RESEARCH DIVISION

Working Paper Series

Out-of-Sample Predictions of Bond Excess Returns and Forward Rates: An Asset-Allocation Perspective

Daniel L. Thornton
and
Giorgio Valente

Working Paper 2010-034B
https://doi.org/10.20955/wp.2010.034

October 2010

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

Out-of-sample Predictions of Bond Excess Returns and Forward Rates: An Asset-Allocation Perspective*

Daniel L. Thornton† and Giorgio Valente‡

This draft: February 2012

*Acknowledgments: This work was partly written while Giorgio Valente was visiting the Federal Reserve Bank of St Louis, whose hospitality is gratefully acknowledged. The views expressed here are the authors' and do not necessarily reflect the views of the Board of Governors of the Federal Reserve System or the Federal Reserve Bank of St. Louis. The authors are grateful to Bob Bliss, John Cochrane, Pasquale della Corte, Eric Engstrom, Massimo Guidolin, Mike McCraken, Monika Piazzesi, Lucio Sarno and Ton Vorst, to participants to the 2010 Workshop on Yield Curve Modelling and Forecasting, Erasmus University Rotterdam, the 6th International Conference on Asia Pacific Financial Markets, Seoul, Korea as well as to seminar participants at the Bank for International Settlements, City University of Hong Kong, Hong Kong Institute for Monetary Research, Hong Kong Lingnan University, KAIST Business School, National Chengchi University for useful conversations and comments on earlier drafts of this paper. The authors alone are responsible for any errors that may remain.

†Federal Reserve Bank of St Louis, Research Division, P.O. Box 442,St. Louis, MO 63166-0442. Phone: +1 (314) 444 8582. Email: thornton@stls.frb.org.

‡Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Phone: +44 1206 872254. Email: gvalente@essex.ac.uk.

Abstract

This paper investigates the out-of-sample predictability of bond excess returns. We assess the economic value of the forecasting ability of empirical models based on long-term forward interest rates in a dynamic asset allocation strategy. The results show that the information content of forward rates does not generate systematic economic value to investors. Indeed, these models do not outperform the no-predictability benchmark. Furthermore, their relative performance deteriorates over time.

JEL classification: G0; G1; E0; E4.

Keywords: bond excess returns, term structure of interest rates, expectations hypothesis, forecasting.

1 Introduction

The predictability of bond excess returns has occupied the attention of financial economists for many years. In past decades, several studies have reported evidence that empirical models based on forward rates or forward spreads are able to generate accurate forecasts of bond excess returns. Since forward rates represent the rate on a commitment to buy a one-period bond at a future date, it is natural to hypothesize that they incorporate information that is useful for predicting bond excess returns. In support of this conjecture, Fama and Bliss (1987, henceforth FB) find that the forward-spot spread has predictive power for the change in the spot rate and excess returns and that the forecasting power increases as the forecast horizon lengthens. Recently, Cochrane and Piazzesi (2005, henceforth CP) extend FB's original work by proposing a framework in which bond excess returns are forecast by the full term structure of forward rates. They find that their specification is able to capture more than 30 percent of the variation of bond excess returns over the period January 1964 - December 2003. More recently, Cochrane and Piazzesi (2008) confirm these results using a larger set of maturities. [1] [2]

We contribute to the existing literature on the predictive ability of forward rates for bond excess returns in two ways: First, since a model's in-sample predictive performance tends to correlate poorly with its ability to generate satisfactory out-of-sample forecasts (Inoue and Kilian, 2004; 2006), we evaluate the forecasting ability of predictive models based on forward rates in a genuine out-of-sample forecasting exercise. Second, given that statistical significance does not mechanically imply economic significance (Leitch and Tanner, 1991; Della Corte *et al.*, 2008; 2009), we assess the economic value of the predictive power of forward rates by investigating the utility gains accrued to investors who exploit the predictability of bond excess returns relative to a no-predictability alternative associated with the validity of the expectations hypothesis.

In the spirit of Fleming *et al.* (2001), Marquering and Verbeek (2004), and Della Corte *et al.* (2008; 2009), we quantify how much a risk-averse investor is willing to pay to switch from a dynamic portfolio strategy based on a model with no predictable bond excess returns to a model that uses either forward spreads (FB) or the term structure of forward rates (CP), with and without dynamic volatility specifications. We consider two volatility specifications: a constant variance consistent with a standard linear regression and a rolling sample volatility model (Foster and Nelson, 1996; Fleming *et al.* 2001; 2003). The latter is computationally efficient and is flexible enough to capture the features of bond excess returns data. In order to take into account the problems arising from potential mispecification and parameter changes in models of conditional mean excess returns, the parameters are estimated over time using all past observations available up to the time of the forecast (recursive scheme) and a selected window of past observations (rolling scheme). In addition, we also allow for parameter uncertainty when constructing optimal portfolios. Specifically, we impose an informative prior to define the distribution of the parameter estimates used to carry out the asset allocation problem, as in Kandel and Stambaugh (1996) and Connor (1997).[3]

We find that none of the predictive models based on forward rates is able to add significant economic value to investors relative to the no-predictability benchmark. However, the extent of the underperformance varies across specifications, especially when parameter uncertainty is taken into account. Also, predictive regressions with conditional volatility show no significant improvement relative to the constant volatility alternative. Finally, we find that the relative performance of the predictive models deteriorates over time. In particular, as suggested by Cochrane (2011), the predictive models seem to be especially unsatisfactory during the recent 2007-2009 financial crisis.

Various studies have attempted to validate, with mixed success, the early empirical findings reported in FB and CP. Rudebush *et al.* (2007) show that the empirical estimates of the term premia implied by CP are less correlated with other available measures and are more volatile. Similarly, using a reverse regression methodology, Wei and Wright (2010) find that ex ante risk premia on long-term bonds are both large and volatile because the underlying parameters appear to be imprecisely estimated.

Other studies have investigated the source of information embedded in forward rates and their genuine predictive power. Radwanski (2010) shows that the one-year-ahead expected inflation extracted from the cross-section of forward rates, together with a level factor capturing the average level of forward rates, is able to attain results similar to CP. Almeida *et al.* (2011) find that term-structure affine models that include interest rate option prices in the estimation are able to generate bond risk premia that better predict excess returns for long-term rates. The R^2 estimates that they obtain are similar in magnitude to those reported in earlier studies. Cieslak and Povala (2011) decompose long-term yields into a persistent inflation component and maturity-related cycles. They show that the CP predictive regressions are special cases of a more general return-forecasting regression where the CP factors are constrained linear combinations of cycles. Using this framework, they obtain in-sample R^2 that are twice those reported by CP. Duffee (2011) criticizes the notion that term structure models ought to rely on bond yields (and linear combinations of them, such as forward rates) to serve as the factors in theoretical and empirical models. He shows that almost half of the variation in bond excess returns can be associated with a (hidden) filtered factor that is not related to the cross-section of bond yields. Furthermore, when the filtered factor is added to the CP regressions, the term structure of forward rates is no longer statistically significant at conventional levels.

The statistical properties of bond yields also give a reason to be skeptical of the forecasting power of predictive regressions based on forward rates. First, the empirical frameworks proposed in this literature implicitly embed long-horizon returns. It is well known that OLS estimations of regressions of bond excess returns on the term structure of yields suffer from small-sample bias and size distortions that exaggerate the degree of predictability. Hence, the estimates of R^2 reported in the existing literature are inadequate measures of the true in-sample predictability (Kirby, 1997; Valkanov, 2003, Campbell and Yogo, 2006; Boudoukh *et al.*, 2008; Wei and Wright, 2010 and the references therein).

Second, bond yields are highly serially correlated and correlated across maturities. If both regressors and regressands exhibit a high serial correlation, the predictive regressions based on forward rates may suffer from a spurious regression problem (Ferson *et al.*, 2003a,b, and the references therein).[4] [5] [6] Consequently, evidence of in-sample predictability need not be a useful indicator of out-of-sample predictive performance. Moreover, the parameters of these empirical models may vary over time (Fama, 2006; Wei and Wright, 2010) and this, in turn, affects the performance of the empirical models when used out of sample.

Our paper is closely related to Duffee (2010) and Barillas (2010), who explore the predictability of bond excess returns from a similar perspective. Duffee (2010) investigates the conditional maximal Sharpe ratios implied by fully flexible term structure models and finds that in-sample model overfitting leads to astronomically high Sharpe ratios. Barillas (2010) investigates the optimal bond portfolio choice of an investor in a model that captures the failure of the expectations hypothesis of interest rates. In an in-sample exercise, the author finds that investors conditioning on bond prices and macroeconomic variables would be willing to give up a sizable portion of their wealth in order to live in a world where the risk premia state variable is observable. Our analysis differs from these studies in three important respects. These studies investigate the in-sample predictability of predictive models while we assess the economic value from using these models out of sample. In addition, we investigate the impact of parameter uncertainty on bond excess return predictions and we also explicitly incorporate estimates of the conditional variances of bond returns into the portfolio allocation problem.[7]

The remainder of the paper is as follows: Section 2 introduces the empirical framework used to model the conditional mean and volatility of bond excess returns. Section 3 discusses the framework for assessing the economic value of bond excess returns predictability for a risk-averse investor with a dynamic portfolio strategy. Section 4 reports the main empirical results and Section 5 explores the performance of the predictive models over time during the past 30 years. Section 6 discusses the results of various robustness checks and a final section concludes.

2 The predictive power of forward rates

In line with the existing literature, we define the log-yield of an n-year bond as

$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}, \tag{1}$$

where $p_t^{(n)}$ is the log price of an n-year zero-coupon bond at time t, i.e., $p_t^{(n)} = \ln P_t^{(n)}$, where $P^{(n)}$ is the nominal dollar-price of a zero coupon bond paying \$1 at maturity. A forward rate with maturity n is then defined as

$$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)}. \tag{2}$$

The excess return of an n-year bond is computed as the log-holding period return from buying an n-year bond at time t and selling it after one year less the yield on a one-year bond at time t,

$$rx_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)}. \tag{3}$$

Recent empirical research has uncovered significant forecastable variations in bond excess returns. More specifically, several studies recorded that bond excess returns vary over time and they are a quantitatively important source of fluctuations in the bond market (see, *inter alia*, Ludvigson and Ng, 2009; Piazzesi and Schneider, 2011). In this empirical study we select two key models that have been proved successful in explaining and forecasting bond excess returns by means of forward rates and forward spreads.

Using monthly data for bond yields with maturities ranging between one and five years, FB estimate the excess return equations[8]

$$rx_{t+12}^{(n)} = \zeta_0 + \zeta_1 (f_t^{(n)} - y_t^{(1)}) + \upsilon_{t+12}^{(n)}, \tag{4}$$

where $n = 2, ..., 5$ denotes the forward rate maturity, expressed in years. Using equation (4), FB find that the forward-spot spread has predictive power for bond excess returns and that the forecasting power increases as the forecast horizon lengthens.

CP propose a modified version of the FB excess returns equation. Specifically, they estimate a general regression where bond excess returns are predicted by the full term structure of forward rates and the one-period bond yield, i.e.,

$$rx_{t+12}^{(n)} = \delta_0 + \delta_1 y_t^{(1)} + \delta_2 f_t^{(2)} + ... + \delta_5 f_t^{(5)} + \varepsilon_{t+12}^{(n)}. \tag{5}$$

They find that their forward rate equation explains between 30 and 35 percent of the variation of bond excess returns over the same bond maturity spectrum investigated by FB.[9]

Note that equations (4) and (5) can be written more generally as

$$rx_{t+12}^{(n)} = c + \boldsymbol{\beta}'\mathbf{Z}_t + \epsilon_{t+12}^{(n)}, \tag{6}$$

where $\mathbf{Z}_t = Z_t^{(n)} = (f_t^{(n)} - y_t^{(1)})$ or $\mathbf{Z}_t = \begin{bmatrix} y_t^{(1)} & f_t^{(2)} & ... & f_t^{(5)} \end{bmatrix}'$ in equations (4) and (5), respectively.

When $\boldsymbol{\beta} = \mathbf{0}$, bond excess returns are not predictable and equal to a constant c. This case is consistent with the expectations hypothesis of the term structure of interest rates, which is frequently used as benchmark against which other empirical bond excess return models are compared. We label this model as EH.

As reported in various studies, and documented later in this paper, there is considerable evidence indicating that the volatility of bond yields and bond excess returns is time-varying and predictable (Gray, 1996; Bekaert *et al.* 1997; Bekaert and Hodrick, 2001). Hence, in addition to equation (6), we model the dynamics of the conditional variance-covariance matrix of bond excess returns with a simple linear regression model and with a rolling sample variance estimator (Foster and Nelson, 1996; Fleming *et al.*, 2001; 2003). More specifically, the linear regression model assumes that the conditional covariance matrix of the residuals $\Sigma_{t+12|t} = E_t\left[\boldsymbol{\epsilon}_{t+12}\boldsymbol{\epsilon}'_{t+12}\right]$ with $\boldsymbol{\epsilon}_{t+12} = \begin{bmatrix} \epsilon_{t+12}^{(2)} & ... & \epsilon_{t+12}^{(5)} \end{bmatrix}$ is constant over time, i.e., $\widehat{\Sigma}_{t+12|t} = \widehat{\Sigma}$. The rolling sample variance estimator is of the general form

$$\widehat{\Sigma}_{t+12|t} = \sum_{l=0}^{\infty} \Omega_{t-l} \odot \boldsymbol{\epsilon}_{t-l}\boldsymbol{\epsilon}'_{t-l}, \tag{7}$$

where Ω_{t-l} is a symmetric 4×4 matrix of weights and $\odot$ denotes element-by-element multiplication. The logic behind this approach is that if Σ_{t+12} is time-varying, then its dynamics are reflected in the sample path of past excess returns. Hence, if a suitable set of weights are applied to squares and cross-products of excess return innovations, it is possible to construct a time series estimate of Σ_{t+12} (Foster and Nelson, 1996). In our empirical application, we follow Fleming *et al.* (2003) and select the optimal weight for a one-sided rolling estimator $\Omega_{t-l} = \alpha \exp(-\alpha l) \mathbf{11}'$, where $\mathbf{1}$ denotes a 4×1 vector of ones and α is the decay rate that governs the relative importance assigned to past excess return innovations. As in Fleming *et al.* (2001, p. 334), we impose that the decay parameter is unique across all cross products of excess return innovations in order to ensure the positivity of the matrix $\Sigma_{t+12.}$. We use this estimation method to compute the conditional covariance matrix Σ_{t+12} since it is not heavily parametrized and it is less difficult to estimate than multivariate ARCH and GARCH models[10] (see, *inter alia*, Bawens *et al.*, 2006 and the references therein). In fact, for certain choices of Ω_{t-l} the rolling sample estimator resembles the Σ_{t+12} process implied by a multivariate GARCH model (Fleming *et al.*, 2003 p. 479).

3 Assessing bond excess returns predictions

3.1 The asset allocation framework

This section explores the economic significance of the predictive information embedded in forward rates and forward spreads relative to the no-predictability alternative. A classic portfolio choice problem is used (Della Corte *et al.*, 2008; 2009). Specifically, we consider an investor who optimally invests in a portfolio comprising $K+1$ bonds similar in all respects but with different maturities: a risk-free one-period bond and K risky n-period bonds. The investor constructs a monthly dynamically rebalanced portfolio that minimizes the conditional portfolio variance subject to achieving a given target of expected return.

Let the conditional expectation and the conditional variance-covariance matrix of the $K \times 1$ vector of bond excess returns, $\mathbf{rx}_{t+12}$, be equal to $\boldsymbol{\mu}_{t+12|t} = E_t(\mathbf{rx}_{t+12})$ and $\Sigma_{t+12|t} = E_t[(\mathbf{rx}_{t+12} -$

$\boldsymbol{\mu}_{t+12|t})(\mathbf{rx}_{t+12} - \boldsymbol{\mu}_{t+12|t})']$, respectively. At the end of each period the investor solves the following problem:

$$\begin{aligned} \min_{\mathbf{w}_t} \quad & \mathbf{w}_t' \Sigma_{t+12|t} \mathbf{w}_t \\ s.t. \quad & \boldsymbol{\mu}_{t+12|t}' \mathbf{w}_t = \zeta_p^*, \end{aligned} \tag{8}$$

where $\mathbf{w}_t = \begin{bmatrix} w_t^{(2)} & \ldots & w_t^{(5)} \end{bmatrix}'$ is the $K \times 1$ vector of portfolio weights on the risky bonds and $\zeta_p^* = \left(\mu_p^* - y_t^{(1)}\right)$ is the target of conditional expected return of the full portfolio returns. The solution to the optimization problem delivers the following weights on the risky n-period bonds,

$$\mathbf{w}_t = \frac{\zeta_p^*}{C_t} \Sigma_{t+12|t}^{-1} \boldsymbol{\mu}_{t+12|t}, \tag{9}$$

where $C_t = \left(\boldsymbol{\mu}_{t+12|t}\right)' \Sigma_{t+12|t}^{-1} \left(\boldsymbol{\mu}_{t+12|t}\right)$ and the weight on the one-period bond is equal to $1 - \mathbf{w}_t' \iota$, where ι is a $K \times 1$ vector of ones.

In the empirical analysis, we winsorize the weights to each of the n-period bonds to $-1 \leq w_t^{(n)} \leq 2$ to prevent extreme investments (Goyal and Welch, 2008; Ferreira and Santa Clara, 2011). These constraints essentially allow for the full proceeds of short sales (see, *inter alia*, Vayanos and Weill, 2008 and the references therein).

3.2 Modelling bond excess returns and their volatility: the role of parameter uncertainty

In order to construct the optimal portfolio weights, $\mathbf{w}_t$, estimates of conditional expected bond excess returns $\boldsymbol{\mu}_{t+12|t}$ and conditional variance-covariance matrices $\Sigma_{t+12|t}$ are required. Three different conditional mean strategies are considered: the benchmark model of no predictability (EH), the FB model, and the CP model. The three models are estimated using both constant volatility (CVOL henceforth) and time-varying volatility (TVOL henceforth) to compute the volatility forecasts.

Given the statistical problems noted in the previous section, it seems likely that there could be uncertainty about the parameter estimates or even the overall parametrization of the datagenerating process.[11] This complicates the asset allocation process because the investor does not know over which parameter set to minimize the portfolio variance's function. In the spirit of the literature on portfolio choice under parameter uncertainty, we follow Kandel and Stambaugh (1996) and Connor (1997), who recommend imposing an informative prior to define the distribution of the parameter estimates used to carry out the asset allocation problem.[12] More specifically, we advocate a weak-form efficiency of the bond markets consistent with the expectations hypothesis, i.e., estimates of $\widehat{\boldsymbol{\beta}}$ are assumed equal to zero. Kandel and Stambaugh (1996) demonstrate that, in a Bayesian regression setup, this prior yields a posterior of the parameter estimates which is the product of the OLS estimates and a shrinking factor that is a function of the precision of the parameter estimates. The smaller the precision of the parameter estimates, the stronger the shrinkage towards their prior mean of zero.

The empirical investigation in this paper takes into account Kandel and Stambaugh's (1996) findings by implementing the procedure introduced by Connor (1997). That is, each of the j parameter estimates in equation (6) are computed as

$$\widehat{\beta}_{j,bayes} = \left[\frac{T}{T + \left(\frac{1}{\rho_j}\right)} \right] \widehat{\beta}_{j,OLS}, \tag{10}$$

where the shrinking factor in brackets is a function of the sample size T and a parameter ρ_j, which represents the marginal degree of predictability of the predictive variable j. ρ_j is computed as

$$\rho_j = E\left[\frac{R_j^2}{1 - R^2} \right], \tag{11}$$

where R_j^2 denotes the marginal R^2 of variable j and R^2 is the coefficient of determination of the full predictive regression (6) using forward rates (or forward spreads) as predictive variables.[13] Hence, estimates based on parameter uncertainty will be closer to the OLS estimates the larger the sample size and the larger is R_j^2.

3.3 The economic value of excess returns predictability

The economic value of the predictability of forward rates is assessed by assuming quadratic utility, as in West *et al.* (1993), Fleming *et al.* (2001), and Della Corte *et al.* (2008; 2009); and the average realized utility, $\overline{U}(\cdot)$, for an investor with initial wealth W_0 is given by

$$\overline{U}(\cdot) = \frac{W_0}{T-12+1} \sum_{t=0}^{T-12} \left\{ R_{p,t+12} - \frac{\lambda}{2(1+\lambda)} (R_{p,t+12})^2 \right\}, \tag{12}$$

where $R_{p,t+12} = 1 + y_t^{(1)} + \mathbf{w}_t' \mathbf{rx}_{t+12}$ is the period $t+12$ gross return on the portfolio and λ denotes the investor's degree of relative risk aversion (RRA). It is also assumed that $W_0 = 1$. [14]

As in Fleming *et al.* (2001), the measure of the economic value of alternative predictive models is obtained by equating average utilities of selected pairs of portfolios. For example, assume that holding a portfolio constructed using the optimal weights based on the EH strategy with constant volatility (EH$_{CVOL}$) yields the same average utility as holding the portfolio implied by the CP strategy with constant volatility (CP$_{CVOL}$). The latter portfolio is subject to management expenses, Φ, expressed as a fraction of wealth invested in the portfolio. If the investor is indifferent between these two strategies, then Φ can be interpreted as the maximum performance fee the investor would be willing to pay to switch from the EH$_{CVOL}$ to the CP$_{CVOL}$ strategy. In general, this criterion measures how much a risk-averse investor is willing to pay for conditioning on the information in the forward rates since the benchmark used implies no predictability in either the conditional mean or the conditional variance. The performance fee is the value of Φ that satisfies

$$\sum_{t=0}^{T-12} \left\{ \left(R_{p,t+12}^{\mathcal{F}} - \Phi\right) - \frac{\lambda}{2(1+\lambda)} \left(R_{p,t+12}^{*} - \Phi\right)^2 \right\} = \sum_{t=0}^{T-12} \left\{ R_{p,t+12}^{EH_{CVOL}} - \frac{\lambda}{2(1+\lambda)} \left(R_{p,t+12}^{EH_{CVOL}}\right)^2 \right\}, \tag{13}$$

where $R_{p,t+12}^{\mathcal{F}}$ denotes the gross portfolio return constructed using the predictions from regression (6) in which forward rates or forward spreads are used as predictors, i.e., $\mathcal{F} = FB, CP$, and $R_{p,t+1}^{EH_{CVOL}}$ is the gross portfolio return implied by the bond excess returns no-predictability benchmark with constant volatility, EH$_{CVOL}$. If there is no predictive power embedded in forward rates or forward spreads, then $\Phi \leq 0$; whereas, if forward rates or forward spreads help to predict bond excess returns, $\Phi > 0$.

In the context of mean-variance analysis, several other measures of performance are routinely employed. A measure frequently used is the Sharpe ratio (SR), which is calculated as the ratio of the average portfolio excess returns to its standard deviation, i.e., $\frac{1}{T-12+1} \sum_{t=0}^{T-12} \left(r_{p,t+12}^{\mathcal{F}} - y_{t+12}^{(1)} \right) / \sigma^{\mathcal{F}}$ for any predictive model, where $r_{p,t+12}^{\mathcal{F}} = R_{p,t+12}^{\mathcal{F}} - 1$ and $\sigma^{\mathcal{F}}$ denotes the standard deviation of $\left(r_{p,t+12}^{\mathcal{F}} - y_{t+12}^{(1)} \right)$. The statistical significance of the difference of the SR from two competing models is tested by using the bootstrap procedure introduced by Ledoit and Wolf (2008). This procedure has been shown to be robust to portfolio returns that are nonnormal and serially correlated. In particular, we construct a studentized time series bootstrap confidence interval for the difference of the SR, using a variant of the circular block bootstrap (Politis and Romano, 1992) and test whether zero is contained in the interval.[15]

While Sharpe ratios are commonly used, they exhibit some drawbacks. Specifically, they do not take into account the effect of nonnormality (Jondeau and Rockinger, 2008), they tend to underestimate the performance of dynamic strategies (Marquering and Verbeek, 2004; Han, 2006 and the references therein), and they can be manipulated in various ways (Goetzmann *et al.,* 2007). In order to take into account these concerns, we follow Goetzmann *et al.* (2007), who suggest a set of conditions under which a manipulation-proof measure exists. This performance measure can be interpreted as a portfolio's premium return after adjusting for risk. We build on Goetzmann *et al.* (2007) and calculate a risk-adjusted abnormal return of the predictive models relative to the EH_{CVOL} strategy as follows:

$$\begin{aligned} GISW &= \frac{1}{(1-\lambda)} \left[\ln \left(\frac{1}{T-12+1} \sum_{t=0}^{T-12} \left[\frac{R_{p,t+12}^{\mathcal{F}}}{\left(1+y_{t+12}^{(1)}\right)} \right]^{1-\lambda} \right) \right. \\ & \left. - \ln \left(\frac{1}{T-12+1} \sum_{t=0}^{T-12} \left[\frac{R_{p,t+12}^{EH_{CVOL}}}{\left(1+y_{t+12}^{(1)}\right)} \right]^{1-\lambda} \right) \right]. \end{aligned} \tag{14}$$

In dynamic investment strategies of the kind used here, portfolio rebalancing entails a significant role for transaction costs. In the U.S. Treasury secondary market, traders charge transaction costs according to counterparty types and trade size. We do not take a specific stand as to how large the transaction costs should be. Instead a break-even transaction cost, τ^{BE}– i.e., the one that renders investors indifferent between two competing strategies (Han, 2006; Della Corte *et al.*, 2009) – is computed. This is accomplished by assuming that transaction costs equal a fixed proportion (τ) of the value traded in the different bonds, V. The average (monthly) transaction cost of a strategy is computed as $\tau \times V$, where

$$V = \frac{1}{T-12+1} \sum_{t=0}^{T-12} \sum_{k=1}^{K} \left| w_t^{(k)} - w_{t-1}^{(k)} \frac{1 + w_t^{(j)} \left(rx_{t+12}^{(n)} + y_t^{(1)} \right)}{R_{p,t+12}} \right|. \tag{15}$$

Following Jondeau and Rockinger (2008), the break-even transaction cost τ^{BE} is computed as

$$\tau^{BE} = \frac{\left(\frac{1}{T-12+1} \sum_{t=0}^{T-12} r_{p,t+12}^{\mathcal{F}} \right) - \left(\frac{1}{T-12+1} \sum_{t=0}^{T-12} r_{p,t+12}^{EH_{CVOL}} \right)}{V^{\mathcal{F}} - V^{EH_{CVOL}}}, \tag{16}$$

where $V^{\mathcal{F}}$ and $V^{EH_{CVOL}}$ denote the value traded in the different bonds associated with the predictive models $\mathcal{F} = FB, CP$ and the benchmark, respectively. In comparing any predictive model $\mathcal{F}$ with EH_{CVOL}, an investor who pays transaction costs lower than τ^{BE} will always prefer model $\mathcal{F}$ to the benchmark. Break-even transaction costs are computed only when they can be meaningfully interpreted, i.e., when the performance fees in equation (13) are positive.

4 Empirical results

4.1 Data and preliminary results

The data set used in this study, consistent with early studies on the predictability of bond excess returns, comprises monthly one- to five-year zero-coupon bond prices from June 1952 through December 2010.

Log-bond excess returns are computed from bond prices as described in Section 2. The summary statistics of the resulting time-series are reported in Table 1, Panels A) and B). Bond excess returns are found to be close to zero on average (ranging between 0.4% and 1% per annum) but all are statistically significant at the 5 percent statistical level. Panel A) also reports the autocorrelation coefficients of order 1 and 12 for the individual time series that show that bond excess returns are highly serially correlated.

Panel B) of Table 1 reports the same summary statistics for absolute bond excess returns used as a proxy for bond excess return volatilities. Absolute bond excess returns exhibit average values that are higher the longer the term to maturity, and they are all statistical significant at the conventional 5 percent significance level. Furthermore, in line with previous studies and the results reported in Panel A), absolute bond excess returns are also highly serially correlated.

In addition to reporting the autocorrelation coefficients, we also compute the correlation between all pairs of bond excess returns. Figures 1 and 2 show the average cross-correlations of bond excess returns and absolute excess returns over different 10-year subperiods of the sample. The correlation coefficients are high for both excess returns and absolute excess returns (larger than 0.8 in all cases) and there is some evidence of time variation over the sample period. In particular, the correlation coefficients across maturities increase between the 1960s and the 1980s and then generally decline over the past two decades. This pattern is also exhibited by the correlation coefficients between absolute bond excess returns. The finding of time-varying correlation among excess returns innovations is also corroborated by the Lagrange Multiplier test developed by Tse (2000) that rejects the null of constant conditional correlations with a p-value of virtually zero.[16] In related contexts, there is evidence that shocks generated by negative news may have greater impact on subsequent volatilities than positive shocks of the same magnitude (Engle and Ng, 1993; de Goeij and Marquering, 2006). In order to investigate this issue, we have estimated the following threshold GARCH(1,1) model (Glosten *et al.*, 1993; Zakoïan, 1994) for each of the four bond excess returns time series:

$$\left[\sigma_t^{(n)}\right]^2 = \psi_0 + \psi_1 \left[\sigma_{t-1}^{(n)}\right]^2 + \psi_2 \left[\epsilon_{t-1}^{(n)}\right]^2 + \psi_3 \left[\epsilon_{t-1}^{(n)}\right]^2 I_{t-1}^{-} + \xi_t,$$

where $\sigma_t^{(n)}$ is the conditional volatility of the n-period bond excess return, $\epsilon_{t-1}^{(n)}$ is the lagged residual from the mean equation[17] and I^{-}_{t-1} is a dummy variable that is equal to 1 if $\epsilon^{(n)}_{t-1} < 0$ and zero otherwise. The symmetry in the excess returns conditional volatility is assessed by testing H_0 : $\psi_3 = 0$. The results of the estimations over the full sample period and for all bond maturities, not reported to save space, suggest that the null hypothesis of symmetry is not rejected at conventional levels.

The preliminary exploration of the data is completed by estimating the parameters of the three candidate models over two sample periods: the full sample period 1952-2010 and the CP's sample period, 1964-2003. The in-sample estimates are reported in Table 2 Panels A)-C). Estimates of FB and CP models computed over the sample period 1964-2003 are similar to those reported in Cochrane and Piazzesi (2005, 2006). A comparison of the FB model over the two sample periods shows that the estimates of $\bar{R}^2$ are somewhat smaller over the full sample period; however, the estimates of the parameters changed little. Nevertheless, the parameters estimated over the full sample period generally lie outside the confidence interval of the ones estimated over the smaller sample period for all n.

The comparison of CP's model for the two samples shows a marked reduction in the estimates of $\bar{R}^2$: the estimates over the full sample are at least 45 percent lower relative to CP's sample period. While the tent shape of the parameter estimates noted by CP is evident in both samples, the estimates of the parameters are considerably different. Specifically, with the exception of estimates of β_4, the estimates are much smaller in absolute value over the full sample period, and in most cases the parameters estimated over the full sample period lie outside the confidence interval of the ones estimated over the smaller sample. This finding is indicative of a considerable time variation in the parameter estimates, which in turn is reflected in the marked reduction in the estimates of $\bar{R}^2$ and potentially affect the model's out-of-sample performance.

4.2 Economic value calculations

This section reports the results of the economic value calculations discussed in Section 3. Forecasts are generated using parameters that are estimated using information only available at the time the forecast is made. More specifically, we also employ two forecasting schemes: 1) a recursive scheme, which uses all the observations available up to the time of the forecast, and 2) a rolling scheme where only a window of past observations is used. We consider a rolling scheme since it is likely that because of changes in the macroeconomic environment (shifts in the Fed policy etc.), parameters estimated using data from very past periods may not be necessarily useful to make current out-of-sample predictions.[18] Furthermore, as outlined in Section 3.2, we incorporate the role of parameter uncertainty in the conditional mean and compute the parameter estimates with and without the correction reported in equations (10) and (11).

The combination of the six models (i.e. FB, CP an EH with and without time-varying volatility) and the four scenarios (i.e. recursive and rolling estimations with and without parameter uncertainty) yields 24 sets of results. The performance measures are calculated for the out-of-sample period January 1970 through December 2010 assuming $\lambda = 5$, in line with the value used in previous studies (Barberis, 2000; Della Corte *et al.*, 2008 and the references therein). We also use two annual targets of portfolio excess returns, $\zeta_p^* = 0.01, 0.02$. The target portfolio excess returns are consistent with reasonable average excess returns obtained by portfolios of Treasury bonds and are higher than the average bond excess returns reported in Table 1, Panel A). The performance measures SR and GISW and the performance fees Φ are reported as annualized. The GISW and Φ measures are expressed in decimals (i.e., 0.01 = 1 annual percentage point). The time-varying variance-covariance matrix of excess returns is computed using $\alpha = 0.05$, a value within the range of those reported in existing studies (Foster and Nelson, 1996; Fleming *et al.*, 2001; 2003). Finally, the rolling forecasting scheme is implemented using a window of the past 120 months.[19]

Table 3 shows the results of these exercises in four panels. Panel A) presents the results on the recursive forecasting scheme with no parameter uncertainty. The results indicate the FB and CP predictive models with constant volatility provide no economic value relative to the EH_{CVOL} benchmark. Their estimated SR are lower than the one of the benchmark model for either choice of ζ_p^*. However, in nearly all instances, the difference is not statistically significant at conventional levels. In the few cases where the difference is statistically significant, the EH_{CVOL} model generates larger economic gains than the competing predictive models. Qualitatively, the results are identical with the GISW measure. All of the estimates are negative and range between -2.2 percent (EH_{TVOL}) and -1.0 percent (CP_{CVOL}).

The conclusions are unchanged when we allow for parameter uncertainty, Panel B). In this case, however, the EH_{CVOL} model is superior to both the FB_{CVOL} and CP_{CVOL} models at the 5 percent significance level using the SR measure in three of the four cases reported. As in the case of no parameter uncertainty, all of the estimates of GISW are again negative and are in a range that is only slightly narrower that of Panel A).

The conclusions are invariant to the rolling forecasting scheme, reported in Panels C) and D). The estimates of SR for the EH_{CVOL} are generally smaller than those with recursive estimation. However, the equality null hypothesis is rejected in only two instances. The pattern based on SR is confirmed by estimates of GISW. Indeed, they are either zero or negative and of similar magnitude to those obtained with the recursive scheme.

The fact that none of the model with time-varying volatility is superior to those with constant volatility is consistent with the findings of Duffee (2002) and Cheridito *et al.* (2007). They find that bond excess returns are best captured by constant volatility models, in spite of the fact that such models cannot match the time-series variation in interest rate volatility.[20]

The results reported in Table 3 are corroborated by the estimates of the performance fees reported in Table 4, Panels A)-D). All of the performance fees are negative, reflecting the fact that none of the alternative models is economically superior to the no-prediction benchmark with constant volatility. The magnitude of the performance fees is relatively unaffected by whether parameter uncertainty is taken into account. The performance fees are frequently less negative, however, when the models are estimated using a rolling scheme rather than recursively. Hence, from the perspective of performance fees, there seems to be a gain from focusing on the most recent data.

5 Sub-sample analysis

This section refines the results reported in the previous section by assessing the predictive ability of predictive models in different sub-sample periods. This exercise is motivated by a recent literature suggesting that macroeconomic variables, and more specifically interest rates, have been more difficult to predict since the Great Moderation (e.g., Clark and McCracken, 2008 and D'Agostino *et al.*, 2006). Specifically, we investigate the economic value of the predictive ability of the FB and CP models over four sub-samples: January 1970 through December 1979, January 1980 through December 1989, January 1990 through December 1999 and January 2000 until December 2010. The results of this exercise are summarized in Table 5 Panels A)-D), which reports the SR from the various models together with the p-values from the Ledoit and Wolf's (2008) test. The measures are computed using a level of annual target of portfolio excess returns $\zeta_p^* = 0.01$ and the other parameters are set equal to the values used to carry out the baseline estimates reported in Tables 3 and 4.

The results for the different sub-periods confirm the conclusion obtained using the full sample. For every period except the decade of the 1970s, for all models and for all forecasting schemes the SR from the benchmark EH_{CVOL} is higher than that from the competing models. However, there are only few instances where the economic performance of the EH_{CVOL} model is statistically superior at the 10 percent significance level. During the 1970s all of the competing models generate a SR larger than that of the EH_{CVOL} model; however, there is only one instance where the difference is statistically significant at the 10 percent level.

It is interesting to note that the performance of the predictive models relative to the EH_{CVOL} model deteriorated over time, i.e., the difference between the SR of the EH_{CVOL} and any competing model generally increased over time. Indeed, the differences are almost always the largest during the 2000s. This performance deterioration is also reflected in the GISW measure for the FB and CP predictive regressions (Figure 3). The deterioration is smaller for the performance fees (Figure 4). Both figures plot the average of the performance measures computed over the decade of reference across all 24 specifications, as well as their across-specification standard deviation. The average GISW performance measures and the average performance fees are positive during the 1970s but they deteriorate quickly during the 1980s, remaining negative until the end of the sample. These results are generally consistent with the notion that forecasting has become more difficult since the Great Moderation. The general pattern of deterioration in these performance measures may be associated with shifts in monetary policy in the late 1980s (see, *inter alia*, Sims and Zha, 2006; and Thornton, 2006, 2010 and the references therein), which generated a greater persistence in the Fed's target and induced less predictability in excess returns.

We also compute the performance measures over the period of the recent financial crisis, January 2007 through December 2009. This analysis is motivated by the evidence that suggests that the predictive ability of the FB model broke down during the recent financial crisis (Cochrane, 2011). Figure 5 plots the GISW performance measures during the crisis period. It is interesting to note that both FB and CP record negative GISW measures and the values are nearly four times larger than the ones recorded over the full sample period across various specifications, especially when time-varying volatility is taken into account. The rolling forecasting scheme seems to mitigate this negative performance – however, only for FB with constant volatility and CP with constant volatility when parameter uncertainty is not taken into account. During this period of high uncertainty, the predictive models did not provide evident economic gains to investors seeking to rebalance their portfolios.

6 Robustness

This section checks the robustness of the baseline results reported in Section 4.2. More specifically, we test whether our results are sensitive to 1) different rolling window sizes, 2) different values of the RRA coefficient, λ, and 3) different values of the decay parameter α used to calculate the rolling sample estimator of the variance-covariance matrix of bond excess returns. We show that our main results are robust to all of these issues.[21]

The first robustness exercise involves the consideration of different window sizes used to carry out the rolling forecasting scheme. Specifically, we consider a rolling window of 240 months. A longer rolling window does not change the conclusions qualitatively. Indeed, the results are quantitatively similar to the ones reported in Table 3. Virtually all of the competing models record SR that are lower than the ones exhibited by the benchmark EH_{CVOL}. The only exception, as in Table 3, is represented by CP_{CVOL} that records (1) SR that are slightly higher than the benchmark and (2) positive but very small performance fees. However, very few of the differences in SR are statistically significant at conventional levels.

As a second robustness test, we consider two alternative values of the RRA coefficient $\lambda = 2, 3$. In all cases, the results are qualitatively and quantitatively similar to the ones reported in Tables 3 and 4. The performance measures tend to increase in absolute value for a lower RRA coefficient. As investors become less risk averse, the evidence against the predictive models strengthens in favor of the EH_{CVOL} benchmark.

The final test considers two alternative values of the decay parameter $\alpha = 0.01, 0.10$. The results show that our baseline findings do not hinge on the selected value of the decay parameter. Indeed, thy are virtually identical to those reported in Tables 3 and 4.

7 Conclusions

This study investigates the economic gains accruing to an investor who exploits the predictability of bond excess returns relative to the no-predictability alternative consistent with the expectations hypothesis. In particular, we quantify how much a risk-averse investor is willing to pay to switch from a dynamic portfolio strategy based on a model with no predictable bond excess returns to a model where the forecasts are based on either forward spreads or the term structure of forward rates.

The results show that the no-predictability benchmark is difficult to beat in economic terms by either of the competing forward-rate models. The evidence in favour of naive models replicates the one already recorded in similar or unrelated contexts (see, among others, Goyal and Welch, 2008; Guidolin and Thornton, 2010 and the references therein). Indeed, the predictive regressions do not record any significant economic value over the no-predictability benchmark. The extent of the underperformance varies across specifications. Generally, it is larger when model parameters are estimated recursively and parameter uncertainty is taken into account. Moreover, the forecasts of the variance-covariance matrix of excess returns computed by a rolling sample estimator generally do not improve upon the performance of the predictive regressions with constant volatility. Importantly, the qualitative conclusions are robust to the sample period as well as the value of the key parameters used in our baseline estimation. We also found that the performance of all predictive models based on forward rates deteriorates over time. Indeed, the relative performance of these models is generally worse in the decade of the 2000s. Overall, our findings confirm that it is very difficult to improve upon a simple naïve benchmark and that the predictability of bond excess returns found in the literature does not necessarily translate into economic gains for investors who rely on forecasts from these models.

Notes

[1]However, not all studies are supportive of the predictive power of forward rates. In fact, Hamburger and Platt (1975), Fama (1984), and Shiller *et al.* (1983) find weak evidence that forward rates predict future spot rates.

[2]The predictability of excess returns recorded in these studies strongly corroborates the well-documented empirical failure of the expectations hypothesis of the term structure of interest rates (Fama, 1984; Stambaugh, 1988; Bekaert *et al.*, 1997, 2001; Sarno *et al.*, 2007), and it is generally assumed to be the consequence of the slow mean reversion of the spot rate toward a time-invariant equilibrium anchor that becomes more evident over longer horizons (Fama, 1984; 2006, and the references therein).

[3]For a comprehensive overview of portfolio choice problems see, Brandt (2010) and the references therein.

[4]This argument is echoed in Dai *et al.* (2004) and Singleton (2006), who show that these predictive regressions are affected by a small-sample bias that causes the R^2 statistics to be substantially higher than their population values.

[5]The evidence of the near unit-root nature of bond yields is strengthened by other studies that record that the slow mean reversion of the spot rate toward a constant is no longer valid after 1986 (Fama, 2006) and its dynamics are better approximated by a mean-reverting process that is anchored to a nonstationary central tendency that stochastically changes over time (Balduzzi *et al.*, 1998). Duffee and Stanton (2008) also show that the high persistence of interest rates has important implications for the preferred method used to estimate term structure models.

[6]Strictly speaking, Ferson *et al.* (2003a,b) assume the regressand does not exhibit high serial correlation but the regressor (predictor) does. However, the spurious regression may occur even without highly autocorrelated regressand if its conditional mean is highly correlated.

[7]It is interesting to note that Cochrane and Piazzesi (2006) also investigate trading rule profits based on the CP predictive model. Their results are supportive of their in-sample evidence; however, the real-time profits are about half of those obtained over the full-sample. Cochrane and Piazzesi (2006, p. 12) point out that 'real trading rules should [...] follow an explicit portfolio maximization problem. They also must incorporate estimates of the conditional variance of returns'. These features are key ingredients of our empirical investigation.

[8]The new indexation in equations (4) and (5) reflects the fact that data are sampled at a monthly frequency while

bond maturities are of one year and above, and hence a multiple of 12 months.

[9] It is instructive to note that the FB regression can be obtained from the CP regression by imposing that $\zeta_1 = \delta_q = -\delta_1$ for $q = n$, $\delta_q = 0$ for $q \neq n$ and $q > 0$.

[10] The estimation problems are strongly exacerbated in the context of bond excess returns, where the high correlations across bond maturities often cause variance-covariance matrices to be near-singular.

[11] In our study we do not implement an analogue parameter uncertainty correction for the conditional volatility model for various reasons. First, it is well known that volatility is highly persistent and therefore predictable. Hence, the approach advocated by Kandel and Stambaugh (1996), which implies a prior of no predictability, is inappropriate. Furthermore, the scope of the paper is to assess the economic value of the predictability of bond excess returns by means of forward rates in the conditional mean. The extensions proposed in this study, which consider the case of predictable bond excess returns volatilities, are included for completeness but they do not represent the main focus of the paper.

[12] See Brandt (2010) and the references therein.

[13] The marginal R_j^2 is defined as the full R^2 from equation (6) minus the R^2 from the equation where the variable j is dropped from the model. For further details, see Connor (1997, p. 50).

[14] A critical aspect of this analysis is that it relies, as in previous studies, on the assumptions of the normality of bond returns and quadratic utility function. Although the quadratic utility assumption is appealing for its tractability properties, it not necessary to justify the use of mean-variance optimization (Della Corte *et al.*, 2009 p. 3501).

[15] Full details of the bootstrap procedure are reported in the appendix to the working paper version of this study (Thornton and Valente, 2012).

[16] In particular, we have estimated over the full sample period a multivariate GARCH(1,1) model using the residuals from the EH model, i.e., where the conditional mean of bond excess returns is equal to a constant. Then we carried out the LM test of conditional correlation by Tse (2000) on the multivariate GARCH model estimated assuming a constant conditional correlation. The result is not reported to save space, but available from the authors upon request.

[17] We have carried out the symmetry tests assuming that the mean equation contains only an intercept and an

intercept and the predictive variables. In both cases, the results lead to the same conclusion.

[18] We thank the anonymous referee for suggesting this to us.

[19] The sensitivity of our baseline results to the choice of the relevant parameters is assessed in the robustness Section 6.

[20] However, a notable exception is represented by Almeida *et al.* (2011) who find, using the information embedded in interest rate options, that the most successful models for predicting excess returns have risk factors with stochastic volatility.

[21] The full set of results is reported in the Appendix to the working paper version of this study (Thornton and Valente, 2012).

References

Almeida, C., J. J. Graveline, and S. Joslin. 2011. Do Interest Rate Options Contain Information About Excess Returns? *Journal of Econometrics* 164:35–44.

Balduzzi, P., S. R. Das, and S. Foresi. 1998. The Central Tendency: A Second Factor in Bond Yields. *Review of Economics and Statistics* 80:62–72.

Barberis, N. 2000. Investing for the Long-run when Returns are Predictable. *Journal of Finance* 55:225–64.

Barillas, F. 2010. Can We Exploit Predictability in Bond Markets. Mimeo, New York University.

Bawens, L., S. Laurent, and J. V. K. Rombouts. 2006. Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics* 21:79–109.

Bekaert, G., R. J. Hodrick, and D. A. Marshall. 1997. On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates. *Journal of Financial Economics* 44:309–48.

Bekaert, G., and R. J. Hodrick. 2001. Expectations Hypothesis Tests. *Journal of Finance* 56:1357–94.

Boudoukh, J., M. Richardson, and R. F. Whitelaw. 2008. The Myth of Long-Horizon Predictability. *Review of Financial Studies* 21:1577–605.

Brandt, M. 2010. Portfolio Choice Problems. In Y. Ait-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics,* vol 1, pp. 269-336. North Holland: Elsevier.

Campbell, J. Y., and M. Yogo. 2006. Efficiency Tests of Stock Return Predictability. *Journal of Financial Economics* 81:27–60.

Cheridito, P., D. Filipovic, and R. L. Kimmel. 2007. Market Price of Risk Specifications for Affine Models: Theory and Evidence. *Journal of Financial Economics* 83:123–70.

Cieslak, A., and P. Povala. 2011. Understanding Bond Risk Premia. Mimeo, Northwestern University and University of Lugano.

Clark, T. E., and M. W. McCracken. 2008. Forecasting with Small Macroeconomic VARs in the Presence of Instabilities. In D. E. Rapach and M. E. Wohar (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty, Fronters of Economics and Globalization*, vol. 3, pp. 93-147. U.K.: Emerald Group Publishing Limited.

Cochrane, J. H., and M. Piazzesi. 2005. Bond Risk Premia. *American Economic Review* 95:138–60.

Cochrane, J. H., and M. Piazzesi. 2006. Appendix to Bond Risk Premia. Mimeo, University of Chicago.

Cochrane, J. H., and M. Piazzesi. 2008. Decomposing the Yield Curve. Mimeo, University of Chicago and University of Stanford.

Cochrane, J. H. 2011. http://faculty.chicagobooth.edu/john.cochrane/research/index.asp

Connor, G. 1997. Sensible Return Forecasting for Portfolio Management. *Financial Analysts Journal*, 53:44–51.

D'Agostino, A., D. Giannone, and P. Surico. 2006. (Un)Predictability and Macroeconomic Stability. ECP Working Paper No. 605.

Dai, Q., K. Singleton, and W. Yang. 2004. Predictability of Bond Excess Returns and Affine Term Structure Models. Mimeo, New York University and University of Stanford.

Della Corte, P., L. Sarno, and D. L. Thornton. 2008. The Expectations Hypothesis of the Term Structure of Very Short-Term Rates: Statistical Tests and Economic Value. *Journal of Financial Economics* 89:158–74.

Della Corte, P., L. Sarno, and I. Tsiakas. 2009. An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies* 22:3491–530.

de Goeij, P., and W. Marquering. 2006. Macroeconomic Announcements and Asymmetric Volatility in Bond Returns. *Journal of Banking and Finance* 30:2659–80.

Duffee, G. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* 52:405–43.

Duffee, G. 2010. Sharpe Ratios in Term Structure Models. Mimeo, John Hopkins University.

Duffee, G. 2011. Information in (and not in) the Term Structure. *Review of Financial Studies* 24:2895–934.

Duffee, G., and R. H. Stanton. 2008. Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation. *Journal of Financial Econometrics* 6:108–42.

Engle, R. F., and V. Ng. 1993. Time-varying Volatility and the Dynamic Behaviour of the Term Structure. *Journal of Money, Credit and Banking* 25:336–49.

Fama, E. F. 1984. The Information in the Term Structure. *Journal of Financial Economics* 13:509–28.

Fama, E. F. 2006. The Behaviour of Interest Rates. *Review of Financial Studies* 19:359–79.

Fama, E. F., and R. R. Bliss. 1987. The Information in Long-Maturity Forward Rates. *American Economic Review* 77:680–92.

Ferreira, M. A., and P. Santa Clara. 2011. Forecasting Stock Market Returns: The Sum of the Parts is More than the Whole. *Journal of Financial Economics* 100:514–37.

Ferson, W. E., S. Sarkissian, and T. T. Simin. 2003a. Spurious Regressions in Financial Economics. *Journal of Finance* 58:1393–413.

Ferson, W. E., S. Sarkissian, and T. T. Simin. 2003b. Is Stock Return Predictability Spurious? *Journal of Investment Management* 1:1–10.

Fleming, J., C. Kirby, and B. Ostdiek. 2001. The Economic Value of Volatility Timing. *Journal of Finance* 56:329–52.

Fleming, J., C. Kirby, and B. Ostdiek. 2003. The Economic Value of Volatility Timing Using Realized Volatility. *Journal of Financial Economics* 67:473–509.

Foster, D. P., and D. B. Nelson. 1996. Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica* 64:139–74.

Glosten, L. R., R. Jagannathan, and E. E. Runkle. 1993. On the Relationship between the Expected Value and the Volatility of Nominal Excess Returns on Stocks. *Journal of Finance* 48:1779–801.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch. 2007. Portfolio Performance Manipulation and Manipulation-proof Performance Measures. *Review of Financial Studies* 20:1503–46.

Goyal, A. and I. Welch. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Predictions, *Review of Financial Studies* 21:1455–508.

Gray, S. F. 1996. Modelling the Conditional Distribution of Interest Rates as a Regime Switching Process. *Journal of Financial Economics* 42:27–62.

Guidolin, M., D. L. Thornton. 2010. Predictions of Short-Term Rates and the Expectations Hypothesis. Federal Reserve Bank of St. Louis Working Paper 2010-013A.

Hamburger, M. J., and E. N. Platt. 1975. The Expectations Hypothesis and the Efficiency of the Treasury Bill Market. *Review of Economics and Statistics* 57:190–99.

Han, Y. 2006. Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *Review of Financial Studies* 19:237–71.

Inoue, A., and L. Kilian. 2004. In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews* 23:371–402.

Inoue, A., and L. Kilian. 2006. On the Selection of Forecasting Models. *Journal of Econometrics* 130:273–306.

Jondeau, E., and M. Rockinger. 2008. The Economic Value of Distributional Timing. Mimeo, Swiss Finance Institute.

Kandel, S., and R. F. Stambaugh. 1996. On the Predictability of Stock Returns: An Asset Allocation Perspective. *Journal of Finance* 51:385–424.

Kirby, C. 1997. Measuring the Predictable Variation in Stock and Bond Returns. *Review of Financial Studies* 10:579-630.

Ledoit, O., and M. Wolf. 2008. Robust Performance Hypothesis Testing with the Sharpe Ratio. *Journal of Empirical Finance* 15:850–59.

Leitch, G., and J. E. Tanner. 1991. Economic Forecast Evaluation: Profits Versus the Conventional Measures. *American Economic Review* 81:580–90.

Ludvigson, S., and S. Ng. 2009. Macro Factors in Bond Risk Premia. *Review of Financial Studies* 22:5027–67.

Marquering, W., and M. Verbeek. 2004. The Economic Value of Predicting Stock Index Returns and Volatility. *Journal of Financial and Quantitative Analysis* 39:407–29.

Piazzesi, M., and Schneider, M. 2011. Trends and Cycle in Bond Premia. Mimeo, Stanford University.

Politis, D. N., and J. P. Romano. 1992. A Circular Block-resampling Procedure for Stationary Data. In R. LePage and L. Billard (eds.), *Exploring the Limits of Bootstrap*, pp. 263-70. New York: Wiley.

Radwanski, J. F. 2010. Understanding Bond Risk Premia Uncovered by the Term Structure. Mimeo, Vienna University of Economics and Business.

Rudebush, G. D., B. Sack, and E. T. Swanson. 2007. Macroeconomic Implications of Changes in the Term Premium. *Federal Reserve Bank of St. Louis Review* 89:241–69.

Sarno, L., D. L. Thornton, and G. Valente. 2007. The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields. *Journal of Financial and Quantitative Analysis* 42:81–100.

Shiller, R. J., J. Y. Campbell, and K. L. Schoenholtz. 1983. Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates. *Brookings Papers on Economic Activity* 1:173–217.

Sims, C. A., and T. Zha. 2006. Were There Regime Switches in US Monetary Policy. *American Economic Review* 96:54–80.

Singleton, K. J. 2006. *Empirical Dynamic Asset Pricing*. Princeton, NJ: Princeton University Press.

Stambaugh, R. F. 1988. The Information in Forward Rates: Implications for Models of the Term Structure. *Journal of Financial Economics* 21:41–70.

Thornton, D. L. 2006. When Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell Us. *Journal of Money, Credit and Banking* 38:2039–71.

Thornton, D. L. 2010. The Unusual Behavior of the Federal Funds Rate and Treasury Yields: A Conundrum or an Instance of Goodhart's Law? Federal Reserve Bank of St. Louis Working Paper 2007-039C.

Thornton, D. L., and G. Valente. 2012. Out of sample Predictions of Bond Excess Returns and Forward Rates: An Asset Allocation Perspective. Federal Reserve Bank of St. Louis Working Paper 2010-034B.

Tse, Y. K. 2000. A Test for Constant Correlations in a Multivariate GARCH Model. *Journal of Econometrics* 98:107–27.

Valkanov, R. 2003. Long-Horizon Regressions: Theoretical Results and Applications. *Journal of Finanical Economics* 68:201–32.

Vayanos, D., and P. O. Weill. 2008. A Search-Based Theory of the On-the-run Phenomenon. *Journal of Finance* 63:1361–98.

Wei, M., and J. H. Wright. 2010. Reverse Regressions and Long-Horizon Forecasting. Mimeo, Federal Reserve Board and Johns Hopkins University.

West, K. D., H. J. Edison, and D. Cho. 1993. A Utility-Based Comparison of Some Models of Exchange Rate Volatility. *Journal of International Economics* 35:23–45.

Zakoïan, J. M. 1994. Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control* 18:931–55.

Table 1. Summary Statistics

The Table reports the descriptive statistics for bond excess returns (Panel A) and absolute bond excess returns (Panel B) computed over the different maturities, n. The data sample ranges from June 1952 until December 2010 for a sample size of 703 monthly observations. *, **, *** denote statistical significance at 10%, 5% and 1% and statistical significance is evaluated using autocorrelation and heteroskedasticity-consistent standard errors (Newey and West, 1987). Mean and Std Dev are reported in decimals per annum (i.e. 0.01 = 1 annual percentage point).

Panel A) Bond excess returns

	$n=2$	$n=3$	$n=4$	$n=5$
Mean $rx_t^{(n)}$	0.004***	0.007***	0.009***	0.010***
Std Dev $rx_t^{(n)}$	0.017***	0.031***	0.043***	0.052***
$Corr(rx_t^{(n)}, rx_{t-1}^{(n)})$	0.929***	0.933***	0.933***	0.923***
$Corr(rx_t^{(n)}, rx_{t-12}^{(n)})$	0.189***	0.135***	0.105***	0.071**

Panel B) Absolute bond excess returns

	$n=2$	$n=3$	$n=4$	$n=5$
Mean $\left\|rx_t^{(n)}\right\|$	0.013***	0.024***	0.033***	0.040***
Std Dev $\left\|rx_t^{(n)}\right\|$	0.011***	0.020***	0.028***	0.034***
$Corr(\left\|rx_t^{(n)}\right\|, \left\|rx_{t-1}^{(n)}\right\|)$	0.860***	0.867***	0.865***	0.847***
$Corr(\left\|rx_t^{(n)}\right\|, \left\|rx_{t-12}^{(n)}\right\|)$	0.130***	0.178***	0.206***	0.214***

Table 2. In-sample Estimates

The table reports the estimates of the no predictability benchmark consistent with the expectations hypothesis (Panel A) and the Fama and Bliss (1987) and Cochrane and Piazzesi (2005) predictive regressions (Panels B,C). All equations are estimated assuming a constant conditional variance of excess returns innovations. Equations in Panels A), B), and C) are estimated over two sample periods: June 1952 - December 2010 and June 1964 - December 2003. n denotes the maturity of forward rates and bond excess returns expressed in years. $\log L$ denotes the log-likelihood value of the regressions and v is the estimated constant conditional volatility of excess returns innovations. Values in parenthesis are asymptotic standard errors computed using least-square estimators. *, **, *** denote statistical significance at 10%, 5% and 1% level. $\overline{R}^2$ denotes the in-sample adjusted coefficient of determination. See also notes to Table 1.

Panel A) No predictability

	$n=2$		$n=3$		$n=4$		$n=5$	
				1952-2010				
β_0	0.004***	(0.001)	0.007***	(0.001)	0.009***	(0.001)	0.010***	(0.002)
v	0.017		0.031		0.043		0.052	
$\log L$	1384.22		1418.81		1193.13		1051.23	
				1964-2003				
β_0	0.005***	(0.001)	0.008***	(0.001)	0.010***	(0.002)	0.010***	(0.002)
v	0.019		0.034		0.048		0.059	
$\log L$	1219.57		929.34		772.90		675.05	

Panel B) Fama and Bliss (1987)

	$n=2$		$n=3$		$n=4$		$n=5$	
				1952-2010				
β_0	0.001*	(0.001)	0.001	(0.001)	-0.001	(0.001)	0.001	(0.002)
β_1	0.759***	(0.089)	1.001***	(0.113)	1.272***	(0.127)	0.995***	(0.147)
v	0.016		0.029		0.040		0.051	
$\log L$	1868.30		1455.97		1239.60		1073.43	
$\overline{R}^2$	0.09		0.10		0.12		0.06	
				1964-2003				
β_0	0.001	(0.001)	-0.001	(0.001)	-0.004	(0.001)	-0.008	(0.003)
β_1	0.975***	(0.104)	1.301***	(0.133)	1.526***	(0.152)	1.181***	(0.183)
v	0.017		0.031		0.044		0.057	
$\log L$	1259.57		973.05		818.62		694.95	
$\overline{R}^2$	0.15		0.16		0.17		0.07	

Panel C) Cochrane and Piazzesi (2005)

	$n=2$		$n=3$		$n=4$		$n=5$	
				1952-2010				
β_0	-0.007***	(0.001)	-0.012***	(0.003)	-0.018***	(0.004)	-0.024v	(0.005)
β_1	-0.649***	(0.105)	-1.118***	(0.192)	-1.633***	(0.261)	-2.035***	(0.325)
β_2	0.378*	(0.416)	0.049	(0.371)	0.130	(0.504)	0.255	(0.628)
β_3	0.416***	(0.168)	1.553***	(0.305)	1.543***	(0.415)	1.550***	(0.517)
β_4	0.328***	(0.127)	0.514**	(0.231)	1.458***	(0.314)	1.491***	(0.391)
β_5	-0.329***	(0.099)	-0.787***	(0.181)	-1.218***	(0.246)	-0.928***	(0.306)
v	0.015		0.028		0.038		0.047	
$\log L$	1900.05		1486.62		1274.48		1123.40	
$\overline{R}^2$	0.16		0.17		0.20		0.18	
				1964-2003				
β_0	-0.015***	(0.002)	-0.026***	(0.004)	-0.037***	(0.006)	-0.047***	(0.008)
β_1	-0.938***	(0.124)	-1.711***	(0.223)	-2.481***	(0.302)	-3.101***	(0.378)
β_2	0.470*	(0.254)	0.326	(0.458)	0.598	(0.620)	0.899	(0.775)
β_3	1.173***	(0.213)	3.010***	(0.385)	3.532***	(0.520)	4.023***	(0.651)
β_4	0.334**	(0.156)	0.461	(0.282)	1.385***	(0.381)	1.374***	(0.477)
β_5	-0.819***	(0.129)	-1.745***	(0.233)	-2.583***	(0.316)	-2.649***	(0.395)
v	0.015		0.028		0.038		0.048	
$\log L$	1310.81		1028.14		883.28		776.01	
$\overline{R}^2$	0.30		0.33		0.36		0.33	

Table 3. Out-of-sample Performance Assessment: Performance Measures

The Table reports summary statistics of the returns from alternative portfolios constructed using the out-of-sample forecasts from the benchmark of no-predictability model of bond excess returns with constant volatility and the other competing models. FB, CP and EH denote Fama and Bliss (1987), Cochrane and Piazzesi (2005) and the no predictability benchmark respectively. Recursive and Rolling Estimation denote forecasts that are generated using all past observations available up to the time of the forecast and a window of past 120 months, respectively. Parameter Uncertainty denotes the case when the expectations hypothesis prior is imposed to define the distribution of parameter estimates (Kandel and Stambaugh, 1996; Connor, 1997). The subscripts CVOL and TVOL denote models whose conditional volatility is estimated as a constant and by means a rolling sample estimator (Foster and Nelson, 1996; Fleming et al., 2001; 2003), respectively. SR denote Sharpe ratios achieved by each strategy and computed as the ratio of the sample average to the sample standard deviation of portfolios' excess returns. Values in brackets are p-values of the null hypothesis that the SR of the model is equal to the one of EH_{CVOL} (Ledoit and Wolf, 2008). The p-values are computed using $V = 1,000$ bootstrap replications. $GISW$ is a variant of the Goetzmann *et al.* (2007) manipulation-proof measure of performance computed as portfolios' premium return above the benchmark after adjusting for risk. The asset allocations for all models are carried out using two annual targets of portfolio excess returns: $\zeta_p^* = 0.01, 0.02$. Time-varying variance-covariance matrices of excess returns are estimated using a decay parameter $\alpha = 0.05$. GISW are computed using a Relative Risk Aversion (RRA) coefficient $\lambda = 5$. The out-of-sample forecasting exercise runs from January 1970 through December 2010. SR are reported as annualized and GISW are reported in decimals per annum (i.e. 0.01 = 1 annual percentage point).

Panel A) Recursive Estimation, No Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\zeta_p^* = 0.01$			
SR	0.464	0.211	0.272	-0.082	0.100	0.187
		[0.01]	[0.07]	[$<$ 0.01]	[0.14]	[0.17]
$GISW$	–	-0.013	-0.010	-0.022	-0.016	-0.014
			$\zeta_p^* = 0.02$			
SR	0.436	0.250	0.254	-0.049	0.074	0.213
		[0.02]	[0.21]	[0.02]	[0.13]	[0.25]
$GISW$	–	-0.011	-0.012	-0.022	-0.018	-0.012

Panel B) Recursive Estimation, Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\zeta_p^* = 0.01$			
SR	0.464	0.202	0.317	–	0.035	0.169
		[< 0.01]	[0.03]	–	[0.25]	[0.10]
$GISW$	–	-0.014	-0.012	–	-0.018	-0.014
			$\zeta_p^* = 0.02$			
SR	0.436	0.235	0.261	–	0.024	0.216
		[0.01]	[0.25]	–	[0.10]	[0.14]
$GISW$	–	-0.011	-0.011	–	-0.020	-0.012

Panel C) Rolling Estimation, No Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\zeta_p^* = 0.01$			
SR	0.304	0.261	0.304	0.144	0.095	0.189
		[0.73]	[0.99]	[0.35]	[0.21]	[0.47]
$GISW$	–	-0.002	0.000	-0.005	-0.008	-0.004
			$\zeta_p^* = 0.02$			
SR	0.355	0.205	0.279	0.128	0.076	0.174
		[0.12]	[0.36]	[0.19]	[0.04]	[0.25]
$GISW$	–	-0.006	-0.002	-0.008	-0.014	-0.007

Panel D) Rolling Estimation, Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\zeta_p^* = 0.01$			
SR	0.304	0.288	0.157	–	0.105	0.139
		[0.90]	[0.31]	–	[0.25]	[0.32]
$GISW$	–	-0.001	-0.005	–	-0.006	-0.005
			$\zeta_p^* = 0.02$			
SR	0.355	0.228	0.156	–	0.081	0.130
		[0.14]	[0.19]	–	[0.02]	[0.16]
$GISW$	–	-0.005	-0.008	–	-0.014	-0.008

Table 4. Out-of-sample Performance Assessment: Performance Fees

The Table reports out-of-sample performance fees Φ based on out-of-sample forecasts of mean and variance from competing models against the benchmark of no bond excess returns predictability with constant volatility. The measures are computed for two levels of target portfolio excess returns $\zeta_p^* = 0.01, 0.02$. The performance fees denote the amount the investor with quadratic utility function and a Relative Risk Aversion (RRA) coefficient $\lambda = 5$ would be willing to pay for switching from the model with no excess return predictability and constant volatility to the alternative model. Performance fees are reported in decimals per annum (i.e. 0.01 = 1 annual percentage point). See also notes to Table 3.

Panel A) Recursive Estimation, No Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			Performance fees Φ		
$\zeta_p^* = 0.01$	-0.010	-0.009	-0.022	-0.016	-0.010
$\zeta_p^* = 0.02$	-0.008	-0.012	-0.022	-0.018	-0.009

Panel B) Recursive Estimation, Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			Performance fees Φ		
$\zeta_p^* = 0.01$	-0.014	-0.007	–	-0.018	-0.011
$\zeta_p^* = 0.02$	-0.011	-0.005	–	-0.020	-0.008

Panel C) Rolling Estimation, No Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			Performance fees Φ		
$\zeta_p^* = 0.01$	-0.001	-0.001	-0.005	-0.008	-0.003
$\zeta_p^* = 0.02$	-0.006	-0.004	-0.008	-0.014	-0.007

Panel D) Rolling Estimation, Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			Performance fees Φ		
$\zeta_p^* = 0.01$	-0.001	-0.004	–	-0.005	-0.004
$\zeta_p^* = 0.02$	-0.005	-0.005	–	-0.011	-0.007

Table 5. Sub-sample Analysis

The Table reports out-of-sample the Sharpe ratios based on out-of-sample forecasts of mean and variance from competing models and the benchmark of bond excess returns no-predictability with constant volatility. The measures are computed for a level of annual target of portfolio excess returns $\zeta_p^* = 0.01$ over the four different subperiods. Values in brackets are p-values of the null hypothesis that the SR of the model is equal to the one of EH_{CVOL} (Ledoit and Wolf, 2008). See also notes to Tables 3 and 4.

Panel A) Recursive Estimation, No Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
1970-1979	-0.137	-0.045	-0.019	-0.035	-0.041	0.165
		[0.78]	[0.67]	[0.78]	[0.77]	[0.35]
1980-1989	0.385	0.199	0.219	-0.067	0.019	0.196
		[0.42]	[0.28]	[0.32]	[0.23]	[0.38]
1990-1999	0.541	0.371	0.405	-0.033	0.059	-0.016
		[0.64]	[0.40]	[0.41]	[0.21]	[0.15]
2000-2010	0.856	0.460	0.426	-0.236	0.404	0.441
		[0.13]	[0.09]	[0.12]	[0.11]	[0.16]

Panel B) Recursive Estimation, Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
1970-1979	-0.137	-0.090	0.053	–	-0.072	0.190
		[0.85]	[0.56]	–	[0.82]	[0.26]
1980-1989	0.385	0.204	-0.026	–	-0.142	0.043
		[0.42]	[0.10]	–	[0.18]	[0.31]
1990-1999	0.541	0.365	0.267	–	0.061	0.140
		[0.66]	[0.09]	–	[0.30]	[0.06]
2000-2010	0.856	0.465	0.671	–	0.396	0.354
		[0.13]	[0.38]	–	[0.07]	[0.21]

Panel C) Rolling Estimation, No Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
1970-1979	-0.243	-0.126	-0.103	0.226	0.124	0.082
		[0.59]	[0.42]	[0.15]	[0.12]	[0.24]
1980-1989	0.481	0.241	0.291	0.009	-0.147	0.366
		[0.06]	[0.30]	[0.25]	[0.02]	[0.77]
1990-1999	0.737	0.462	0.525	0.291	0.333	-0.026
		[0.17]	[0.18]	[0.24]	[0.62]	[0.19]
2000-2010	0.709	0.455	0.462	0.155	0.359	0.285
		[0.27]	[0.18]	[0.10]	[0.25]	[0.13]

Panel D) Rolling Estimation, Parameter Uncertainty

	EH_{CVOL}	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
1970-1979	-0.243	-0.129	0.304	–	0.040	0.177
		[0.64]	[0.04]	–	[0.32]	[0.13]
1980-1989	0.481	0.263	-0.054	–	-0.132	0.172
		[0.06]	[0.01]	–	[0.04]	[0.46]
1990-1999	0.737	0.519	0.021	–	0.403	0.158
		[0.69]	[0.06]	–	[0.69]	[0.29]
2000-2010	0.709	0.452	0.253	–	0.368	0.049
		[0.20]	[0.11]	–	[0.23]	[0.08]

Legends to the Figures

Figure 1. Sub-sample Correlations: Excess Returns

The Figure shows the correlation coefficient computed for all excess return pairs over the sample period June 1952 - December 2010. The correlation coefficients are computed using monthly data sampled over each decade. $corr(er\_i, er\_j)$ denotes correlation coefficients computed between excess returns with maturity i and j, respectively.

Figure 2. Sub-sample Correlations: Absolute Excess Returns

The Figure shows the correlation coefficient computed for all absolute excess return pairs over the sample period June 1952 - December 2010. The correlation coefficients are computed using monthly data sampled over each decade. $corr(|er\_i|, |er\_j|)$ denotes correlation coefficients computed between absolute excess returns with maturity i and j, respectively.

Figure 3. Sub-sample Analysis: GISW Performance Measure

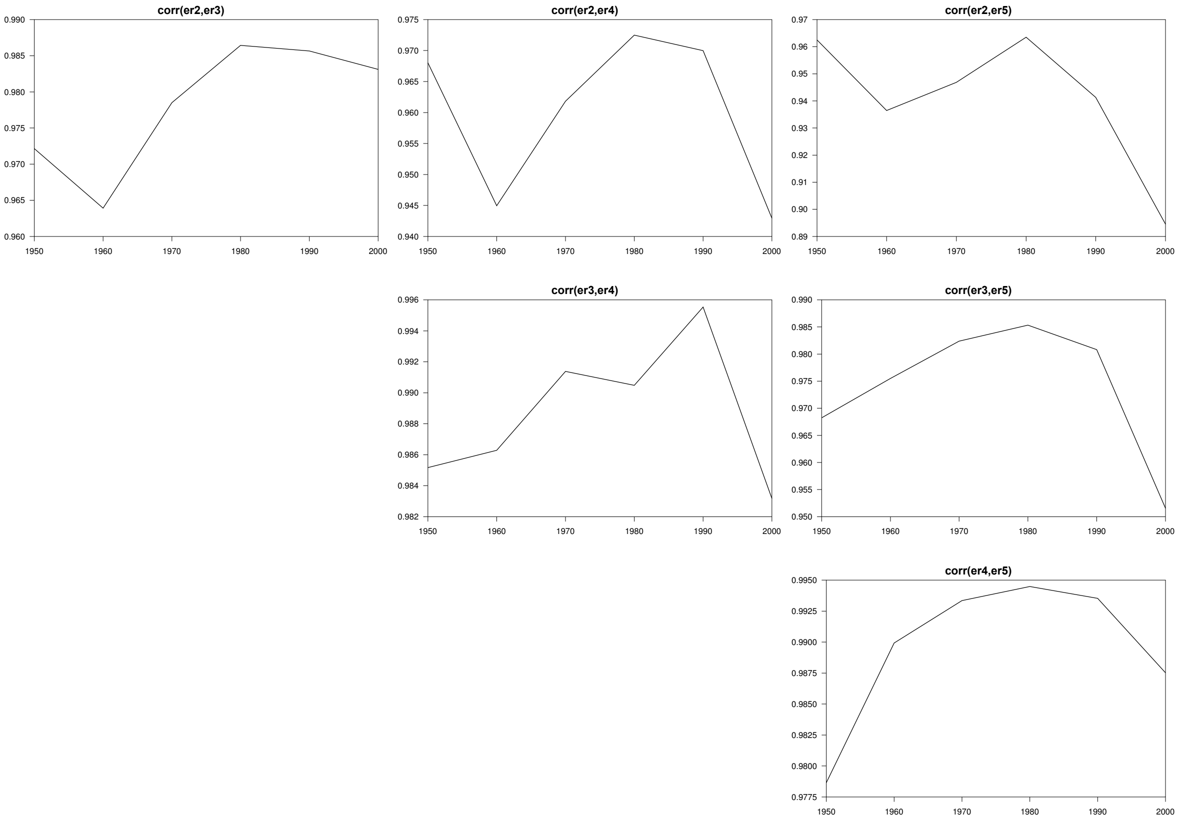
The Figure shows GISW measures of performance computed for each decade in the out-of-sample forecasting period (January 1970 - December 2010). For each sub-sample GISW measures are computed for all predictive models and all scenarios, as discussed in Sections 3 and 4 of the main text. The solid square and the relative vertical lines in each of the two graphs denote the average and standard deviation of GISW measures computed across all specification. GISW measures are reported in decimals per annum (i.e. 0.01 = 1 annual percentage point).

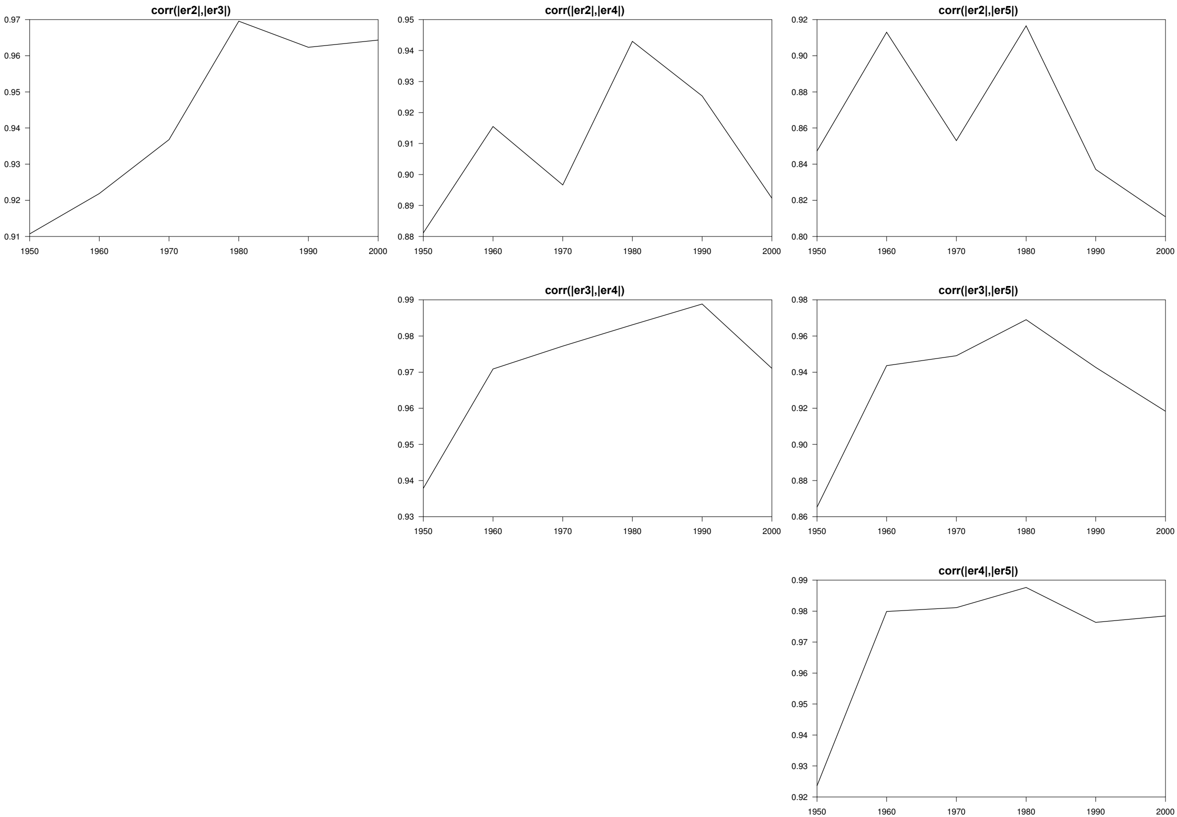
Figure 4. Sub-sample Correlations: Performance Fees, Φ

The Figure shows performance fees Φ computed for each decade in the out-of-sample forecasting period (January 1970 - December 2010). For each sub-sample Φ are computed for all predictive models and all scenarios, as discussed in Sections 3 and 4 of the main text. The solid square and the relative vertical lines in each of the two graphs denote the average and standard deviation of Φ computed across all specification. Φ are reported in decimals per annum (i.e. 0.01 = 1 annual percentage point).

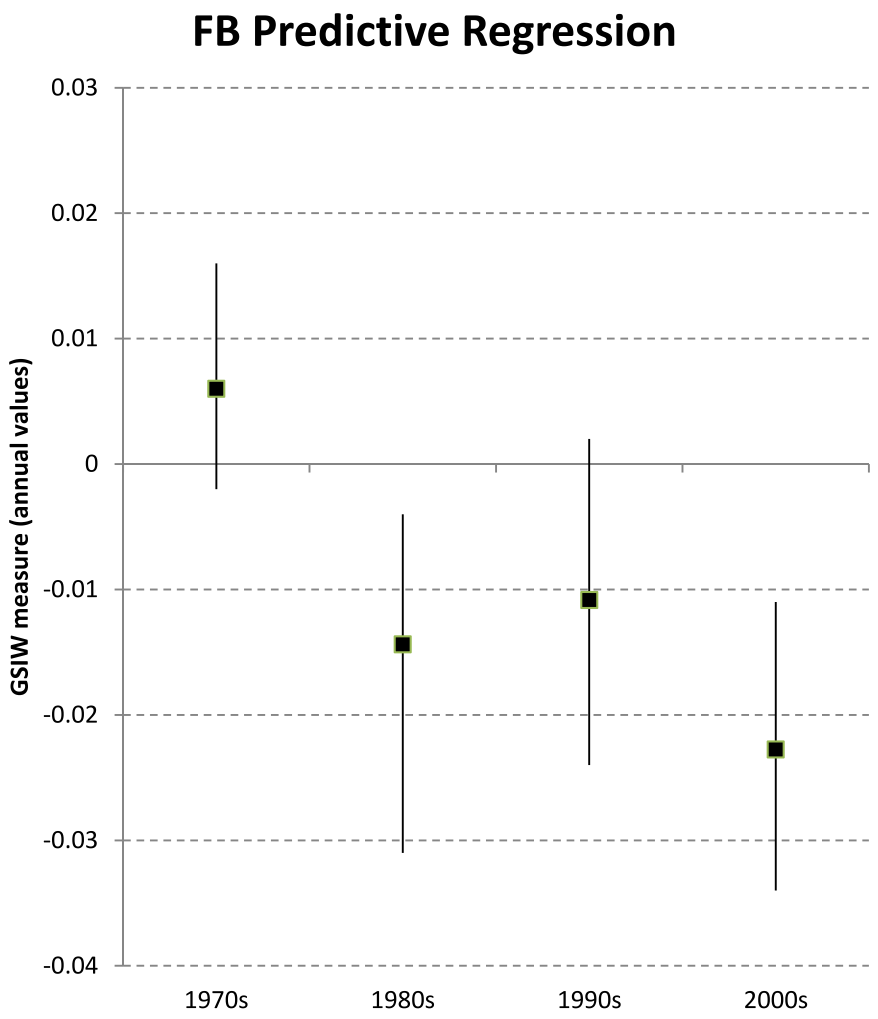
Figure 5. Predictive Performance During the Crisis

The Figure shows GISW measures computed for all predictive models and all scenarios over the period January 2007 through December 2009. Rec and Roll denote recursive and rolling forecasting schemes, respectively, as discussed in Section 4.2 of the main text. PU and NPU denote estimations with and without correction for parameter uncertainty, respectively, as discussed in Section 3.2. See also notes to Figure 3 and Tables 3 and 4.

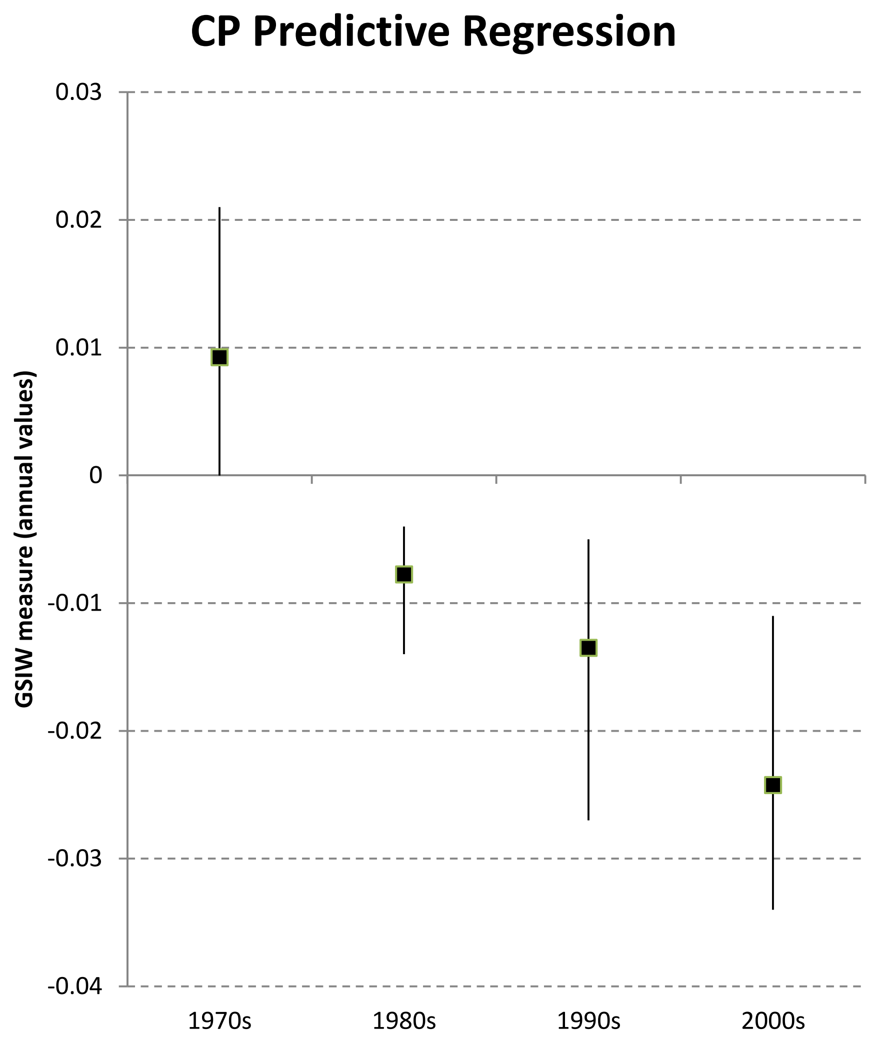

corr(er2,er3)
0.990
0.985
0.980
0.975
0.970
0.965
0.960
corr(er2,er4)
0.975
0.970
0.965
0.960
0.955
0.950
0.945
0.940
corr(er2,er5)
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90
0.89
corr(er3,er4)
0.996
0.994
0.992
0.990
0.988
0.986
0.984
0.982
corr(er3,er5)
0.990
0.985
0.980
0.975
0.970
0.965
0.960
0.955
0.950
corr(er4,er5)
0.9950
0.9925
0.9900
0.9875
0.9850
0.9825
0.9800
0.9775
1950
1960
1970
1980
1990
2000



corr(|er2|,|er3|)
0.97
0.96
0.95
0.94
0.93
0.92
0.91
1950
1960
1970
1980
1990
2000
corr(|er2|,|er4|)
0.95
0.94
0.93
0.92
0.91
0.90
0.89
0.88
1950
1960
1970
1980
1990
2000
corr(|er2|,|er5|)
0.92
0.90
0.88
0.86
0.84
0.82
0.80
1950
1960
1970
1980
1990
2000
corr(|er3|,|er4|)
0.99
0.98
0.97
0.96
0.95
0.94
0.93
1950
1960
1970
1980
1990
2000
corr(|er3|,|er5|)
0.98
0.96
0.94
0.92
0.90
0.88
0.86
1950
1960
1970
1980
1990
2000
corr(|er4|,|er5|)
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
1950
1960
1970
1980
1990
2000

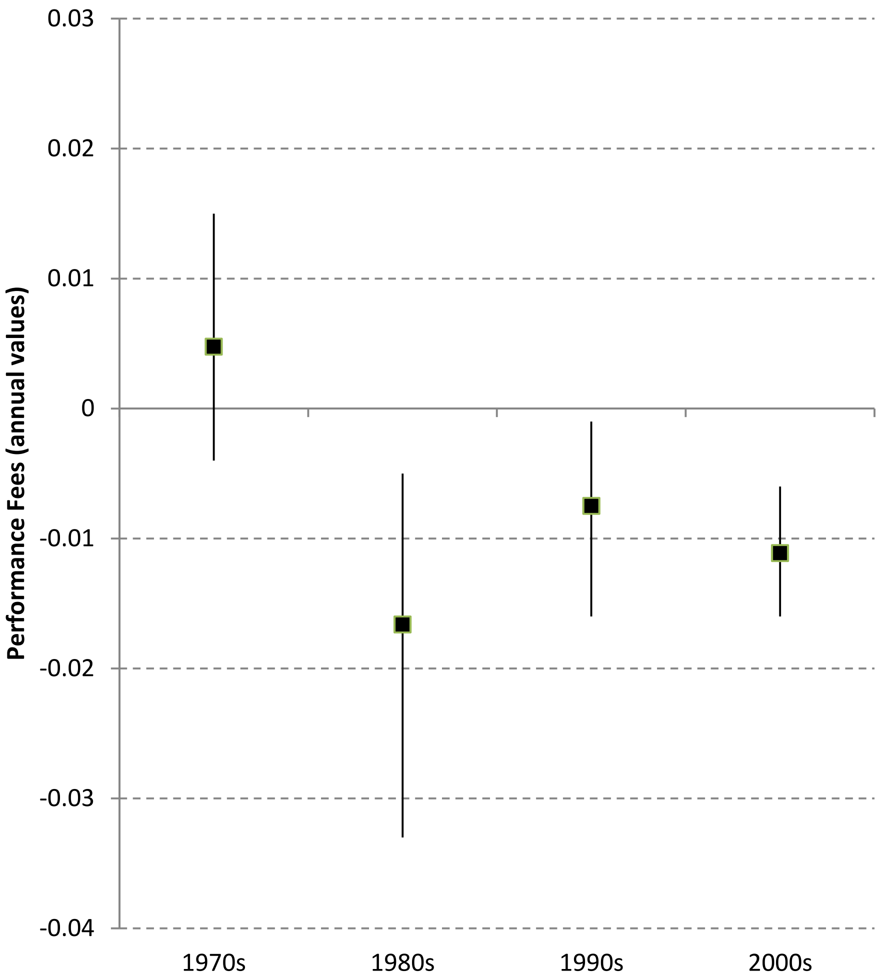


FB Predictive Regression
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
GSIW measure (annual values)
1970s
1980s
1990s
2000s

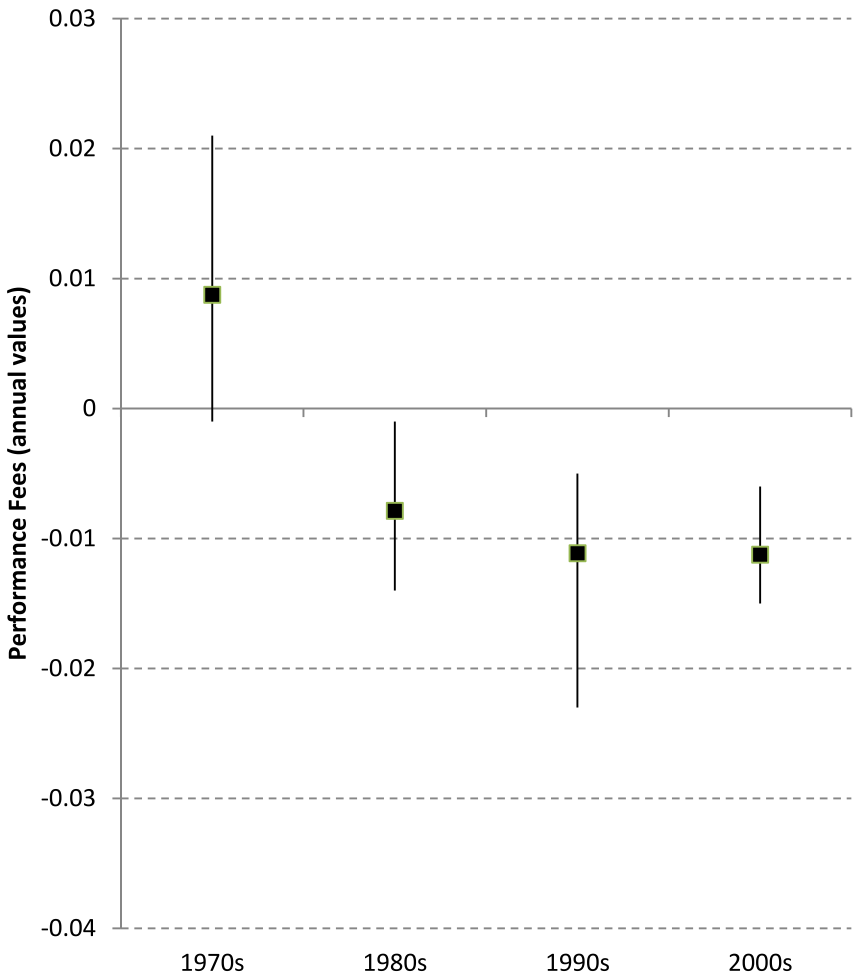


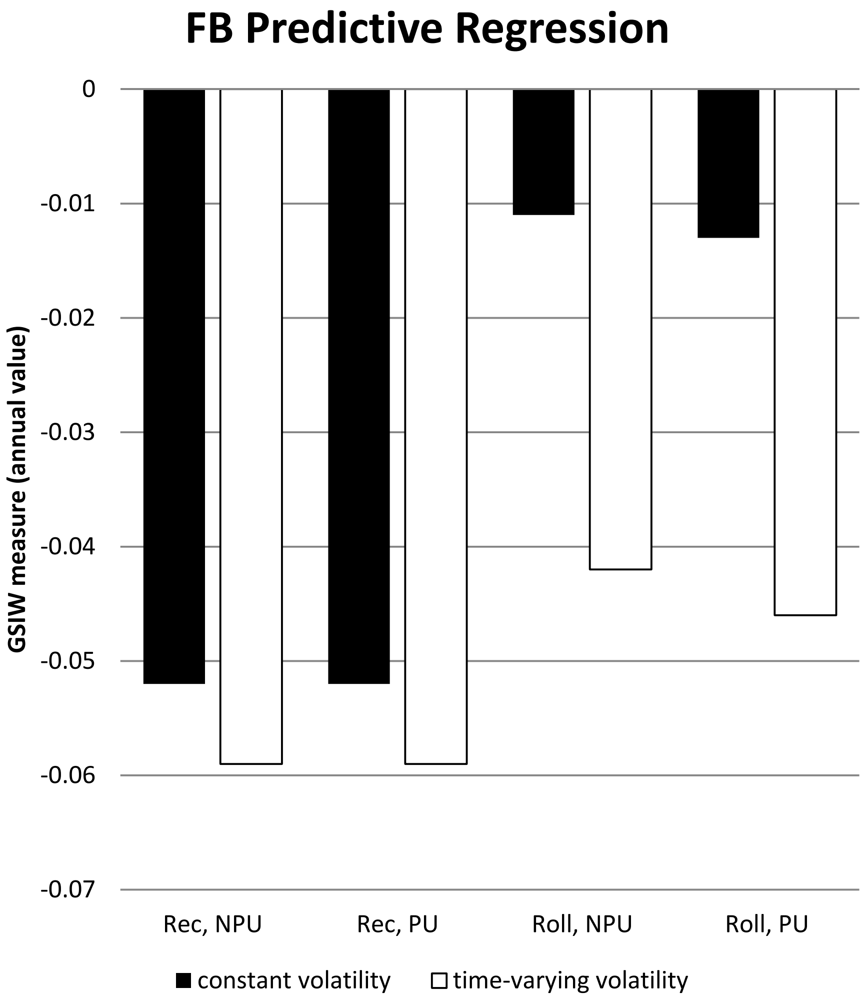
CP Predictive Regression
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
GSIW measure (annual values)
1970s
1980s
1990s
2000s

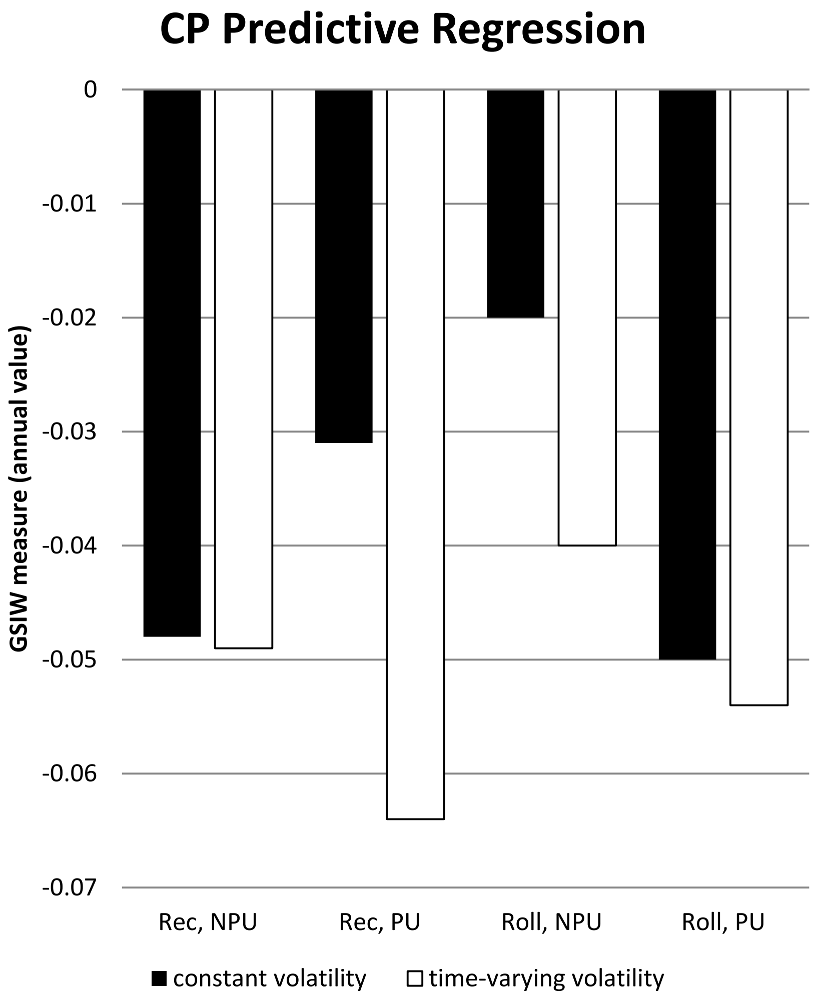
FB Predictive Regression



CP Predictive Regression




FB Predictive Regression
GSIW measure (annual value)
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
Rec, NPU
Rec, PU
Roll, NPU
Roll, PU
constant volatility
time-varying volatility


CP Predictive Regression
GSIW measure (annual value)
0
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
Rec, NPU
Rec, PU
Roll, NPU
Roll, PU
constant volatility
time-varying volatility

Appendix to Out-of-Sample Predictions of Bond Excess Returns and Forward Rates: An Asset Allocation Perspective

This version: February 2012

A Bootstrap Test for the Equality of SR

We employ the boostrap procedure introduced by Ledoit and Wolf (2008) to test for the null hypothesis that the difference between the SR obtained from portfolio returns based on the forecast of a given predictive model $\mathcal{F} = FB, CP$ is equal to the one implied by the forecasts of the benchmark EH$_{CVOL}$. More specifically, given the returns from the two portfolios $r_t^{\mathcal{F}}$ and $r_t^{EH_{CVOL}}$over the forecasting period $t = 1, ..., T$, we compute the two sample means, $m^{\mathcal{F}}, m^{EH_{CVOL}}$ and the two uncentered second moments $s^{\mathcal{F}} = E\left[\left(r_t^{\mathcal{F}}\right)^2\right]$, $s^{EH_{CVOL}} = E\left[\left(r_t^{EH_{CVOL}}\right)^2\right]$. Let $\upsilon = \left[\begin{array}{cccc} m^{EH_{CVOL}} & m^{\mathcal{F}} & s^{EH_{CVOL}} & s^{\mathcal{F}} \end{array}\right]'$ and define

$$\Delta = f(\upsilon) = \frac{m^{EH_{CVOL}}}{\sqrt{s^{EH_{CVOL}} - m^{EH_{CVOL}}}} - \frac{m^{\mathcal{F}}}{\sqrt{s^{\mathcal{F}} - m^{\mathcal{F}}}}.$$

where $\widehat{\Delta} = f(\widehat{\upsilon})$. Ledoit and Wolf (2008) propose to test $H_0 : \Delta = 0$ by inverting a bootstrap confidence interval (with nominal level 1-p) for Δ. If this interval does not contain zero, then H_0 is rejected at the nominal level p. The null hypothesis is tested by bootstrapping the original series in order to obtain the estimate of standard error of Δ, denoted as $\varsigma\left(\widehat{\Delta}^*\right)$. Given that our portfolio returns are serially correlated, we generate our bootstrap data by means of the circular block bootstrap by Politis and Romano (1992). The algorithm consists of the following steps:

1. We first select a set of reasonable block sizes b,

2. We generate L boostrapped sequences $\left(r_t^{\mathcal{F}}\right)^*$ and $\left(r_t^{EH_{CVOL}}\right)^*$ and for each sequence L and for each b we compute a confidence interval $CI_{q,b}$, $q = 1, ..., L$ with nominal level $1 - 0.05$ for $\widehat{\Delta}$

3. We then compute $g(b)$ as the number of times $\widehat{\Delta} \in CI_{q,b}$ divided by the number of sequences L. We compute $\widetilde{b}$ as the value of b that minimizes $|\widehat{g}(b) - 0.05|$

4. Once we have selected the optimal block size $\widetilde{b}$, we compute $h =$int$\left(\widetilde{b}/T\right)$ where int$(\cdot)$ denotes the integer part.

5. We then bootstrap the data series and compute

$$
\begin{aligned}
z_t^* &= \begin{bmatrix} \left(r_t^{EH_{CVOL}}\right)^* - \left(m^{EH_{CVOL}}\right)^* \\ \left(r_t^{\mathcal{F}}\right)^* - \left(m^{\mathcal{F}}\right)^* \\ \left[\left(r_t^{EH_{CVOL}}\right)^*\right]^2 - \left(s^{EH_{CVOL}}\right)^* \\ \left[\left(r_t^{\mathcal{F}}\right)^*\right]^2 - \left(s^{\mathcal{F}}\right)^* \end{bmatrix}, t = 1, .., T \\
\zeta_j &= \frac{1}{\sqrt{\widetilde{b}}} \sum_{t=1}^{\widetilde{b}} z^*_{(j-1)b+t}, \quad t = 1, ..., h \\
\widehat{\Psi}^* &= \frac{1}{h} \sum_{j=1}^{h} \zeta_j \zeta_j',
\end{aligned}
$$

6. We compute the bootstrap estimate of the standard error of $\widehat{\Delta}$ as

$$
\varsigma\left(\widehat{\Delta}^*\right) = \sqrt{\frac{\nabla' f(\upsilon^*)\, \widehat{\Psi}^* \nabla f(\upsilon^*)}{T}},
$$

where

$$
\nabla f(\upsilon^*) = \begin{bmatrix} \left(s^{EH_{CVOL}}\right)^* / \left\{\left(s^{EH_{CVOL}}\right)^* - \left[\left(m^{EH_{CVOL}}\right)^*\right]^2\right\}^{1.5} \\ -\left(s^{\mathcal{P}}\right)^* / \left\{\left(s^{\mathcal{P}}\right)^* - \left[\left(m^{\mathcal{P}}\right)^*\right]^2\right\}^{1.5} \\ -\frac{1}{2}\left(s^{EH_{CVOL}}\right)^* / \left\{\left(s^{EH_{CVOL}}\right)^* - \left[\left(m^{EH_{CVOL}}\right)^*\right]^2\right\}^{1.5} \\ \frac{1}{2}\left(s^{\mathcal{P}}\right)^* / \left\{\left(s^{\mathcal{P}}\right)^* - \left[\left(m^{\mathcal{P}}\right)^*\right]^2\right\}^{1.5} \end{bmatrix}.
$$

7. Finally, we compute the centered studentized statistics over the $\upsilon = 1, .., V$ bootstrap replications

$$
d^{*,\upsilon} = \frac{\left|\widehat{\Delta}^{*,\upsilon} - \widehat{\Delta}\right|}{\varsigma\left(\widehat{\Delta}^{*,\upsilon}\right)},
$$

8. The p-values reported in the Tables are computed as

$$\frac{1}{V+1}\sum_{v=1}^{V} I\left(d^{*,v} \geq d\right) + 1$$

where $I(\cdot)$ denotes an indicator function that is equal to one if its argument is true and zero otherwise.

The p-values reported in the main text are computed using a grid of block sizes $b = \left[\begin{array}{ccccc} 1 & 3 & 6 & 10 & 15 \end{array}\right]$, in line with Ledoit and Wolf's (2008) suggestions, and we set the number of bootstrap replications $V = 1,000$.

B Additional Results

Table B1. Sensitivity Analysis: Rolling Moving Windows

The Table reports summary statistics of the returns and performance fees Φ from alternative portfolios constructed using the out-of-sample forecasts from the benchmark of no-predictability model of bond excess returns with constant volatility and the other competing models. The values are computed a moving windows for the rolling forecasting scheme of 240 months. The asset allocations for all models are carried out an annual target of portfolio excess returns $\zeta_p^* = 0.01$. Time-varying variance-covariance matrices of excess returns are estimated using a decay parameter $\alpha = 0.05$. GISW and Φ are computed using a Relative Risk Aversion (RRA) coefficient $\lambda = 5$. Values in brackets are p-values of the null hypothesis that the SR of the model is equal to the one of EH$_{CVOL}$ (Ledoit and Wolf, 2008). The p-values are computed using $V = 1,000$ bootstrap replications. The out-of-sample forecasting exercise runs from January 1974 through December 2010. See also notes to Tables 3 and 4.

	EH$_{CVOL}$	FB$_{CVOL}$	CP$_{CVOL}$	EH$_{TVOL}$	FB$_{TVOL}$	CP$_{TVOL}$
			No Parameter Uncertainty			
SR	0.291	0.284	0.306	-0.005	0.034	0.110
		[0.96]	[0.88]	[0.27]	[0.16]	[0.31]
$GISW$	–	-0.002	-0.001	-0.011	-0.011	-0.007
Φ	–	0.002	0.003	-0.004	-0.003	-0.002
			Parameter Uncertainty			
SR	0.291	0.301	0.340	–	0.021	0.099
		[0.93]	[0.91]	–	[0.15]	[0.28]
$GISW$	–	-0.002	-0.004	–	-0.012	-0.007
Φ	–	0.003	0.001	–	-0.004	-0.002

Table B2. Sensitivity Analysis: RRA Coefficients

The Table reports the manipulation-proof measure of performance, $GISW$ and performance fees Φ from alternative portfolios constructed using the out-of-sample forecasts from the benchmark of no-predictability model of bond excess returns with constant volatility and the other competing models. The values are computed using two alternative Relative Risk Aversion (RRA) coefficients, i.e. $\lambda = 2, 3$. The asset allocations for all models are carried out an annual target of portfolio excess returns $\zeta_p^* = 0.01$. Time-varying variance-covariance matrices of excess returns are estimated using a decay parameter $\alpha = 0.05$. The out-of-sample forecasting exercise runs from January 1970 through December 2010. See also notes to Tables 3 and 4.

Panel A) Recursive Estimation, No Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\lambda = 2$		
$GISW$	-0.016	-0.011	-0.025	-0.018	-0.016
Φ	-0.016	-0.011	-0.024	-0.018	-0.016
			$\lambda = 3$		
$GISW$	-0.015	-0.011	-0.024	-0.018	-0.015
Φ	-0.014	-0.010	-0.022	-0.017	-0.014

Panel B) Recursive Estimation, Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\lambda = 2$		
$GISW$	-0.016	-0.015	–	-0.021	-0.013
Φ	-0.016	-0.014	–	-0.020	-0.013
			$\lambda = 3$		
$GISW$	-0.015	-0.014	–	-0.020	-0.013
Φ	-0.014	-0.012	–	-0.018	-0.012

Panel C) Rolling Estimation, No Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\lambda = 2$		
$GISW$	-0.003	-0.001	-0.006	-0.007	-0.005
Φ	-0.003	-0.001	-0.006	-0.007	-0.005
			$\lambda = 3$		
$GISW$	-0.003	-0.001	-0.006	-0.007	-0.005
Φ	-0.003	-0.001	-0.006	-0.006	-0.005

Panel D) Rolling Estimation, Parameter Uncertainty

	FB_{CVOL}	CP_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
			$\lambda = 2$		
$GISW$	-0.002	-0.007	–	-0.007	-0.010
Φ	-0.002	-0.007	–	-0.006	-0.010
			$\lambda = 3$		
$GISW$	-0.002	-0.006	–	-0.007	-0.010
Φ	-0.002	-0.006	–	-0.006	-0.009

Table B3. Sensitivity Analysis: Decay Parameter, α

The Table reports summary statistics of the returns and performance fees Φ from alternative portfolios constructed using the out-of-sample forecasts from the benchmark of no-predictability model of bond excess returns with constant volatility and the other competing models. The asset allocations for all models are carried out using an annual target of portfolio excess returns $\zeta_p^* = 0.01$. Time-varying variance-covariance matrices of excess returns are estimated using two alternative decay parameters $\alpha = 0.01, 0.10$. GISW and Φ are computed using a Relative Risk Aversion (RRA) coefficient $\lambda = 5$. Values in brackets are p-values of the null hypothesis that the SR of the model is equal to the one of EH_{CVOL} (Ledoit and Wolf, 2008). The p-values are computed using $V = 1,000$ bootstrap replications. The out-of-sample forecasting exercise runs from January 1970 through December 2010. See also notes to Tables 3 and 4.

Panel A) Recursive Estimation, No Parameter Uncertainty

	EH_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
		$\alpha = 0.01$		
SR	0.464	-0.082	0.100	0.187
		[< 0.01]	[0.13]	[0.17]
$GISW$	–	-0.022	-0.016	-0.014
Φ	–	-0.018	-0.012	-0.010
		$\alpha = 0.10$		
SR	0.464	-0.082	0.100	0.187
		[< 0.01]	[0.13]	[0.17]
$GISW$	–	-0.022	-0.016	-0.014
Φ	–	-0.018	-0.012	-0.010

Panel B) Recursive Estimation, Parameter Uncertainty

	EH_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
		$\alpha = 0.01$		
SR	0.464	–	0.035	0.169
		–	[0.26]	[0.09]
$GISW$	–	–	-0.018	-0.014
Φ	–	–	-0.013	-0.011
		$\alpha = 0.10$		
SR	0.464	–	0.035	0.169
		–	[0.29]	[0.10]
$GISW$	–	–	-0.018	-0.014
Φ	–	–	-0.013	-0.011

Panel C) Rolling Estimation, No Parameter Uncertainty

	EH_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
		$\alpha = 0.01$		
SR	0.304	0.144	0.095	0.189
		[0.39]	[0.23]	[0.48]
$GISW$	–	-0.005	-0.008	-0.004
Φ	–	-0.005	-0.005	-0.003
		$\alpha = 0.10$		
SR	0.304	0.144	0.095	0.189
		[0.38]	[0.24]	[0.51]
$GISW$	–	-0.005	-0.008	-0.004
Φ	–	-0.005	-0.005	-0.003

Panel D) Rolling Estimation, Parameter Uncertainty

	EH_{CVOL}	EH_{TVOL}	FB_{TVOL}	CP_{TVOL}
		$\alpha = 0.01$		
SR	0.304	–	0.105	0.139
		–	[0.25]	[0.32]
$GISW$	–	–	-0.007	-0.005
Φ	–	–	-0.005	-0.004
		$\alpha = 0.10$		
SR	0.304	–	0.105	0.139
		–	[0.25]	[0.31]
$GISW$	–	–	-0.007	-0.005
Φ	–	–	-0.005	-0.004