

Bond return predictability in expansions and recessions

Tom Engsted, Stig V. Møller and Magnus Sander

CREATES Research Paper 2013-13

Bond return predictability in expansions and recessions*

April 2013

Tom Engsted[†] Stig V. Møller[§] Magnus Sander[‡]

*We thank Lasse Bork and Eric Hillebrand for useful comments. We acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

[†]CREATES and Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus, Denmark. Phone: (45) 8716 5572 and e-mail tengsted@creates.au.dk.

[§]CREATES and Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus, Denmark. Phone: (45) 8716 4825 and e-mail svm@asb.dk.

[‡]CREATES and Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus, Denmark. Phone: (45) 8716 6021 and e-mail mdsj@creates.au.dk.

Abstract

We document that over the period 1953-2011 US bond returns are predictable in expansionary periods but unpredictable during recessions. This result holds in both in-sample and out-of-sample analyses and using both univariate regressions and combination forecasting techniques. A simulation study shows that our tests have power to reject unpredictability in both expansions and recessions. To judge the economic significance of the results we compute utility gains for a mean-variance investor who takes the predictability patterns into account and show that utility gains are positive in expansions but negative in recessions. The results are also consistent with tests showing that the expectations hypothesis of the term structure holds in recessions but not in expansions. However, the results for bonds are in sharp contrast to results for stocks showing that stock returns are predictable in recessions but not in expansions. Thus, our results indicate that there is not a common predictive pattern of stock and bond returns associated with the state of the economy.

JEL Codes: C53, G12

Keywords: Return predictability, expansions and recessions, out-of-sample tests, power properties, mean-variance investor, expectations hypothesis.

1 Introduction

Whether asset returns can be predicted is one of the most debated questions in finance. In recent years, forecastability of stock returns has been questioned by e.g. Goyal and Welch (2003, 2008) and, on the other hand, defended by e.g. Campbell and Thompson (2008), Cochrane (2008), Rapach et al. (2010), Henkel et al. (2011) and Dangl and Halling (2012). Rapach and Zhou (2012) survey the literature on stock return forecasting and find that simple univariate regressions seem to have limited out-of-sample forecasting power, but if forecasters apply more advanced techniques such as combination forecasting, use economically based priors, or take time-varying effects into account, out-of-sample performance can be significant.

Although bond returns have received less attention in the literature, both in-sample (**IS**) and out-of-sample (**OoS**) predictability have been documented by several authors e.g. Fama and Bliss (1987), Ilmanen (1995), Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009). Thus, there is a vast amount of empirical evidence against the expectation hypothesis, which says that excess returns on bonds are unpredictable. However, recent research questions the ability to predict bond returns OoS. Thornton and Valente (2012) find that forward rates do not add economic value relative to a no-predictability benchmark. Similarly, Sarno et al. (2013) find that affine term structure models forecast returns poorly OoS relative to a no-predictability benchmark.

In this paper, we comprehensively investigate the empirical evidence on bond return forecasting using the most popular predictors from the literature. We use both simple regressions as well as incorporate more recent advances in forecasting methodology. Our results show that bond returns are predictable in expansions but unpredictable in recessions. This result is robust using both statistical and economic criteria, using both IS and OoS analysis, and using both univariate regressions and combination forecasting. In expansions, the forecasting models consistently beat the no-predictability benchmark, while the forecasting models fail to beat the no-predictability benchmark in recessions. Due to the relatively low number of recessionary observations in the sample one might worry about power. However, a simulation study shows that our tests have power to reject unpredictability in both expansions and recessions.

We compute utility gains for mean-variance investors from optimally trading on the predictors relative to using a no-predictability benchmark. The utility gains are positive in expansions but strongly negative in recessions. Thus, the result of Thornton and Valente (2012) that there is no economic gain from bond return forecasting holds only in recessions.

We find the complete opposite predictive pattern for bond returns than what Henkel et al. (2011), Rapach et al. (2010) and Rapach and Zhou (2012) find for stock returns. They show that predictability of stock returns is concentrated around recessions. Based on the same list of predictors that we use for bond return forecasting, we confirm that stock return predictability is mainly a recessionary phenomenon. Fama and French (1989) provide empirical evidence that there is a common predictive pattern of stock and bond returns associated with the state of the economy. In contrast, we find that bond returns are predictable in expansions but not in recessions, whereas stock returns are predictable in recessions but not in expansions. Our results indicate that previous analyses that have found stock and bond returns to be predictable from the same set of variables over identical time periods (e.g. Engsted et al. (2010)) may be sensitive to distinguishing between expansions and recessions.

Finally, we relate our results to tests of the expectations hypothesis of the term structure. We run Campbell and Shiller (1991) predictive regressions of changes in long and short rates onto the lagged yield spread, and where we allow the spread coefficient to take different values under expansions and recessions. According to the expectations hypothesis the spread coefficient should equal one. We find that indeed the spread coefficient is closer to one in recessions than in expansions; formally the hypothesis is statistically rejected during expansions but not during recessions. We also find that interest rate volatility is much higher in recessions than in expansions. Thus, the tests of the expectations hypothesis are consistent with our return predictability findings, and also consistent with the earlier literature showing that there is more support to the expectations hypothesis when interest rates are highly volatile (e.g. Mankiw and Miron (1986)).

The rest of the paper is organized as follows. Section 2 describes the methodology. Section 3 presents our empirical results. Section 4 concludes. The Appendix provides further detailed information about the data, describes a bootstrap procedure we use for inference, and presents additional robustness checks.

2 Methodology

2.1 Out-of-sample regressions

The most simple regressions we run are

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + u_{i,t+1}, \quad (1)$$

where r_{t+1} is the log excess return on a bond from time t to $t + 1$,¹ defined as the log of the holding period gross return minus the log of 1 plus the risk free rate, and $x_{i,t}$ is some ex ante regressor based only on the information set at time t . We estimate (1) for the full sample period of T observations to obtain IS estimates $(\hat{\alpha}_i, \hat{\beta}_i)$ as well as the R^2 in order to assess the IS fit of the different models. The main focus of this paper is, however, on the OoS performance of the models we consider.² In OoS studies, it is important that the forecaster only uses information available at time t when predicting the excess return in period $t + 1$. Thus, in (1), $x_{i,t}$ must be observable at time t , and (α_i, β_i) can only be estimated using observations obtainable at time t . Following the notation in Rapach et al. (2010), we reserve the first m observations for the initial estimates that generate the first forecast. The first prediction becomes

$$\hat{r}_{i,m+1} = \hat{\alpha}_{i,m} + \hat{\beta}_{i,m} x_{i,m}, \quad (2)$$

where $(\hat{\alpha}_{i,m}, \hat{\beta}_{i,m})$ are the estimates of α_i and β_i in (1) from the regression of $\{r_t\}_{t=2}^m$ on a constant and $\{x_{i,t}\}_{t=1}^{m-1}$. The next forecast is generated from

$$\hat{r}_{i,m+2} = \hat{\alpha}_{i,m+1} + \hat{\beta}_{i,m+1} x_{i,m+1}, \quad (3)$$

where $(\hat{\alpha}_{i,m+1}, \hat{\beta}_{i,m+1})$ are the estimates from the regression of $\{r_t\}_{t=2}^{m+1}$ on a constant and $\{x_{i,t}\}_{t=1}^m$. We continue generating $T - m = q$ forecasts, $\{\hat{r}_{i,t+1}\}_{t=m}^{T-1}$, for each of the N predictors $x_{i,t}$, $i = 1, \dots, N$. This implies that we use an expanding (recursive) estimation window.

We consider a number of extensions of this simple framework, namely a ‘kitchen sink’ model (**ks**), forecasts based on the “best” model available as measured by the Schwarz Information

¹From now on, unless otherwise stated, the term return will refer to log excess return.

²There is some theoretical controversy regarding OoS versus IS predictability studies. On one hand, you lose power by throwing away observations (Inoue and Kilian (2004) and Cochrane (2008)). On the other hand, OoS studies are a more realistic test of the practical usefulness of predictability in real-time portfolio management (Goyal and Welch (2008)).

Criterion (**SIC**), and forecast combinations. The ks model is simply based on including all N predictors in the regression

$$r_{t+1} = \alpha + \mathbf{x}'_t \boldsymbol{\beta} + u_{t+1}, \quad (4)$$

where $\mathbf{x}_t = [x_{1,t} \ x_{2,t} \ \dots \ x_{N,t}]'$ and $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_N]'$. In addition to just using all predictors each period, we let the SIC decide on the best model. We allow up to three regressors of any combination to enter a particular specification, meaning that we in our case have 175 different models to consider each period since we have $N = 10$ predictor variables.³ At each time point, we calculate the SIC for each of the alternatives and choose the one with the smallest value. For the first prediction, the SIC of the j 'th alternative becomes

$$SIC_{j,m} = \ln(\hat{\sigma}_{m,u}^2) + (K + 1) \cdot \ln(m - 1) / (m - 1), \quad (5)$$

$$\hat{\sigma}_{m,u}^2 = \sum_{t=2}^m (r_t - \hat{r}_{j,t})^2 / (m - 1). \quad (6)$$

To get the formula in the next period, simply replace m with $m + 1$. We note that for each time period, a new model, possibly with a different number of predictors K , can have the lowest SIC, and the preferred model can thus vary over time. For both the ks model and the model selected by the SIC, forecasts are computed using the same approach as for the univariate model in (1).

Rapach et al. (2010) show that combination forecasts can provide significant OoS performance for stock returns. Therefore, it should be interesting to investigate whether this is also true for bond returns. Generally, the combination forecast for period $t + 1$ using method c is

$$\hat{r}_{c,t+1} = \sum_{i=1}^N \omega_{i,t}^c \hat{r}_{i,t+1}, \quad (7)$$

where $\{\omega_{i,t}^c\}_{i=1}^N$ are the forecast weights based only on information available at time t . We consider two different weighting schemes. The first is the mean combination forecast (**POOLA**) where $\omega_{i,t}^c = 1/N$, so that the weights are constant over time and equal across regressors. The second scheme (**POOLW**) computes time-varying weights based on the performance of the regressors in an initial hold-out period of the OoS period. We reserve q_0 initial observations leaving $q - q_0$ forecasts from this method.⁴ We follow Stock and Watson (2004) and compute

³ $N! / ((N - K)!K!)$, $N = 10$, $K = (1, 2, 3)$.

⁴The OoS results we report in Section 3 are only based on the $q - q_0$ observations where we have forecasts

weights based on the mean square prediction error (**MSPE**)

$$\omega_{i,t}^c = \phi_{i,t}^{-1} / \sum_{j=1}^N \phi_{j,t}^{-1}, \quad (8)$$

where

$$\phi_{i,t} = \sum_{s=m}^{t-1} \lambda^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1})^2, \quad (9)$$

and λ is a discount factor that determines how much recent performance is overweighted given that $\lambda < 1$. Like Rapach and Zhou (2012) we set $\lambda = 0.75$. If the MSPE of a forecast is high, indicating bad performance, the forecast receives a low weight and vice versa. Rapach et al. (2010) argue that forecast combinations stabilize forecasts of excess returns in the same way diversification reduces portfolio variance. The individual predictors tend to produce forecasts that are too volatile, while the historical average return benchmark is too smooth as it fails to reflect the temporal variation in required risk compensation. One could argue that the ks model also incorporates the information in all predictors. However, in empirical applications, it tends to perform badly, possibly due to IS overfitting of the many parameters that need estimation (e.g. Goyal and Welch (2008); Rapach et al. (2010)).

2.2 Goodness of fit

Following the existing literature on return predictability, we compare the forecasts of the models we propose with the historical average return as a benchmark. For the IS estimates, we compute the standard R^2 . For the OoS study, the benchmark is recomputed each period. The first benchmark prediction is

$$\bar{r}_{m+1} = \frac{1}{m} \sum_{t=1}^m r_t. \quad (10)$$

We evaluate the performance of each model using the Campbell and Thompson (2008) OoS R^2

$$R_{OoS}^2 = 1 - \frac{\sum_{t=m+q_0}^{T-1} (r_{t+1} - \hat{r}_{t+1})^2}{\sum_{t=m+q_0}^{T-1} (r_{t+1} - \bar{r}_{t+1})^2}. \quad (11)$$

If $R_{OoS}^2 > 0$ the MSPE of the model we consider is below the MSPE of the benchmark. The R_{OoS}^2 has a similar interpretation to conventional R^2 s.

from every model, so that the numbers are comparable.

To take sampling error into account, we also conduct a more formal statistical test of predictability. We wish to test the null that $R_{OoS}^2 \leq 0$ (no predictability) against the alternative that $R_{OoS}^2 > 0$ (predictability). One approach could be the Diebold and Mariano (1995) and West (1996) statistic. However, Clark and McCracken (2001) and McCracken (2007) emphasize that this statistic has a nonstandard distribution when comparing nested models. This is exactly the case here, as we get the historical average benchmark if $\beta_i = 0$ in (1). Clark and West (2007) propose the MSPE-adjusted statistic for which the standard normal distribution is a good approximation asymptotically when comparing forecasts from nested models. They show in a realistic simulation study with a persistent regressor setup that the size and power properties of this test are satisfactory. We first define

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - \left[(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2 \right]. \quad (12)$$

To compute the MSPE-adjusted statistic we simply regress $\{f_{s+1}\}_{s=m+q_0}^{T-1}$ on a constant and calculate the t -statistic of this constant. We then evaluate the one-sided upper-tail p -value using the normal distribution.

We emphasize that especially for monthly regressions, we expect to see very low R^2 s. Zhou (2010) derives the theoretical bound implied by asset pricing models. For stock returns, he shows that even R^2 s below 1% actually exceed the bound consistent with popular asset pricing models such as the habit-based model of Campbell and Cochrane (1999). Thus even a very modest predictive power can imply market inefficiencies from the viewpoint of modern asset pricing models. We further note that predictability is not necessarily inconsistent with market efficiency. To the extent that time-varying expected returns reflect temporal variation in risk compensation due to e.g. the state of the economy, forecastability is entirely consistent with an efficient market. Only in the case that the risk-adjusted return (incl. adjustments for trading costs, liquidity and borrowing constraints, research costs etc.) is not zero, we can conclude that the market is inefficient (Rapach and Zhou (2012)).

2.3 Utility gains

Campbell and Thompson (2008) note that even though R^2 s from predictive regressions of stock returns seem low, a mean-variance investor can obtain sizable economic gains using

these predictions rather than the no-predictability benchmark. To investigate whether the predictive power of bond return regressions can be exploited for economic gains, we follow Campbell and Thompson (2008), among others, and employ mean-variance analysis. In our case, we compute utility gains for a mean-variance investor who allocates his capital between a long risky government bond and a short T-bill. The following share of the portfolio is placed in the risky asset at the end of time t :

$$w_{i,t} = \left(\frac{1}{\gamma} \right) \left(\frac{\hat{r}_{i,t+1}}{\hat{\sigma}_{t+1}^2} \right), \quad (13)$$

where γ is the coefficient of relative risk aversion, $\hat{r}_{i,t+1}$ is the forecast from the i 'th predictor, and $\hat{\sigma}_{t+1}^2$ is the variance forecast.⁵ The variance forecast is based on a rolling window of the past 5 years of monthly observations. We then compute the average utility from trading on the information in regressor i as

$$\hat{v}_i = \hat{\mu}_i - \frac{\gamma}{2} \hat{\sigma}_i^2, \quad (14)$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the sample mean and variance of the resulting portfolio returns. We set $\gamma = 5$ and use simple returns so that portfolio returns are computed as weighted averages of individual asset returns. We further restrict that there be no short sales and at most 50% leverage ($0\% \leq w_{i,t} \leq 150\%$).⁶

Finally, we compute the average utility for a mean-variance investor who uses the historical mean return to make his portfolio allocation decisions. The utility gain is then given by the difference in average utility between using predictive regressions or combination forecasting relative to using the historical mean.

⁵The portfolio return in period $t + 1$ thus becomes $w_{i,t} \times r_{t+1} + (1 - w_{i,t}) \times r_{f,t+1}$, where r_{t+1} is the realized simple return on the risky asset in period $t + 1$, and $r_{f,t+1}$ is the return on the risk free asset in that period.

⁶Thornton and Valente (2012) also use a coefficient of relative risk aversion of 5 while Sarno et al. (2013) use a value of 3. Both papers measure the economic gain using the procedure of Goetzmann et al. (2007). Campbell and Thompson (2008) and Rapach and Zhou (2012) use values of 3 and 5 respectively employing the same performance measure as we do. Setting $\gamma = 3$ leaves the overall conclusions of this paper unchanged.

3 Empirical analysis

3.1 Data

In this study we use long term US government bond returns based on bonds with an approximate duration of 20 years. The bond returns are from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook.⁷ The risk free rate is the return on the 30-day T-bill. The dependent variable is the monthly excess return calculated as the log rate of return on bonds minus the log risk free rate of return.

3.1.1 Predictive variables

We use the following ten predictive variables. In the Appendix we report descriptive statistics for the list of predictor variables.

1. Forward spread (fws): The difference between the 5-year forward rate and the 1-year yield, see e.g. Fama (1976) and Fama and Bliss (1987).

2. Term spread (tms): The difference between the long term yield and the T-bill rate, see e.g. Fama and French (1989) and Campbell and Shiller (1991).

3. Forward tent (ftent): Tent-shaped linear combination of forward rates, which we compute following the procedure in Cochrane and Piazzesi (2005). We use weights that are estimated using the entire sample period, and so it is principally not available in real time.

4. Yield tent (ytent): Inspired by Cochrane and Piazzesi (2005), we also include a tent-shaped linear combination of yields. It is calculated as $w'y$ where $w = (-0.5, 1.0, -0.5)'$ and $y = (y^{(1)}, y^{(5)}, y^{(10)})'$. $y^{(1)}$, $y^{(5)}$ and $y^{(10)}$ are yields of 1, 5 and 10 years maturity.

5. Credit spread (cs): The difference between Moody's BAA rated corporate bond yield and the T-bill rate, see e.g. Keim and Stambaugh (1986).

6. Default spread (dfy): The difference between Moody's BAA- and AAA-rated corporate bond yields, see e.g. Fama and French (1989).

⁷We have tried with other popular government bond return series and found results similar to the ones reported in this paper.

7. Real short rate (srr): The inflation adjusted T-bill rate, see e.g. Ilmanen (1995).

8. The dividend-price ratio (dp): We compute the dividend-price ratio of the S&P 500 index as the difference between the log of trailing 12 month dividends and the log of the price (Fama and French (1989)).

9. Employment growth (emp): The 12 month change in the log of seasonally adjusted non-farm employment (Piazzesi and Swanson (2008)).⁸

10. Inflation (infl): We compute the inflation rate based on the Consumer Price Index for all urban consumers, see e.g. Ludvigson and Ng (2009).

To reduce the computational burden, the employment and inflation series are not based on vintage data, and so they are principally not available in real-time. However, we do take into account announcement delays by lagging employment growth, inflation and the real short rate an additional period in the predictive regressions.

3.1.2 Data sources and sample period

The returns on the 30 day T-bill and the Fama-Bliss data set for calculating the forward spread and the Cochrane and Piazzesi (2005) factor are collected from the CRSP database. Inflation and the data for calculating the term spread, the credit spread, the default spread, the short real rate, the dividend-price ratio as well as the government bond return series are all taken from the updated Goyal and Welch (2008) data set provided on Amit Goyal's web page. The employment series as well as the yields for ytent are all collected from the database of the St. Louis FED (FRED). To identify recessions and expansions, we use the NBER business cycle dummy also collected from FRED.

The sample period is from 1953:04 to 2011:12 giving a total of 704 observations for the IS regressions.⁹ There are ten recessions in the total sample corresponding to 111 months of economic contraction (16% of the sample). For the OoS study, we choose an initial estimation period of 20 years (1953-1973) allowing for the inclusion of four recessions. We choose a

⁸There is a voluminous literature on the relation between the yield curve and macro factors (going both ways causally). Relevant to our case is especially the literature on yield curve modeling in which time-varying risk premia are a function of macro factors. Examples are Ang and Piazzesi (2003) and Hördahl et al. (2004).

⁹The start of the sample is dictated by the availability of the constant maturity yields for computing the yield tent (ytent).

hold-out period of five years for computing combination forecast weights. This results in an OoS period with 405 observations running from 1978 to 2011 including five recessions corresponding to 56 observations.

3.2 Forecasting bond returns

Table 1 reports IS and OoS results. For the IS estimates of β and the R^2 , we calculate p -values using a bootstrap procedure, which we describe in the Appendix. For the OoS R^2 values, the p -values are from the Clark and West (2007) test.¹⁰ The upper panel of Table 1 considers one predictor at a time, while the lower panel considers forecasts based on more than one predictor at a time. The table shows results for the total period as well as separate results for recessionary and expansionary periods. Similar to Rapach and Zhou (2012), we use the NBER recession dates to define the state of the economy. We divide the forecast errors between expansions and recessions based on the state of the economy at the time of the forecast. In the following we start by discussing the results based on the total period, and then afterwards discuss the results obtained for expansions and recessions.

For the total period, we find that four out of ten predictors generate statistically significant IS and OoS predictive power based on a 5% significance level. The four variables are the forward spread, term spread, the yield tent and the credit spread. The Cochrane and Piazzesi (2005) factor also gives a positive OoS R^2 , but it is not statistically significant. For the models based on more than one predictor, the kitchen sink approach and the SIC-based approach both perform poorly OoS, whereas the forecast combination methods beat the no-predictability benchmark in terms of a positive R^2_{OoS} but not in a statistically significant way.

When the various forecasting models are analyzed during expansions and recessions separately, we obtain a distinct predictive pattern. For all ten predictors, we find no evidence of predictability in recessions. Likewise, combination forecasting fails to beat the no-predictability benchmark in recessions. On the other hand, in expansions five predictors generate statistically significant IS predictive power and seven predictors generate statistically significant OoS predictive power. Furthermore, combination forecasting strongly beats the historical mean benchmark. The predictors that generate both statistically significant IS and OoS predictive

¹⁰The program we use for computing OoS statistics is partly based on the MATLAB code of Rapach and Zhou (2012) provided on David Rapach's web page.

power in expansions are the forward spread, the term spread, the forward term, the yield term and the credit spread. Interestingly, however, none of these predictors work in recessions, neither IS nor OoS. The difference in predictive power between expansions and recessions is high: The IS and OoS R^2 s for these five predictors are around -3% to -1% in recessions and around 2% to 4% in expansions.

3.2.1 Power and size properties

Due to the low frequency of recessions one might worry about power issues. In Table 2 we investigate the power of the Clark and West (2007) OoS statistic for the five predictors that have significant IS and OoS forecasting power. We compute the power both for the total OoS evaluation period as well as for recessions and expansions separately. We generate 10,000 samples using the total sample OLS estimates from predictive regressions and the bootstrap system described in the Appendix. Implicitly, we therefore assume a stable DGP with no differences between expansions and recessions. Thus, we investigate how likely it is that we find predictability differences across states of the economy when there in fact are none.

With a significance level of 5%, we find powers around 50-90% for the total period and in expansions. When only looking at recession-based forecast errors, we find power levels of around 30% even though there are only 56 observations. With a significance level of 10%, the power increases to a minimum of 40% in recessions. To get a feeling for the level of power we can expect, Clark and West (2007) find power levels *up to* approximately 40% for the *total* OoS period using a realistic stock return DGP with dividend yields and a nominal size of 10%. Thus, if there is a stable DGP generating forecastability both in expansions *and* recessions, it seems that we have reasonable power to detect it. Second, it should be noted that we do not observe positive and high R^2 s in recessions and conclude no predictability due to high p -values. On the contrary, we see a change in sign in the point estimate of the R^2 between expansions and recessions both IS and OoS. This is unlikely to be explained by power issues.

To check whether we obtain satisfactory power at the cost of inflated size, we also investigate the size of the Clark and West (2007) OoS statistic. Table 3 shows the size results for the same five predictors. All the empirical sizes are relatively close to their nominal 5% and 10% level respectively, and thus the normality approximation does not seem to be too far off. This is

also what Clark and West (2007) find in their simulation studies.¹¹

As explained in the Appendix, the bootstrap procedure we use for calculating the empirical size and power is based on a fixed business cycle structure. The above conclusions are robust to relaxing this assumption. As a robustness check, we have tried simulating a new realisation of the NBER recession dummy prior to each bootstrap run using the Markov Chain of Artis et al. (2004). Both power and size are overall unaffected by this change. The results are available from the authors upon request.

3.2.2 Graphical illustrations

Goyal and Welch (2003, 2008) recommend plotting the difference between the cumulative squared forecast error for the no-predictability benchmark and the cumulative squared forecast error for the candidate predictor. The purpose is to assess the stability of the candidate predictor's forecasting performance over time. Figure 1 shows the results of this graphical approach for both the IS and OoS regressions. A positive slope means that the candidate predictor has outperformed the benchmark in a given month, and a positive value means that the candidate predictor has outperformed the benchmark so far. The figure confirms the result from Table 1 that the term spread, the yield curvature factor, the forward spread and the credit spread all have predictive ability for bond returns, both IS and OoS. The predictive power of these variables is not just reflecting a few outliers.¹²

Next we use the graphical approach of Goyal and Welch (2003) to examine expansions and recessions separately. In Figure 2, we focus on the IS forecast errors in the 593 expansion months, and in Figure 3 on the IS forecast errors in the 111 recession months. Figure 2 illustrates that several of the predictors consistently outperform the no-predictability benchmark in expansions. In contrast, Figure 3 illustrates that the same predictors that work well in expansions have unstable performance in recessions.

¹¹We have confirmed this by computing bootstrap p -values of the Clark and West (2007) statistics for Table 1. They are rather close to the p -values from the normal distribution we have reported in Table 1.

¹²According to Goyal and Welch (2003), the OoS predictive power of the dividend-price ratio for stock returns is mainly driven by two years, 1973 and 1974.

3.3 Utility gains

Thornton and Valente (2012) and Sarno et al. (2013) conclude that predictability of bond risk premia cannot be exploited for economic profit. In Table 4 we report annualized OoS utility gains for each of the ten predictors as well as for the multivariate methods. The gain is computed as the difference between the average utility from using the predictive regressions and the utility from using the historical average benchmark. The numbers in Table 4 thus measure how much a risk averse mean-variance investor is willing to pay in yearly %-point fees to access the information in the regressions rather than to assume no predictability.

Overall, this exercise yields similar conclusions to those reached based on the R_{OoS}^2 . For the first five predictors and combination forecasts, we see sizable economic profits for the total OoS period. Further, we see that the economic gains are almost entirely earned from predictions made in periods where the economy is doing well as measured by the NBER dummy. Conversely, in periods of economic contraction, substantial utility losses are generally incurred when trading on the information in the predictive regressions. Thus, while Thornton and Valente (2012) conclude that the informational content in forward rates cannot be transformed into economic profit, we argue that such total sample results can be blurred by differences between states of the economy.¹³ All in all, the conclusion of significant predictability in expansions and no predictability in recessions carry over when predictive power is measured in economic rather than statistical terms.

3.4 Forecasting stock returns

Henkel et al. (2011), Rapach et al. (2010) and Rapach and Zhou (2012) find that stock return predictability is concentrated in a few recession periods. We find the opposite for bond returns, namely that bond returns are predictability during expansions but not during recessions.

The opposite predictive patterns for stocks and bonds are not just a result of different choices of predictors. Henkel et al. (2011) consider four classic predictors; the dividend yield, the term spread, the short rate and the default spread. These four predictors are also in our list of

¹³Their OoS period runs from 1970:01 to 2010:12, while ours runs from 1978:04 to 2011:12. Thus, there is a rather high degree of overlap between the evaluation periods.

predictors. We also have a high degree of overlap with the set of predictors that Rapach et al. (2010) and Rapach and Zhou (2012) apply.

In Table 5 we show the results of predicting stock returns using our list of predictors.¹⁴ We find absolute no evidence of stock return predictability in expansions. The R_{OoS}^2 is negative for all ten predictors in expansions. In contrast, many of the predictors generate a positive R_{OoS}^2 in recessions. Thus, our results for stock returns are overall in line with the findings of Henkel et al. (2011), Rapach et al. (2010) and Rapach and Zhou (2012). We note that we use exactly the same forecasting methodology as Rapach et al. (2010) and Rapach and Zhou (2012), but find slightly less predictive power in recessions than they do. This is because we have not included the recession of the mid-1970s in our OoS period. In the Appendix we reproduce Table 1 and 5 with a pre-estimation period of 10 years so that the OoS period starts in 1968. While the results for bond returns are very similar to the ones reported in Table 1, stock return predictability is now markedly higher in recessions but still absent in expansions. Thus, the differences in predictability between stocks and bonds across states of the economy become even clearer when including the 1970s.

3.5 The expectation hypothesis

The finding that bond returns are predictable in expansions but not in recessions should have its mirror image in tests of the Expectations Hypothesis of the Term structure (EHT). This hypothesis says that long-term interest rates are determined as the average of expected future short-term rates plus a time-invariant - but maturity-dependent - term premium. According to the EHT, excess returns on long bonds are unpredictable. Thus, given the results in the previous sections we should expect to find more support for the EHT during recessions than during expansions.

Empirical tests of the EHT on US data generally reject the hypothesis, see Gürkaynak and Wright (2012) for a recent survey. However, it is also well-known that the results for the EHT are highly sample dependent. For the US, the EHT fares much better under money supply targeting by the Fed or when the yield spread is used as an explicit monetary policy indicator. By contrast, under interest rate targeting where the Fed attempts to smooth interest rates,

¹⁴We use the log excess return on the S&P500 index collected from CRSP.

there is not much evidence in support of the EHT (see e.g. Mankiw and Miron (1986) and Hsu and Kugler (1997)). Similar results have been reported for other countries, e.g. Kugler (1988), Engsted (1996), and Engsted and Tanggaard (1995). A general finding is that there is much more support to the EHT when interest rates are volatile, e.g. when the central bank targets monetary aggregates and let short rates fluctuate freely in a market determined fashion, or under currency crises where short rates are set to support the domestic currency. Under interest rate smoothing, however, the volatility of short rates is reduced and the support for the EHT vanishes.

Tables 6 and 7 show the results of running Campbell and Shiller's (1991) long-rate and short-rate predictive regressions on US zero-coupon bond yields over the period 1952-2011, and where the yield spread regressor coefficient is allowed to take different values in expansions and recessions.¹⁵ (Below we investigate whether interest rates are in general more volatile in recessions). The regressions are

$$y_{t+m}^{(n-m)} - y_t^{(n)} = \alpha_{REC} I_t + \beta_{REC} \frac{m}{n-m} (y_t^{(n)} - y_t^{(m)}) I_t + \alpha_{EXP} (1 - I_t) + \beta_{EXP} \frac{m}{n-m} (y_t^{(n)} - y_t^{(m)}) (1 - I_t) + u_{t+m}, \quad (15)$$

and

$$\sum_{i=1}^{k-1} \left(1 - \frac{i}{k}\right) (y_{t+im}^{(m)} - y_{t+(i-1)m}^{(m)}) = \alpha_{REC} I_t + \beta_{REC} (y_t^{(n)} - y_t^{(m)}) I_t + \alpha_{EXP} (1 - I_t) + \beta_{EXP} (y_t^{(n)} - y_t^{(m)}) (1 - I_t) + u_{t+(k-1)m}, \quad (16)$$

where $k = n/m$ and I_t is the NBER recession dummy at time t . In the regressions (15) and (16), $y_t^{(n)}$ and $y_t^{(m)}$ are the n -month and m -month yields, respectively, where $n > m$. In all regressions we let the short rate be the 1-year yield, i.e. $m = 12$, and set $n = 24, 36, 48$, and 60 , i.e. the maturity of the long bond ranges from 2 to 5 years. The slope coefficients, β_{REC} and β_{EXP} , should equal 1 under the EHT.

¹⁵We use the Fama and Bliss (1987) dataset from CRSP.

The upper panel in Table 6 shows that over the whole period, and without distinguishing between expansions and recessions, the estimated slope coefficient in (15) is consistently negative and significantly different from 1. This is the well-known result that in the US the slope of the term structure in general predicts long rates in the 'wrong' direction. However, when we allow the slope coefficient to be different in expansionary and recessionary periods (lower panel), we see that during recessions the beta estimates are positive and not significantly different from 1. In recessions it seems that the term structure predicts future long rates in the direction implied by the EHT, in contrast to what happens during expansions. Table 7 shows that when predicting future short rates (regression (16)), the evidence in support of the EHT is also stronger under recessions than under expansions. The slope estimates are positive in both periods, but closer to 1 in recessions than during expansionary periods. Taken together, the results in Tables 6 and 7 are consistent with our earlier findings that bond returns are predictable in expansions but not in recessions.

Table 8 reports the annual volatility obtained by fitting a GARCH(1,1) process to the residuals of an AR(1) model for the 1-year yield also used in the above regressions. Short rate volatility is almost double as high during recessions than during expansions, a phenomenon that is also clearly visible in Figure 4 which shows the time-varying volatility from the estimated GARCH model. In Table 8 and Figure 4 we also show volatility estimates using a 12-month rolling window of the 1-year bond yield, and the conclusions are the same. Thus, our results are consistent with the earlier literature - cited above - showing that there is more support to the EHT in times of high interest rate volatility compared to periods with low volatility.

4 Conclusion

In this paper we have documented - using both in-sample and out-of-sample analyses - that US bond returns are predictable in expansionary periods but unpredictable during recessions. This result is robust to a number of changes to the modeling framework, and a mean-variance portfolio analysis reveals that there are positive utility gains by utilizing the predictability during expansions while the utility gains are negative during recessions. We also document that the insignificant predictability in recessions is not the result of lack of power due to a limited number of recessionary observations. Finally, we have shown that the predictability

results for bond returns are consistent with results for the expectations hypothesis of the term structure which rejects the hypothesis in expansions but not in recessions.

Our results for bonds stand in contrast to results for stocks showing that stock returns are predictable in recessions but not in expansionary periods, cf. Henkel et al. (2011). Thus, the latest evidence does not support Fama and French's (1989) - and several subsequent studies' - finding that stock and bond returns are related in the same way to business cycle patterns. Our results cast doubt on the prevalent explanation for return predictability, namely that it is due to time-varying risk premia and, hence, that over identical time periods stock and bond returns are related to the same set of predictive variables summarizing this time-variation in risk premia.

References

- Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* **50**: 745–787.
- Artis, M. J., Marcellino, M. and Proietti, T. (2004). Dating business cycles: A methodological contribution with an application to the euro area, *Oxford Bulletin of Economics and Statistics* **66**(4): 537–565.
- Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* **107**(2): 205–251.
- Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* **58**: 495–514.
- Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* **21**(4): 1509–1531.
- Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* **105**(1): 85–110.
- Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* **138**(1): 291–311.
- Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability, *Review of Financial Studies* **21**(4): 1533–1575.
- Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia, *American Economic Review* **95**(1): 138–160.
- Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients, *Journal of Financial Economics* **106**: 157–181.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**(3): 253–263.
- Engsted, T. (1996). The predictive power of the money market term structure, *International Journal of Forecasting* **12**: 289–295.

- Engsted, T., Hyde, S. and Møller, S. V. (2010). Habit formation, surplus consumption and return predictability: International evidence, *Journal of International Money and Finance* **29**: 1237–1255.
- Engsted, T. and Tanggaard, C. (1995). The predictive power of yield spreads for future interest rates: Evidence from the danish term structure, *Scandinavian Journal of Economics* **97**(1): 145–159.
- Fama, E. F. (1976). Forward rates as predictors of future spot rates, *Journal of Financial Economics* **3**: 361–377.
- Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates, *American Economic Review* **77**(4): 680–692.
- Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* **25**: 23–49.
- Goetzmann, W., Ingersoll, J., Spiegel, M. and Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures, *Review of Financial Studies* **20**: 1503–1546.
- Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios, *Management Science* **49**(5): 639–654.
- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**(4): 1455–1508.
- Gürkaynak, R. S. and Wright, J. H. (2012). Macroeconomics and the term structure, *Journal of Economic Literature* **50**(2): 331–367.
- Henkel, S. J., Martin, J. S. and Nardari, F. (2011). Time-varying short-horizon predictability, *Journal of Financial Economics* **99**(3): 560–580.
- Hördahl, P., Tristani, O. and Vestin, D. (2004). A joint econometric model of macroeconomic and term structure dynamics, *Working Paper November*. ECB Working Paper No. 405; AFA 2005 Philadelphia Meetings Paper. Available at SSRN: <http://ssrn.com/abstract=601025>.
- Hsu, C. T. and Kugler, P. (1997). The revival of the expectations hypothesis for the us term structure, *Economic Letters* **55**: 115–120.

- Ilmanen, A. (1995). Time-varying expected returns in international bond markets, *Journal of Finance* **50**(2): 481–506.
- Inoue, A. and Kilian, L. (2004). In-sample or out-of-sample tests of predictability: Which one should we use?, *Econometric Reviews* **23**(4): 371–402.
- Keim, D. B. and Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets, *Journal of Financial Economics* **17**: 357–390.
- Kugler, P. (1988). An empirical note on term structure and interest rate stabilization policies, *Quarterly Journal of Economics* **103**(4): 789–792.
- Lewellen, J. (2004). Predicting returns with financial ratios, *Journal of Financial Economics* **74**: 209–235.
- Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia, *Review of Financial Studies* **22**(12): 5027–5067.
- Mankiw, N. G. and Miron, J. A. (1986). The changing behavior of the term structure of interest rates, *Quarterly Journal of Economics* **101**: 211–228.
- McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality, *Journal of Econometrics* **140**(2): 719–752.
- Morley, J., Piger, J. and Tien, P.-L. (2013). Reproducing business cycle features: Are nonlinear dynamics a proxy for multivariate information?, *Studies in Nonlinear Dynamics and Econometrics* . Forthcoming.
- Piazzesi, M. and Swanson, E. T. (2008). Futures prices as risk-adjusted forecasts of monetary policy, *Journal of Monetary Economics* **55**: 677–691.
- Rapach, D. E., Strauss, J. K. and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* **23**(2): 821–862.
- Rapach, D. E. and Zhou, G. (2012). Forecasting stock returns, In: Elliott, G., Timmermann, A. (Eds.), *Handbook of Economic Forecasting* **2**. Forthcoming.
- Sarno, L., Schneiderz, P. and Wagner, C. (2013). The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis?, *Working Paper* .

- Stambaugh, R. F. (1999). Predictive regressions, *Journal of Financial Economics* **54**: 375–421.
- Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* **23**: 405–430.
- Thornton, D. L. and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective, *Review of Financial Studies* **25**(10): 3141–3168.
- West, K. D. W. (1996). Asymptotic inference about predictive ability, *Econometrica* **64**(5): 1067–1084.
- Zhou, G. (2010). How much stock return predictability can we expect from an asset pricing model?, *Economic Letters* **108**: 184–186.

Table 1: Forecasting bond returns

	In-Sample				Out-of-Sample		
	β	R^2			R^2_{OoS}		
	Total period	Total period	Expansions	Recessions	Total period	Expansions	Recessions
fwds	0.24 (0.03)	1.41 (0.00)	2.71 (0.00)	-1.67 (1.00)	1.07 (0.05)	2.22 (0.00)	-1.60 (0.63)
tms	0.27 (0.01)	2.01 (0.00)	4.07 (0.00)	-2.83 (1.00)	1.72 (0.01)	3.05 (0.00)	-1.35 (0.58)
ftent	0.18 (0.07)	1.02 (0.02)	2.22 (0.00)	-1.81 (1.00)	0.22 (0.21)	1.75 (0.00)	-3.30 (0.79)
ytent	0.01 (0.01)	1.82 (0.00)	3.45 (0.00)	-2.01 (1.00)	1.38 (0.03)	3.16 (0.00)	-2.74 (0.68)
cs	0.21 (0.01)	1.85 (0.00)	3.31 (0.00)	-1.60 (1.00)	1.46 (0.04)	2.39 (0.00)	-0.69 (0.42)
dfy	0.22 (0.58)	0.13 (0.35)	-0.13 (0.91)	0.74 (0.16)	-1.07 (0.77)	-0.91 (0.94)	-1.44 (0.60)
srr	-0.02 (0.69)	0.05 (0.55)	0.40 (0.09)	-0.75 (0.99)	-0.89 (0.75)	0.90 (0.05)	-5.04 (0.95)
dp	0.00 (0.58)	0.05 (0.55)	0.40 (0.10)	-0.77 (0.99)	-0.98 (0.83)	-1.89 (0.97)	1.13 (0.08)
empl	-0.07 (0.18)	0.29 (0.17)	0.17 (0.25)	0.55 (0.20)	0.01 (0.35)	-0.20 (0.52)	0.52 (0.29)
infl	-0.07 (0.89)	0.01 (0.83)	0.15 (0.25)	-0.33 (0.94)	-0.75 (0.77)	0.64 (0.03)	-3.94 (0.92)
KS					-4.03 (0.70)	-1.42 (0.27)	-10.08 (0.84)
SIC					-2.50 (0.68)	-1.75 (0.45)	-4.23 (0.77)
POOLA					0.71 (0.09)	1.66 (0.00)	-1.50 (0.73)
POOLW					0.70 (0.10)	1.74 (0.00)	-1.73 (0.75)

All R^2 s are in %. For the IS β estimates, we report bootstrap p -values of the Newey-West t -statistic in parantheses. For the IS R^2 s, we also report bootstrap p -values in parantheses. The bootstrap procedure is described in the Appendix. For the OoS R^2 s, we report p -values from the normal distribution of the Clark and West (2007) statistic in parantheses. The IS period is 1953:04-2011:12, and the OoS period is 1978:04-2011:12. We report results for the total period as well as separate results for expansions and recessions. We use the NBER recession dates to define the state of the economy.

Table 2: Power of Clark and West (2007) OoS test

	Total period	Expansions	Recessions
Asymptotic size: 5%			
fwds	0.757	0.725	0.302
tms	0.933	0.920	0.371
ftent	0.603	0.551	0.270
ytent	0.896	0.877	0.362
cs	0.867	0.878	0.281
Asymptotic size: 10%			
fwds	0.841	0.823	0.417
tms	0.964	0.959	0.504
ftent	0.726	0.682	0.393
ytent	0.942	0.933	0.484
cs	0.920	0.925	0.396

The power is computed from 10,000 bootstrap samples where the total sample OLS estimates are used for generating the samples. See the Appendix for details.

Table 3: Size of Clark and West (2007) OoS test

	Total period	Expansions	Recessions
Asymptotic size: 5%			
fwds	0.039	0.041	0.050
tms	0.050	0.043	0.068
ftent	0.039	0.037	0.051
ytent	0.040	0.044	0.060
cs	0.033	0.039	0.044
Asymptotic size: 10%			
fwds	0.077	0.076	0.102
tms	0.095	0.084	0.126
ftent	0.074	0.077	0.097
ytent	0.083	0.081	0.113
cs	0.067	0.074	0.091

The size is computed from 10,000 bootstrap samples where the null of no predictability is imposed when generating the samples. See the Appendix for details.

Table 4: Utility gains from forecasting bond returns

	Δ (annual %)	EXP Δ (annual %)	REC Δ (annual %)
fwds	0.60	1.24	-3.44
tms	1.62	2.06	-1.19
ftent	1.82	2.16	-0.31
ytent	2.26	2.36	1.61
cs	0.79	1.46	-3.53
dfy	-1.59	-1.15	-4.36
srr	-0.27	-0.25	-0.39
dp	-1.60	-1.89	0.20
empl	0.08	0.36	-1.71
infl	-0.49	0.00	-3.58
ks	-1.17	-0.12	-7.84
SIC	-1.69	-1.27	-4.40
POOLA	0.79	1.30	-2.48
POOLW	0.79	1.38	-2.96

Utility gains are the annualized difference between the utility obtained from trading based on predictive regressions and the utility obtained from trading on the predictions of the historical average benchmark. We assume a mean-variance investor with a coefficient of relative risk aversion $\gamma = 5$.

Table 5: Forecasting stock returns

	In-Sample				Out-of-Sample		
	β	R^2			R^2_{OoS}		
	Total period	Total period	Expansions	Recessions	Total period	Expansions	Recessions
fwds	0.22 (0.07)	0.51 (0.07)	0.10 (0.33)	1.63 (0.05)	-0.87 (0.11)	-1.88 (0.23)	2.35 (0.15)
tms	0.21 (0.07)	0.48 (0.07)	0.37 (0.11)	0.79 (0.14)	-1.05 (0.15)	-1.57 (0.21)	0.61 (0.24)
ftent	0.12 (0.22)	0.19 (0.28)	-0.02 (0.69)	0.77 (0.21)	-1.10 (0.50)	-1.25 (0.62)	-0.61 (0.31)
ytent	0.01 (0.11)	0.37 (0.12)	-0.12 (0.93)	1.73 (0.05)	-1.12 (0.32)	-2.26 (0.61)	2.54 (0.07)
cs	0.14 (0.18)	0.33 (0.14)	0.23 (0.20)	0.61 (0.19)	-1.19 (0.24)	-1.27 (0.29)	-0.91 (0.34)
dfy	0.22 (0.63)	0.05 (0.56)	0.01 (0.60)	0.18 (0.41)	-0.76 (0.69)	-0.64 (0.64)	-1.13 (0.63)
srr	-0.09 (0.15)	0.31 (0.13)	0.33 (0.12)	0.25 (0.32)	-2.57 (0.51)	-1.99 (0.44)	-4.39 (0.57)
dp	0.01 (0.16)	0.51 (0.13)	0.15 (0.41)	1.49 (0.07)	-0.92 (0.32)	-1.49 (0.46)	0.93 (0.20)
empl	-0.21 (0.01)	1.05 (0.01)	0.70 (0.03)	2.00 (0.02)	-1.82 (0.52)	-2.19 (0.63)	-0.63 (0.36)
infl	-0.19 (0.76)	0.02 (0.71)	0.21 (0.20)	-0.51 (0.96)	-1.89 (0.82)	-0.58 (0.37)	-6.06 (0.93)
ks					-7.80 (0.86)	-6.12 (0.61)	-13.14 (0.91)
SIC					-4.73 (0.87)	-5.88 (0.95)	-1.05 (0.35)
POOLA					-0.16 (0.27)	-0.34 (0.35)	0.41 (0.31)
POOLW					-0.22 (0.29)	-0.39 (0.37)	0.32 (0.32)

See notes to Table 1.

Table 6: Campbell & Shiller long rate regressions

Total period							
n	β	se	p -value	\bar{R}^2			
24	-0.52	0.44	0.00	1.08			
36	-0.86	0.52	0.00	2.35			
48	-1.28	0.57	0.00	4.55			
60	-1.31	0.61	0.00	4.17			
Expansions vs. recessions							
n	β_{REC}	se	p -value	β_{EXP}	se	p -value	\bar{R}^2
24	0.54	0.50	0.36	-0.75	0.48	0.00	8.19
36	0.32	0.55	0.22	-1.07	0.58	0.00	7.71
48	0.10	0.79	0.25	-1.51	0.62	0.00	8.77
60	0.73	1.15	0.81	-1.65	0.64	0.00	8.66

The upper panel reports results from the regression:

$$y_{t+m}^{(n-m)} - y_t^{(n)} = \alpha + \beta \frac{m}{n-m} \left(y_t^{(n)} - y_t^{(m)} \right) + u_{t+m}$$

while the lower panel reports results from the specification in (16).

Newey-West standard errors (se) are computed with 11 lags. The p -values correspond to the EHT that the slope estimate is 1.

Table 7: Campbell & Shiller short rate regressions

Total period							
n	β	se	p -value	\bar{R}^2			
24	0.24	0.22	0.00	0.88			
36	0.41	0.29	0.04	3.22			
48	0.59	0.27	0.13	7.29			
60	0.75	0.23	0.28	12.47			
Expansions vs. recessions							
n	β_{REC}	se	p -value	β_{EXP}	se	p -value	\bar{R}^2
24	0.77	0.25	0.36	0.13	0.24	0.00	8.01
36	0.65	0.21	0.10	0.37	0.32	0.05	8.11
48	0.74	0.20	0.20	0.55	0.31	0.15	11.42
60	0.99	0.24	0.97	0.70	0.27	0.27	14.35

The upper panel reports results from the regression:

$$\sum_{i=1}^{k-1} \left(1 - \frac{i}{k}\right) \left(y_{t+im}^{(m)} - y_{t+(i-1)m}^{(m)}\right) = \alpha + \beta \left(y_t^{(n)} - y_t^{(m)}\right) + u_{t+(k-1)m}$$

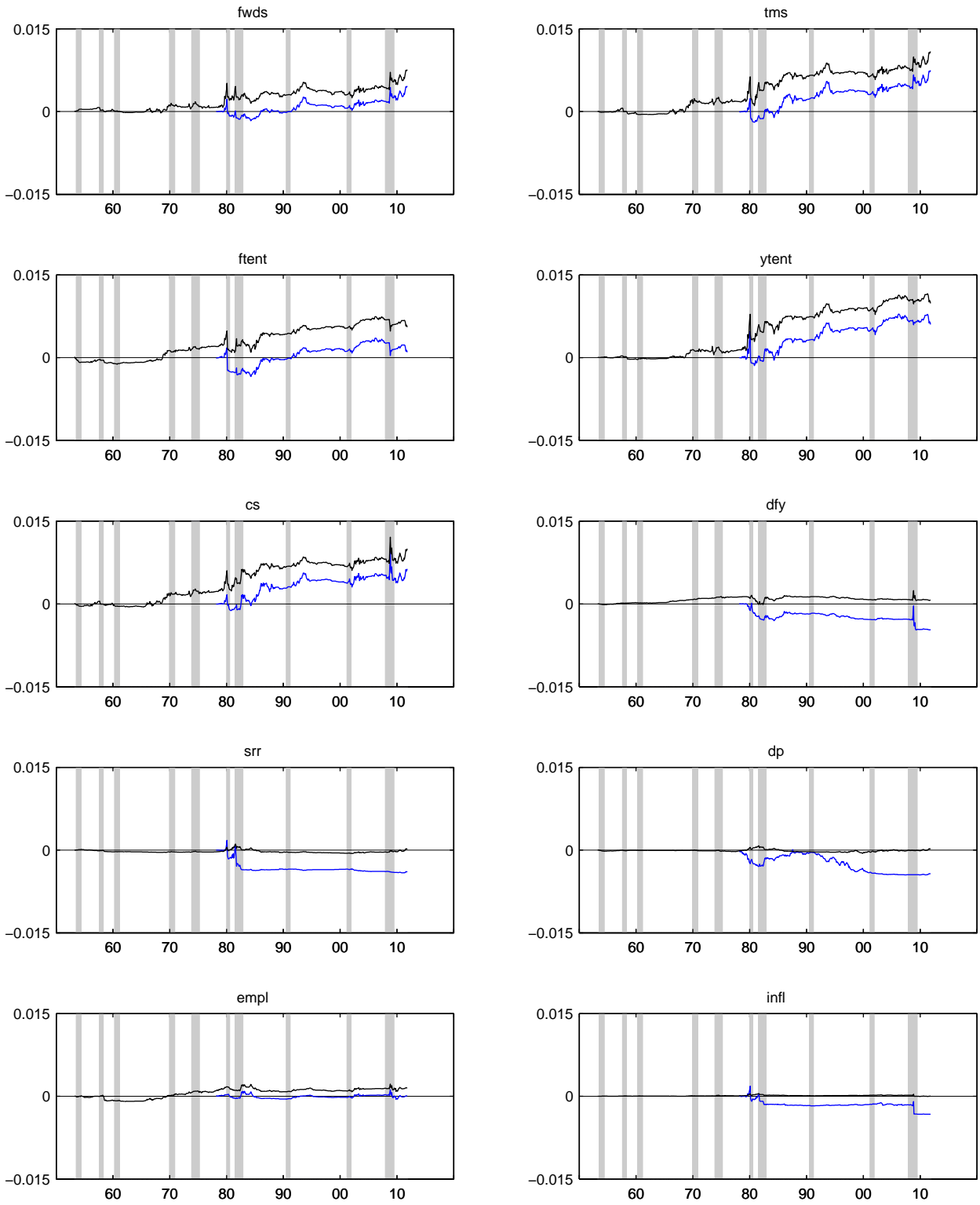
while the lower panel reports results from the specification in (17). Newey-West standard errors (se) are computed with $n - m - 1$ lags. The p -values correspond to the EHT that the slope estimate is 1.

Table 8: Average annualized volatility of short rate

	Total	Expansions	Recessions
GARCH(1,1)	0.014	0.012	0.022
Realized 12 m.	0.020	0.018	0.031

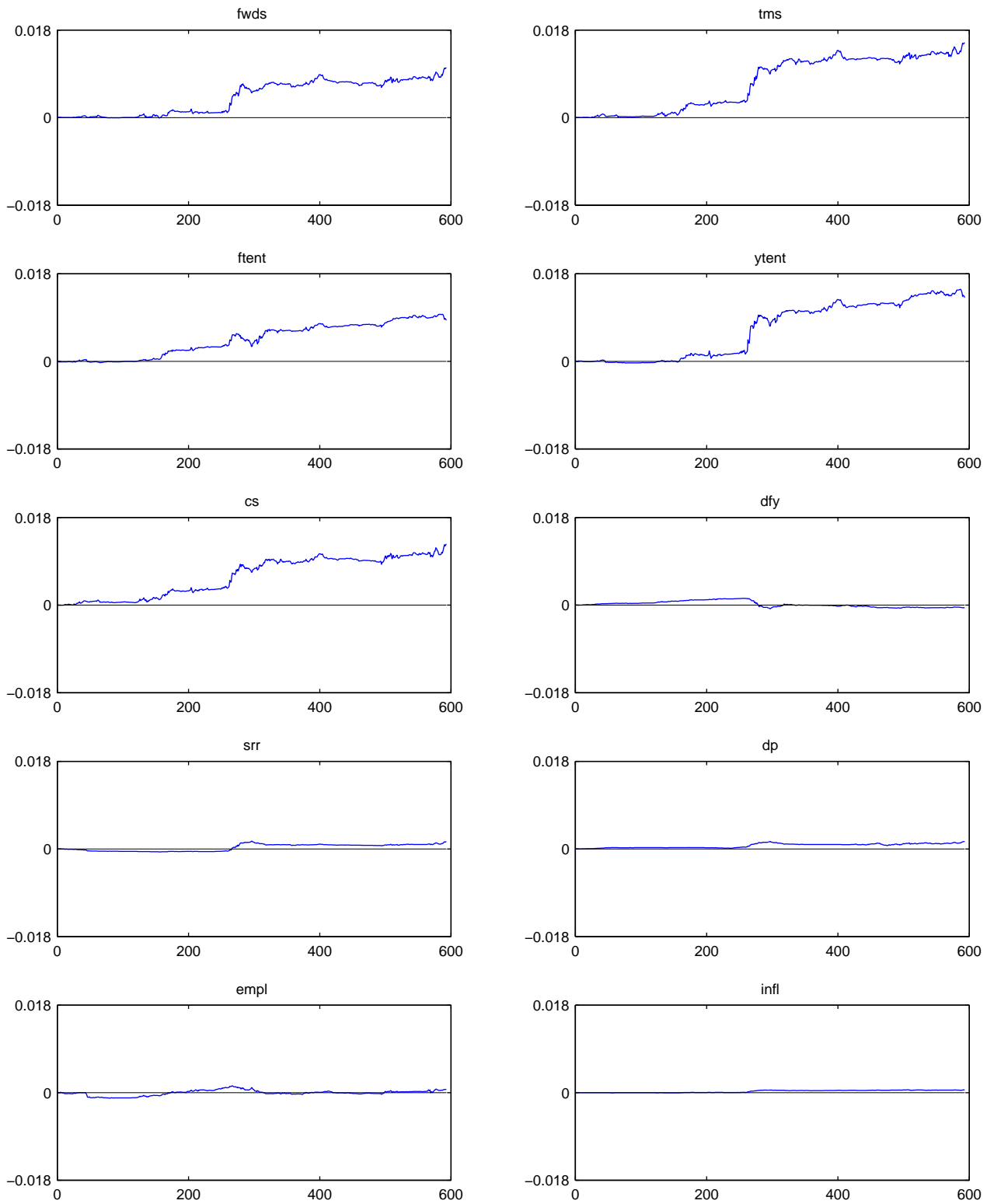
The first row is the average annualized volatility of the 1 year bond as measured by a GARCH(1,1) model fitted to the residuals from an AR(1) model. The second row uses a 12 month rolling window of the 1 year bond to estimate the volatility.

Figure 1: Difference in cumulative squared forecast errors



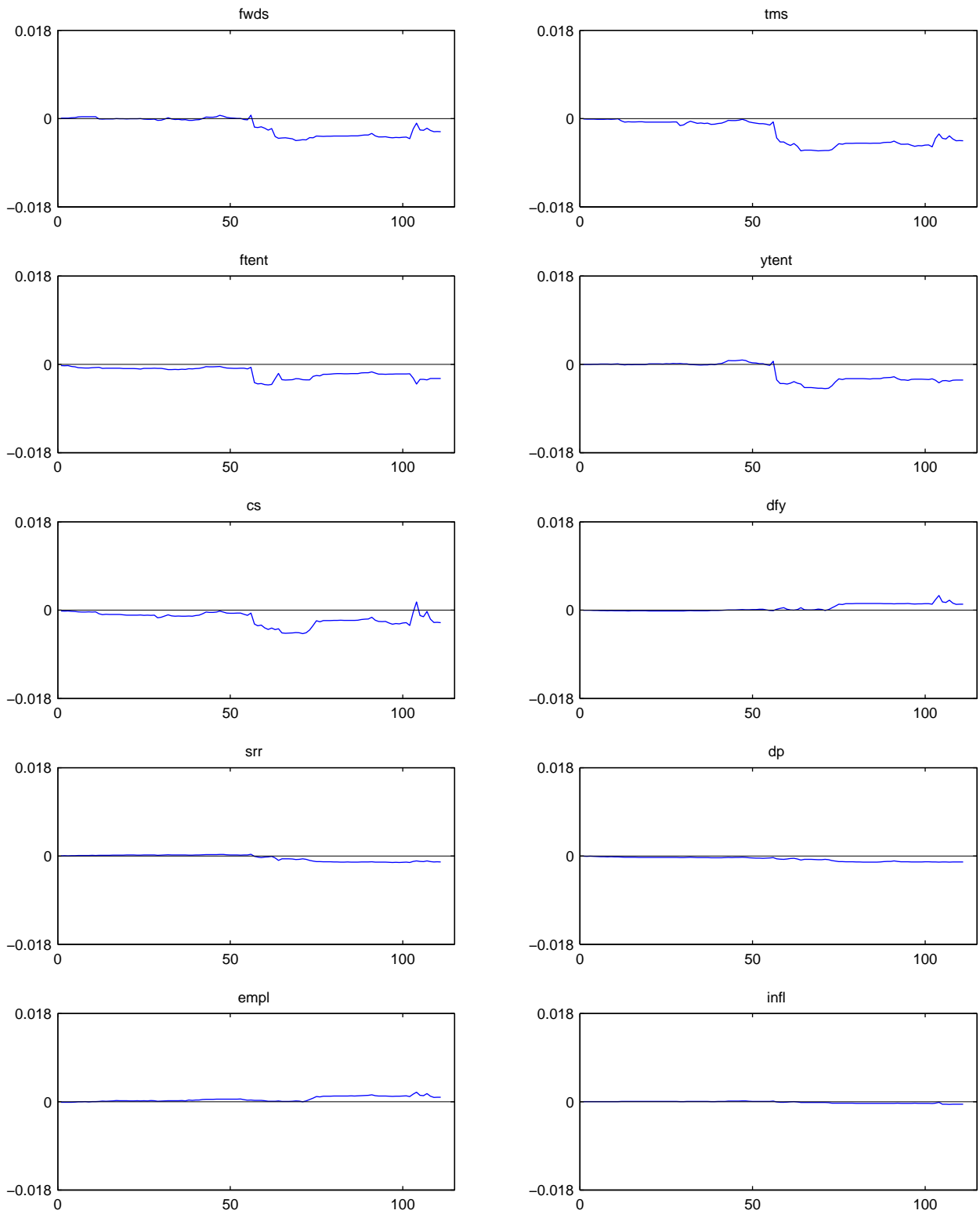
The black curve is based on IS forecast errors while the blue curve is based on OoS forecast errors. The gray bars indicate recession periods as reported by the NBER.

Figure 2: Difference in cumulative squared forecast errors in expansions



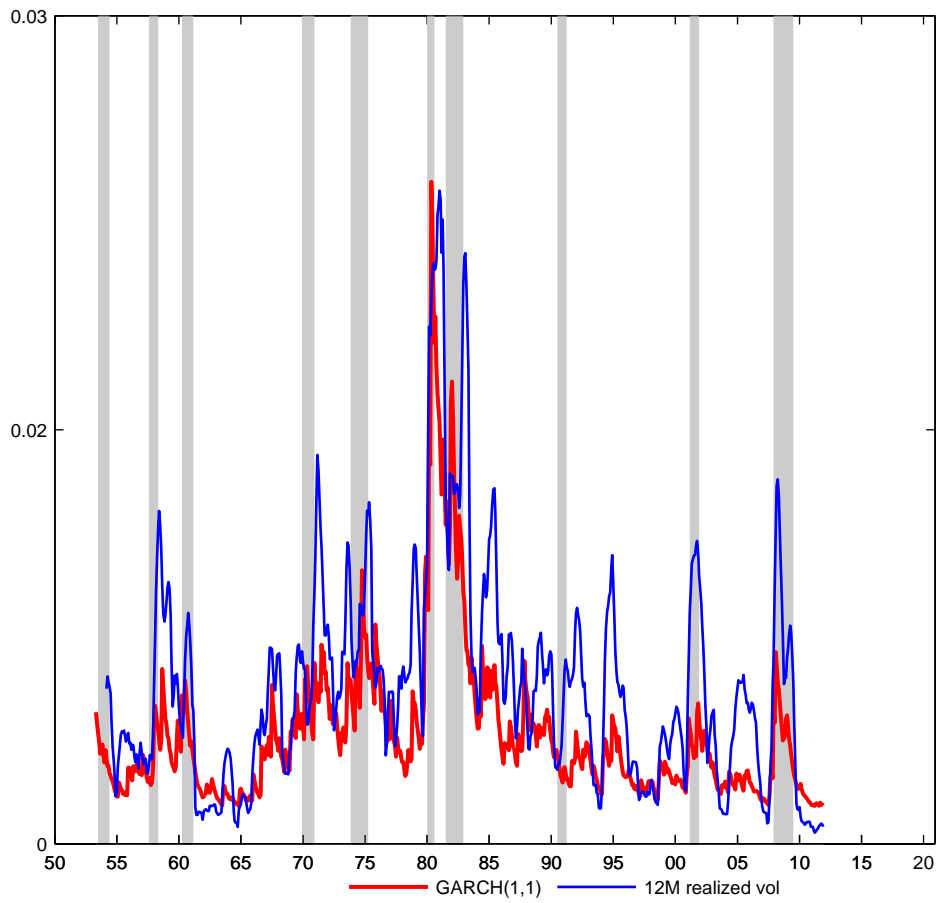
The curves are based on IS forecast errors from expansions.

Figure 3: Difference in cumulative squared forecast errors in recessions



The curves are based on IS forecast errors from recessions.

Figure 4: Volatility of short rate



The blue curve is the volatility predictions (square root of variance predictions) from a GARCH(1,1) model fitted to the residuals from an AR(1) model of the 1-year bond. The red curve is the realized standard deviation of the 1-year bond calculated using a 12-month rolling window.

Appendix

A. Regressor descriptives

We report the AR(1) coefficient θ , the corresponding Dickey-Fuller test, the correlation coefficient between shocks to the regressor and shocks to returns ρ_{uv} , the implicit Stambaugh (1999) bias $E(\hat{\beta} - \beta)$, as well as the Augmented-Dickey Fuller test with lag length selected by the SIC. We compute the implied Stambaugh bias as $E(\hat{\beta} - \beta) \approx -[(1 + 3\theta) / T] \sigma_{uv} / \sigma_v^2$, where σ_{uv} is the covariance between shocks to the predictor and shocks to the return regression, σ_v^2 is the variance of the error in the DGP of the predictor, and T is the sample size.

Table A.1: Regressor descriptives

	θ	DF	ρ_{uv}	$E(\hat{\beta} - \beta)$	ADF
fwds	0.915	-5.572	0.071	-0.002	-4.829
tms	0.956	-3.970	-0.315	0.011	-4.468
ftent	0.860	-7.312	-0.101	0.002	-4.803
ytent	0.936	-4.771	-0.084	0.000	-5.352
cs	0.975	-2.948	0.143	-0.006	-3.376
dfy	0.969	-3.343	0.246	-0.034	-3.855
srr	0.983	-2.362	0.044	-0.001	-2.362
dp	0.993	-1.795	-0.125	0.000	-1.795
empl	0.984	-2.442	-0.071	0.003	-3.264
infl	0.605	-13.124	-0.129	0.005	-2.813

Critical values for the DF/ADF test are (10%, 5%, 1%) = (-2.594, -2.871, and -3.458).

Most of the regressors are highly persistent, and we cannot reject the null of a unit root in dp and srr at any conventional significance level. As nonstationarity is the null, there could be problems with power in the DF/ADF test. However, unreported results show that the KPSS test where the alternative is a unit root rejects for both variables at least at the 5% level.

Generally, the contemporaneous correlations between shocks to the regressors and shocks to returns are not too high. Combined with the high number of observations due to the monthly sampling frequency, this makes the implied Stambaugh bias in the predictive regressions very small.¹⁶ There are several ways to correct for the bias (see e.g. Lewellen (2004)), but since the goal of this study is assessment of out-of-sample predictive power, we refrain from it here.

¹⁶Note that in computing $E(\hat{\beta} - \beta)$, we have just plugged in the estimate of the AR(1) coefficient, $\hat{\theta}$, even though we should use the true value. We do this just to get an idea of the magnitude of the bias, but of course if the estimate $\hat{\theta}$ is downward biased, this will reduce the Stambaugh bias that we report.

We do, however, compute bootstrap p -values taking the persistent regressors as well as the cross-correlation structure into account. For details see the next section.

B. The bootstrap

B.1. IS p -values

We conduct a residual-based bootstrap using the following system

$$r_{t+1} = \alpha + \beta x_t + u_{t+1}, \quad (17)$$

$$x_{t+1} = \delta + \theta x_t + v_{t+1}. \quad (18)$$

We first use OLS to estimate the system equation by equation. This provides us with a matrix of (unrestricted) residuals. We then classify each pair of residuals $(\hat{u}_{t+1}, \hat{v}_{t+1})$ as being either from a recession or an expansion based on the value of the NBER dummy in period t , where the prediction was made. This results in a ‘recession matrix’ of residuals and an ‘expansion matrix’ of residuals.¹⁷ We then generate 10,000 samples of returns and predictors by simulating the above system while imposing the null of no predictability by letting $\beta = 0$. The residuals for each sample are drawn in the following way. Each period, we randomly draw with replacement residuals $(\hat{u}_{t+1}, \hat{v}_{t+1})$ from either the recession matrix or the expansion matrix based on the value of the NBER dummy in period t . We thus make a simplifying assumption by fixing the original business cycle structure. To randomize the starting point, we initiate the system using the mean of the regressor followed by a pre-run of 200 time periods we afterwards throw away. Based on the 10,000 bootstrap samples, we compute two-sided p -values of the Newey-West t -test with 0 lags and one-sided upper-tail p -values of the IS R^2 s.

B.2. Size and power properties of the OoS tests

We also use the system in (17)-(18) to compute size and power properties of the Clark and West (2007) OoS test. The way we construct bootstrap samples follow the procedure described

¹⁷Morley et al. (2013) implement a similar bootstrap for simulating from a Markov-Switching model.

above. When computing the size of the test, we impose $\beta = 0$. When computing the power of the test, we do not impose $\beta = 0$.

C. Including the 1970s in the OoS period

Table A.2: Forecasting bond returns: OoS period starts in 1968

	In-Sample				Out-of-Sample		
	β	R^2			R^2_{OoS}		
	Total period	Total period	Exp.	Rec.	Total period	Exp.	Rec.
fwds	0.24 (0.03)	1.41 (0.00)	2.71 (0.00)	-1.67 (1.00)	0.86 (0.06)	1.88 (0.00)	-1.43 (0.64)
tms	0.27 (0.01)	2.01 (0.00)	4.07 (0.00)	-2.83 (1.00)	1.46 (0.01)	2.75 (0.00)	-1.42 (0.62)
ftent	0.18 (0.07)	1.02 (0.02)	2.22 (0.00)	-1.81 (1.00)	0.20 (0.19)	1.59 (0.00)	-2.92 (0.80)
ytent	0.01 (0.01)	1.82 (0.00)	3.45 (0.00)	-2.01 (1.00)	1.22 (0.02)	2.90 (0.00)	-2.54 (0.70)
cs	0.21 (0.01)	1.85 (0.00)	3.31 (0.00)	-1.60 (1.00)	1.24 (0.03)	2.10 (0.00)	-0.71 (0.43)
dfy	0.22 (0.58)	0.13 (0.35)	-0.13 (0.91)	0.74 (0.16)	-0.95 (0.66)	-0.63 (0.68)	-1.68 (0.60)
srr	-0.02 (0.69)	0.05 (0.55)	0.40 (0.09)	-0.75 (0.99)	-0.99 (0.81)	0.70 (0.07)	-4.76 (0.96)
dp	0.00 (0.58)	0.05 (0.55)	0.40 (0.10)	-0.77 (0.99)	-0.62 (0.53)	-1.61 (0.91)	1.58 (0.02)
empl	-0.07 (0.18)	0.29 (0.17)	0.17 (0.25)	0.55 (0.20)	0.09 (0.27)	-0.07 (0.39)	0.45 (0.28)
infl	-0.07 (0.89)	0.01 (0.83)	0.15 (0.25)	-0.33 (0.94)	-0.96 (0.74)	0.61 (0.04)	-4.47 (0.94)
ks					-5.22 (0.65)	-1.82 (0.16)	-12.84 (0.88)
SIC					-2.56 (0.70)	-1.66 (0.42)	-4.58 (0.81)
POOLA					0.67 (0.07)	1.60 (0.00)	-1.40 (0.75)
POOLW					0.65 (0.08)	1.66 (0.00)	-1.60 (0.76)

Table A.3: Forecasting stock returns: OoS period starts in 1968

	In-Sample				Out-of-Sample		
	β	R^2			R^2_{OoS}		
	Total period	Total period	Exp.	Rec.	Total period	Exp.	Rec.
fwds	0.22 (0.07)	0.51 (0.07)	0.10 (0.33)	1.63 (0.05)	-0.16 (0.03)	-2.18 (0.24)	4.79 (0.02)
tms	0.21 (0.07)	0.48 (0.07)	0.37 (0.11)	0.79 (0.14)	-1.04 (0.03)	-2.64 (0.12)	2.87 (0.04)
ftent	0.12 (0.22)	0.19 (0.28)	-0.02 (0.69)	0.77 (0.21)	-0.54 (0.14)	-1.23 (0.35)	1.15 (0.09)
ytent	0.01 (0.11)	0.37 (0.12)	-0.12 (0.93)	1.73 (0.05)	-0.75 (0.08)	-2.27 (0.30)	2.96 (0.03)
cs	0.14 (0.18)	0.33 (0.14)	0.23 (0.20)	0.61 (0.19)	-1.17 (0.04)	-2.85 (0.20)	2.95 (0.03)
dfy	0.22 (0.63)	0.05 (0.56)	0.01 (0.60)	0.18 (0.41)	-0.77 (0.57)	-1.11 (0.71)	0.06 (0.29)
srr	-0.09 (0.15)	0.31 (0.13)	0.33 (0.12)	0.25 (0.32)	-1.48 (0.11)	-0.68 (0.07)	-3.42 (0.35)
dp	0.01 (0.16)	0.51 (0.13)	0.15 (0.41)	1.49 (0.07)	-0.81 (0.25)	-1.16 (0.31)	0.06 (0.33)
empl	-0.21 (0.01)	1.05 (0.01)	0.70 (0.03)	2.00 (0.02)	-0.24 (0.07)	-1.11 (0.21)	1.88 (0.08)
infl	-0.19 (0.76)	0.02 (0.71)	0.21 (0.20)	-0.51 (0.96)	-0.98 (0.30)	0.17 (0.04)	-3.78 (0.58)
ks					-5.66 (0.42)	-4.79 (0.23)	-7.80 (0.65)
SIC					-3.23 (0.37)	-4.77 (0.72)	0.54 (0.17)
POOLA					0.81 (0.02)	0.15 (0.10)	2.42 (0.04)
POOLW					0.78 (0.02)	0.05 (0.11)	2.55 (0.04)

Research Papers 2013



- 2012-54: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperature Trends and Co-Shifting: A Shifting Mean Vector Autoregressive Analysis
- 2012-55: Daniel J. Nordman, Helle Bunzel and Soumendra N. Lahiri: A Non-standard Empirical Likelihood for Time Series
- 2012-56: Robert F. Engle, Martin Klint Hansen and Asger Lunde: And Now, The Rest of the News: Volatility and Firm Specific News Arrival
- 2012-57: Jean Jacod and Mark Podolskij: A test for the rank of the volatility process: the random perturbation approach
- 2012-58: Tom Engsted and Thomas Q. Pedersen: Predicting returns and rent growth in the housing market using the rent-to-price ratio: Evidence from the OECD countries
- 2013-01: Mikko S. Pakkanen: Limit theorems for power variations of ambit fields driven by white noise
- 2013-02: Almut E. D. Veraart and Luitgard A. M. Veraart: Risk premia in energy markets
- 2013-03: Stefano Grassi and Paolo Santucci de Magistris: It's all about volatility (of volatility): evidence from a two-factor stochastic volatility model
- 2013-04: Tom Engsted and Thomas Q. Pedersen: Housing market volatility in the OECD area: Evidence from VAR based return decompositions
- 2013-05: Søren Johansen and Bent Nielsen: Asymptotic analysis of the Forward Search
- 2013-06: Debopam Bhattacharya, Pascaline Dupasand Shin Kanaya: Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets
- 2013-07: Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns
- 2013-08: Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation
- 2013-09: Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox
- 2013-10: Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models
- 2013-11: Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis
- 2013-12: Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications
- 2013-13: Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions