponential function.

[2]This is partly similar to (14) and (10). However, they applies clustering methods for stock picking instead of cointegration, and trajectory tracking without explicitly considering transaction costs as part of the objective function.

stability.

We first introduce the general problem of index tracking, and then describe the portfolio selection algorithm. We suggest an alternative, efficient use of genetic algorithms in a model selection logic, combined with a trajectory based portfolio allocation loss function solved through a standard nonlinear optimization algorithm.

2 The Partial Index Tracking problem

Let us consider a market with N securities with price $p_{i,t} \in (0, \infty)$ at each time *t* with $i = 1, ..., N$. By $c_{j,t}$ we indicate the dollar amount invested in the j_{th} asset such that

$$w_{j,t} = \frac{c_{j,t} p_{j,t}}{\sum_{j=1}^{K} c_{j,t} p_{j,t}} \qquad \text{with} \qquad w_{j,t} \in (0, 1) \tag{1}$$

is the relative proportion of the wealth invested in the j_{th} risky constituent stock, with $j = 1, ..., K$ and $K < N$. The level I_t of the benchmark index and the dollar amount of the tracking portfolio V_t are defined as

$$I_t = \sum_{i=1}^{N} \tilde{c}_{i,t} \times p_{i,t} \qquad \text{and} \qquad V_t = \sum_{i \in \Xi} c_{i,t} \times p_{i,t} \tag{2}$$

with Ξ the subset $K < N$ of selected stocks in the tracking portfolio. Time is discrete and the dollar amount of each stock in the index $\tilde{c}_{i,t}$ evolves dynamically. Therefore, the aim of the portfolio manager is to (1) select the optimal subset of stocks $K < N$, according to some risk-measure criterium, and (2) subsequently rebalance the portfolio weights consistently with the stock picking rationale. Although most of the literature treats them as separate concepts, we argue that, especially for Index Tracking, portfolio selection (stock picking) and allocation (rebalancing) are intimely related. This feature is carefully considered in the two-steps Index Tracking scheme we designed.

Several risk measures have been proposed in the literature. We can separate them in two main categories, namely correlation based measures and fundamentals based ones.

The former involves to select those stocks with higher in-sample correlation with the benchmark. As we mentioned before, stock picking based on fundamentals in high-dimensional indexes is likely prohibitive. Likewise, correlation based portfolio selection, although its simplicity, usually leads to unstable out-of-sample portfolios. Indeed, as noted in (3), (14), (10) and (11), correlation is unstable over time, limiting its out-of-sample usefulness. Yet, selecting stocks based on their correlation with the index could be somehow misleading. Indeed, suppose the index and the tracking fund trajectories move together, with a constant positive spread. They are likely going to be highly positively correlated anyway regardless their absolute distance. This is of course an undesirable feature. Finally, correlation is measured upon log-returns which are computed as the first order difference of the levels. This entails losing information about both the long-run mean of the stocks and the benchmark. This information is likely to be valuable for the manager especially to infer the long-run behavior of the index we would replicate.

In the following we define a proper risk measure for portfolio selection based on cointegration. The latter could be interpreted as a generalization of long-run correlation for non-stationary variables. In doing that we allow the manager to select the stocks based on trajectories, i.e. log-prices, keeping all the information about long-run behaviors.

2.1 Portfolio Selection and the Genetic Algorithm

2.1.1 Portfolio Selection

In the first step of the index tracking procedure the portfolio manager is asked to select the optimal subset $K < N$ of stocks tracking the trajectory of the benchmark composite index. Extending (3) and (11) and similarly to (14) and (10), we suggest to consider the first step as a high-dimensional model selection problem. The key point is cointegration in a univariate framework as in (12), (18) and (17). Let us consider a linear model linking the tracking fund and the benchmark levels as follows

$$\log(I_t) = \sum_{i=0}^{N} z_i \beta_i \log(p_{i,t}) + \epsilon_t \quad \epsilon_t \sim \pi(0, \sigma^2) \tag{3}$$

where I_t and $p_{i,t}$ represent respectively the index and the i_{th} constituent prices, and $\pi(.)$ a general distribution function. The error term ϵ_t is interpreted as the well-known tracking error. Notice that $\log(I_t)$ and $\log(p_{i,t})$ are non-stationary processes. The standard definition of univariate cointegration implies that if ϵ_t in (3) is stationary, then $\log(p_{i,t})$ and $\log(I_t)$ are cointegrated (see (12), (18) and (17) for more details). By exploiting the standard strict exogeneity of Ordinary Least Squares [OLS] estimates and the aforementioned cointegration property, if benchmark and tracking funds are cointegrated, then the tracking error ϵ_t is a zero-mean-reverting process. This is definitively a desirable property.

From an operational point of view, the manager is asked to select the cardinally fixed subset $\mathbf{z}^*$ such that $\beta = \{\beta_1, ..., \beta_K\}$ representing the strongest long-run equilibrium relationship (cointegration vector) between the index and the subset of constituent stocks.

What we propose is to use the standard Augmented Dickey Fuller [ADF] test methodology developed in (12) as a loss function to select $\mathbf{z}^*$, endowing later the manager with an efficient tool to address the model selection problem. The loss function is defined as

$$\varrho(\epsilon) = \frac{\hat{\rho} - 1}{se(\hat{\rho})} \tag{4}$$

with ρ a first order autoregressive coefficient in the auxiliary regression of the residuals in (3)

$$\epsilon_t = \rho\epsilon_{t-1} + \sum_{i=1}^{D} \Delta\epsilon_{t-i} + u_t \equiv \mathbf{x}_t\beta + u_t \quad \text{with} \quad \Delta\epsilon_t = \epsilon_t - \epsilon_{t-1} \tag{5}$$

and $se(\hat{\rho})$, *D* are respectively the relative standard error and the lag-order of the auxiliary regression. The null hypothesis is $H_0 : \rho = 1$, i.e. there is a unit root. The autoregressive univariate model in (5) is neither with intercept nor with a stationary trend since $E(u_t) = 0$.

Summarizing, the manager has to select those stocks minimizing (4), since the lower the ADF test statistics, the higher the stationarity of residuals, i.e. the stronger the long-run equilibrium relationship between the benchmark and the tracking portfolio. The binary model selection problem reads as

$$\begin{aligned} \mathbf{z}^* &\in \arg\min_{\mathbf{z}\in\mathcal{Z}} \varrho(\epsilon) \\ \text{s.t.} & \quad \sum_{i=1}^{K} z_i = K, \qquad \text{with } K < N \end{aligned} \tag{6}$$

It is quite intuitive that (4) is highly non-linear. Yet, as pointed out in (3), using the ADF in a standard brute force algorithm is just unfeasible due to the computational burden. We address the latter by using a Genetic Algorithm. This kind of population heuristics procedure is relatively standard in stochastic optimization literature and is becoming largely used in finance. Several meaningful examples of population heuristics used for portfolio selection can be found in (30), (6), (15), (26), (9), (29) and (16) among the others. A more detailed description of the Genetic Algorithm used can be found in the Appendix.

As far as we know, this cointegration-based stock-picking algorithm is applied on a large-scale to pairwise trading strategies, but is not considered for high-dimensional index tracking frameworks. This application is quite a novelty and represents one of the main contributions of the paper.

2.1.2 The Genetic Algorithm

The model selection issue is to find out most cointegrated subset of stocks, where the *degree* of cointegration is a functional of the linear model residuals in (3). The optimal indicator vector $\mathbf{z}^*$ is clearly highly non-linear. The Genetic Algorithm [GA] used for the portfolio selection problem depicted is developed as follows.

The P' and P'' are $n \times k$ matrices, where each row rapresent a portfolio, i.e. a vector of integers from 1 to N, where N is the number of stocks in the index. Intuitively, each integer represents a stock. Similarly, P' is a sub matrix of

Algorithm 1 Genetic Algorithm

Generate initial population P, initialize p_{mut} and p_{cross}
while stopping criteria not met **do**
  Select $P' \subset P$ (mating pool) set $P'' = \oslash$ (set of child)
  for i=1 to n_P (population size) **do**
    Select individuals m^a and m^b at random from P'
    if $u(0,1) < p_{cross}$ **then**
      cross-over: set $m^c = m^a \wedge m^b$
    else
      do not cross-over: set $m^c = m^a \vee m^b$
    end if
    if $u(0,1) < p_{mut}$ **then**
      mutate m^c into μ
    else
      do not mutate: set $\mu = m^c$
    end if
    set $P'' = P'' \cup \mu$
  end for
  set $P = P''$
end while
Return first row of P

P whose row size depends on the parameters discussed below. When the solutions are taken individually as $1 \times k$ vectors, they are labeled with m.

The asymptotic convergence to global solution relies on the schema theorem by (20) and on (27). They use the possibility of constructing Markov transition matrix from one generation to the other in order to apply standard Markov chain theory. In other words, as the population or the number of generations increase (i.e. higher problem difficulty), convergence is harder to reach in a reasonable amount of time.

The hyperparameters of the GA are often chosen according to some general guidelines presented in literature and by calibrating, mainly by trial and error, the algorithm to the problem (see (8)). Following (24) and (19) the value for p_{cross} and p_{mut} have been set to 0.9 and 0.3. This allow the algorithm to explore a wide set of initial solutions. As suggested by (31), they decrease over time, in order to make the algorithm to converge faster. Given its optimal properties, uniform crossover is adopted. When a solution is mutated $n < N$ genes (stocks) are substituted by n randomly selected genes (stocks). However, none of the replacing elements belongs to the initial chromosome (portfolio); it would generate a non invertible regressor matrix. In order to improve the algorithm performance, elitism operator is used by setting the row size of P' 0.5 times the one of P; it implies that only the better half of the current population is used to breed the next generation. The algorithm stops whether at least half of the P matrix is populated by the same model, or the best model is the same for more than half of the total number of generations. In both cases keep running the algorithm would be useless, because it has, most probably, already converged towards a solution. On the light of these considerations, the only parameter to be tuned by trial and error is n_P (Population Size). This is crucial because if it is too small, there would be not enough genetic diversification at damage of the quality of the final solution. If it is too big, too many iterations will be performed before it eventually converge.

A set of different population sizes are chosen; for each value in this range the algorithm is run ten times and the median fit function computed. Finally, these values are plotted against the population sizes. The value for which the median fit function start converging, is chosen. The last point to be discussed is how we select the fifty portfolios. Instead of running the algorithms fifty times and picking the best solutions we run it once, rank all unique solutions and select the first 50. This is done not only to reduce the computational burden but also to have a more diversified set of final portfolios.

2.2 Portfolio Allocation, Trajectories and Transaction Costs

Once the portfolio composition is selected the manager is asked to allocate proportionally the wealth in each of the stocks, according to properly defined distance/risk metrics.

Several measures have been proposed in the literature. Most of them are based on tracking error deviations. As noted in (6) and (4), these measures are however flawed. Indeed, if the distance between the index path and the tracking portfolio is constant over time, the tracking error volatility would be zero. This, of course, is an undesirable result because it does not take into account the tracking bias.

The distance measure we propose is inspired to (14), (10) and (15). Their risk measure were based on normalized trajectories of the benchmark and the tracking fund. We found this kind of approach particularly appealing, being consistent with the rationale of the cointegration-based stock picking strategy[3].

In addition we considered transaction costs directly in the objective function just as in (9) and (1). This allows the fund manager both to discriminate those stocks with high trans-

[3]Cointegration allowed to keep all of the information in the price levels, without disregarding time trends and unconditional long-run levels. In order to be consistent with this rationale we used a trajectories based risk function, which allows to keep the price levels in the non-linear optimization problem.

action costs, e.g. high bid-ask spreads, and to regularize the optimization algorithm (see (7) for more details). Once the $\mathbf{z}^*$ optimal portfolio composition is chosen from the genetic algorithm, the fraction $\mathbf{w}_t(\mathbf{z}^*)$ of wealth invested in each of the K stocks is the solution of

$$\mathbf{w}_t^* = \min_{\mathbf{w}_t \in \mathcal{W}} \left\{ \lambda \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\tilde{\mathbf{p}}_t \mathbf{w}_t}{\tilde{\mathbf{p}}_{t-1} \mathbf{w}_t} - \frac{I_t}{I_{t-1}} \right)^2} \right. \quad (7)$$

$$\left. + (1-\lambda) \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\tilde{\mathbf{p}}_t \mathbf{w}_t}{\tilde{\mathbf{p}}_{t-1} \mathbf{w}_t} - \frac{I_t}{I_{t-1}} \right) \right\}$$

$$\sum_{i=1}^{N} s_i \left| w_{i,t} - w_{i,t-1} \right|$$

$$\text{s.t. } \mathbf{w}_t \in C(\mathbf{w}_t) \quad (8)$$

where I_t is the level of the index, s_i the transaction costs in basis points, and

$$C(\mathbf{w}_t) = \left\{ \mathbf{w}_t \middle| \sum_{i=1}^{N} w_{i,t} = 1, \sum_{i=1}^{N} |w_{i,t}| = 1 \right\} \quad (9)$$

with K representing the user-defined maximum cardinality of the portfolio, $\tilde{\mathbf{p}}_t = \mathbf{p}_t/\mathbf{p}_0$ the normalized prices, and $\lambda \in (0,1)$ an implicit trade-off between tracking error and excess return over the benchmark[4]. In the empirical experiment we solve the optimization problem for different values of λ in his computational domain. It is easy to see that the optimal allocation $\mathbf{w}_t^*$ turns out to be a function of $\mathbf{z}^*$.

The set $C(\mathbf{w}_t)$ defines the portfolio constraints, restricting the set $\mathcal{W}$ of all potential portfolios to a smaller set of *feasible* solutions[5]. We impose a no-short sales constraint as in (22) and (3)[6].

[4]Notice $\lambda = 1$, corresponds to minimizing tracking error, i.e. *pure index tracking*, while $\lambda = 0$ implies maximising the excess returns over the benchmark.

[5]From a mathematical perspective the source of constraint is irrelevant, what matters is their structure. The most basic constraint is the budget constraint, i.e. $\mathbf{e}'\mathbf{w} = 1$. There are several other usually considered manager-specific contraints. The most simple is the floor/ceiling trading constraint, defined as $L_i \leq w_{i,t} \leq U_i$ with L_i, U_i representing respectively the lower and upper bounds for each portfolio weight. With the same rationale we could define bundle static constraints, by defining the lower and upper bound of trading in a certain industry or geographical area, i.e $L_B \leq \sum_{i \in B} w_{i,t} \leq U_B$.

[6]This kind of gross-exposure constraints not only have economic relevance but also helps in regularize the optimization algorithm, minimizing the portfolio risk exposure (see (21) and (13) for more details). Notice, this is not true in general, however holds if any dynamics or estimation is used in the optimization problem, as we do.

2.3 Portfolio simulation and Performance Measure

We assume that the stock selected at time *t* are still available at time $t + h$, with *h* the investment horizon. This is kind of reasonable if *h* is fairly small. We consider a $h = 252$, i.e. one-year of trading days as investment horizon, with different rebalancing periods within. In other words the portfolio composition selected through the Genetic Algorithm at time *T* is assumed to be constant till time $T + h$, while $\mathbf{w}_t^*$ is rebalanced at different frequencies according to (8). This is done on purpose to point out the reliability of cointegration as a stock picking strategy to get more out-of-sample long-run stable portfolios.

We implicitly applied an historical look-back approach as in (6), (3), (22), (9) and (29) among the others. The assumption is that the past contains enough information to get the future potential dynamics of the index. This is reasonable as far as we do not make estimates or predictions.

For the sake of robustness we construct $G = 50$ optimal portfolios, representing the first 50 sorted solutions of the Genetic Algorithm, such that we get $\mathbf{w}_{g=1}^{50}$ optimal allocations. The analysis rely on a *rolling sample* approach. This is a relative standard and straightforward way to impose some sort of simple dynamics in the asset allocation. Given a training dataset with $T = 504$ daily observations, we choose the optimal composition according to (6). The first allocation is defined in T and kept constant till $t0 = T + m$ with *m* the number of trading days defining the rebalancing period, i.e. $m = 21$ trading days for monthly relancing. Then, at *t0* the portfolio is rebalanced and kept constant till $t1 = t0 + m$. Rolling over this procedure for h/m times, we get *h* monthly *out-of-sample* returns generated by each of the *G* portfolios. The rebalancing horizons considered are $m = 21, 63, 126$ trading days plus a buy-and-hold strategy which, i.e. $m = 252$. Finally we consider different portfolio sizes $K = p * N$ with $p = [0.05, 0.1, 0.15]$, and different $\lambda = [1, 0.75, 0.5]$.

We compute a set of performance measures which are averaged out across the G portfolios for the sake of robustness. As a first measure of closeness between the tracking portfolios and the benchmark index we compute the average beta. This is obtained by regressing the index returns r_t^I on each of the *G* tracking returns r_t^g as follows

$$\hat{\beta} = \frac{1}{G}\beta_g \text{with} \qquad r_t^I = \beta_g r_t^g + \eta_t \tag{10}$$

The $\hat{\beta}$ is interpreted as an average measure of sensitivity of tracking portfolio returns with respect to the index movements. The closer $\hat{\beta}$ is to 1, the more reactive is the tracking funds to the benchmark path. Just as in (26) and (6) the average Mean Squared Error is a second measure of closeness

$$\widehat{MSE} = \frac{1}{G}\sum_{g=1}^{G} MSE_g \tag{11}$$

with

$$MSE_g = \frac{1}{h}\sum_{t=T+1}^{T+h} (r_t^g - r_t^I)^2 \tag{12}$$

The average Tracking Error Volatility is computed to be consistent with most of the reference literature (see (3), (22) and (25))

$$\widehat{TEV} = \frac{1}{G}\sum_{g=1}^{G} TEV_g \tag{13}$$

with

$$TEV_g = \frac{1}{h}\sum_{t=T+1}^{T+h} \left[(r_t^I - r_t^g) - (\hat{r}^I - \hat{r}^g)\right]^2 \tag{14}$$

where $\hat{r}^I$ and $\hat{r}^g$ respectively the index and g_{th} trackers mean returns out-of-sample. Other than closeness we argued at the beginning that cointegration-based stock picking algorithms allow the manager to get more out-of-sample stable portfolios. This a priori should imply lower average trading volumes. The average Turn Over for $g = 1, .., G$ is averaged out across portfolios

$$\hat{TO} = \frac{1}{G}\sum_{g=1}^{G} TO_g$$

with

$$TO_g = \frac{1}{h} \sum_{t=T}^{T+h-1} \sum_{j=1}^{N} \left(|w_{j,t+1}^{g} - w_{j,t}^{g}| \right) \tag{16}$$

where $w_{j,t+1}^{g}$ is the relative weight of the j_{th} stock in the g_{th} portfolio at time $t+1$. Finally in order to get some flavour about the profitability of the tracking procedure we consider a non-parametric perfomance measure. This is useful to avoid misleading results due to outliers. Yet, what we are interested in is the probability to outperform the index, keeping index tracking as a priority. The measure we adopted is the average probability of a positive excess returns with respect to the index

$$\widehat{ER} = \frac{1}{G} \sum_{g=1}^{G} P(r^g > r_I) \tag{17}$$

with

$$P(r^g > r_I) = \frac{\sum_{t=T+1}^{T+h} \mathbf{1}_{\{r_t^g > r_t^I\}}}{h}$$

(18)with r_t^g and r_t^I respectively the returns at time *t* of the g_{th} tracking portfolio and the benchmark. This measure is meant to be fair in the comparison across methodologies since does not depend on any parametric for of the returns dynamics.

3 Comparing Methodologies

We test the cointegration-based index tracking scheme proposed above, against four different methodologies. We have chosen them to disentangle the benefit in using (6) and (8) jointly as Index Tracking scheme. The four alternative methodologies are obtained combining either a correlation-based or a random portfolio selection strategy with either (8) or standard returns-based objective function for the portfolio allocation.

With reference to the correlation-based scheme the "optimal" subset of stocks for the tracker fund is obtained by taking the most *K* correlated stocks with the index. Then to generate the $G = 50$ portfolios we used a stationary non-parametric bootstrap as in (28) such that

$$\mathbf{z}_g^* = rank \left\{\rho_{i,g}\right\}_{i=1}^{K} \tag{19}$$

with

$$\rho_{i,g} = corr(r_i^g, r^I) \qquad \text{and} \qquad g = 1, ..., G \tag{20}$$

where r_i^g is the returns of the i_{th} stock in the g_{th} bootstraped sample, and $rank(.)$ is a ranking operator sorting in descending order the most *K* correlated stocks with respect to the index. The second stock picking strategy is a simple random selection of $G = 50$ portfolios[7]. Each of these two alternative stock picking strategies is combined with either (8) or a returns-based objective function defined as

$$\begin{aligned} \mathbf{w}_t^* = & \min_{\mathbf{w}_t \in \mathcal{W}} \left\{ \lambda \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{r}_t \mathbf{w}_t - I_t\right)^2} + \right. \\ & \left. (1-\lambda) \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{r}_t \mathbf{w}_t - I_t\right) \right\} \\ + & \sum_{i=1}^{K} s_i \left| w_{i,t} - w_{i,t-1} \right| \end{aligned} \tag{21}$$

$$\text{s.t.} \qquad \mathbf{w}_t \in C(\mathbf{w}_t) \tag{22}$$

[7]This method is taken as benchmark since taking a large subset of stocks in a broadly based indexes, likely implies the possibility to get cointegration between the stock selected and the index. Thus, comparing this method with the one we propose helps in pointing out the advantages of the genetic algorithm with the ADF as loss function.

where

$$C(\mathbf{w}_t) = \left\{ \mathbf{w}_t \middle| \sum_{i=1}^{N} w_{i,t} = 1, \sum_{i=1}^{N} |w_{i,t}| = 1 \right\} \tag{23}$$

with $\mathbf{r}_t$ the K-dimensional vector of returns. table (1) provides a short summary of the crossing schemes used for comparison.

[Insert Table 1 here]

4 Data

We consider three different broadly based indexes. They are selected based on increasing dimensions. The aim is (1) to point out the increasing usefulness of the Genetic Algorithm for the stock picking, and (2) to test the usefulness of cointegration as the ratio K/T gets close to 1. Indeed, as the number of stocks gets close to the number of insample observations, the way standard correlation is measured introduces increasing noise getting less stable out-of-sample portfolios.

The datasets refer to the *FTSE 100*, the *NIKKEI 225* and the *S&P500*. We have chosen these indexes on purpose, since not only the dimension but also the way they are constructed differs. The S&P500 and the FTSE are value-weighted composite indexes while the Nikkei is one of the few example of price-weighted indexes. As we will see this classification helps in providing some interesting results.

Tha data are collected from CRSP, Datastream and Bloomberg, referring to daily closing price of components stocks over the period September 2005 - August 2008. The descriptive statistics are reported in Table (2).

[Insert Table 2 here]

The sample period has been chosen on purpose to get bull and bear markets respectively in-sample and out-of-sample. This likely is going to penalize our results in terms of excess returns on the index, but on the other hand shed some light on the out-of-sample robustness of cointegration instead of correlation-based and random stock picking. Figure (1) reports the entire sample period of the indexes.

[Insert Figure 1 here]

The grey vertical line indicates the timing of stock picking. Therefore the left part represents the insample calibration period, while the right hand side is the out-of-sample testing period of the indexing schemes.

5 Empirical Results

We report the empirical results separately for each of the three indexes considered, for $\lambda = [1, 0.75, 0.5]$ in (8), and $K = p * N$ with $p = [0.05, 0.1, 0.15]$. Figure (2) reports the results for the average portfolio obtained through the Genetic Algorithm together with the benchmark

[Insert Figure 2 here]

The penalizing, strongly bear, out-of-sample period results in an overall negative cumulative return of the tracking portfolio. However the latter constantly outperform the benchmark both for the S&P500 and the FTSE100. This is not true as a whole for the Nikkei225, since the cumulative return of the tracking fund is underperforming across the end of 2007.

The competing methodologies are classified as in Table (1). We labeled GAcoint the indexing scheme proposed in this paper. In each of the Exhibits the first column reports the absolute value of the cointegration-based method for each of the perfomance metrics. Yet, from the second to the last column we report the ratio of the GAcoint procedure with respect the alternative measure[8]. Notice the βs are compared considering the ratio $|1-\hat{\beta}|/|1-\beta'|$ where β' is from the alternative portfolios schemes.

[8]A value greater than 1 means the GAcoint has a greater value for that measure and the other way round

5.1 FTSE 100

Table (3) reports the experimental results for the FTSE100. The $\hat{\beta}$ span from 0.782 to 0.808. All of the alternative procedures reports significantly lower values, i.e. ratios less than one. This is true for all the four investment horizons considered and all of the lambdas. Yet, randomly selected portfolios perform better than the correlation-based ones. This provides some insight about the out-of-sample unreliability of correlation-based stock picking strategies. The reason why the randomly selected portfolios perform comparably well is because taking a large subset of stocks in the index likely implies getting some weak cointegration relationship.

[Insert Table 3 here]

The relationship between correlation-based and random selection persists also with reference to the average Mean Squared Error and the average Tracking Error Volatility. Overall the GAcoint scheme significantly outperform the others.

This is true across lambdas and investment horizons. The probability of getting positive returns is still in favour of the GAcoint scheme, even though is slightly less than 50% in absolute terms. This sounds compromising, however, as we showed in Figure (1), the out-of-sample path of the benchmarks is strongly penalizing.

[Insert Table 4 here]

Table (4) Panel A, shows the average trading volume, in percentages, for the FTSE case. The GAcoint scheme is slightly outperformed by the RC scheme at the one-month rebalancing horizon. However the GAcoint starts to sensibly outperform the others procedures as the rebalancing horizon increases, i.e. 6 and 12 months, pointing out the long-term reliability of cointegration to build out-of-sample stable portfolios.

5.2 NIKKEI 225

Table (5) reports the results for the Nikkei 225 index. The βs are sensibly higher since the portfolio sizes $K = p * N$ increases overall. Apart from the randomly selection RC all of the alternatives schemes are strongly outperformed. Let us recall that for large sizes, RC likely entails some weak cointegration relationship. Yet the portfolio allocation is done consistently tracking the index trajectory. Therefore, considering also the price weighted property of the Nikkei index, the good performance of RC is not really surprising. However, the competitiveness of the randomly selection strategies disappear in all of the other measures. The average Mean Squared Error and the average Tracking Error Volatility are clearly in favour of the GAcoint scheme.

[Insert Table 5 here]

The probability to get a positive excess returns shows an interesting picture. All of the procedures are quite equally meaningful, without any clear outperforming scheme, neither looking at the lambdas, nor the rebalancing horizons. However, generally speaking, the cointegrated GAcoint gets a fairly comparable probability of realizing an excess returns with less risk.

Panel B of Table (4) shows the average trading volume. Yet, the RC scheme shows a fairly good performance. This is partly puzzling and no reliable explanation seems to be fairly sustainable. The only thing we can refer to, is the price weighted feature of the Nikkei index which could in principle mislead the algorithm in taking the mostly cointegrated stocks, in an economically meaningful manner. As for the FTSE case all of the scheme performs pretty equivalently in terms of excess returns probabilities.

5.3 S&P 500

The S&P500 is the most relevant example. Both its value-weighting construction and dimension represent a good benchmark, not only for the stock picking algorithm, but also for the portfolio allocation objective function. Table (6) reports the results. The βs are strongly in favour of GAcoint, and higher in absolute terms. This is because of the increasing sizes of the average tracking funds. The average Mean Squared Error and the average Tracking Error Volatility strongly support the reliability of the GAcoint scheme without exceptions.

[Insert Table 6 here]

This is true across lambdas, portfolio sizes and rebalancing horizons. Yet, the probability of getting positive excess returns is clearly higher than the alternative strategies even though the randomly selection-based schemes are fairly comparable on average. The correlation based schemes are, however, definitively outperformend.

The same argument applies to the average Trading Volume showed in Panel C of Table (4). Even though the RC scheme is pretty valuable, on average the GAcoint delivers the more stable portfolios out-of-sample. This is strongly true especially compared to the correlation-based schemes.

6 Concluding remarks

The theoretically benefit of cointegration portfolios in an Index Tracking setting have been extensively investigated in the literature. However, two factors represented a limit. The index dimensions usually small, to keep computational burden under control, and the opaqueness of the stock picking procedure. This paper addresses both the issues. We empirically investigate the benefit of a cointegration relationship between a benchmark index and its tracking portfolio in an high-dimensional Partial Index Tracking setting, developing a two-steps benchmark replication scheme. The first step develops the stock picking framework getting the optimal subset of stocks as a byproduct of the Augmented Dickey Fuller test minimization. The computational burden implied by the model selection issue is solved efficiently through a Genetic Algorithm. The second step entails to obtain the optimal vector of weights through standard non-linear optimization with a trajectories-based objective function. This particular objective function, links the cointegration-based stock picking rationale to the portfolio allocation issue.

We test this two-step procedure against four different alternative methods. These testing methods are constructed combining standard correlation-based and random stock picking strategies with either trajectories- or returns-based loss function in the portfolio allocation step. This is done on purpose to disentagle the joint benefit of using the cointegration portfolios with a consistent rebalancing setup.

In the empirical exercise we considered three broadly based indexes, four different investment horizons and three different portfolio sizes. Yet, all of the performances are averaged out on $G = 50$ portfolios generated for each of Index Tracking scheme. The two-step scheme proposed outperform the alternative methodologies as a whole. This is true with reference to several distance measures, portfolio turnover and excess returns with few exceptions. The mostly penalizing approach turns out to be the correlation-based stock picking, regardless the way we compute the optimal portfolio allocation. On the other hand the random selected portfolios, even though underperform on average are surprisingly comparably satisfying.

We would thank Paolo Colla, Fabrizio Leisen and Eva Besada for helpful comments and suggestions. All of the errors are our own.

References

[1] Adcock, C., Meade, N.: A simple algorithm to incorporate transaction costs in quadratic optimization. European Journal of Operational Research **79**, 85–94 (1994)

[2] Alexander, C.: Optimal hedging using cointegration. Phiosophical Transactions of the Royal Society **Series A**, 2039–2058 (2001)

[3] Alexander, C., Dimitru, A.: Indexing and statistical arbitrage. Journal of Portfolio Management **Winter**, 50–63 (2005)

[4] Ammann, M., Zimmermann, H.: Tracking error and tactical asset allocation. Financial Analysts Journal **57(2)**, 32–43 (2001)

[5] Barber, B., Odean, T.: Trading is hazardous to your wealth: the common stock investment performance of individual investors. Journal of Finance **55(2)**, 773–806 (2000)

[6] Beasley, J.E., Meade, N., Chang, T.J.: An evolutionary heuristic for the index tracking problem. European Journal of Operations Research **148(3)**, 621–643 (2003)

[7] Brodie, J., Daubechies, I., De Mol, C., Giannone, D., Loris, I.: Sparse and stable markowitz portfolios. ECB, Working Paper Series **No. 936**, 4–19 (2008)

[8] Chen, S. (ed.): Genetic Algorithms and Genetic Programming in computational finance (2002)

[9] Derigs, U., Nickel, N.H.: On a local-search heuristic for a class of tracking error minimization problems in portfolio management. Annals of Operations Research **131**, 45–77 (2004)

[10] Dose, C., Cincotti, S.: Clustering of financial time series with application to index and enhanced index tracking portfolio. Physica A **355**, 145–151 (2005)

[11] Dunis, C., Ho, R.: Cointegration portfolios of european equities for index tracking and market neutral strategies. Journal of Asset Management **6**, 33–52 (2005)

[12] Engle, R.F., Granger, C.W.J.: Cointegration and error correction: Representation, estimation and testing. Econometrica **55**, 251–276 (1987)

[13] Fan, J., Zhang, J., Yu, K.: Asset allocation and risk assessment with gross exposure constraints for vast portfolios. Preprint, Princeton University (2008)

[14] Focardi, S., Fabozzi, F.: A methodology for index tracking based on time-series clustering. Quantitative Finance **4**, 417–425 (2004)

[15] Gaivoronski A., K.S., Van der Wijst, N.: Optimal portfolio selection and dynamic benchmark tracking. European Journal of Operational Research **163**, 115–131 (2005)

[16] Gilli, M., Kellezi, E.: Threshold accepting for index tracking. computing in economics and finance. Society for Computational Economics **72** (2001)

[17] Granger, C.W.J.: Some properties of time series data and their use in econometric model specification. Journal of Econometrics **16**, 121–30 (1981)

[18] Granger, C.W.J.: Cointegrated variables and error correction models. Discussion Paper, UCSD (1983)

[19] Haupt, R., Haupt, S.: Practical genetic algorithms. Wiley (2004)

[20] Holland, J.: Adaptation in natural and artificial systems. University of Michigan press (1974)

[21] Jagannathan, R., Ma, T.: Risk reduction in large portfolios: Why imposing the wrong constraint helps. Journal of Finance **58(4)**, 1651–1684 (2003)

[22] Jansen, R., Van Dijk, R.: Optimal benchmark tracking with small portfolios. The Journal of Portfolio Management **Winter 2002**, 33–39 (2002)

[23] Jensen, M.: The performance of mutual funds in the period 1945-1964. Journal of Finance **23(2)**, 389–416 (1968)

[24] Jong, J.D.: An analysis of the behavior of a class of genetic adaptive systems (1975)

[25] Jorion, P.: Portfolio optimization with tracking-error constraints. Financial Analysts Journal **September/October 2003**, 70–82 (2003)

[26] Maringer, D., Oyewumi, O.: Index tracking with constrained portfolios. Intelligent Systems in Accounting, Finance and Management **15**, 57–71 (2007)

[27] Moral, P.D., Miclo, L.: On the convergence and applications of generalised simulated annealing. SIAM Journal of Controp Optimization **2**, 109–135 (1999)

[28] Politis, N., Romano J., P.: The stationary bootstrap. Journal of the American Statistical Association **89** (1994)

[29] Ruiz-Torrubiano, R., Surez, A.: An hybrid optimization approach to index tracking. Annals of Operations Research **166**, 57–71 (2009)

[30] Shapcott, J.: Index tracking: genetic algorithms for investment portfolio selection. Technical Report, EPCC-SS92-24. Edinburgh, Parallel Computing Centre. (1992)

[31] Sivanandam, S., Deepa, S.: Introduction to Genetic Algorithms (2008)

TABLES AND FIGURES

Table 1: Competing Strategies Classification.This table summarizes the compteting strategies investigated in the paper plus the one we propose. We classify the strategies according to the stock and weights selection procedures used.

Label	Stock Picking Procedure	Portfolio Allocation Procedure
GAcoint	cointegration-based + GA	Trajectories-based
CC	correlation-based + bootstrap	Trajectories-based
CR	correlation-based + bootstrap	Returns-based
RC	random selection	Trajectories-based
RR	random selection	Returns-based

Table 2: Descriptive Statistics. This table reports descriptive statistics of the considered indexes. Period from Sep 2005 to Aug 2008, $T = 756$ daily observations. Notice ValueW means Value-Weighted index composition. The same is true for Price-Weights. Data are not express in percentages.

	Num. Stocks	Mean	Median	St. Dev.	Skew	Kurt	Type
FTSE100	81	-0.0001	0.0000	0.0117	0.1304	8.7012	ValueW
NIKKEI225	208	0.0002	0.0000	0.0131	-0.4368	4.6387	PriceW
S&P500	442	0.0001	0.0004	0.0094	-0.1624	5.2076	ValueW

Table 3: FTSE100. This table reports the results for the FTSE 100 index. The sample comes from Sep 2005 to Aug 2008 with $T + h = 756$ daily observations. The in-sample lenght is $T = 504$ daily observations with a testing period of $h = 252$ daily observations. A month is represented by 21 trading days, investment horizons are reported in months. Models classifications are according to Table (1). **Beta** represents the OLS coefficient of regressing the index returns on the tracker returns, **Mse** is the Mean Squared Error, **Tev** represents the Tracking Error Volatility while **Prob** is the probability to get a positive excess return.

		Out-of-Sample Performances									
		Beta					**Mse**				
λ	Horizon	GAcoin	CC	CR	RC	RR	GAcoint	CC	CR	RC	RR
1	1	0.805	0.535	0.569	0.702	0.778	8.35e-005	0.698	0.759	0.696	0.846
	3	0.803	0.549	0.590	0.742	0.816	8.68e-005	0.730	0.819	0.714	0.867
	6	0.793	0.559	0.593	0.735	0.837	9.24e-005	0.699	0.763	0.694	0.867
	12	0.787	0.520	0.559	0.758	0.819	9.89e-005	0.567	0.606	0.667	0.836
0.75	1	0.808	0.512	0.555	0.684	0.735	7.81e-005	0.631	0.704	0.679	0.732
	3	0.800	0.546	0.601	0.732	0.803	8.85e-005	0.740	0.846	0.775	0.856
	6	0.789	0.566	0.624	0.759	0.797	9.57e-005	0.739	0.827	0.770	0.816
	12	0.782	0.525	0.580	0.784	0.786	1.04e-004	0.590	0.644	0.750	0.784
0.5	1	0.807	0.528	0.556	0.679	0.794	8.07e-005	0.680	0.723	0.699	0.865
	3	0.800	0.553	0.600	0.734	0.829	8.77e-005	0.750	0.827	0.758	0.878
	6	0.792	0.562	0.607	0.739	0.829	9.01e-005	0.706	0.757	0.712	0.834
	12	0.784	0.522	0.569	0.761	0.819	9.72e-005	0.552	0.598	0.681	0.805

		Tev					**Prob**				
λ	Horizon	GAcoin	CC	CR	RC	RR	GAcoint	CC	CR	RC	RR
1	1	3.40e-002	0.799	0.847	0.820	0.892	49.54	1.189	1.161	1.041	1.041
	3	3.42e-002	0.813	0.870	0.824	0.895	49.05	1.149	1.131	1.029	1.020
	6	3.54e-002	0.798	0.838	0.818	0.896	49.22	1.200	1.159	1.030	1.032
	12	3.60e-002	0.707	0.745	0.807	0.875	49.13	1.180	1.136	1.033	1.030
0.75	1	3.34e-002	0.768	0.826	0.816	0.848	49.26	1.180	1.153	1.040	1.043
	3	3.45e-002	0.811	0.884	0.849	0.890	49.02	1.159	1.130	1.026	1.029
	6	3.59e-002	0.813	0.876	0.851	0.871	49.21	1.192	1.161	1.030	1.041
	12	3.68e-002	0.715	0.769	0.842	0.853	48.92	1.181	1.142	1.030	1.036
0.5	1	3.37e-002	0.793	0.831	0.822	0.904	49.07	1.173	1.138	1.034	1.034
	3	3.45e-002	0.819	0.878	0.843	0.904	49.07	1.153	1.131	1.025	1.023
	6	3.53e-002	0.806	0.851	0.833	0.891	49.09	1.185	1.151	1.029	1.033
	12	3.62e-002	0.706	0.752	0.820	0.874	49.01	1.175	1.130	1.033	1.034

Table 4: Average Trading Volumes. This table reports the results for the trading volumes. The results are in percentages. The sample comes from 15-Sep-2005 to 6-Aug-2008 with $T + h = 756$ daily observations. The in-sample lenght is $T = 504$ daily observations with a testing period of $h = 252$ daily observations. A month is represented by 21 trading days, investment horizons are reported in months. Models classifications are according to Table (1).

Panel A

		Ftse 100				
λ	Horizon	GAcoint	CC	CR	RC	RR
1	1	8.73	0.948	0.874	1.027	0.851
	3	18.39	0.898	0.933	1.010	1.001
	6	18.63	0.567	0.464	0.697	0.916
0.75	1	7.89	0.871	0.766	0.881	0.767
	3	17.59	0.815	0.869	0.854	0.897
	6	18.19	0.552	0.520	0.643	0.844
0.5	1	7.23	0.883	0.711	0.806	0.728
	3	16.51	0.829	0.839	0.832	0.911
	6	15.64	0.483	0.456	0.518	0.782

Panel B

		Nikkei 225				
λ	Horizon	GAcoint	CC	CR	RC	RR
1	1	4.89	0.763	0.681	0.999	0.890
	3	7.46	0.614	0.630	0.817	0.840
	6	14.60	0.858	0.884	1.142	1.304
0.75	1	5.26	0.805	0.745	1.067	0.950
	3	7.46	0.610	0.659	0.814	0.777
	6	14.89	0.894	1.123	1.186	1.515
0.5	1	5.03	0.753	0.697	1.017	0.880
	3	7.63	0.633	0.735	0.843	0.784
	6	15.18	0.946	1.352	1.215	1.435

Panel C

		S&P 500				
λ	Horizon	GAcoint	CC	CR	RC	RR
1	1	2.40	0.603	0.514	1.032	0.738
	3	3.47	0.618	0.467	0.943	0.708
	6	3.33	0.315	0.209	0.647	0.500
0.75	1	2.17	0.531	0.463	0.986	0.718
	3	3.72	0.714	0.512	1.086	0.861
	6	3.83	0.363	0.239	0.753	0.607
0.5	1	2.20	0.555	0.471	1.047	0.743
	3	3.53	0.674	0.460	1.064	0.840
	6	3.51	0.333	0.223	0.649	0.579

Table 5: $Nikkei$. This table reports the results for the Nikkei 225 index. The sample comes from 15-Sep-2005 to 6-Aug-2008 with $T + h = 756$ daily observations. The in-sample lenght is $T = 504$ daily observations with a testing period of $h = 252$ daily observations. A month is represented by 21 trading days. A month is represented by 21 trading days, investment horizons are reported in months. Models classifications are according to Table (1). **Beta** represents the OLS coefficient of regressing the index returns on the tracker returns, **Mse** is the Mean Squared Error, **Tev** represents the Tracking Error Volatility while **Prob** is the probability to get a positive excess return.

		Out-of-Sample Performances									
		Beta					**Mse**				
λ	Horizon	GAcoint	CC	CR	RC	RR	GAcoint	CC	CR	RC	RR
1	1	0.932	0.486	0.508	0.901	0.700	2.45e-005	0.841	0.727	0.759	0.707
	3	0.929	0.537	0.546	0.925	0.774	2.45e-005	0.865	0.727	0.744	0.714
	6	0.931	0.552	0.602	1.012	0.827	2.46e-005	0.854	0.759	0.753	0.755
	12	0.931	0.569	0.606	1.095	0.824	2.36e-005	0.838	0.739	0.761	0.713
0.75	1	0.930	0.495	0.526	0.948	0.730	2.44e-005	0.869	0.732	0.761	0.760
	3	0.927	0.528	0.584	0.944	0.758	2.43e-005	0.867	0.757	0.749	0.731
	6	0.929	0.560	0.640	1.039	0.800	2.44e-005	0.886	0.793	0.753	0.750
	12	0.929	0.577	0.645	1.109	0.818	2.34e-005	0.863	0.768	0.763	0.724
0.5	1	0.931	0.485	0.524	0.934	0.682	2.44e-005	0.863	0.764	0.764	0.693
	3	0.927	0.542	0.602	0.971	0.754	2.45e-005	0.898	0.829	0.764	0.707
	6	0.929	0.564	0.646	1.046	0.766	2.44e-005	0.889	0.823	0.754	0.696
	12	0.929	0.570	0.646	1.104	0.784	2.34e-005	0.862	0.795	0.764	0.681

		Tev					**Prob**				
λ	Horizon	GAcoint	CC	CR	RC	RR	GAcoint	CC	CR	RC	RR
1	1	1.94e-002	0.926	0.869	0.885	0.876	42.44	0.933	0.918	0.925	0.918
	3	1.95e-002	0.939	0.876	0.881	0.881	45.83	0.991	0.998	0.995	0.987
	6	1.95e-002	0.935	0.892	0.886	0.899	46.58	1.030	1.010	1.004	1.005
	12	1.91e-002	0.926	0.881	0.891	0.881	47.33	1.050	1.027	1.021	1.024
0.75	1	1.94e-002	0.938	0.872	0.887	0.888	46.61	1.037	1.001	0.994	0.983
	3	1.94e-002	0.937	0.892	0.884	0.879	46.12	1.030	0.989	1.003	0.999
	6	1.94e-002	0.946	0.906	0.886	0.888	46.47	1.032	0.996	1.006	1.015
	12	1.91e-002	0.935	0.894	0.892	0.878	48.67	1.081	1.046	1.054	1.053
0.5	1	1.93e-002	0.935	0.888	0.889	0.858	46.12	1.030	0.989	1.003	0.999
	3	1.94e-002	0.952	0.921	0.891	0.870	47.21	1.050	1.014	0.971	1.014
	6	1.94e-002	0.948	0.918	0.887	0.865	44.95	1.000	0.961	1.032	1.037
	12	1.91e-002	0.935	0.905	0.893	0.857	48.67	1.088	1.047	1.049	1.054

Table 6: $S\&P500$. This table reports the results for the S&P 500 index. The sample comes from 15-Sep-2005 to 6-Aug-2008 with $T + h = 756$ daily observations. The in-sample lenght is $T = 504$ daily observations with a testing period of $h = 252$ daily observations. A month is represented by 21 trading days. A month is represented by 21 trading days, investment horizons are reported in months. Models classifications are according to Table (1). **Beta** represents the OLS coefficient of regressing the index returns on the tracker returns, **Mse** is the Mean Squared Error, **Tev** represents the Tracking Error Volatility while **Prob** is the probability to get a positive excess return.

		Out-of-Sample Performances									
		Beta					**Mse**				
λ	Horizon	GAcoint	CC	CR	RC	RR	GAcoint	CC	CR	RC	RR
1	1	0.947	0.221	0.218	0.661	0.504	1.77e-005	0.271	0.271	0.863	0.666
	3	0.959	0.212	0.202	0.698	0.527	1.66e-005	0.212	0.195	0.826	0.637
	6	0.963	0.211	0.197	0.697	0.521	1.64e-005	0.203	0.179	0.812	0.628
	12	0.962	0.194	0.177	0.751	0.515	1.66e-005	0.163	0.133	0.871	0.613
0.75	1	0.950	0.210	0.208	0.661	0.493	1.70e-005	0.257	0.259	0.868	0.625
	3	0.956	0.200	0.192	0.695	0.511	1.65e-005	0.210	0.197	0.870	0.613
	6	0.962	0.200	0.188	0.700	0.513	1.64e-005	0.162	0.134	0.919	0.594
	12	0.960	0.182	0.166	0.759	0.498	1.91e-002	0.935	0.894	0.892	0.878
0.5	1	0.952	0.219	0.218	0.681	0.558	1.67e-005	0.253	0.258	0.840	0.723
	3	0.961	0.210	0.201	0.728	0.589	1.61e-005	0.206	0.192	0.845	0.719
	6	0.965	0.207	0.195	0.726	0.588	1.60e-005	0.197	0.176	0.830	0.721
	12	0.965	0.191	0.175	0.806	0.569	1.61e-005	0.158	0.133	0.905	0.688
		Tev					**Prob**				
λ	Horizon	GAcoint	CC	CR	RC	RR	GAcoint	CC	CR	RC	RR
1	1	1.66e-002	0.528	0.528	0.951	0.856	49.30	1.158	1.181	1.025	1.018
	3	1.61e-002	0.470	0.449	0.929	0.837	49.27	1.168	1.195	1.023	1.018
	6	1.60e-002	0.461	0.431	0.922	0.829	49.26	1.174	1.202	1.023	1.018
	12	1.61e-002	0.417	0.374	0.952	0.824	49.17	1.184	1.221	1.013	1.015
0.75	1	1.63e-002	0.514	0.516	0.952	0.835	49.15	1.152	1.180	1.019	1.015
	3	1.61e-002	0.468	0.451	0.952	0.831	49.38	1.173	1.200	1.021	1.017
	6	1.59e-002	0.458	0.429	0.941	0.829	49.07	1.170	1.198	1.017	1.021
	12	1.60e-002	0.415	0.374	0.978	0.822	49.07	1.183	1.217	1.011	1.015
0.5	1	1.61e-002	0.511	0.515	0.940	0.875	49.23	1.156	1.177	1.022	1.014
	3	1.59e-002	0.464	0.446	0.941	0.871	49.31	1.169	1.189	1.017	1.015
	6	1.58e-002	0.454	0.426	0.935	0.871	49.21	1.172	1.197	1.018	1.021
	12	1.58e-002	0.411	0.373	0.972	0.858	49.15	1.183	1.220	1.012	1.023

Figure 1: Indexes

This figure reports the three indexes considered. The grey vertical line represents the in-sample end for the GA stock picking procedure. As we can see there is a regime switch from bull to bear market at the beginning of the out-of-sample period.

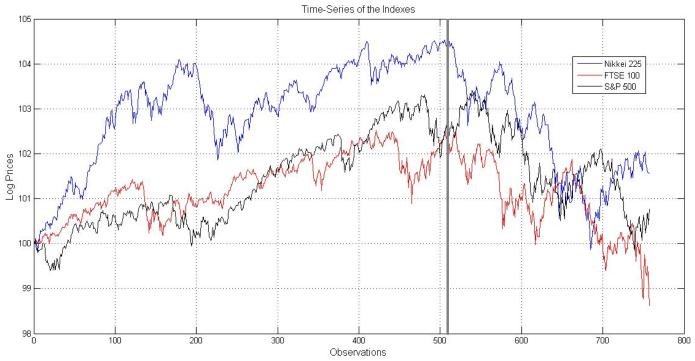


Figure 2: Indexes and the Cointegrated GA procedure

This figure reports the three indexes considered together with the average trakcer fund got by our proposed approach. The series reported is the $T - M$ out-of-sample period.