MPRA
Munich Personal RePEc Archive

When does the yield curve contain predictive power? Evidence from a data-rich environment

Hännikäinen, Jari

University of Tampere

4 April 2016

Online at https://mpra.ub.uni-muenchen.de/70489/
MPRA Paper No. 70489, posted 06 Apr 2016 05:01 UTC

TAMPERE ECONOMIC WORKING PAPERS

WHEN DOES THE YIELD CURVE CONTAIN PREDICTIVE POWER?
EVIDENCE FROM A DATA-RICH ENVIRONMENT

Jari Hännikäinen

Working Paper 103
April 2016

SCHOOL OF MANAGEMENT

FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 978-952-03-0108-8

When does the yield curve contain predictive power? Evidence from a data-rich environment*

Jari Hännikäinen†

School of Management, University of Tampere, Tampere, Finland

Abstract

This paper analyzes the predictive content of the level, slope and curvature of the yield curve for U.S. real activity in a data-rich environment. We find that the slope contains predictive power, but the level and curvature are not successful leading indicators. The predictive power of each of the yield curve factors fluctuates over time. The results show that economic conditions matter for the predictive ability of the slope. In particular, inflation persistence emerges as a key variable that affects the predictive content of the slope. The slope tends to forecast output growth better when inflation is highly persistent.

Keywords: yield curve, factor model, data-rich environment, forecasting, macroeconomic regimes, conditional predictive ability

JEL codes: C53, C55, E43, E44, E47, E52

*I would like to thank Gene Ambrocio, Matti Hovi, Jani-Petri Laamanen, Juuso Vataja and participants of the XXXVIII Annual Meeting of the Finnish Economic Association for helpful comments and suggestions and Koen Inghelbrecht for providing the data for the macroeconomic regimes. Financial support from the OP-Pohjola Group Research Foundation is gratefully acknowledged.

†School of Management, University of Tampere, Kanslerinrinne 1, FI-33014 Tampere, Finland. E-mail: jari.hannikainen@uta.fi

1. Introduction

Economists have long understood that the behavior of the yield curve changes across the business cycle. In recessions, short-term interest rates tend to be low because the Federal Reserve lowers the policy rate in order to boost economic activity. The long-term rates tend to be high relative to the short-term rates because the Fed is expected to raise the short-term rate in the future when the economic conditions get better. The slope of the yield curve, or the term spread, is thus positive in recessions. In contrast, the Fed raises the short-term rate when the economy is overheating or facing inflationary pressures. A slowdown in real activity typically follows such a policy with a lag. Monetary policy tightening will raise both short- and long-term interest rates. If monetary policy is expected to ease once economic activity or inflation declines, the short-term rate is likely to rise more than the long-term rate. Therefore, the yield curve tends to flatten or even invert before slowdowns. This discussion suggests that the short-term rate tends to be procyclical, and the slope of the yield curve tends to be countercyclical. Based on this observation, economists have argued that the yield curve might tell us something about future real activity.

Since the late 1980s, a large amount of literature has analyzed the predictive content of the yield curve (see, e.g., Abdymomunov, 2013; Aguiar-Conraria et al., 2012; Bernanke, 1990; Bernanke and Blinder, 1992; Estrella and Hardouvelis, 1991; Hamilton and Kim, 2002; Harvey, 1988; Mody and Taylor, 2003).[1] This literature has found that the yield curve contains substantial predictive power. In particular, the slope of the yield curve has been identified as one of the most informative leading indicators for U.S. real economic activity (see, e.g., Stock and Watson, 2003). The relationship between the slope of the yield curve and future real activity is positive; i.e., a high slope precedes periods of strong growth, while a low slope indicates weak activity in the fu-

[1] For a comprehensive survey of the literature, see Wheelock and Wohar (2009).

ture. Other elements of the yield curve also contain information about subsequent real activity. For instance, Ang et al. (2006) find that the short-term rate predicts U.S. real GDP growth.

Today, much evidence shows that the predictive power of the yield curve fluctuates over time (Estrella et al., 2003; Gertler and Lown, 1999; Mody and Taylor, 2003; Rossi and Sekhposyan, 2011; Stock and Watson, 2003). For example, many studies have found that the ability of the slope of the yield curve to predict U.S. real growth has largely disappeared since the mid-1980s. There exists no universally agreed-upon explanation why the predictive power of the yield curve varies over time. However, most researchers point out that monetary policy and the yield curve are closely connected. For instance, Giacomini and Rossi (2006) argue that the changes in the predictive content of the yield curve can be linked to changes in the monetary policy behavior of the Fed. They show that the reliability of the yield curve as a predictor of output growth has changed during the Burns-Miller and Volker monetary policy regimes. In a sequence of papers, Bordo and Haubrich (2004, 2008a, 2008b) suggest that the credibility of the monetary policy is the key determinant of the predictive power of the yield curve. Using a very long data sample from 1875 to 1997, they find that the slope of the yield curve tends to forecast output growth particularly well when the credibility of the monetary policy is low, i.e., when inflation is highly persistent. However, using the same data sample as Bordo and Haubrich (2004, 2008a, 2008b) but more flexible methods, Benati and Goodhart (2008) confirm that the predictive power of the slope of the yield curve fluctuates over time, but these changes do not closely match changes in inflation persistence. In a recent paper, Hännikäinen (2015) shows that the real-time predictive content of the slope of the yield curve for U.S. industrial production growth has changed since the beginning of the zero lower bound (ZLB) and unconventional monetary policy period in December 2008. The beginning of the ZLB/unconventional monetary policy era represents a fundamental change in U.S. monetary policy. Thus, the results reported in Hännikäinen (2015) provide evidence supporting the view that changes in the monetary policy regime affect the predictive ability of the yield curve. There are also other explanations for the apparent changes in the predictive power. D'Agostino et al. (2006) argue that the reduced informativeness of the yield curve in recent years is due to the increased stability of U.S. output growth and other key macroeconomic variables since the mid-1980s. When the macroeconomic variable to be forecast is not volatile, simple benchmark forecasting models, like low order autoregressive (AR) models, produce accurate forecasts. In such a case, it is very challenging to find leading indicators that contain marginal predictive power over and above that already encoded in the lagged values of the series to be forecast.

In this paper, we examine the predictive power of the entire yield curve for U.S. industrial production growth. We extract the level, slope and curvature of the yield curve using the dynamic Nelson-Siegel model developed by Diebold and Li (2006). Unlike the vast majority of previous studies, we explore the out-of-sample predictive content of each of the three components of the yield curve in a data-rich environment using factor models. The standard practice in the extant literature is to analyze the predictive power of the yield curve over and above that in the past values of output growth using AR models. There are two reasons why we prefer factor models to AR models. First, factor models provide a parsimonious way to study the crucial issue of whether the components of the yield curve contain predictive information which is not already encoded in other macroeconomic variables. Second, factor models produce substantially more accurate industrial production forecasts than simple AR models (see, e.g., Bernanke and Boivin, 2003; Clements, 2015; Stock and Watson, 2002a, 2002b). We also pay attention to time variations in the predictive power over time. Finally, and most importantly, we investigate whether the forecasting ability of each of the components of the yield curve can be linked to economic conditions. Following recent papers by Dotsey et al. (2015), Hännikäinen (2015) and Ng and Wright (2013), we employ the test of equal conditional predictive ability developed by Giacomini and White (2006). The novelty of the Giacomini and White (2006) test is that it allows tests of forecast accuracy conditional on a set of possible explanatory variables. Thus, it enables us to analyze, for instance, how inflation persistence, output volatility, recessions and monetary policy regimes affect the reliability of the yield curve as a predictor of real activity. To the best of our knowledge, no other paper has linked the predictive power of the yield curve to economic conditions in a systematic way. Our paper is intended to bridge this gap.

Our main findings can be summarized as follows. First, the slope of the yield curve is a better predictor of real activity than the level or curvature of the yield curve. We find that the slope contains predictive power for industrial production growth in a data-rich environment. In contrast, the level and curvature perform poorly in the out-of-sample forecasting exercise, and they are typically uninformative about future output growth. Second, the results reveal that the slope estimated from the entire yield curve outperforms the standard empirical slope (i.e., the difference between the 10-year and 3-month yields). This finding implies that the additional information in the entire yield curve helps improve forecast accuracy. Third, the predictive power of the yield curve components varies over time. For instance, the slope was a particularly informative leading indicator in the late 1970s and early 1980s, but it performed poorly in the latter half of the 1980s and the early 1990s. Fourth, the predictive ability of the slope seems to depend on the state of the economy. The slope tends to forecast output growth more accurately when inflation is highly persistent or when inflation is not highly volatile.

The remainder of the paper is organized as follows. Section 2 describes the methodology we employ in the out-of-sample forecasting exercise. Section 3 introduces the data, and Section 4 presents the empirical results. Section 5 contains concluding remarks. Appendix A provides a detailed description of the dataset.

2. Methodology

In this section, we describe the econometric methodologies used in the out-of-sample forecasting exercise. The purpose of this study is to examine whether different elements of the yield curve contain predictive power for U.S. real economic activity.

2.1. Dynamic Nelson-Siegel yield curve model

We extract the latent level, slope and curvature of the yield curve using the dynamic Nelson-Siegel yield curve model introduced by Diebold and Li (2006):

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right), \tag{1}$$

where $y_t(\tau)$ is the yield of a zero-coupon bond with maturity τ; β_{1t}, β_{2t} and β_{3t} are three time-varying latent factors, and λ_t is the exponential decay rate responsible for fitting the yield curve at different maturities. A central feature of the dynamic Nelson-Siegel model is that the latent factors β_{1t}, β_{2t} and β_{3t} can be interpreted as the level, slope and curvature of the yield curve, respectively. An increase in β_{1t} increases all yields by the same amount regardless of their maturity. Thus, β_{1t} determines the level of the yield curve. In contrast, β_{2t} is related to the slope of the yield curve because a change in β_{2t} leads to an unequal change in short- and long-term yields. Finally, an increase in β_{3t} increases the medium-term yields to a greater extent than short- and long-term yields. Therefore, β_{3t} affects the curvature of the yield curve. We estimate the latent factors using the two-step procedure proposed by Diebold and Li (2006). That is, we first fix λ_t at 0.0609[2] for all t and then estimate the factors by OLS.

[2] As discussed in Diebold and Li (2006), λ_t determines the maturity at which the loading on the curvature factor reaches its maximum. It is often assumed that the curvature is maximized either at 2- or 3-year maturity. The loading on the curvature factor is maximized at exactly 30 months, which is the average between 2- and 3-year maturities, if λ_t is set to 0.0609.

2.2. Forecasting models

We examine the predictive content of the yield curve using factor models. Factor models are particularly well-suited to deal with a large number of candidate predictors. The key insight of this method is that predictors are often strongly correlated, and thus, the information in a large set of predictors can be summarized by a handful of unobserved factors. As shown by Stock and Watson (2002a), these factors can be consistently estimated by principal components. Factor models typically produce more accurate macroeconomic forecasts than alternative forecasting models, such as AR and vector autoregressive models. As a consequence, factor models have become popular in macroeconomic forecasting in recent years.[3]

Our forecasting model is the following linear, h-step-ahead factor model, augmented with an element of the yield curve:

$$y_{t+h}^{h} = \alpha_h + \sum_{j=1}^{m}\sum_{i=1}^{k} \beta_{hij} \hat{F}_{i,t-j+1} + \sum_{j=1}^{p} \gamma_{hj} y_{t-j+1} + \phi_h z_t + \varepsilon_{t+h}^{h}, \tag{2}$$

where the dependent variable and the lagged dependent variable are $y_{t+h}^{h} = (1200/h)$ $ln(IP_{t+h}/IP_t)$ and $y_t = 400ln(IP_t/IP_{t-1})$, respectively, IP_t is the industrial production at month t, $\hat{F}_{i,t}$ is the ith principal component constructed from the large set of predictors, z_t is either the level, slope or curvature of the yield curve, and ε_{t+h}^{h} is the forecast error. The constant term is explicitly included in the forecasting model (2), and the subscripts h indicate that the parameters are forecast horizon specific.

2.3. Out-of-sample forecasting exercise

We evaluate the forecasting performance of the components of the yield curve in a pseudo-out-of-sample forecasting exercise. In this exercise, the forecast horizon h is

[3] See Luciani (2014) and Stock and Watson (2011) for recent surveys of factor models.

chosen such that we forecast economic activity one, two, three, and four quarters ahead (i.e., h = 3, 6, 9, and 12). At each forecast origin, the factors are extracted by principal components using the whole data sample available at that date. In contrast, the parameters of the forecasting model (2) are re-estimated at each forecast origin by OLS using a rolling window of 120 observations, corresponding to 10 years of monthly data.[4] We restrict the forecasting model such that it contains only contemporaneous values of factors $\hat{F}_t$ (i.e., $m = 1$) and the components of the yield curve. The number of autoregressive lags p and the number of factors k are determined by the data. We consider several variants of the forecasting model (2). The first, denoted by DIAR, includes contemporaneous factors and an element of the yield curve and lags of y_t, with k and p selected by minimizing the Bayesian information criterion (BIC), with $1 \leq k \leq 4$, and $0 \leq p \leq 6$. Thus, the smallest model that BIC can choose includes only a single contemporaneous factor and a contemporaneous value of the yield curve. The second variant, denoted by K1 and K2, includes a fixed number of factors (k = 1 or 2) and a contemporaneous value of the yield curve.[5] The third variant is a simple AR model augmented with yield curve information. Thus, the model includes the contemporaneous value of an element of the yield curve and lags of y_t. Again, the number of autoregressive lags is selected by BIC, with $0 \leq p \leq 6$. We denote this variant by AR in the following tables.

2.4. Forecast accuracy

A standard way to quantify out-of-sample performance is to compare the forecasting accuracy of a candidate forecast model relative to that of a benchmark model. In our framework, natural benchmark models are obtained by excluding the yield curve

[4] The Giacomini and White (2006) tests (discussed below) require limited memory estimators and thus rule out the recursive estimation scheme.

[5] We also considered models with three or four factors. The results for these two specifications are qualitatively similar to those reported for the K1 and K2 models. However, the K1 and K2 models produce more accurate out-of-sample forecasts (cf. Stock and Watson, 2002b). To save space, we report the results for the K1 and K2 specifications only.

information from the forecasting model (2). Therefore, by comparing the accuracy of the forecasting model that includes a component of the yield curve and the benchmark model, we investigate the marginal predictive power of that element of the yield curve. To facilitate comparisons between the yield curve model and the benchmark, we report the results in terms of their relative mean squared forecast error (MSFE), which is the ratio of the MSFE from the yield curve forecasting model over the MSFE from the benchmark. Values of the relative MSFE below (above) one indicate that the forecasts produced by the yield curve model are more (less) accurate than the forecasts produced by the benchmark model. The statistical significance is evaluated using the one-sided Giacomini and White (2006) test of equal unconditional predictive ability.

We also report the fraction of observations for which the yield curve model generates a smaller absolute forecast error than the benchmark. The reason for this exercise is twofold. First, it allows us to consider whether the yield curve forecasting model qualitatively outperforms the benchmark model. Second, because the MSFE measure gives more weight to large errors, a few extreme forecast errors might bias the MSFE results. The sign statistic, on the other hand, puts smaller weight on outliers and, thus, provides usefull robustness check. We evaluate the statistical significance using the Diebold and Mariano (1995) sign test.

2.4.1. Fluctuation tests

The Giacomini and White (2006) unconditional test tells us whether the forecasts are statistically significantly different from one another on average over the whole out-of-sample period. The unconditional test implicitly assumes that the relative performance of the forecasting models remains constant over time. Giacomini and Rossi (2010) point out that the relative forecasting performance may change over time in an unstable environment. In such a case, average relative performance over the whole out-of-sample period may hide important information and even lead to incorrect conclusions.

In practice, the relative performance of the forecasting models often fluctuate over time (see, e.g., D'Agostino and Surico, 2012; Ng and Wright, 2013; Rossi, 2013). We analyze time variations in the relative forecasting performance using methods developed by Giacomini and Rossi (2010). Their fluctuation test examines whether the local relative performance of two forecasting methods is equal at each point in time. The fluctuation test is equivalent to the Giacomini and White (2006) unconditional test computed over a rolling out-of-sample window. To be more specific, the unconditional test statistic is computed for each rolling window. If the maximum test statistic exceeds the critical value computed by Giacomini and Rossi (2010), the null of equal accuracy between the two forecasting methods at each point in time is rejected.

2.4.2. Conditional predictive ability tests

Finally, we investigate whether the forecasting ability of each of the components of the yield curve can be linked to economic factors. We employ the Giacomini and White (2006) test of equal conditional predictive ability, which provides a simple method for analyzing whether the state of the economy affects the accuracy of the yield curve forecasting model relative to that of the benchmark model. It is important to emphasize that the conditional Giacomini and White test is a marginal test. This means that the conditional test tells us only whether conditioning on a certain variable significantly improves the accuracy of one forecast relative to another, not whether the forecast is actually more accurate. However, we can infer which model produces better forecasts from an auxiliary regression. In this regression, the difference in the squared h-step-ahead forecast errors between the benchmark model and the yield curve model, denoted by δ_{t+h}, is regressed on the conditioning variable x_t:

$$\delta_{t+h} = \beta_0 + \beta_1 x_t + e_{t+h}. \tag{3}$$

The size and sign of the coefficients of this regression determine which forecasting model yields more accurate forecasts. We are particularly interested in the β_1 coefficient. This coefficient tells us how conditioning on variable x_t affects the relative forecasting performance of the two models. For instance, a positive and statistically significant β_1 indicates that conditioning on x_t improves the yield curve forecast relative to the benchmark. Otherwise stated, the yield curve contains more marginal predictive power when the variable x_t is large.

3. Data

The data used in this paper come from two sources. The yield data are taken from the Gürkaynak et al. (2007) database. This database is publicly available on the Federal Reserve's webpage,[6] and it is updated regularly. The macroeconomic data are obtained from the FRED-MD database, which is compiled and maintained by the Federal Reserve Bank of St. Louis.[7] The FRED-MD database is a new publicly available monthly database for macroeconomic research. A detailed description of this dataset can be found in McCracken and Ng (2015).

3.1. Yield curve data

We use monthly zero-coupon U.S. Treasury security yields with maturities of 3, 6, 9, 12, 24, 36, 48, 60, 72, 84, 96, 108, and 120 months covering the period from June 1961 to April 2015. The data refer to the yields on the last day of each month. All yields are continuously compounded. Summary statistics of the yields are presented in Table 1. A number of important features of the yield curve can be seen from this table. First, the yield curve is typically upward sloping; i.e., the long-term rates are typically higher than the short-term rates. Second, the short end of the yield curve is more volatile

[6] http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html

[7] https://research.stlouisfed.org/econ/mccracken/fred-databases

than the long end. Third, the long-term rates are more persistent than the short-term rates.

We extract the latent level, slope and curvature of the yield curve from the raw yield data using the dynamic extension of the Nelson-Siegel (1987) model introduced by Diebold and Li (2006). We also consider widely used empirical proxies for the three components of the yield curve. In particular, we use the 10-year yield as a proxy for the level of the yield curve. The empirical slope is the difference between the 10-year and 3-month yields. The empirical proxy for the curvature is defined as twice the 2-year yield minus the sum of the 3-month and 10-year yields. By comparing the forecasting performance of the latent factors and their empirical proxies, we are able to analyze whether the additional information encoded in the entire yield curve helps improve forecast accuracy.

The level, slope and curvature of the yield curve and their empirical proxies play a prominent role in the sequel. Thus, we focus on them now in some detail. Table 2 presents descriptive statistics for the three components of the yield curve. The estimated factors and their empirical proxies display similar features. Furthermore, the correlations between the estimated factors and their empirical proxies are very high. The correlations are $\rho(\hat{\beta}_{1t}, l_t) = 0.981$, $\rho(-\hat{\beta}_{2t}, s_t) = 0.983$, and $\rho(\hat{\beta}_{3t}, c_t) = 0.994$, where $\left\{\hat{\beta}_{1t}, -\hat{\beta}_{2t}, \hat{\beta}_{3t}\right\}$ are the factors corresponding to the level, slope and curvature of the yield curve and (l_t, s_t, c_t) are the empirical level, slope and curvature. Figure 1 plots the level, slope and curvature of the yield curve and 12-month-ahead industrial production growth. For instance, in January 2000, we plot industrial production growth from January 2000 to January 2001 together with a component of the yield curve at January 2000. Two aspects of the graph are particularly interesting. First, the level of the yield curve moves in the opposite direction to future economic activity. Second, periods with high (low) slope are typically followed by high (low) industrial production growth. The informal picture, then, is that there is a positive relation between the slope of the yield curve and future economic activity. The evidence for the curvature is mixed. If anything, the curvature and industrial production growth seem to be negatively correlated in the earlier part of the sample but positively correlated in the latter part. This preliminary evidence suggests that the predictive ability of the curvature might have changed over time.

3.2. Macroeconomic data

The macroeconomic factors are extracted from the FRED-MD dataset. This dataset contains 134 monthly time series representing different facets of the U.S. macroeconomy (e.g., production, consumption, employment and price inflation). The principal components estimation of the factors require a balanced panel of data. We form a balanced panel by dropping series 64 (New Orders for Consumer Goods), 66 (New Orders for Non-defense Capital Goods), 83 (S&P PE ratio), 101 (Trade Weighted U.S. Dollar Index: Major Currencies) and 130 (Consumer Sentiment Index) from the original dataset. To avoid multicollinearity between the estimated factors and the components of the yield curve, we exclude the federal funds rate (series 84), Treasury bill rates (series 86–87), Treasury bond rates (series 88–90), and Treasury spreads (series 94–98) from the FRED-MD dataset. After these modifications, we are able to work with a balanced panel of 118 monthly series dating from 1961:M6 to 2015:M4. We use the August 2015 vintage values of data.[8] The data are made stationary and standardized to have zero sample mean and unit sample variance. A complete list of the series and transformations applied to each series are reported in Appendix A.

[8]Most macroeconomic time series are published with a lag and are subject to important revisions (see Croushore, 2011). In this paper, we follow the majority of the literature and do not take into account the real-time nature of many macroeconomic time series.

4. Empirical results

Next, we present the results of the out-of-sample forecasting exercise outlined in Section 2. The aim of this exercise is to analyze (i) whether the elements of the yield curve contain predictive power in a data-rich environment, (ii) whether the predictive ability fluctuates over time and (iii) whether the predictive power can be linked to economic conditions.

4.1. Out-of-sample forecasting results

4.1.1. MSFE results

We start our analysis by considering the whole out-of-sample period running from 1972:M5 to 2015:M7. The MSFE results for this period are summarized in Table 3. This table shows the MSFE value of the model augmented with a yield curve element relative to the MSFE value of the benchmark model. Values below (above) unity indicate that the model augmented with the yield curve element has produced more (less) accurate forecasts than the benchmark, implying that the yield curve element contains (does not contain) marginal predictive power. The statistical significance is evaluated using the one-sided Giacomini and White (2006) test of equal unconditional predictive ability.

Three main results emerge from Table 3. First, the slope of the yield curve contains predictive power for U.S. industrial production growth in a data-rich environment. The model augmented with the slope produces more accurate forecasts than the benchmark model irrespective of which model specification or forecast horizon is considered. The improvements in forecast accuracy are quite large. However, the Giacomini and White (2006) test rejects the null of equal accuracy only in approximately one fourth of the cases. Second, neither the level nor the curvature of the yield curve are useful predictors of industrial production growth over the full out-of-sample period. In particular, the model augmented with the level of the yield curve performs poorly. Inclusion of the curvature factor increases forecast accuracy in a few cases. In these cases, however, the curvature makes only a very slight improvements over the benchmark. Third, the predictive ability of the estimated factors and their empirical counterparts seem to be alike. This finding is not surprising. As discussed in Section 3, the estimated factors and their empirical proxies are strongly correlated and display similar properties. If anything, the results in Table 3 suggest that the estimated slope factor outperforms the empirical slope. Furthermore, the empirical level and curvature seem to produce better forecasts than the estimated level and curvature factors, respectively.

4.1.2. Sign predictability

Table 4 reports the fraction of observations for which the model augmented with a component of the yield curve generates a smaller absolute forecast error than the benchmark model over the 1972:M5–2015:M7 period. We test the statistical significance using the Diebold and Mariano (1995) sign test. The results in Table 4 indicate that the forecasts from the model containing the slope are qualitatively superior to those from the benchmark model. The model with the slope provides more accurate forecasts for more than 50% of the observations in the clear majority of the model specification/forecast horizon combinations. This finding provides further evidence supporting the view that the slope of the yield curve has predictive power for real activity when the predictive information encoded in a large number of macroeconomic variables is already taken into account. The level of the yield curve performs somewhat better when we quantify out-of-sample forecast accuracy with a qualitative measure. Still, the forecasting performance of the level is not very convincing. The model that includes the level factor outperforms the benchmark in approximately half of the cases. Consistent with the results in Table 3, the model augmented with the curvature performs poorly relative to the benchmark. Thus, there is only very limited evidence that the curvature factor helps improve forecast accuracy in a data-rich environment. Again, the estimated factors and their empirical counterparts seem to perform equally well.

4.2. Comparison between estimated and empirical yield curve factors

We proceed by comparing more formally the relative performance of the estimated factors and their empirical proxies. The goal of this exercise is to shed light on whether the additional information encoded in the entire yield curve helps improve forecast accuracy. To analyze the relative ranking of the methods, we compute the MSFE of the model augmented with the estimated factor relative to the MSFE of the model augmented with the corresponding empirical factor. In addition, we compute the fraction of observations for which the model that includes the estimated factor produces a smaller absolute forecast error than the model that includes the empirical factor. The results are summarized in Table 5. The most important finding is that the estimated slope produces systematically smaller MSFE values than the empirical slope, often by quite a large margin. This implies that the slope estimated from the entire yield curve contains more predictive power for output growth than the empirical slope typically used in the previous literature. This finding is consistent with the results presented in Abdymomunov (2013). The results are more mixed when we consider whether the model with the estimated slope qualitatively outperforms the model with the empirical slope. In general, the empirical slope performs better than the estimated slope. For the level and curvature of the yield curve, the empirical factors dominate the estimated ones regardless which measure of forecast performance is used. Therefore, for these two components of the yield curve, using information from the entire yield curve does not improve forecast accuracy.

4.3. Rolling relative MSFE values

The results reported in Tables 3–5 focus on the average predictive power over the whole out-of-sample period. There is a lot of evidence that shows that the predictive content of leading indicators often fluctuate over time (see, e.g., D'Agostino and Surico, 2012; Ng and Wright, 2013; Rossi, 2013). If the predictive ability is not stable over time, the average performance over the full out-of-sample period may hide important information and even lead to incorrect conclusions. For this reason, we next analyze whether the predictive ability of the elements of the yield curve remains stable over time. To this end, we plot the relative MSFE values computed over a rolling out-of-sample window of 150 observations in Figure 2.[9] To save space, we report the results only for the DIAR model at $h = 3$ months. The results for the other model specifications and forecast horizons are qualitatively similar.

Inspection of Figure 2 reveals interesting details concerning the predictive ability of the yield curve. The evidence suggests that none of the three elements is a robust predictor of output growth as the MSFE ratio fluctuates around one in all cases. The rolling relative MSFE values are consistent with the observation that the slope contains more predictive power than the level or curvature of the yield curve. We find that the mid-1980s is a clear watershed for the predictive power of the slope. In the late 1970s and early 1980s, the model augmented with the slope produces substantially smaller MSFE values than the benchmark, implying that the slope was a particularly useful leading indicator during that period. Despite the large differences in the predictive ability, the Giacomini and Rossi (2010) fluctuation test rejects the null of equal accuracy at conventional significance levels only for the first few rolling windows. The predictive ability of the slope has substantially weakened since the mid-1980s. In particular, the slope produces very inaccurate forecasts in the latter half of the 1980s and early 1990s.

[9] As recommended by Giacomini and Rossi (2010), we select the length of the rolling out-of-sample window such that it corresponds roughly to 30% of the full out-of-sample period.

The timing of the break in the forecasting ability of the slope—the mid-1980s—is in line with the results from earlier contributions (see, e.g., Rossi and Sekhposyan, 2011; Stock and Watson, 2003). The results for the level of the yield curve are also fascinating. At the beginning of the out-of-sample period, the model with the level performs poorly and produces substantially less accurate forecasts than the benchmark model. However, in the period running from the late 1980s to early 2000s, inclusion of the level helps reduce forecast errors, and thus, the level adds incremental predictive information during this period. At the end of the sample, the level has again been less successful at predicting output growth. For the curvature, the results of the rolling relative MSFE exercise corroborate the findings in Tables 3 and 4. The rolling MSFE ratio fluctuates widely around one. We conclude from this evidence that the curvature of the yield curve has not been a useful forecasting tool for output growth over the last 43 years.

4.4. Rolling sign predictability

Figure 3 plots the rolling fraction of observations for which the forecasting model augmented with a yield curve element produces more accurate forecasts than the benchmark model. The fraction is computed using a rolling window of 150 out-of-sample observations. Again, we focus on the DIAR model and the 3-month-ahead forecast horizon. Figure 3 confirms by and large the findings in Figure 2. The predictive ability of each of the three yield curve elements varies over time. The slope was a useful leading indicator in the late 1970s and early 1980s. The predictive power of the slope disappeared for a short period in the mid-1980s. Similarly, the slope has not been an informative indicator at the end of the sample. The level of the yield curve was a poor predictor until the late 1980s but has performed well since then. The most notable difference between the results in Figures 2 and 3 is that the curvature contains predictive power in the 1980s when we use the qualitative measure of forecast accuracy.

4.5. Conditional predictive ability

As a final exercise, we investigate whether the predictive ability of the components of the yield curve can be linked to economic conditions. To do this, we employ the Giacomini and White (2006) test of equal conditional predictive ability. This test allows us to examine whether the state of the economy affects the accuracy of the yield curve model relative to the benchmark model. This means that we are able to analyze, say, how inflation persistence affects the marginal predictive power of the yield curve. Following Dotsey et al. (2015) and Ng and Wright (2013), we conduct the test by regressing the difference in the squared h-step-ahead forecast errors between the benchmark model and the model augmented with a yield curve component on conditioning variables. We also consider a different specification where the dependent variable is a binary variable that takes the value of one when the model with the yield curve factor produces a more accurate forecast than the benchmark and zero otherwise.

4.5.1. Conditioning variables

We consider several conditioning variables. The first variable is the NBER recession dates. Faust et al. (2013) find that the credit spread discussed in their paper contains more predictive power in recessions than during normal times. In a similar vein, the predictive power of the yield curve could differ in recessions and expansions. The results in the previous literature suggest that changes in the predictive content of the yield curve often correspond closely to major changes in the conduct of monetary policy (Bordo and Haubrich, 2008a, 2008b; Giacomini and Rossi, 2006; Hännikäinen, 2015). To evaluate formally whether changes in the monetary policy affect the predictive ability, we divide our out-of-sample period into three monetary policy regimes. The first regime, called the Great Inflation, runs from 1972:M5 to 1983:M12. The second regime, the Great Moderation, spans from 1984:M1 to 2008:M11.[10] The last regime covers the period from 2008:M12 to the end of the sample. During this last period, short-term interest rates have been stuck at the ZLB, and the Fed has used unconventional monetary policy. In what follows, we use the first regime as a benchmark and create dummy variables for the last two regimes. The third conditioning variable is inflation persistence. Bordo and Haubrich (2004) find that the slope of the yield curve forecasts output growth particularly well when inflation is highly persistent. Following the common practice in the literature, persistence is measured by the sum of the AR coefficients (see, e.g., Andrews and Chen, 1994; Benati, 2008; Clark, 2006; Pivetta and Reis, 2007).[11] We estimate an AR(5) model for monthly inflation.[12] At each forecast origin, the parameters of the model are re-estimated using a rolling window of 60 observations, and inflation persistence is computed as the sum of the AR coefficients. We obtain a measure for output persistence using exactly the same procedure. Finally, D'Agostino et al. (2006) argue that the reduced informativeness of the yield curve in recent years is due to the increased stability of U.S. output growth and other key variables since the mid-1980s. To examine the plausibility of this argument, we condition the forecasting performance of the yield curve components on U.S. output growth and inflation volatility. The measure of output volatility is the 5-year rolling standard deviation of annual industrial production growth (cf. Blanchard and Simon, 2001). Similarly, we use the 5-year rolling standard deviation of annual inflation as a measure of inflation volatility (cf. Lamla and Maag, 2012). The behavior of the conditioning

[10] D'Agostino and Surico (2012) also use these two periods when they investigate inflation predictability in the U.S. across the monetary regimes of the 20th century. For an in-depth discussion of U.S. monetary regimes, see Bordo and Schwartz (1999).

[11] We also considered the non-parametric measure of inflation persistence introduced by Marques (2005). The results are similar for both measures of inflation persistence and hence we report the results only for the sum of the AR coefficients measure.

[12] We select the lag order of the AR model using the 1972:M5–2015:M4 sample. The possible lag lengths are $p = 1, \dots, 6$, and we choose the model that minimizes the BIC and AIC. The BIC selects the model with three lags, and the AIC recommends the model with five lags. Overall, the preliminary analysis indicates that the AR(5) model fits the data the best. However, other model specifications yield similar estimates of inflation persistence.

variables is depicted in Figure 4.

4.5.2. Conditional predictive ability tests

The results of the conditional predictive ability tests are summarized in Table 6. For the sake of exposition, we concentrate on the DIAR model at $h = 3$ and $h = 12$ forecast horizons. The results for the slope of the yield curve are especially interesting. The null of equal conditional predictive ability between the model augmented with the slope and the benchmark model is rejected at the shortest forecast horizon. This means that the state of the economy affects the accuracy of the model with the slope relative to the benchmark model in a statistically significant way.

The most intriguing finding is that the performance of the slope depends on inflation persistence. The slope tends to forecast output growth better during periods when inflation is highly persistent (cf. Bordo and Haubrich, 2004). This observation helps explain why the slope is a particularly informative leading indicator in the late 1970s and early 1980s (see Figures 2–4). Theories of intertemporal consumption smoothing state that the real yield curve contains information about future real activity (Harvey, 1988). It is worth emphasizing that the extent to which changes in the nominal yield curve studied in this paper reflect changes in the real yield curve depends on inflation persistence. Consider first the case where inflation is a random walk (i.e., very persistent). In this case, inflation shocks shift expected inflation at all horizons by an equal amount. Inflation shocks have no effect on the slope of the nominal yield curve, as short- and long-term rates move up by the amount of the permanently higher inflation rate. Because the nominal yield curve fluctuates one-for-one with the real yield curve, the predictive power intrinsic to the slope of the real yield curve translates to the slope of the nominal yield curve. However, when inflation has little persistence, inflation shocks increase short-term inflation expectations more than long-term inflation expectations. As a consequence, short-term rates rise relative to the long-term rates. This means that, for a given real yield curve, inflation shocks tend to flatten the nominal yield curve. Because the real yield curve and the nominal yield curve do not move hand-in-hand, the nominal yield curve is a less reliable predictor of future real activity. Thus, according to the real yield curve explanation, a decrease in inflation persistence should lead to a decrease in the predictive power of the slope of the nominal yield curve. Our findings are consistent with this argument. Hence, our results can be seen as providing empirical support to Harvey's (1988) theoretical conjecture.

Another important finding from Table 6 is that the coefficient of inflation volatility is negative and statistically significant. This implies that the slope produces less accurate forecasts when inflation is highly volatile. Our analysis reinforces the finding that regime shifts in monetary policy play a role for the predictive content of the slope (see, e.g., Bordo and Haubrich, 2004, 2008a, 2008b; Giacomini and Rossi, 2006; Hännikäinen, 2015). The dummy for the Great Moderation period (MPR2) is negative and statistically significant. Thus, the slope of the yield curve was a less reliable indicator of future real activity during the Great Moderation period. The coefficient on the NBER recession dates is insignificant, suggesting that whether the economy is in a recession or not does not matter for the predictive power of the slope. Finally, we find no evidence supporting the conjecture in D'Agostino et al. (2006) that output volatility affects the predictive ability of the slope.

The Giacomini and White (2006) test rejects the null of equal conditional predictive ability between the model augmented with the level and the benchmark model at the 3-month-ahead horizon. The predictive power of the level depends on inflation volatility. The level tends to produce less accurate forecasts when inflation is highly volatile. In addition, at the longer 12-month-ahead horizon, the model augmented with the level performs better relative to the benchmark when output persistence is low and output volatility is high. The conclusions are substantially different for the curvature factor. Examination of Table 6 leads us to conclude that none of the conditioning variables systematically affects the predictive ability of the curvature across all specifications and forecast horizons.[13]

4.5.3. Binary dependent variable specification

Table 7 shows the results for the binary dependent variable specification. These results confirm our main findings from Table 6. Most importantly, inflation persistence and inflation volatility emerge as key variables that affects the predictive ability of the slope of the yield curve. The higher the persistence of inflation or the smaller the volatility of inflation, the more likely the model augmented with the slope yields qualitatively superior forecasts relative to the benchmark model. The most notable difference between the results in Tables 6 and 7 is that when qualitative differences are considered, regime shifts in monetary policy seem to matter less for the predictive ability of the slope. The results for the level and curvature factors are qualitatively similar. The probability that the model augmented either with the level or curvature factor outperforms the benchmark model appears unrelated to any of the conditioning variables.[14]

4.5.4. Conditional predictive ability tests with quarterly data

As a sensitivity check, we repeat the above analysis using alternative conditioning variables. We replace the monetary policy regime dummies with three macroeconomic regimes discussed in Baele et al. (2015). These regimes are derived from a New-Keynesian model that accommodates regime switches in systematic monetary policy and macroeconomic shocks. The macroeconomic regimes are the high inflation shock volatility regime, the high output shock volatility regime and the active monetary policy

[13] The adjusted R^2s for the slope regressions (not reported) range from 6.1% to 17.1%. The adjusted R^2s for the level and curvature regressions are typically substantially lower than those for the slope regressions.

[14] The adjusted R^2s for the slope specifications range between 5.9% and 6.8%. Again, the adjusted R^2s for the slope specifications are larger than those for the level and curvature specifications.

regime in which the Fed aggressively stabilizes inflation. The data for the macroeconomic regimes run from 1972:Q2 to 2008:Q2. Figure 5 shows the smoothed probabilities of the regimes. Because macroeconomic regimes are available on a quarterly frequency, we use quarterly (rather than monthly) observations in our exercise. Monthly forecasts and conditioning variables are aggregated to the quarterly frequency by using the value of the last month of each quarter.[15]

The results of this sensitivity analysis, reported in Tables 8 and 9, confirm by and large the findings in Tables 6 and 7. In particular, the positive relationship between inflation persistence and the predictive ability of the slope is robust to using quarterly specification and an alternative set of conditioning variables. Similarly, the negative link between inflation volatility and the predictive power of the slope is robust to using quarterly observations and alternative conditioning variables. An important finding from Table 8 is that the coefficient of active monetary policy regime is negative and statistically significant for the slope at the longer forecasting horizon. Thus, there is some evidence supporting the view that monetary policy activism matters for the predictive power of the slope. The results suggest that the slope forecasts less accurately when the Fed concentrates on controlling inflation (cf. Estrella, 2005).

5. Conclusions

This paper examines the predictive power of the level, slope and curvature of the yield curve for U.S. real activity in a data-rich environment. Our analysis leads us to four main conclusions. First, of the three yield curve factors, the slope is the most informative leading indicator. We find that the slope contains predictive power for industrial production growth when the information encoded in a large set of macroeconomic predictors is already taken into account. In general, the level and curvature of the yield

[15]To check whether the results in Tables 6 and 7 are overturned by using quarterly data, we estimate the same model specifications (excluding MPR3 dummy) over the 1972:Q2–2008:Q2 period. The results of this exercise are very similar to those in Tables 6 and 7.

curve are not good predictors of future real activity. Second, the slope estimated from the entire yield curve produces more accurate forecasts than the empirical slope (i.e., the difference between the 10-year and 3-month yields). This is an important finding because the empirical slope is used extensively in the literature. In short, our results suggest that one should use the slope extracted from the entire yield curve rather than the empirical slope when forecasting subsequent real activity. Third, the predictive power of each of the components of the yield curve fluctuates over time. The slope was a particularly good leading indicator in the late 1970s and early 1980s, but it performed poorly in the late 1980s and early 1990s. On the other hand, the level contains predictive power only from the late 1980s to early 2000s. Fourth, and most importantly, the predictive power of the slope depends on economic conditions. The slope tends to forecast output growth better when inflation is highly persistent and when inflation is not highly volatile. The finding that the performance of the slope as a predictor of output growth depends on the persistence of inflation is an intriguing one because it provides empirical support for the real yield curve explanation for the predictive power suggested by Harvey (1988).

Our results could be extended in several ways. We have considered the predictive power of the yield curve factors for U.S. industrial production growth. It would be interesting to know whether our results hold for other measures of real activity. In addition, evidence from other countries may lead to a better understanding of how economic conditions affect the predictive content of the yield curve. Therefore, analyzing the predictive ability further using, e.g., the international dataset of 10 countries studied by Wright (2011) might be a fruitful area for future research.

References

Abdymomunov, A. (2013). Predicting output using the entire yield curve. *Journal of Macroeconomics*, 37, 333–344.

Aguiar-Conraria, L., Martins, M. M. F. & Soares, M. J. (2012). The yield curve and the macro-economy across time and frequencies. *Journal of Economic Dynamics and Control*, 36, 1950–1970.

Andrews, D. W. K. & Chen, H.-Y. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business and Economic Statistics*, 12, 187–204.

Ang, A., Piazzesi, M. & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, 131, 359–403.

Baele, L., Bekaert, G., Cho, S., Inghelbrecht, K. & Moreno, A. (2015). Macroeconomic regimes. *Journal of Monetary Economics*, 70, 51–71.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *Quarterly Journal of Economics*, 123, 1005–1060.

Benati, L. & Goodhart, C. (2008). Investigating time-variation in the marginal predictive power of the yield spread. *Journal of Economic Dynamics and Control*, 32, 1236–1272.

Bernanke, B. S. (1990). On the predictive power of interest rates and interest rate spreads. *New England Economic Review*, 51–68.

Bernanke, B. S. & Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review*, 82, 901–921.

Bernanke, B. S. & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50, 525–546.

Blanchard, O. J. & Simon, J. A. (2001). The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity*, 135–164.

Bordo, M. D. & Haubrich, J. G. (2004). The yield curve, recessions and the credibility of the monetary regime: long run evidence 1875–1997. NBER Working Paper, no. 10431.

Bordo, M. D. & Haubrich, J. G. (2008a). Forecasting with the yield curve; level, slope, and output 1875–1997. *Economics Letters*, 99, 48–50.

Bordo, M. D. & Haubrich, J. G. (2008b). The yield curve as a predictor of growth: long-run evidence, 1875–1997. *Review of Economics and Statistics*, 90, 182–185.

Bordo, M. D. & Schwartz, A. J. (1999). Monetary policy regimes and economic performance: The historical record. In J. B. Taylor, & M. Woodford (Eds), *Handbook of Macroeconomics*, vol. 1, North-Holland, Amsterdam. 149–234.

Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics*, 21, 563–587.

Clements, M. P. (2015). Real-time factor model forecasting and the effects of instability. *Computational Statistics and Data Analysis*, forthcoming.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, 49, 72–100.

D'Agostino, A., Giannone, D. & Surico, P. (2006). (Un)predictability and macroeconomic stability. European Central Bank Working Paper, no. 605.

D'Agostino, A. & Surico, P. (2012). A century of inflation forecasts. *Review of Economics and Statistics*, 94, 1097–1106.

Diebold, F. X. & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130, 337–364.

Diebold, F. X. & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–265.

Dotsey, M., Fujita, S. & Stark, T. (2015). Do Phillips curves conditionally help to forecast inflation? Federal Reserve Bank of Philadelphia Working Paper, no. 15-16.

Estrella, A. & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, 46, 555–576.

Estrella, A., Rodrigues, A. P. & Schich, S. (2003). How stable is the predictive power of the yield curve? Evidence from Germany and the United States. *Review of Economics and Statistics*, 85, 629–644.

Faust, J., Gilchrist, S., Wright, J. H. & Zakrajšek, E. (2013). Credit spreads as predictors of real-time economic activity: a Bayesian model-averaging approach. *Review of Economics and Statistics*, 95, 1501–1519.

Gertler, M. & Lown, C. S. (1999). The information in the high-yield bond spread for the business cycle: evidence and some implications. *Oxford Review of Economic Policy*, 15, 132–150.

Giacomini, R. & Rossi, B. (2006). How stable is the forecasting performance of the yield curve for output growth? *Oxford Bulletin of Economics and Statistics*, 68, 783–795.

Giacomini, R. & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25, 595–620.

Giacomini, R. & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.

Gürkaynak, R. S., Sack, B. & Wright, J. H. (2007). The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54, 2291–2304.

Hamilton, J. D. & Kim, D. H. (2002). A re-examination of the predictability of the yield spread for real economic activity. *Journal of Money, Credit, and Banking*, 34, 340–360.

Harvey, C. R. (1988). The real term structure and consumption growth. *Journal of Financial Economics*, 22, 305–333.

Hännikäinen, J. (2015). Zero lower bound, unconventional monetary policy and indicator properties of interest rate spreads. *Review of Financial Economics*, 26, 47–54.

Lamla, M. J. & Maag, T. (2012). The role of media for inflation forecast disagreement of households and professional forecasters. *Journal of Money, Credit and Banking*, 44, 1325–1350.

Luciani, M. (2014). Large-dimensional dynamic factor models in real-time: a survey. Working Paper, Université libre de Bruxelles.

Marques, C. R. (2005). Inflation persistence: facts or artefacts? *Bank of Portugal Economic Bulletin*, 69–79.

McCracken, M. W. & Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, forthcoming.

Mody, A. & Taylor, M. P. (2003). The high-yield spread as a predictor of real economic activity: evidence of a financial accelerator for the United States. *IMF Staff Papers*, 50, 373–402.

Nelson, C. R. & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60, 473–489.

Newey, W. K. & West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–708.

Ng, S. & Wright, J. H. (2013). Facts and challenges from the Great Recession for forecasting and macroeconomic modeling. *Journal of Economic Literature*, 51, 1120–1154.

Pivetta, F. & Reis, R. (2007). The persistence of inflation in the United States. *Journal of Economic Dynamics and Control*, 31, 1326–1358.

Rossi, B. (2013). Advances in forecasting under instability. In G. Elliott, & A. Timmermann (Eds), *Handbook of Economic Forecasting*, vol. 2, North-Holland, Amsterdam. 1203–1324.

Rossi, B. & Sekhposyan, T. (2011). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting*, 26, 808–835.

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

Stock, J. H. & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147–162.

Stock, J. H. & Watson, M. W. (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature*, 41, 788–829.

Stock, J. H. & Watson, M. W. (2011). Dynamic factor models. In M. P. Clements, & D. F. Hendry (Eds), *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press. 35–60.

Wheelock, D. C. & Wohar, M. E. (2009). Can the term spread predict output growth and recessions? A survey of the literature. Federal Reserve Bank of St. Louis *Review*, 91, 419–440.

Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review*, 101, 1514–1534.

Appendix A

Table 10: Data description

id	Mnemonic	Trans. code	Description
1	RPI	5	Real Personal Income
2	W875RX1	5	Real Personal Income ex transfer receipts
3	DPCERA3M086SBEA	5	Real personal consumption expenditures
4	CMRMTSPLx	5	Real Manu. and Trade Industries Sales
5	RETAILx	5	Retail and Food Services Sales
6	INDPRO	5	IP Index
7	IPFPNSS	5	IP: Final Products and Nonindustrial Supplies
8	IPFINAL	5	IP: Final Products (Market Group)
9	IPCONGD	5	IP: Consumer Goods
10	IPDCONGD	5	IP: Durable Consumer Goods
11	IPNCONGD	5	IP: Nondurable Consumer Goods
12	IPBUSEQ	5	IP: Business Equipment
13	IPMAT	5	IP: Materials
14	IPDMAT	5	IP: Durable Materials
15	IPNMAT	5	IP: Nondurable Materials
16	IPMANSICS	5	IP: Manufacturing (SIC)
17	IPB51222s	5	IP: Residential Utilities
18	IPFUELS	5	IP: Fuels
19	NAPMPI	1	ISM Manufacturing: Production Index
20	CUMFNS	2	Capacity Utilization: Manufacturing
21	HWI	2	Help-Wanted Index for United States
22	HWIURATIO	2	Ratio of Help Wanted/No. Unemployed
23	CLF16OV	5	Civilian Labor Force
24	CE16OV	5	Civilian Employment
25	UNRATE	2	Civilian Unemployment Rate
26	UEMPMEAN	2	Average Duration of Unemployment (Weeks)
27	UEMPLT5	5	Civilians Unemployed - Less Than 5 Weeks
28	UEMP5TO14	5	Civilians Unemployed for 5-14 Weeks
29	UEMP15OV	5	Civilians Unemployed - 15 Weeks & Over
30	UEMP15T26	5	Civilians Unemployed for 15-26 Weeks
31	UEMP27OV	5	Civilians Unemployed for 27 Weeks and Over
32	CLAIMSx	5	Initial Claims
33	PAYEMS	5	All Employees: Total nonfarm
34	USGOOD	5	All Employees: Goods-Producing Industries
35	CES1021000001	5	All Employees: Mining and Logging: Mining
36	USCONS	5	All Employees: Construction
37	MANEMP	5	All Employees: Manufacturing
38	DMANEMP	5	All Employees: Durable goods
39	NDMANEMP	5	All Employees: Nondurable goods
40	SRVPRD	5	All Employees: Service-Providing Industries
41	USTPU	5	All Employees: Trade, Transportation & Utilities
42	USWTRADE	5	All Employees: Wholesale Trade
43	USTRADE	5	All Employees: Retail Trade
44	USFIRE	5	All Employees: Financial Activities
45	USGOVT	5	All Employees: Government
46	CES0600000007	1	Avg Weekly Hours: Goods-Producing
47	AWOTMAN	2	Avg Weekly Overtime Hours: Manufacturing
48	AWHMAN	1	Avg Weekly Hours: Manufacturing
49	NAPMEI	1	ISM Manufacturing: Employment Index
50	HOUST	4	Housing Starts: Total New Privately Owned
51	HOUSTNE	4	Housing Starts, Northeast
52	HOUSTMW	4	Housing Starts, Midwest
53	HOUSTS	4	Housing Starts, South
54	HOUSTW	4	Housing Starts, West
55	PERMIT	4	New Private Housing Permits (SAAR)
56	PERMITNE	4	New Private Housing Premits, Northeast (SAAR)
57	PERMITMW	4	New Private Housing Permits, Midwest (SAAR)
58	PERMITS	4	New Private Housing Permits, South (SAAR)
59	PERMITW	4	New Private Housing Permits, West (SAAR)
60	NAPM	1	ISM: PMI Composite Index

(Continued)

Table 10 – *(Continued)*

id	Mnemonic	Trans. code	Description
61	NAPMNOI	1	ISM: New Orders Index
62	NAPMSDI	1	ISM: Supplier Deliveries Index
63	NAPMII	1	ISM: Inventories Index
65	AMDMNOx	5	New Orders for Durable Goods
67	AMDMUOx	5	Unfilled Orders for Durable Goods
68	BUSINVx	5	Total Business Inventories
69	ISRATIOx	2	Total Business: Inventories to Sales Ratio
70	M1SL	6	M1 Money Stock
71	M2SL	6	M2 Money Stock
72	M2REAL	5	Real M2 Money Stock
73	AMBSL	6	St. Louis Adjusted Monetary Base
74	TOTRESNS	6	Total Reserves of Depository Institutions
75	NONBORRES	7	Reserves of Depository Institutions, Nonborrowed
76	BUSLOANS	6	Commercial and Industrial Loans, All Commercial Banks
77	REALLN	6	Real Estate Loans at All Commercial Banks
78	NONREVSL	6	Total Nonrevolving Credit Owner and Securitized Outstanding
79	CONSPI	2	Nonrevolving Consumer Credit to Personal Income
80	S & P 500	5	S&P's Common Stock Price Index: Composite
81	S & P: indust	5	S&P's Common Stock Price Index: Industrials
82	S & P div yield	2	S&P's Composite Common Stock: Dividend Yield
85	CP3Mx	2	3-Month AA Financial Commercial Paper Rate
91	AAA	2	Moody's Seasoned Aaa Corporate Bond Yield
92	BAA	2	Moody's Seasoned Baa Corporate Bond Yield
93	COMPAPFFx	1	3-Month Commercial Paper Minus FEDFUNDS
99	AAAFFM	1	Moody's Aaa Corporate Bond Minus FEDFUNDS
100	BAAFFM	1	Moody's Baa Corporate Bond Minus FEDFUNDS
102	EXSZUSx	5	Switzerland / U.S. Foreign Exchange Rate
103	EXJPUSx	5	Japan / U.S. Foreign Exchange Rate
104	EXUSUKx	5	U.S. / U.K. Foreign Exchange Rate
105	EXCAUSx	5	Canada / U.S. Foreign Exchange Rate
106	PPIFGS	6	PPI: Finished Goods
107	PPIFCG	6	PPI: Finished Consumer Goods
108	PPIITM	6	PPI: Intermediate Materials
109	PPICRM	6	PPI: Crude Materials
110	OILPRICEx	6	Crude Oil, spliced WTI and Cushing
111	PPICMM	6	PPI: Metals and Metal Products
112	NAPMPRI	1	ISM Manufacturing: Prices Index
113	CPIAUCSL	6	CPI: All Items
114	CPIAPPSL	6	CPI: Apparel
115	CPITRNSL	6	CPI: Transportation
116	CPIMEDSL	6	CPI: Medical Care
117	CUSR0000SAC	6	CPI: Commodities
118	CUUR0000SAD	6	CPI: Durables
119	CUSR0000SAS	6	CPI: Services
120	CPIULFSL	6	CPI: All Items Less Food
121	CUUR0000SA0L2	6	CPI: All Items Less Shelter
122	CUSR0000SA0L5	6	CPI: All Items Less Medical Care
123	PCEPI	6	Personal Cons. Expend.: Chain Price Index
124	DDURRG3M086SBEA	6	Personal Cons. Expend.: Durable Goods
125	DNDGRG3M086SBEA	6	Personal Cons. Expend.: Nondurable Goods
126	DSERRG3M086SBEA	6	Personal Cons. Expend.: Services
127	CES0600000008	6	Avg Hourly Earnings: Goods-Producing
128	CES2000000008	6	Avg Hourly Earnings: Construction
129	CES3000000008	6	Avg Hourly Earnings: Manufacturing
131	MZMSL	6	MZM Money Stock
132	DTCOLNVHFNM	6	Consumer Motor Vehicle Loans Outstanding
133	DTCTHFNM	6	Total Consumer Loans and Leases Outstanding
134	INVEST	6	Securities in Bank Credit at All Commercial Banks

Notes: The transformation code (column 3) denotes the transformation applied to the variable before principal components are calculated. The transformation codes are 1 = no transformation, 2 = first difference, 3 = second difference, 4 = natural logarithm, 5 = first difference of logarithms, 6 = second difference of logarithms. The data sample is 1961:M6–2015:M4. The data source is the FRED-MD database.

Table 1: Summary statistics for U.S. Treasury yields

Maturity (Months)	Mean	Std. dev.	Min.	Max.	$\hat{\rho(1)}$	$\hat{\rho(12)}$
3	4.867	3.107	0.010	15.520	0.986	0.825
6	5.193	3.260	0.089	16.219	0.986	0.832
9	5.263	3.256	0.079	16.182	0.987	0.841
12	5.330	3.244	0.099	16.110	0.988	0.848
24	5.557	3.168	0.188	15.782	0.989	0.867
36	5.738	3.078	0.306	15.575	0.990	0.878
48	5.890	2.990	0.453	15.350	0.990	0.885
60	6.020	2.909	0.627	15.178	0.990	0.888
72	6.133	2.838	0.815	15.061	0.990	0.889
84	6.231	2.775	1.007	14.987	0.990	0.890
96	6.318	2.722	1.197	14.940	0.990	0.890
108	6.393	2.676	1.380	14.911	0.990	0.889
120	6.459	2.638	1.552	14.892	0.990	0.888

Notes: The table presents summary statistics for monthly U.S. Treasury yields at different maturities. The last two columns contain sample autocorrelations at displacements of 1 and 12 months. The sample period runs from 1961:M6 to 2015:M4.

Table 2: Summary statistics for the level, slope and curvature of the yield curve

	Mean	Std. dev.	Min.	Max.	$\hat{\rho(1)}$	$\hat{\rho(12)}$
Level	6.771	2.401	2.121	13.919	0.989	0.883
Level (empirical)	6.459	2.638	1.552	14.892	0.990	0.888
Slope	1.841	1.873	-5.106	5.632	0.958	0.534
Slope (empirical)	1.592	1.362	-2.555	4.395	0.950	0.504
Curvature	-0.861	2.708	-8.181	6.764	0.934	0.645
Curvature (empirical)	-0.211	1.005	-2.690	2.907	0.924	0.603

Notes: We fit the Diebold–Li (2006) model using monthly yield data from 1961:M6 to 2015:M4. We fix λ_t at 0.0609. The three estimated factors $\hat{\beta}_{1t}$,$-\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$ are the level, slope and curvature of the yield curve. Following Diebold and Li (2006), we define the empirical level as the 10-year yield, the empirical slope as the difference between the 10-year and 3-month yields, and the empirical curvature as twice the 2-year yield minus the sum of the 3-month and 10-year yields.

Table 3: Out-of-sample MSFE values

	Model	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Level					
	DIAR	1.089	1.089	1.112	1.096
	K1	1.058	1.100	1.128	1.152
	K2	1.029	1.030	1.015	0.987
	AR	1.063	1.088	1.116	1.133
Level (empirical)					
	DIAR	1.135	1.163	1.126	1.106
	K1	1.039	1.077	1.114	1.119
	K2	1.017	1.014	1.006	0.969
	AR	1.055	1.084	1.125	1.124
Slope					
	DIAR	0.947	0.973	0.948	0.950
	K1	0.933*	0.880*	0.868*	0.855
	K2	0.982	0.960	0.953	0.945
	AR	0.929*	0.869**	0.857*	0.861
Slope (empirical)					
	DIAR	0.968	1.000	0.988	0.982
	K1	0.954	0.911	0.893*	0.878
	K2	0.997	0.989	0.974	0.963
	AR	0.952	0.909*	0.877*	0.892
Curvature					
	DIAR	1.012	0.986	1.010	1.034
	K1	1.010	1.009	1.009	1.021
	K2	0.995	0.978	0.982	1.009
	AR	1.023	1.029	1.036	1.057
Curvature (empirical)					
	DIAR	1.009	0.982	1.003	1.023
	K1	1.002	1.000	0.997	1.006
	K2	0.988	0.972	0.972	0.995
	AR	1.014	1.015	1.021	1.036

Notes: The out-of-sample forecasting period runs from 1972:M5 to 2015:M7. Each row reports the ratio of the MSFE of a forecasting model augmented with an element of the yield curve relative to the MSFE of the benchmark model. Asterisks mark rejection of the one-sided Giacomini and White (2006) test at the 1%(***), 5%(**) and 10%(*) significance levels, respectively. The truncation lag for the Newey-West (1987) HAC estimator is h-1, where h is the forecast horizon.

Table 4: Sign predictability

	Model	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Level					
	DIAR	0.516	0.519	0.516	0.509
	K1	0.502	0.515	0.494	0.475
	K2	0.519	0.540**	0.498	0.464
	AR	0.457	0.480	0.425	0.412
Level (empirical)					
	DIAR	0.500	0.505	0.524	0.513
	K1	0.514	0.552***	0.533*	0.533*
	K2	0.523	0.554***	0.531*	0.521
	AR	0.450	0.472	0.461	0.473
Slope					
	DIAR	0.512	0.499	0.527	0.507
	K1	0.527	0.530*	0.524	0.523
	K2	0.490	0.530*	0.514	0.497
	AR	0.496	0.511	0.490	0.501
Slope (empirical)					
	DIAR	0.488	0.509	0.512	0.501
	K1	0.529*	0.548**	0.533*	0.523
	K2	0.502	0.550**	0.512	0.491
	AR	0.512	0.526	0.510	0.503
Curvature					
	DIAR	0.488	0.480	0.480	0.454
	K1	0.521	0.487	0.520	0.487
	K2	0.488	0.491	0.518	0.467
	AR	0.459	0.450	0.457	0.438
Curvature (empirical)					
	DIAR	0.481	0.480	0.480	0.456
	K1	0.516	0.489	0.520	0.485
	K2	0.483	0.487	0.506	0.462
	AR	0.455	0.444	0.461	0.436

Notes: The out-of-sample forecasting period runs from 1972:M5 to 2015:M7. Each row reports the fraction of observations for which the forecasting model augmented with the yield curve element produces more accurate out-of-sample forecasts than the benchmark model. Asterisks mark rejection of the Diebold and Mariano (1995) sign test at the 1%(***), 5%(**) and 10%(*) significance levels, respectively.

Table 5: Out-of-sample performance of estimated versus empirical factors

	Model	$h=3$	$h=6$	$h=9$	$h=12$
Level					
	DIAR	0.958*	0.952	0.971**	0.967*
	K1	1.019	1.022	1.013	1.029
	K2	1.012	1.016	1.009	1.019
	AR	0.993	1.012	0.997	1.014
Slope					
	DIAR	0.981	0.972*	0.961*	0.977*
	K1	0.978**	0.965*	0.973	0.973
	K2	0.985	0.970	0.978	0.981
	AR	0.976*	0.967	0.986	0.964
Curvature					
	DIAR	1.005	1.001	1.008	1.011
	K1	1.008	1.009	1.012	1.016
	K2	1.007	1.007	1.010	1.014
	AR	1.010	1.014	1.027	1.024
Level (sign)					
	DIAR	0.496	0.528	0.514	0.469
	K1	0.500	0.489	0.478	0.473
	K2	0.504	0.485	0.455	0.489
	AR	0.548**	0.552***	0.525	0.499
Slope (sign)					
	DIAR	0.508	0.561***	0.553***	0.544**
	K1	0.519	0.497	0.478	0.458
	K2	0.483	0.474	0.453	0.422
	AR	0.492	0.485	0.447	0.471
Curvature (sign)					
	DIAR	0.506	0.468	0.465	0.503
	K1	0.496	0.483	0.471	0.442
	K2	0.488	0.499	0.484	0.485
	AR	0.490	0.485	0.496	0.456

Notes: The out-of-sample forecasting period runs from 1972:M5 to 2015:M7. The upper panel reports the MSFE of the model that includes the estimated factor relative to that of the model that includes the corresponding empirical factor. Asterisks mark rejection of the one-sided Giacomini and White (2006) test at the 1%(***), 5%(**) and 10%(*) significance levels, respectively. The truncation lag for the Newey-West (1987) HAC estimator is h-1, where h is the forecast horizon. The lower panel reports the fraction of observations for which the forecasting model augmented with the estimated factor produces more accurate out-of-sample forecasts than the model augmented with the empirical factor. Asterisks mark rejection of the Diebold and Mariano (1995) sign test at the 1%(***), 5%(**) and 10%(*) significance levels, respectively.

Table 6: Test of conditional predictive ability

	Intercept	NBER	MPR2	MPR3	Infl. pers.	Infl. volatility	Output pers.	Output volatility	p-value
$\underline{h=3}$									
Level	6.384	11.351	3.344	6.338	53.857	-7.671*	-55.234	-0.359	0.061
	(27.416)	(11.816)	(7.956)	(10.847)	(38.706)	(4.394)	(40.538)	(1.802)	
Level (empirical)	-1.711	12.805	7.170	10.179	49.944	-8.883*	-43.511	-0.823	0.022
	(28.737)	(12.374)	(9.633)	(13.378)	(41.867)	(5.365)	(39.419)	(1.963)	
Slope	12.365	16.599*	-10.004**	-9.086*	37.006**	-6.653**	-40.910	1.826	0.075
	(21.233)	(9.492)	(4.438)	(5.108)	(17.923)	(2.901)	(32.371)	(1.719)	
Slope (empirical)	4.197	12.894	-2.709	-5.037	34.783**	-5.370***	-39.559	2.343	0.056
	(17.411)	(8.020)	(2.499)	(3.590)	(15.361)	(1.947)	(28.459)	(1.441)	
Curvature	-3.809	0.248	-0.521	0.907	-11.438	1.145	17.617*	-0.942*	0.417
	(7.381)	(3.322)	(2.897)	(2.229)	(10.452)	(1.611)	(9.219)	(0.508)	
Curvature (empirical)	-3.092	0.554	-1.631	0.667	-7.533	0.829	14.579	-1.025**	0.397
	(7.144)	(3.518)	(2.951)	(2.174)	(10.609)	(1.572)	(9.502)	(0.510)	
$\underline{h=12}$									
Level	67.422	17.405*	0.356	-13.509	28.102	-9.471*	-103.695***	3.140*	0.105
	(41.814)	(8.874)	(9.714)	(14.012)	(45.272)	(5.709)	(33.099)	(1.831)	
Level (empirical)	70.320	15.184	-2.226	-19.114	46.388	-10.279*	-124.208***	3.771*	0.304
	(56.336)	(10.436)	(10.271)	(16.139)	(54.589)	(6.091)	(45.621)	(2.119)	
Slope	0.920	7.764	-17.991***	-5.742	52.932***	-8.986***	-26.445	-0.123	0.166
	(14.599)	(7.217)	(5.712)	(5.133)	(19.395)	(3.471)	(19.337)	(1.510)	
Slope (empirical)	1.254	5.015	-12.302**	-3.919	43.569***	-8.084**	-26.422*	0.574	0.230
	(14.386)	(5.593)	(5.237)	(4.349)	(16.720)	(3.138)	(15.690)	(1.330)	
Curvature	-16.680	-5.382	-1.538	-1.443	4.270	-0.236	17.372	-0.506	0.290
	(12.046)	(4.006)	(3.487)	(3.440)	(13.820)	(2.043)	(11.658)	(0.646)	
Curvature (emprical)	-18.014	-4.145	-2.424	-1.675	2.428	-0.547	21.990*	-0.532	0.305
	(12.640)	(3.802)	(3.452)	(3.505)	(13.898)	(1.936)	(12.411)	(0.643)	

Notes: The difference in the squared h-step-ahead forecast errors between the benchmark model and the yield curve model, denoted by δ_{t+h}, is regressed on the conditioning variables x_t. 'MPR2' and 'MPR3' denote the Great Moderation and the ZLB/unconventional monetary policy periods, respectively. Newey-West standard errors are reported in parentheses. Asterisks denote statistical significance at the 1%(***), 5%(**) and 10%(*) levels, respectively. The p-value of the Giacomini and White (2006) test of conditional predictive ability is reported in the last column. The forecasting model is DIAR. The sample period runs from 1972:M5 to 2015:M7.

Table 7: Conditional predictive ability for the binary model

	Intercept	NBER	MPR2	MPR3	Infl. pers.	Infl. volatility	Output pers.	Output volatility
$h = 3$								
Level	1.788***	-0.034	-0.114	-0.157	-1.151*	-0.009	0.029	-0.031
	(0.681)	(0.098)	(0.101)	(0.108)	(0.591)	(0.060)	(0.659)	(0.028)
Level (empirical)	2.052***	0.036	-0.001	-0.257**	-1.314**	-0.057	-0.306	0.016
	(0.632)	(0.098)	(0.106)	(0.109)	(0.566)	(0.061)	(0.563)	(0.029)
Slope	-1.323*	0.042	-0.220***	-0.087	1.883***	-0.158***	0.550	-0.012
	(0.693)	(0.090)	(0.084)	(0.101)	(0.619)	(0.045)	(0.498)	(0.029)
Slope (empirical)	-1.625**	0.077	-0.113	-0.075	2.117***	-0.179***	0.483	0.008
	(0.665)	(0.084)	(0.080)	(0.100)	(0.573)	(0.045)	(0.545)	(0.028)
Curvature	0.300	-0.006	-0.128	-0.013	-0.124	0.064	0.567	-0.059**
	(0.837)	(0.090)	(0.093)	(0.115)	(0.682)	(0.057)	(0.484)	(0.024)
Curvature (empirical)	-0.315	0.021	-0.143	-0.007	0.701	0.045	0.395	-0.050**
	(0.717)	(0.096)	(0.091)	(0.103)	(0.576)	(0.057)	(0.535)	(0.025)
$h = 12$								
Level	2.676***	0.208*	0.108	-0.355**	-1.084	-0.052	-1.432**	0.047
	(0.761)	(0.114)	(0.171)	(0.149)	(0.794)	(0.096)	(0.405)	(0.043)
Level (empirical)	2.331**	0.124	0.115	-0.196	-0.964	-0.067	-1.269**	0.071
	(1.021)	(0.152)	(0.160)	(0.186)	(0.916)	(0.094)	(0.602)	(0.043)
Slope	-0.297	0.022	-0.190	0.153	2.359***	-0.150**	-1.162*	0.001
	(0.990)	(0.089)	(0.140)	(0.143)	(0.730)	(0.064)	(0.653)	(0.043)
Slope (empirical)	-0.501	0.099	-0.107	0.179	2.379***	-0.177***	-1.042	0.011
	(1.033)	(0.096)	(0.135)	(0.148)	(0.772)	(0.062)	(0.739)	(0.042)
Curvature	-0.680	-0.065	-0.022	-0.106	1.191*	0.008	0.045	0.003
	(0.676)	(0.107)	(0.144)	(0.132)	(0.617)	(0.070)	(0.591)	(0.037)
Curvature (empirical)	-0.311	-0.076	-0.070	-0.085	1.002	0.023	-0.099	-0.010
	(0.682)	(0.101)	(0.135)	(0.123)	(0.646)	(0.066)	(0.575)	(0.035)

Notes: The dependent variable is an indicator variable taking the value of one when the model augmented with a yield curve element produces more accurate forecast than the benchmark model and zero otherwise. 'MPR2' and 'MPR3' denote the Great Moderation and the ZLB/unconventional monetary policy periods, respectively. Newey-West standard errors are reported in parentheses. Asterisks denote statistical significance at the 1%(***), 5%(**) and 10%(*) levels, respectively. The forecasting model is DIAR. The sample period runs from 1972:M5 to 2015:M7.

Table 8: Test of conditional predictive ability (Quarterly variables)

	Intercept	NBER	ActiveMP	Infl. reg.	Output reg.	Infl. pers.	Infl. volatility	Output pers.	Output volatility	p-value
$h = 1$ quarter										
Level	-34.545	9.155	11.655	19.140	-5.897	26.866	-9.238	25.896	-4.315	0.571
	(65.328)	(13.989)	(14.902)	(12.093)	(10.036)	(46.859)	(6.591)	(71.458)	(2.743)	
Level (empirical)	-12.011	6.556	17.039	18.276	-7.895	42.219	-10.507	-14.858	-4.241	0.625
	(82.624)	(18.135)	(19.631)	(16.172)	(12.476)	(61.492)	(8.886)	(92.726)	(3.176)	
Slope	0.701	19.810	-9.425	-9.228	13.399	61.945	-5.961	-65.265	3.632	0.203
	(82.436)	(18.353)	(7.977)	(15.572)	(14.982)	(48.345)	(3.905)	(132.639)	(3.445)	
Slope (empirical)	44.488	15.890	-5.430	-11.101	7.282	49.263	-4.884*	-104.598	4.550*	0.351
	(62.933)	(15.974)	(7.515)	(12.302)	(12.046)	(39.884)	(2.796)	(101.243)	(2.425)	
Curvature	-10.319	0.496	0.505	1.272	1.820	3.603	1.982	9.030	-1.480	0.628
	(23.065)	(6.422)	(3.892)	(5.427)	(5.133)	(15.418)	(2.610)	(21.243)	(0.985)	
Curvature (empirical)	-22.992	1.771	-0.491	3.117	0.983	9.645	1.253	17.588	-1.496	0.636
	(24.556)	(7.512)	(4.614)	(5.486)	(5.289)	(17.051)	(2.522)	(22.590)	(0.998)	
$h = 4$ quarters										
Level	-91.777	17.201**	-4.608	44.089***	-37.590***	-4.102	-17.567**	106.295	2.121	0.058
	(74.663)	(7.765)	(8.215)	(16.496)	(13.701)	(45.473)	(6.845)	(71.558)	(3.095)	
Level (empirical)	-123.707	15.121	-8.965	45.129***	-38.910***	10.878	-18.162**	126.776*	2.527	0.152
	(80.839)	(9.204)	(9.277)	(17.151)	(14.383)	(51.884)	(7.432)	(72.265)	(3.413)	
Slope	-118.015*	9.108	-10.253*	27.625*	-12.849	74.258**	-18.146***	63.411	2.307	0.266
	(59.874)	(10.387)	(5.482)	(14.060)	(11.960)	(37.093)	(6.506)	(68.987)	(2.563)	
Slope (empirical)	-45.480	6.650	-7.302*	13.410	-7.775	67.090***	-15.134***	-14.326	3.727**	0.141
	(44.797)	(6.275)	(4.052)	(11.319)	(9.016)	(25.580)	(5.486)	(42.969)	(1.871)	
Curvature	-18.157	-3.303	-6.069**	2.616	-4.233	20.157	-1.315	0.375	0.549	0.115
	(29.601)	(4.648)	(2.867)	(7.583)	(5.163)	(12.237)	(3.419)	(27.304)	(1.207)	
Curvature (emprical)	-27.799	-1.753	-7.215***	5.527	-5.392	26.982**	-2.779	4.530	0.798	0.113
	(27.283)	(4.695)	(2.685)	(6.588)	(4.341)	(11.533)	(2.836)	(28.382)	(1.161)	

Notes: The difference in the squared h-step-ahead forecast errors between the benchmark model and the yield curve model, denoted by δ_{t+h}, is regressed on the conditioning variables x_t. 'ActiveMP', 'Infl. reg.', 'Output reg.' denote the active monetary policy, high inflation shock volatility, and high output shock volatility regimes, respectively. Newey-West standard errors are reported in parentheses. Asterisks denote statistical significance at the 1%(***), 5%(**) and 10%(*) levels, respectively. The p-value of the Giacomini and White (2006) test of conditional predictive ability is reported in the last column. The forecasting model is DIAR. The sample period runs from 1972:Q2 to 2008:Q2.

Table 9: Conditional predictive ability for the binary model (Quarterly variables)

	Intercept	NBER	ActiveMP	Infl. reg.	Output reg.	Infl. pers.	Infl. volatility	Output pers.	Output volatility
$h = 1$ quarter									
Level	-0.825	0.008	0.042	0.300	-0.122	-1.598	-0.014	3.359**	-0.099**
	(1.283)	(0.128)	(0.138)	(0.203)	(0.180)	(0.983)	(0.085)	(1.574)	(0.045)
Level (empirical)	1.354	0.134	0.183	0.163	-0.087	-1.174	-0.123	0.261	0.014
	(1.242)	(0.124)	(0.131)	(0.200)	(0.190)	(0.949)	(0.085)	(1.626)	(0.047)
Slope	-1.241	0.081	-0.114	0.104	-0.179	2.633***	-0.169**	-0.634	0.041
	(1.365)	(0.136)	(0.144)	(0.198)	(0.186)	(1.005)	(0.081)	(1.609)	(0.055)
Slope (empirical)	0.151	0.168	-0.148	0.018	-0.321*	2.899***	-0.223***	-2.538	0.103**
	(1.366)	(0.131)	(0.132)	(0.206)	(0.163)	(0.991)	(0.082)	(1.604)	(0.045)
Curvature	1.041	-0.035	0.168	0.067	0.183	-0.482	0.145	-0.052	-0.102*
	(1.309)	(0.121)	(0.144)	(0.211)	(0.192)	(1.004)	(0.089)	(1.690)	(0.054)
Curvature (empirical)	-0.445	-0.153	0.181	0.143	0.343**	0.466	0.120	0.751	-0.150***
	(1.077)	(0.116)	(0.134)	(0.202)	(0.158)	(0.965)	(0.080)	(1.507)	(0.045)
$h = 4$ quarters									
Level	2.122	0.167	-0.055	0.222	-0.403	-1.025	-0.063	-0.766	0.047
	(1.596)	(0.148)	(0.204)	(0.294)	(0.247)	(1.095)	(0.133)	(2.025)	(0.080)
Level (empirical)	1.877	0.124	-0.082	0.208	-0.405	-1.711*	-0.059	0.186	0.051
	(1.528)	(0.159)	(0.208)	(0.293)	(0.247)	(0.963)	(0.131)	(1.856)	(0.075)
Slope	1.034	0.071	-0.027	0.077	0.029	2.550**	-0.196**	-3.224*	0.070
	(1.446)	(0.120)	(0.189)	(0.217)	(0.222)	(1.086)	(0.094)	(1.869)	(0.069)
Slope (empirical)	1.194	0.169	-0.041	-0.157	0.072	2.486**	-0.147	-3.318	0.072
	(1.642)	(0.121)	(0.202)	(0.247)	(0.227)	(1.172)	(0.106)	(2.037)	(0.072)
Curvature	0.458	-0.027	-0.176	-0.268	0.042	2.882***	-0.004	-2.930*	0.056
	(1.476)	(0.130)	(0.140)	(0.258)	(0.208)	(0.674)	(0.111)	(1.657)	(0.063)
Curvature (empirical)	1.147	-0.066	-0.035	-0.242	0.188	2.747***	0.054	-3.598**	0.024
	(1.342)	(0.128)	(0.126)	(0.233)	(0.180)	(0.707)	(0.092)	(1.600)	(0.051)

Notes: The dependent variable is an indicator variable taking the value of one when the model augmented with a yield curve element produces more accurate forecast than the benchmark model and zero otherwise. 'ActiveMP', 'Infl. reg.', 'Output reg.' denote the active monetary policy, high inflation shock volatility, and high output shock volatility regimes, respectively. Newey-West standard errors are reported in parentheses. Asterisks denote statistical significance at the 1%(***), 5%(**) and 10%(*) levels, respectively. The forecasting model is DIAR. The sample period runs from 1972:Q2 to 2008:Q2.

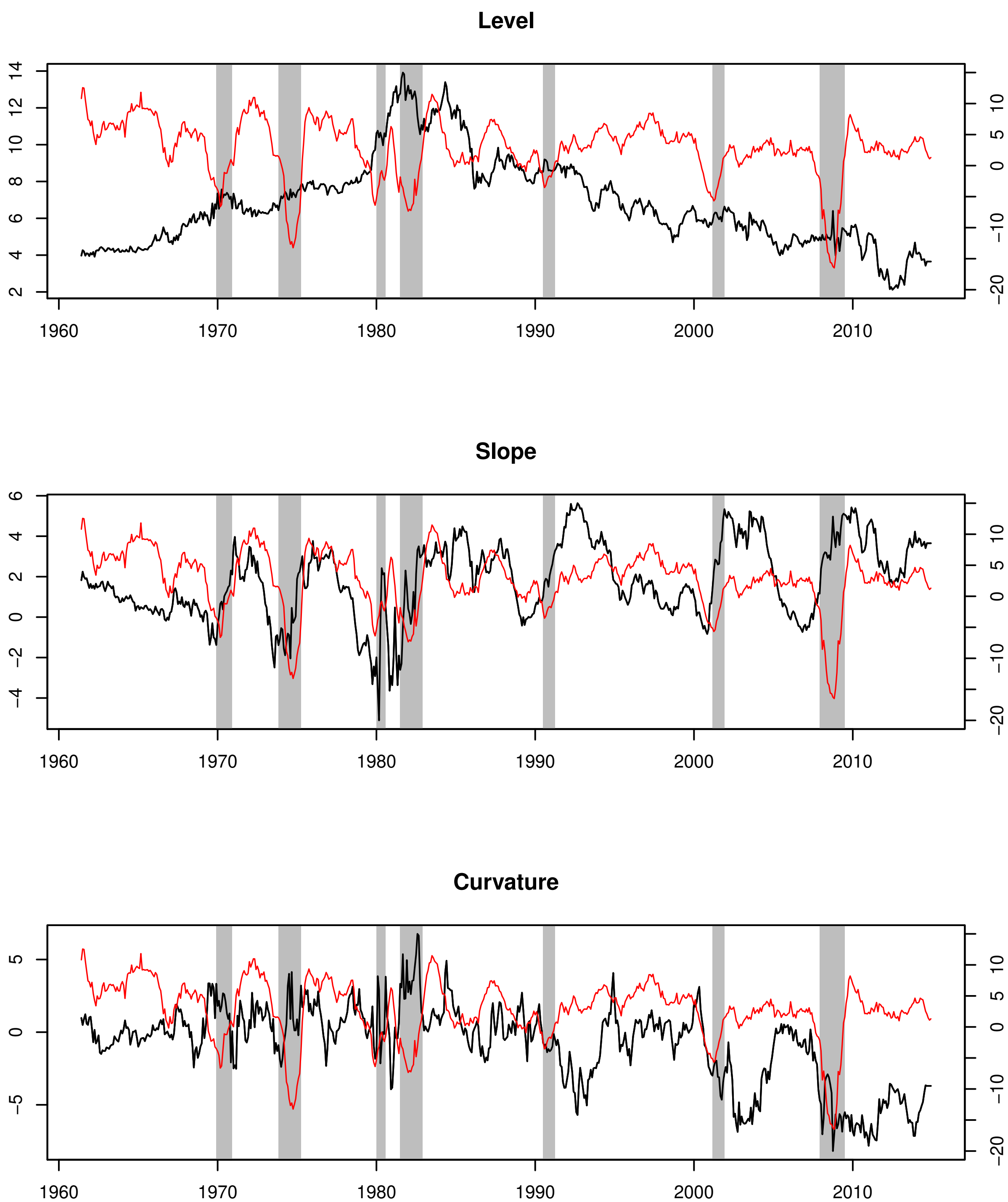

Figure 1: **Yield curve and industrial production growth**
Notes: The figure displays the level, slope and curvature of the yield curve (black line, left scale) and the subsequent 12-month-ahead industrial production growth (red line, right scale) from 1961:M6 to 2014:M7. The latent factor extractions are based on full-sample parameter estimates. Shaded areas indicate NBER recession dates.

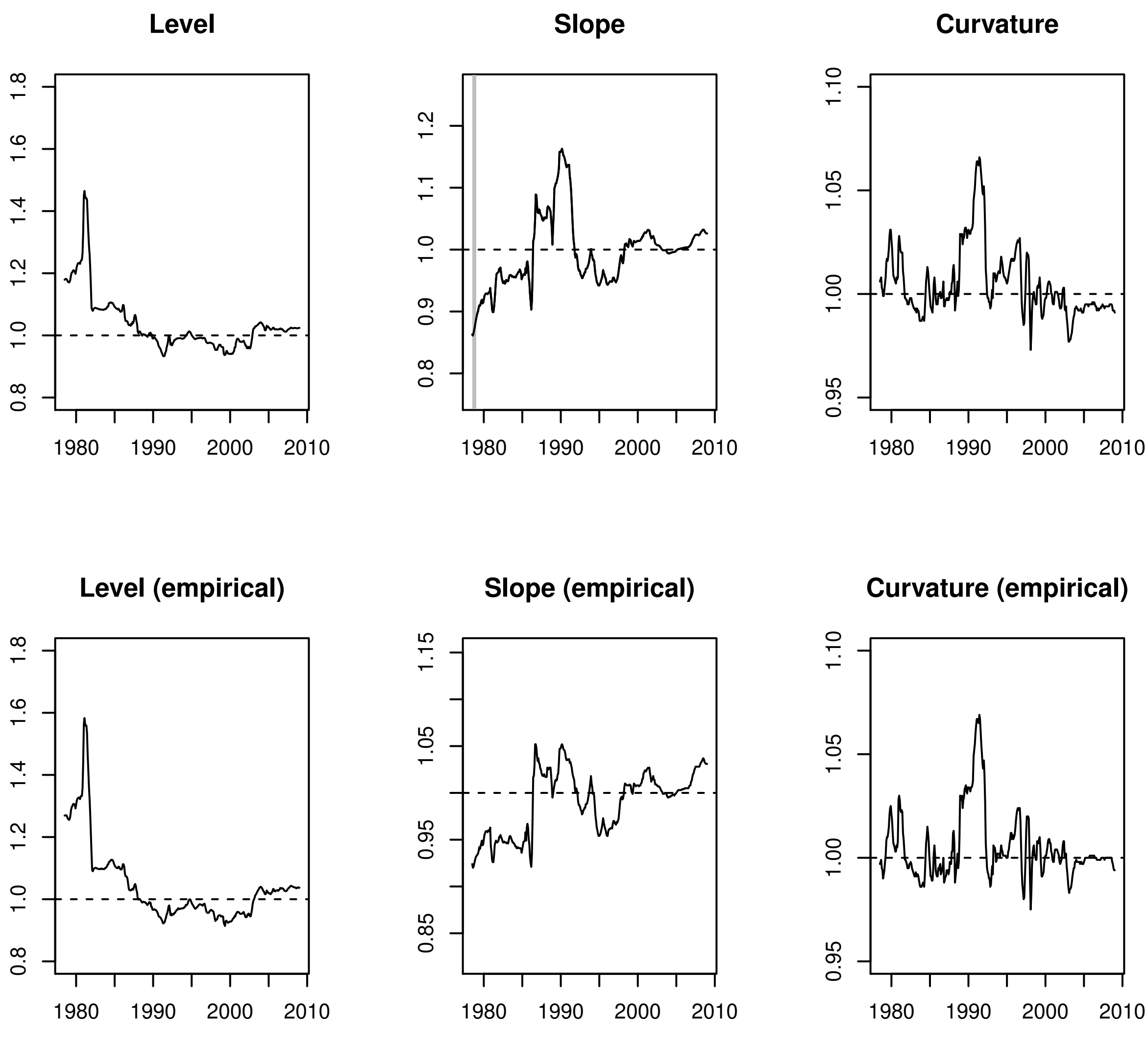

Figure 2: **Rolling relative MSFE values**

Notes: The figure plots relative MSFE values computed over a rolling window of 150 out-of-sample observations. The shaded areas denote the midpoints of windows in which the Giacomini and Rossi (2010) fluctuation test rejects the null of equal accuracy at the 10% significance level. The forecasting period runs from 1972:M5 to 2015:M7. The forecasting model is DIAR, and the forecast horizon is $h = 3$ months.

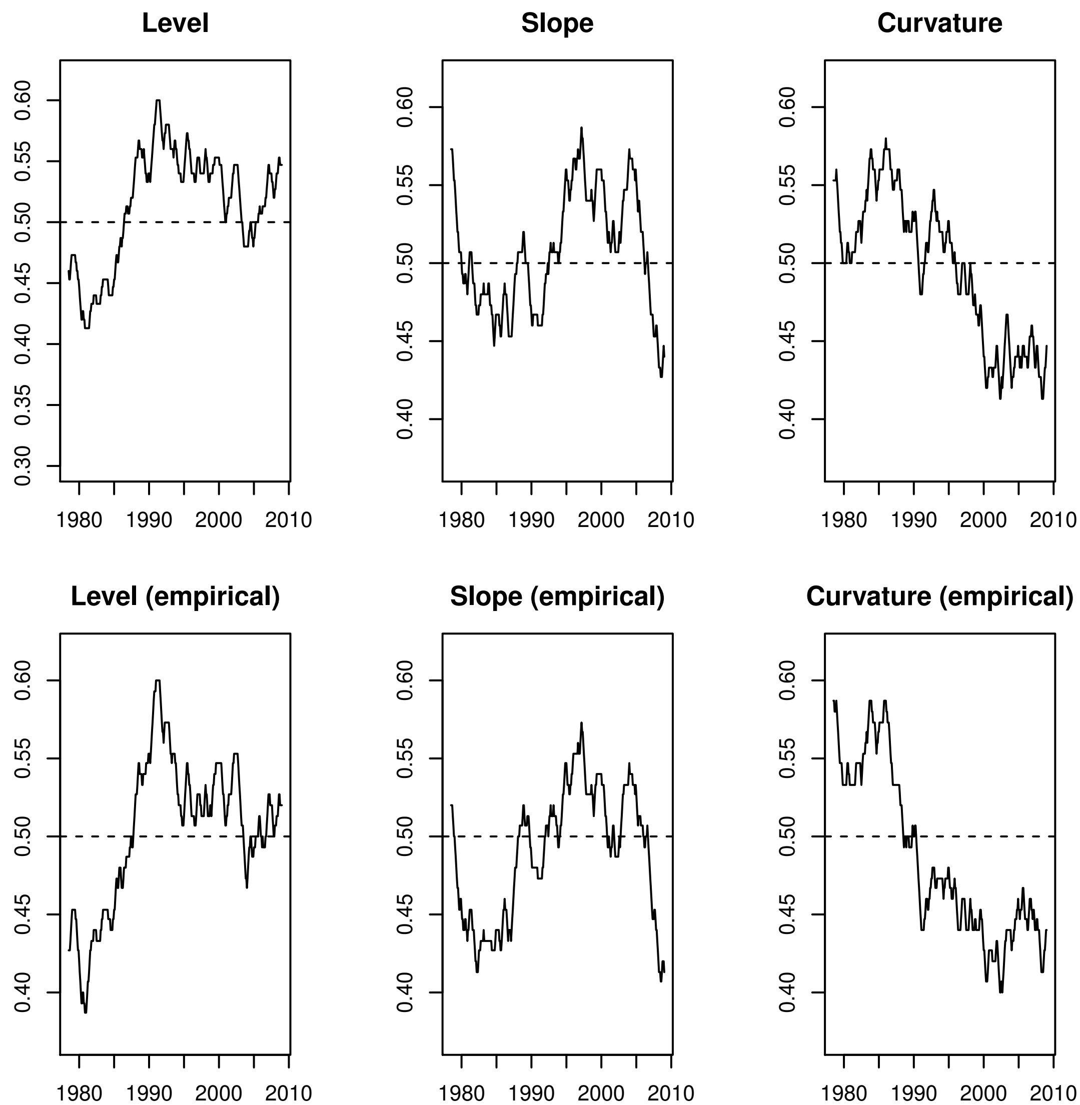

Figure 3: **Rolling sign predictability**

Notes: The figure plots the fraction of observations for which the forecasting model augmented with a yield curve element produces more accurate out-of-sample forecasts than the benchmark model. The fraction is computed using a rolling window of 150 out-of-sample observations. The forecasting period runs from 1972:M5 to 2015:M7. The forecasting model is DIAR, and the forecast horizon is $h = 3$ months.

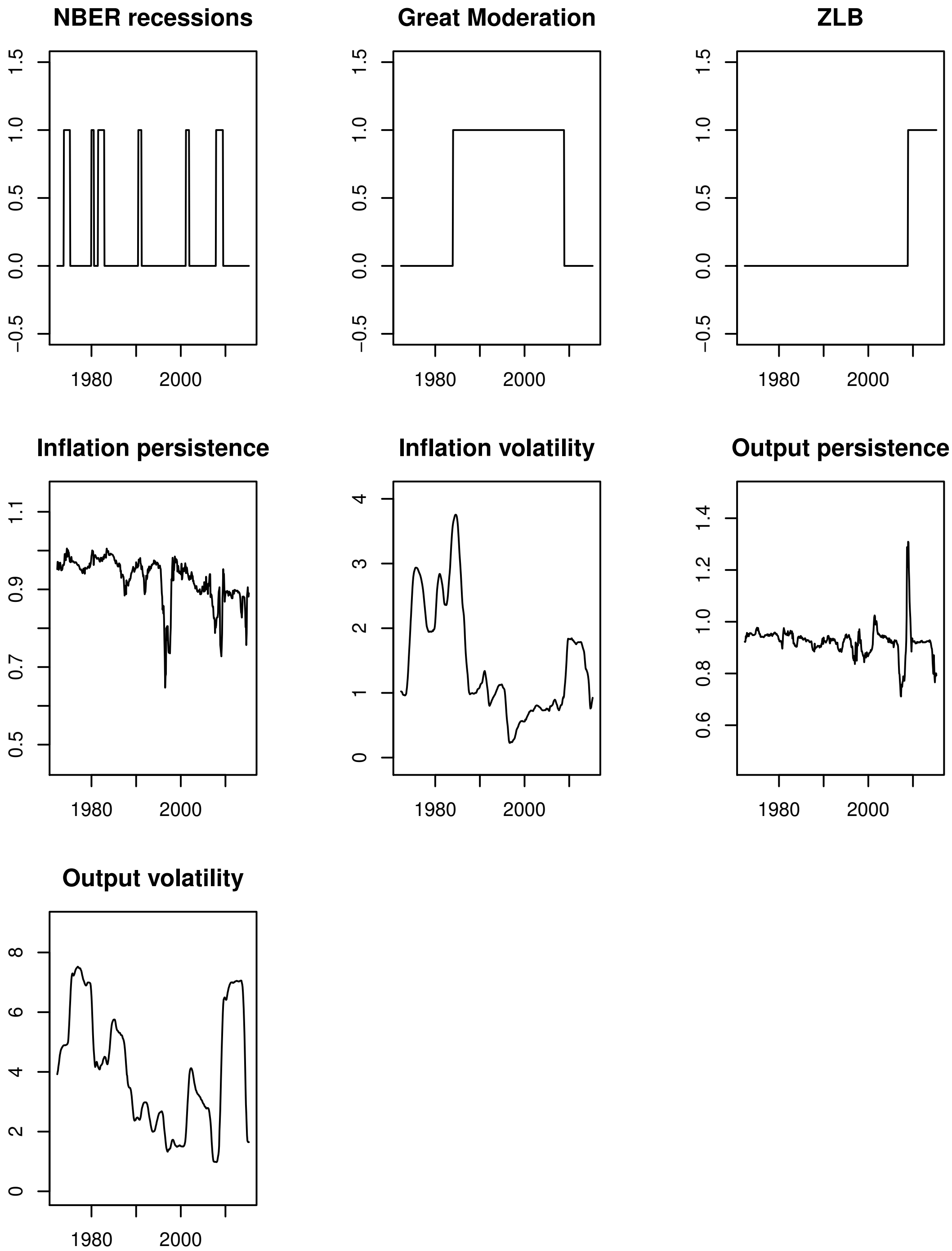

Figure 4: **Conditioning variables**

Notes: The figure depicts the conditioning variables used in the Giacomini and White (2006) test of equal conditional predictive ability. The sample period runs from 1972:M5 to 2015:M7.

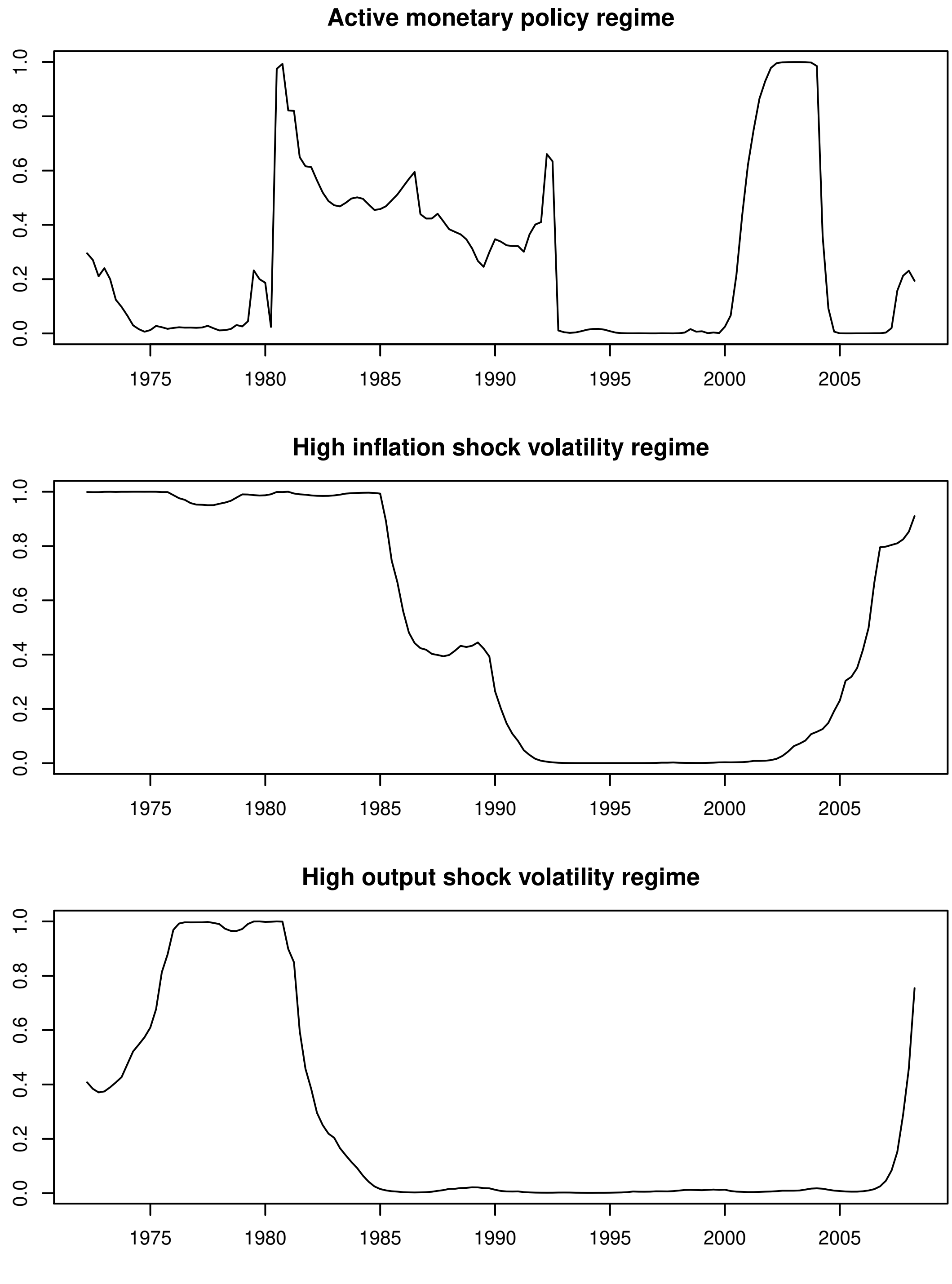

Figure 5: **Quarterly conditioning variables**

Notes: The figure depicts the conditioning variables used in the Giacomini and White (2006) test of equal conditional predictive ability. The sample period runs from 1972:Q2 to 2008:Q2.