Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis Financial Economics

Time variation in the term spread's predictive power: Evidence from the U.S. Treasury market

Name student: T.A. Kruik

Student ID number: 358759

Supervisor: Dr. K. Ji

Second assessor: Dr. J. Lemmen

Date final version: 19th of November 2018

ABSTRACT

In economic literature, scholars agree that the U.S. term spread can predict future real GDP growth. However, no such agreement exists with respect to its underlying mechanism. This paper contributes to the vast body of research by examining the instability of the predictive term spread and the mechanism behind it. The paper examined whether monetary policy conduct is the confounding factor of the forecast relation. In essence, this thesis investigated whether major monetary policy events coincide with significant changes in the term spread's predictive power. The reported findings do not support that monetary policy is the crucial factor in the yield curve's predictive power, nor do they strengthen the case of this relation is confounded by the consumption-smoothing behavior of Treasury market participants.

TABLE OF CONTENTS

I. INTRODUCTION

A considerable share of past literature examines the link between the yield curve and the real economy. This area of research has accumulated around U.S. Treasuries because of their essential role in global financial markets e.g. as a benchmark for currency and debt instruments. Within this strand of research, scholars investigate the positive relation between the spread on long- and short-term Treasury securities[1] and future U.S. real gross domestic product (GDP) growth. Initially, researchers found ample evidence suggesting that the term spread has the ability to predict future output changes.[2] Various economic actors adopted this variable as a forecasting tool because of its accuracy, simplicity, and timely availability. The term spread is especially relevant for economists who benefit from a clear view of future real activity: government agents determining next year's fiscal budget, investors assessing their business opportunities, and central bankers setting monetary policy. More recent literature provides evidence suggesting that this forecasting relation varies over time. The cause of this instability is as yet unknown, which lessens the reliability of the term spread. Nowadays, the term spread is widely used as a financial indicator supplementing the macroeconomic forecast models.

Despite the scholarly consensus on the term spread's predictive content regarding future GDP growth, two competing theories exist. On the one hand, proponents of the monetary-based view argue that this forecasting relation is spurious. They regard monetary policy conduct as the confounding factor of Treasury bond market behavior and real economic activity. The yield curve, they argue, summarizes investors' expected path of future interest rates. The bond market bases its predictions on its view on future actions of the Federal Reserve Bank (Fed). At the same time, the Fed can affect the actual future real GDP growth by conducting an expansionary or restrictive policy. On the other hand, proponents of the consumption-based view argue that the consumers' decision-making process is the driving force behind the predictive power of the term spread. That is, the bond market investor primarily makes an isolated judgment of the future economic environment. In contrast, monetary policy merely reacts to the current economic conditions. Behavior of the term spread is thus driven by the need for investors to invest in short- or long-term bonds based on their expectations of future output. Expectations of Treasury market participants regarding future monetary policy is only relevant for inflation. If bond investors expect future monetary contraction, expected inflation is low as future output is not stimulated. The decline of the yield on long-term bonds follows that of future GDP growth. In sum, both views attribute the positive relation to different mechanisms.

The following thesis aims to find evidence in favor of the monetary-based view by examining major monetary policy events in the post-GFC period. Only a small part of literature discusses the time-varying nature of the forecasting relation during this time span. This is a caveat as large shifts in monetary policy

[1] Also referred to as: Treasury term spread or slope of the Treasury yield curve.
[2] This ability is also known as the forecasting or predictive relation.

have occurred. These developments render earlier methods used in research less informative. More precisely, the monetary policy environment has changed within this time span, making earlier proxies for bond market expectations less accurate. The inability of the Fed to meet policy goals by altering its federal funds rate (FFR) characterizes the post-GFC period. In earlier literature, bond market investors base their prediction on a linear relation between the FFR and macro-economic variables. These so-called policy reaction functions are unable to describe the link between monetary policy conduct and the yield curve in the post-GFC period. In a speech at the Stanford Institute for Economic Policy Research, Yellen was explicit in the fact that these rule-based functions serve as a benchmark but "should not be followed mechanically" (2017, p. 16). This illustrates that the bond market should not base its expectations on policy reaction functions in the post-GFC period. More specifically, this time span encompasses changes in monetary policy circumstances of the Fed, including, for example, the full aftermath of the global financial crisis (GFC), the entrapment of interest rates around the zero lower bound (ZLB), and the conduct of unconventional monetary policy (UMP). In short, facing an evolving monetary policy environment, Treasury market participants find it hard to predict future monetary policy stance. The following thesis circumvents this problem by departing from policy reaction functions. Instead, this paper examines whether significant changes in forecast performance coincide with major monetary policy events. This method circumvents the inapplicability of policy reaction functions to represent bond market investor expectations. The central question in this paper is thus:

"*Does time-variation in the term spread's predictive power in the post-GFC period provide evidence in favor of the monetary-based view?"*

To answer the main question, I apply a two-step analysis. The first step is an examination of the time-varying nature of the forecasting relation over three decades using time-varying parameter models. The second step is an analysis of the timing of forecast breaks and major monetary policy events in the post-GFC period by using forecast break tests. The forecast break tests produce the exact months in which significant changes in the term spread's predictive power occur. In turn, these dates are easily compared to monetary policy events. Forecast breaks occur when the out-of-sample forecast performance is significantly different than its past performance. Major monetary policy events are occurrences that fundamentally shift bond market investors' expectations of future monetary policy. In essence, I argue that forecast breaks coinciding with major monetary policy events provide evidence in favor of the monetary-based view. The motivation for this two-tiered methodology lies within ensuring that the data examined is in line with previous research. Time-variation of the predictive power of the yield curve must be confirmed before embarking on the investigation of significant changes in forecast performance.

The following research contributes to current literature, firstly, by augmenting the limited amount of research on the monetary-based view in the post-GFC period. Answering the central research question can further improve our understanding of the underlying mechanism of a yet unexplained stylized fact.

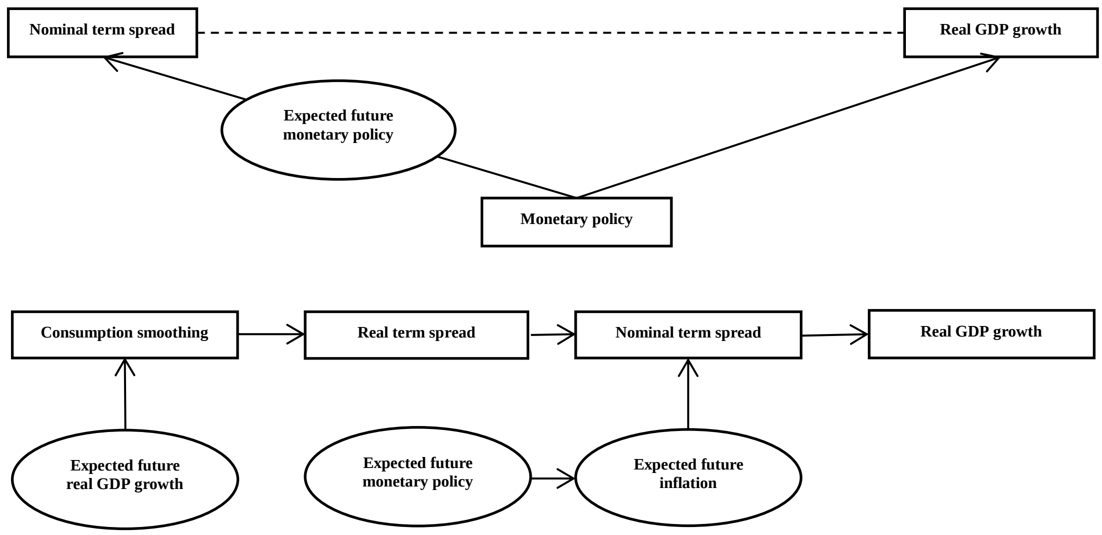
Secondly, this examination circumvents the post-GFC inapplicability of earlier used econometric methods. Thirdly, the current examination uses a methodology that accounts for multiple sources of forecast breaks. This is in contrast with literature mainly focused on the instability of the forecast coefficient. The current examination has relevance for the real-world application of the term spread as well. First, it will feed the debate on the usefulness of the term spread as a predictive tool. A useful economic predictor must at least be able to forecast consistently. A comprehension of the sources of instability enhances the reliability of the term spread as a macroeconomic predictor. Namely, prediction errors are minimized as variation of the forecasting relation can be factored in. For policymakers, the relevance of the following study is somewhat different. Finding evidence validating the monetary-based view cause the term spread to be an uninformative predictor. The key issue is assessing whether the findings extracted from the term spread were not previously processed by the Fed. If the yield curve is just a summary of what the Fed already knows, it is unfit to be an economic indicator in the toolbox of policy makers. Secondly, the current thesis examines the predictive power of the term spread from the perspective of its actual users. The examination of the difference between expected and actual out-of-sample performance mimics their every-day challenge to gain a reliable view of future economic output.

The remainder of this thesis is organized as follows: Section II reviews related literature and presents the first hypothesis, Section III presents the major monetary policy events expected to influence the forecasting relation and provides the second hypothesis, Section IV discusses data used, Section V elaborates on the methods applied, Section VI presents empirical results and confirms or rejects hypotheses, and Section VII answers the main question.

II. LITERATURE REVIEW

The link between the U.S. Treasury term spread and future real GDP growth has been widely studied in literature. Several scholars pioneered at examining this forecasting relation in isolation. All of their seminal papers used U.S. post-WWII to late-1980’s data, regressed in a linear OLS regression, and applied benchmarked in-sample and/or out-of-sample forecasts. To name a few: Laurent (1988) sets up a horse race between several monetary policy indicators. He finds the regression containing the spread between 20-year bond rate and the Federal Funds Rate (FFR) to have the lowest RMSE over the period 1961Q2-1986Q4, particularly at a 1- to 2-quarter horizon. Harvey (1988) shows that the spread on 9- and 3-month Treasury notes has more predictive power than lagged consumption and lagged stock returns at a 1- to 3-quarter horizon over the timespan 1947Q2 through 1987Q1. In a multivariate setting, spanning from 1955Q1 to 1988Q4, Estrella and Hardouvelis (1991) find that the difference between 10-year and 3-month bond yields can forecast real GDP change at a 4- to 6-quarter horizon more accurately than lagged inflation and the FFR, amongst other predictors.

Fig. 1: The upper flowcharts depicts the mechanism underlying the forecasting relation according to the monetary-based view. The lower flowchart depicts the mechanism underlying the forecasting relation according to the consumption-based view.



Theories underlying the predictive power of the term spread

While there is ample evidence of the existence of the term spread's predictive power, no such clarity exists about the underlying mechanism. The basis of explanatory theories is the expectations hypothesis (EH). This proposition implies that interest rates of long-term bonds are composed of the expected future short-term rates combined with a term premium. If investors expect a decline in the short rate in the course of the coming years, they would want to substitute their short-term bonds against their long-term equivalents. In effect, the yield on a five-year bond is equal to the one-year short rate expected four years from now, with an added reward for lending money for an extended period. Thus, the expectations of the bond market are said to be the driving force behind the slope of the yield curve. However, the EH is merely a starting point as it does not explain how the term spread is linked to real GDP growth.

Using the EH as a foundation, two competing theories describe how bond market participants drive the forecasting behavior of the yield curve. Firstly, according to the monetary-based view, the Fed's monetary policy is responsible for variation in both the term spread and future GDP growth. To illustrate this, the top panel of Fig. 1 shows two processes. The left-hand channel of effect is through the bond market's expectations of future monetary policy conduct. If, for example, the FFR is lowered today, bond yields would decrease over the entire maturity spectrum of the yield curve. Indeed, the EH argues that this curve represents the expected future path of policy rates. On long and short-term Treasury yields, however, investors react asymmetrically to FFR-alterations. The long-term yield decreases relatively less dramatically due to increased future inflation prospects, thus causing the term spread to increase. The right-hand channel of effect represents the effect monetary policy has on the real economy. Needless to say, the same decrease in the policy rate sparks investments fueled by cheaper credit. Thus, a positive term spread is correlated with an increase in real economic activity. An essential feature of the monetary-based theory is a spurious forecasting relation. That is, the variation in the term spread and future GDP growth is induced by a separate confounding factor. Monetary policy affects the extent to which these two variables are correlated. This is in contrast with the consumption-based view, which connects these two metrics through a microeconomic process.

Secondly, in the consumption-based view, the deep structure of the economy drives the forecasting relation. The lower panel of Fig. 1 provides a visual depiction of this theory. Harvey (1988) introduces this perspective and argues that cyclicality in expected real returns of bonds is expressed in cycles of per capita consumption. According to this view, a utility-maximizing forward-looking risk-averse individual prefers consumption to be constant. So, if individuals expect economic downturn in the coming decade, they will buy long-term bonds to stabilize consumption in the face of lower future yields on bonds. Investors swap short-term bonds for long-term securities, causing the term spread to decrease. In essence, consumption-smoothing is the main explanation behind the forecasting relation. The expected monetary regime is only relevant to indicate future inflation. That is, the consumption-based view applies to the real term structure while most evidence suggesting the forecasting relation exists is based on the nominal yield curve. In order to transcend to nominal terms, inflation should be stable and expected to be so in the future. Thus, the ultimate linkage of the consumption-based view is dependent on monetary policy. As such, changes in predictive power of the yield curve spread should coincide with changes in inflation persistence.

Findings in literature do not consistently point to either the monetary- or the consumption-based view. Regarding the former perspective, Estrella and Hardouvelis (1991) find that a substantial portion of the term spread’s predictive power operates independently from current and future monetary policy proxied by the nominal and real FFR as well as the 3-month Treasury bill rate. Moreover, Plosser and Rouwenhorst (1994) find that the slope of the yield curve provides information about future output growth in excess of current and expected future monetary policy actions, represented by historical twelve-month and realized M1-growth rate over the forecast horizon. Regarding the latter perspective, Rendu de Lint and Stolin (2003) use a dynamic equilibrium asset pricing model to show that an individual’s freedom of consumption-smoothing is essential in explaining the forecasting relation. Still, Benati and Goodhart (2008) find evidence suggesting that the predictive power of the term spread does not hinge on inflation persistence, which is an essential precondition for the validity of the consumption-based view.

The instability of the Treasury term spread’s predictive power and the underlying sources

Evidence shown in in-sample and out-of-sample forecast results suggests varying degrees of the predictive power of the U.S. Treasury term spread during different timespans. For instance: Haubrich and Dombrosky (1996) were among the first to document a deterioration of the yield curve’s forecasting ability across the timespan 1985-1995. In contrast, Chinn and Kucko (2015) find a strengthening of the predictive power of the yield curve in the same period. Moreover, Bordo and Haubrich (2008) find evidence suggesting that the term spread has more predictive power for the post-WWII than for the span of fifty years pre-WWII. De Pace (2013) finds the predictive relation to be weak or absent in the timespan 1995-2008. In general, several underlying sources of this time-variation are theorized. The relation between the yield curve and the real economy can be fundamentally changed by technological advances. Expected real GDP growth is more volatile in an innovative environment. A certain technological advancement can render a specific product or production method obsolete. In the same vein, globalization can play a key role. Production growth rates of one country are more volatile when it hinges partially on the conduct of its competitors and its exposure to them. Also, fundamental changes in market behavior can also be an underlying source. As bond investors grow more sophisticated in their trading behavior, bond prices reflect more accurately the expectations of future GDP growth of Treasury market participants. The other scholars investigate whether monetary policy conduct is a source of the time-variation in the forecasting relation. To name a few of these scholars: Bordo and Haubrich (2004), analyzing the timespan 1875-1997, find the forecasting performance of the term spread to differ when credibility of monetary regimes varies. D'Agostino, Giannone, and Surico (2006) report a deterioration in the predictive power of the Treasury term spread when comparing pre- to post-1985 data. They suggest that this is due to non-transparent and suboptimal monetary policy. Giacomini and Rossi (2006) find forecast breakdowns to coincide with Fed chairmen taking office, from 1979 to 2006. This paper argues that if major monetary policy events coincide with variation in the term spread's predictive power, it poses evidence in favor of the monetary-based view. Further details are provided in Section III.

Methods of examining the instability of the forecasting relation

In general, three methods on how to empirically investigate the instability of term spread's forecasting relation exist. Firstly, a tendency exists to use breakpoint tests to uncover structural breaks in the forecast coefficient. Estrella, Rodrigues, and Schich (2003), for example, investigate the existence of a single known breakpoint by extracting the asymptotic distribution statistics of the forecast coefficient produced by different probability models. These researchers use a Chow-test for parameter homogeneity that incorporates principal statistics dependent on the structural break's characteristics. The occurrence of parameter inconsistency is pinpointed when the resulting F-test on the distribution rejects the null-hypothesis of no breakdown on a specific date. It is proposed that one weakness of a breakpoint analysis is the fact that it does not fully allow for the possibility of more than one breakpoint in the forecasting relation. For example, Bai and Perron (1998) argue that singular structural break tests lack power in the presence of multiple structural breaks. To circumvent this problem, De Pace (2013) employs tests for unknown single as well as multiple breakdowns, although he only examines the term structure in isolation and does not apply this test to the predictive relation. Secondly, scholars investigate instability in the forecasting relation by studying the evolution of the forecast-coefficient over time. De Pace

(2013), for example, formulates basic predictive models that contain time-varying regression coefficients. This method can be applied to different types of parameter heterogeneity. In their literature review, Wheelock and Wohar (2009) report that the coefficient is conventionally theorized to follow either a random walk or a stationary AR(1) process. A common critique of this form of estimation is that it utilizes future data points as well as past ones. Indeed, in-sample predictive performance is less relevant in the practical forecasting framework; policy makers can only use data presently available.

To bypass the concerns of the former two methods, this thesis focuses on the third way, which examines the sources of the forecasting relation's time-variation. Structural break tests and time-varying parameter models cannot be used to examine the timing of major monetary policy events. Neither of these methods necessarily indicate whether significant shifts in the predictive ability occur. Time-varying parameters remain open to interpretation when predictive power is significantly diminished. Of course, earlier literature signaling periods of weakened forecasting relation were sound in inferring their findings. Yet to recognize a forecast break, one needs a less arbitrary gauge than a visual inspection of the time evolution. In the same vein, structural breaks in parameters are not the only sources of forecast breakdowns. Giacomini and Rossi (2009) show that structural breaks can occur in the variance of errors as well. They further emphasize that breaks at multiple higher moments could have countervailing effects on the forecast performance. This plurality of sources of structural breaks hampers the usefulness of structural break tests in uncovering sources of the forecasting relation.

To circumvent these drawbacks, this thesis uses a two-sided version of Giacomini and Rossi's (2009) forecast breakdown test. This alternative framework allows the forecasting relation's instability to be linked to possible sources thereof. Giacomini and Rossi (2009) argue that a forecast breakdown occurs when the average in-sample predictive ability of past data is significantly better than a single subsequent out-of-sample prediction. To circumvent the critique of time-varying parameters, this alternative method provides a clear statistical threshold extracted from the distribution of the differences between expected and actual forecast performance. In other words, their methodology enables significant shifts in predictive content in the term spread to be pinpointed in time, with little space of arbitrariness. By addressing forecast breakdowns directly, the occurrence of structural breaks at higher moments is bypassed. Additionally, Giacomini and Rossi (2009) also circumvent the necessity of a large amount of data in the sub-samples of the structural break tests. This is especially relevant for the following thesis as the examined timespan is expected to contain relatively recent forecast breaks.

Another important advantage of the forecast break test is its circumvention of the absence of a sound post-GFC proxy for expected monetary policy conduct. Some scholars argue that bond market expectations of future monetary policy conduct represent, in fact, a linear relationship between the FFR and monetary policy goals. That is, monetary policy can be characterized by their FFR adjustments in reaction to gaps in expected and targeted inflation and output. Two types of monetary policy reaction functions are generally observed to proxy these expectations. On the one hand, researchers use the backward-looking version as defined by Taylor (1993). This function states that the FFR behaves countercyclically with respect to lagged inflation and output gap. Using this function, Estrella (2005) find alterations in inflation aversion to coincide with a weakened forecasting relation. On the other hand, scholars link the instability to the forward-looking monetary policy reaction function as defined by Clarida, Galí, and Gertler (2003). For example, Estrella, Rodrigues, and Schich (2003) and Feroli (2004) find evidence suggesting that a structural break, which occurred at the appointment of Volcker, is caused by an increased stringency in inflation targeting. On a more profound level, Benati and Goodhart (2008) find evidence suggesting the forecasting ability hinges on the predictability of future monetary policy. If the bond market faces an increasingly unpredictable central bank, they argue, it takes into account an additional risk premium to their expected future interest rates. In the post-GFC period, however, the FFR ceases to be the primary tool of the Fed. The policy rate is stuck around zero, so any increased tendency to combat inflation or output gaps are not reflected by changes in the FFR. In a ZLB and UMP period, the FFR is less informative and renders the policy reaction function as an unreliable proxy for the bond market's expectation-making process. Instead of assuming a linear expectations model, the forecast break test examines disruptions in the accuracy of bond market participants' expectations.

Against this background, this thesis recognizes that the search for the underlying mechanism of the term spread's predictive power of the yield curve spread demands a view of the instability of the forecasting relation over time. Therefore, it is sensible to ascertain whether the used dataset shows the same time-variation in the term spread's predictive power. All in all, the first step in the two-step procedure entails the examination of the following hypothesis:

$\mathrm{H1}_0$: The predictive power of the yield curve is unstable over time.

III. MONETARY POLICY EVENTS AFFECTING THE FORECASTING RELATION

This paper argues that forecast breaks occurring at major monetary policy events provide evidence in favor of the monetary-based explanation of the forecasting relation. Each event is either theorized to invoke a large shift in the expectation-making process of Treasury market participants or is expected to do so in light of earlier evidence found in literature. This is a high threshold and results in the exclusion of some monetary policy events that are generally noteworthy.[3] Furthermore, breaks can either be an amelioration or a deterioration of the forecasting ability of the term spread. Major events can either

[3] Some of the omitted events pertain to speeches at Jackson Hole given by Fed Chairmen. These occasions are indeed forward-looking and widely anticipated by financial markets. However, the imbedded hints lack the gravity which characterizes the announcements of UMP actions. In the same vein, spillover effects of the ECB's monetary policy conduct are unlikely to significantly change the Treasury market's view on the Fed's future monetary policy. The implementation of the ECB's QE or Draghi's forward guidance is not surprising to Treasury market participants. Moreover, the spillover effects of the eruption of the European Sovereign Debt Crisis on the Treasury market is likely to change the expectations forming behavior of the bond market participants, but not the mechanism itself. Thus, while omitted events are important to the Treasury market participants, it cannot be said to steer the manner in which future monetary policy is considered in bond prices.

signal a strengthening or a weakening of monetary policy. Only an event signaling an increase in expected effectiveness of monetary policy should be accompanied by a positive structural change in the predictive power of the term spread. Thus, a forecast break that is in line with (contrary to) the theorized effect of an event, provides evidence in favor of (against) the monetary-based theory.

Central in the following thesis are the following five major monetary policy events that may coincide with forecast breaks in the post-GFC period. Firstly, the occurrence of the GFC could affect the predictive power of the yield curve. That is, a reappraisal of expected future monetary policy can cause a forecast disruption. As Cukierman (2013) noted, the GFC roused anxiety regarding financial cohesion and emphasized the role of a central bank as a lender-of-last-resort. From a monetary perspective, this event signals an increase in the predictability of policy makers' response to past output growth data and inflation rates. The bond market is thus more able to predict the future path of interest rates, as affirmative monetary action is needed. Previous literature recognizes a subdivision between pre- and post-Lehman Brothers collapse when examining how monetary policy shocks can induce changes in the term spread (see, e.g., Claus and Dungey, 2016). Thus, there is likely to be a break in the forecasting relation in September 2008. Furthermore, there is evidence suggesting that the predictive content is likely to increase. Benati and Goodhart (2008) find the term spread to contain increased predictive content during the Volcker-recession and the dot-com bubble. Similarly, De Pace (2013) shows the GFC coincides with an amelioration of the forecasting ability. In short, the predictive power of the term spread is expected to increase at September 2008. Secondly, past evidence suggests that the appointment of a new Chair of the Fed's Board of Governors is related to breaks in the forecasting relation. For instance, Giacomini and Rossi (2006) find forecasting performance to deteriorate under the Burns-Miller and Volcker monetary policy regime. These researchers also find an improvement of this relation at the start of Greenspan's term. Thus, the following paper expects a forecast break to occur at Yellen's appointment as chairwoman in February 2014. Whether this break constitutes an improvement or deterioration of the predictive power, is less clear. Arguably, Yellen's nomination speech would provide an indication of this. In reaction to the actions of her predecessor as Chairman, Yellen states: "I have strongly supported his commitment to openness and transparency and will continue to do so if I am confirmed and serve as Chair" (2013, p. 5). Based on this, there is reason to believe that predictability has at least not weakened, though such an assumption remains arbitrary without hard evidence.

Thirdly, another major monetary policy event is the entrapment of short-term interest rates at the ZLB. In December 2008, the Federal Open Market Committee (FOMC) declared it would set the target spectrum of the FFR between zero and a quarter percent. As Estrella and Hardouvelis (1999) first elaborated, the Fed generally aims to change the short rate directly, while influencing the long-term yields via expected future short rates, inflation, and output growth. The phenomenon of short rates hovering at the ZLB is a worrying development. Because further economic deterioration was expected, credit constraints had to be lowered even more. Against this background, Hamilton and Wu (2012) find evidence suggesting that the yield curve slope is less susceptible to traditional short-term open market operations and more so to monetary policy aimed at the long end of the yield curve in the ZLB-environment. In the same vein, Hännikäinen (2015) finds that the introduction of the ZLB-environment restricts the movement of the yield curve spread. That is, the term spread moves in parallel with the long-term interest rate and short rates are fixated and cannot turn negative. Thus, the fact that the short-rate is stuck around zero can imply that it remains relatively stable compared to future GDP growth. As the primary toolbox of monetary policy makers is depleted, bond market investors experience more difficulty in predicting future monetary policy. An observed deterioration in the predictive content coinciding with the FOMC ZLB-declaration in December 2008 is therefore expected.

Fourthly, the implementation of QE's, also known as large-scale asset purchases (LSAPs), could pose as major monetary policy events affecting the term spread's predictive power.[4] The Fed, stripped from the ability to affect the market interest rates through the FFR, engaged in the purchase of hundreds of millions of medium- and long-term Treasury securities in order to lower credit constraints in the private sector.[5] The first QE program announcement was made in November 2008, but did not include Treasuries. In the following month, Fed chairman Ben Bernanke and the FOMC hinted at expanding the purchase program to longer-term Treasury securities. In March 2009, the Fed announced it would expand QE1 to long-term Treasuries, making it a purchase of $300 billion. In November 2010, QE2 was announced, which constituted buying up $600 billion in Treasury bonds. In September 2012, the Fed signaled QE3, which excluded Treasuries. To ascertain on which dates one can expect a forecast break in the following analysis, Krishnamurthy and Vissing-Jorgensen (2011) postulate two considerations. Firstly, the announcement of LSAPs affect bond yields through the signaling channel. Publication dates are therefore more relevant than implementation dates. Furthermore, publication dates of every piece of information on QE are not likely to independently impact the yield curve. Jarrow and Li (2014) find the effects of QE to be persistent enough to bridge the time gaps in-between announcement dates. Secondly, QE2 had a superfluously negative impact on Treasury yields, due to a shift in expected future FFR decreases. Additionally, in their event study, Gagnon et al. (2011) use a one-day window and find the largest deterioration of the Treasury yield spread on two- and ten-year securities to occur in November 2008 (20 basis points), December 2008 (28 basis points), and March 2009 (25 basis points). The Treasury market participants' predictions are most strongly affected by the introduction than by the inclusion of Treasury securities. Thus, the monetary-based view predicts that significant forecast breaks will occur in the months of the initial announcement of QE1 (November 2008) and not at the occurrence

[4] In contrast, Hännikäinen (2015) combines the start of the ZLB-era with forward guidance and QE. I would argue that their implications for the term spread are fundamentally different. Namely, the former relates to the short-term rates and the latter relates to the long-term rates. As such, these events should be considered in separation.

[5] Gagnon et al. (2011) were among the first to document and explain QE's channels of effect. These researchers argue the large asset purchases conducted by the Fed has reduced the aggregate duration risk through the removal of high duration assets in the market. Investors demand a lower term premium for holding long-term assets, as their portfolio has relatively less duration risk.

Table 1: summary of the five major monetary policy events. Dates of the events are noted, as well as their expected effect on the forecasting relation.

Major monetary policy event	Date	Expected effect on forecasting relation
Great Financial Crisis	September 2008	(+)
Quantitative Easing	November 2008	(+)
Zero Lower Bound	December 2008	(-)
Forward Guidance	March 2009	(+)
Appointment Yellen	February 2014	(+)/(-)

of the hint of expanding QE1 to Treasury securities (December 2008) nor at the announcement of QE2 (November 2010). Furthermore, the forecast breaks are expected to be an improvement of the forecasting ability as the predictability of future monetary policy is increased. Moreover, Hännikäinen (2015) finds QE in a ZLB environment to have a strengthening effect on the forecasting relation in respect to industrial production growth. In short, an improvement of the forecasting relation is expected to occur in November 2008.

Fifthly, the use of forward guidance, the second unconventional monetary policy instrument, is another major monetary policy event relevant for the forecasting relation. Borio and Zabai (2016) and others articulate that after reaching the ZLB, the Fed attempts to guide market expectations regarding future monetary policy actions by means of open communication about the future path of the FFR. In March 2009, the Fed expressed the view that interest rates were expected to be substantially lower for a prolonged period. This type announcement is known as qualitative forward guidance. In September 2009 and August 2011, the FOMC was more precise and specified a time span of low interest rates. These announcements are also known as calendar-based forward guidance. Next to this, there is also threshold-based forward guidance, which occurred on December 2012 and December 2013. It is hard to say how the act of forward guidance causes variation in the forecasting relation through bond market expectations. This thesis argues that the use of forward guidance underlines the credibility of the Fed's capability of influencing the economy. This monetary policy event increases the predictability of future monetary policy and strengthens the forecasting relation as a result. Thus, if the predictive power of the yield spread increases (decreases) in March 2009, this event will supply evidence in favor of (against) the monetary-based perception. Table 1 summarizes all predictions in line with the monetary-based view. The second hypothesis with regard to the second step of the analysis is therefore:

$H2_0$: The instability of the forecasting relation is linked to the five US major monetary policy events.

IV. DATA

The investigated timespan contains monthly observations ranging from January 1986 to December 2016. The rationale behind the start date is twofold. First, Taylor (1999), among others, regards the period 1986-1997 as a distinct monetary era characterized by economic stability and responsiveness vis-à-vis inflation and output gaps. He further states that monetary policy conduct has matured from the start of this period. It can therefore be expected that during the investigated timespan there is some degree of stability of the bond market's perception of future monetary policy. The following thesis, therefore, acknowledges Taylor's (1999) starting date of the Fed's sophisticated monetary policy conduct and extends this period until the end of 2016.[6] Furthermore, a long span of data is needed for a sound application of the used methods. A time span of 31 years is particularly useful for the second step. It is of importance to base the in-sample predictions on a sufficient amount of data points, while leaving enough room to construe the distribution of the metric used to assess the changes in forecast performance.

The data sample includes 3-month Treasury bill rates derived from H.15 releases expressed on a bond equivalent basis.[7] Monthly-averaged observations of 3-month Treasuries are not used, as Estrella and Trubin (2006) find that interpolated values of Treasury yields are less accurate. Furthermore, 10-year Treasury note constant maturity yield and quarterly real GDP data extracted from the Federal Reserve Economic Data (FRED). It is of importance to note that the latter variable is originally observed on a quarterly basis but are interpolated so as to turn them into monthly observations. More precisely, spaces between known data points are estimated through cubic spline interpolation. This method connects two quarterly observations by a smooth line that allows the estimated data points to have equal first and second derivatives as the preceding and succeeding gaps. In layman's terms, instead of drawing a straight line between points, this method allows for a more natural path. Furthermore, the GDP observations are chained in 2009 dollars and seasonally adjusted.

V. METHODOLOGY

To examine whether the post-GFC period provides evidence in favor of the monetary-based view, the current thesis applies a two-step analysis inspired by De Pace (2013) and Giacomini and Rossi (2009). Preliminary to these steps are the examination of the data and the specification of the main benchmark models. The first step is the construction of time-varying parameter models. The second step is an analysis of whether forecast breaks occur in line with the monetary-based view. There are some limitations worth mentioning. Firstly, the following thesis does not embark on a horse race between the term spread and other leading financial indicators. The scope of this thesis is not to build linear models with the strongest forecasting relation nor to assess the overall predictive power. Instead, the emphasis is on the stability of the predictive content over time, no matter the absolute forecast performance. Secondly, the current examination does not discuss all manners in which the predictive power of the

[6] Moreover, while the dot-com bubble is contained in the dataset, Mussa (2005) considers it to be irrelevant for monetary policy or the real economy. Without this linkage it goes thus beyond the scope of this thesis, although Benati and Goodhart (2008) find the term spread to contain increased predictive content during the dot-com bubble.
[7] The Board of Governors of the Federal Reserve System publish daily observations of selected interest rates in their H. 15 releases.

yield curve can be displayed through time. For example, Benati and Goodhart (2008) examine the size of the adjusted R^2 to uncover time-variation in the term spread.

Three types of benchmark models

The raw data is conversed into the main variables used in the methodology using the following formulas:

$$s_t = y_t^{(120)} - y_t^{(3)} \tag{1}$$

$$g_{t\to t+h} \equiv \left(\frac{1{,}200}{h}\right) * \left[\ln\left(\frac{G_{t+h}}{G_t}\right)\right] \tag{2}$$

Eq. (1) defines the spread as the difference between the 10-year and 3-month Treasury yields.[8] Eq. (2) describes the dependent variable as a monthly adaptation of Estrella and Hardouvelis' (1991) formula for annualized cumulative percentage change in real GDP. G_t is the real GDP-level observed at time t. A natural logarithm is used because small changes in the fracture can be directly interpreted as percentages. Following De Pace (2013), the starting point of the following investigation are three basic benchmark models that characterize the term spread's ability to predict future real GDP growth:

$$g_{t\to t+h} = \alpha + \beta * s_t + \varepsilon_{t+h,h}, \qquad \varepsilon_{t+h,h} \sim N(0, \delta_\varepsilon^2) \tag{3}$$

$$g_{t\to t+h} = \alpha + \beta * s_t + \gamma g_{t-h\to t} + \varepsilon_{t+h,h}, \qquad \varepsilon_{t+h,h} \sim N(0, \delta_\varepsilon^2) \tag{4}$$

$$g_{t\to t+h} = \alpha + \beta * s_{t-1} + \gamma g_{t-h\to t} + \varepsilon_{t+h,h}, \qquad \varepsilon_{t+h,h} \sim N(0, \delta_\varepsilon^2) \tag{5}$$

The first benchmark model, represented in Eq. (3), describes the relation between the growth of real GDP from time t until a period h months later, $g_{t\to t+h}$, and the slope of the yield curve observed at time t, s_t. The second benchmark model, defined in Eq. (4), adds lagged real GDP growth rate as an explanatory variable to Eq. (3). It is widely known that the cyclical nature of the economy entails that periods of growth, as well as period of decline, tend to occur in series. Past values are thus expected to have some predictive power regarding future real GDP growth. The third benchmark model, represented by Eq. (5), is similar to Eq. (4) except that the former uses the term spread of one month ago. In the monetary-based view this would mean that the bond market expects monetary policy has a one-month lagged effect on the real economy.

The intuition of the forecast coefficient is similar across Eq.'s (3), (4), and (5). The sign and magnitude forecast coefficient β reflect the historically found predictive content of the yield curve spread. That is, a positive (negative) value indicates that a positive term spread will signal future real GDP growth (decline). Furthermore, the term spread, s_t, is in nominal terms, while the real GDP growth, $g_{t\to t+h}$, is

[8] Similar to Estrella and Hardouvelis (1991), among others, the following thesis does not consider spreads using other maturities as their marginal added value is considered to be minimal.

expressed in percentages. Thus, the interpretation of the magnitude of the forecast coefficient is such that a one percent nominal spread in the term spread corresponds with a factor-β percentage increase of real GDP over the next h months.[9] Each of the three types of benchmark models will be used to compute OLS-regressions across different forecast horizons. The linear OLS estimates that show a statistically significant forecast relation are used in the two-step procedure used to answer the main research question.

Two-step methodology

Step one: time-varying parameter model

After choosing the benchmark models, the focus is on the methods that produce evidence that enables the validation or rejection of the hypotheses. The first step entails plotting the time-varying parameters across the models containing a statistically significant forecast coefficient. The method of constructing time-varying parameters carries two formal assumptions. First, sample T contains T-1 parameter breaks. In contrast with methods focusing on one-time structural changes in parameters, this time-varying parameter model assumes a break in every consecutive observation. Second, considering the finite number of possible breaks, the degree of change in the coefficient following a break is limited by an indicated stochastic process. This paper identifies one type of autoregressive stochastic processes: a random walk model without drift.

$$g_{t\to t+h} = \alpha + \beta_t * s_t + \varepsilon_{t+h,h}, \qquad \varepsilon_{t+h,h} \sim N(0, \delta_\varepsilon^2) \tag{6}$$

$$g_{t\to t+h} = \alpha + \beta_t * s_t + \gamma g_{t-h\to t} + \varepsilon_{t+h,h}, \qquad \varepsilon_{t+h,h} \sim N(0, \delta_\varepsilon^2) \tag{7}$$

$$g_{t\to t+h} = \alpha + \beta_t * s_{t-1} + \gamma g_{t-h\to t} + \varepsilon_{t+h,h}, \qquad \varepsilon_{t+h,h} \sim N(0, \delta_\varepsilon^2) \tag{8}$$

$$\beta_t = \beta_{t-1} + \epsilon_t, \qquad \epsilon_t \sim N(0, \delta_\epsilon^2) \tag{9}$$

The time-varying parameter model starts with the assumption that the coefficient in the benchmark models are time-varying. Eq. (6), (7), and (8) respectively describe Eq. (3), (4), and (5) with the added feature that β is time-dependent. Subscript t, thus, signals that the predictive content of the term spread varies over time. Eq. (9) describes the stochastic process of the time dependence by imposing a Markov structure. Namely, the value of β_t is dependent on its value one month earlier, β_{t-1}, plus a random value, ϵ_t. This is a non-stationary process as there is no stable data generating process. That is, shocks, which are channelized by the error term, are permanent. The following thesis expects that the data generating process of the forecast coefficient is non-stationary without a drift. A drift, represented by a constant, would imply the coefficient would be less likely to decrease over time than to increase. Literature does

[9] On a side note, the forecast coefficient represents the predictive content of the yield curve in excess of other predictors in the second and third type of benchmark models. In literature, this is known as the marginal predictive content of the term spread.

not provide any evidence suggesting such data generating process for the forecast coefficient. Prolonged periods of weakened and strengthened forecast ability would be the best match to a random walk model without a drift. In the same vein, Sarris (1973) argues that if parameters are expected to show large time-variation, a characterization of the coefficient as a random walk is most appropriate. Also, to reach an unbiased computation of the forecast regression, Young et al. (2007) suggest the application of a fixed-interval smoother, as well as a Kalman Filter.[10]

Step two: forecast break test

The second step is the application of a two-sided version of the forecast breakdown test designed by Giacomini and Rossi (2009). The central metric is the surprise forecast error (SFE). This variable represents the mismatch between expected and actual out-of-sample forecast performance. Expected out-of-sample performance is based on the average in-sample performance. The following thesis introduces recursive and moving-window regressions for these predictions. These two methods mimic the decision-making process of economic actors. It considers the amount of information available when determining their actions.[11] To calculate a certain data point, a recursive regression uses all observations prior to it. Each subsequent observation is based on an additional data point compared to the preceding observation. By contrast, the rolling-window regression uses a fixed vintage.

The investigation of the second hypothesis consists of four stages. In the first stage, a series of in-sample and out-of-sample forecast performances are generated. These require the data sample to be divided into three parts: in-sample window (m), forecast horizon (h), and out-of-sample window (n). The following equations specify in-sample and out-of-sample forecast errors:

$$FE_{t+h}(\widehat{\beta_t}) = \left(g_{t\to t+h} - f_t(\widehat{\beta_t})\right)^2 \qquad \text{for } t = m, m+1, \dots, T-h, \qquad (10)$$

$$\overline{FE}_t(\widehat{\beta_t}) = t^{-1} \sum_{k=1}^{t-h} \left(g_{k-h\to k} - f_{k-h}(\widehat{\beta_t})\right)^2 \qquad \text{for } t = m, m+1, \dots, T-h, \qquad (11)$$

$$\overline{FE}_t^*(\widehat{\beta_t}) = m^{-1} \sum_{k=t-m+1}^{t-h} \left(g_{k-h\to k} - f_{k-h}(\widehat{\beta_t})\right)^2 \qquad \text{for } t = m, m+1, \dots, T-h. \qquad (12)$$

Eq. (10) describes the manner in which the sequence of out-of-sample forecast errors of the benchmark model is produced. $f_t(\widehat{\beta_t})$ is the out-of-sample forecast of the real GDP growth rate at time t estimated by one of the benchmark models. Each forecast is generated by the forecast coefficient $\widehat{\beta_t}$ derived from the in-sample window m. The variable $g_{t\to t+h}$ is the actual value of real GDP growth. $FE_{t+h}(\widehat{\beta_t})$ is derived by subtracting the forecasted value from the observed value, and subsequently squaring it to gain purely positive values. Eq. (11) and (12) describe the manner in which the sequence of in-sample

[10] The MATLAB add-in "Captain Toolbox" is used to compute the evolution of β_t over time.

[11] Arguably, unrevised data can be used to show the actual information available. However, it goes beyond the scope of this thesis to incorporate a factor of estimation error.

forecast errors of the benchmark model is produced. The equations relate to recursive and rolling-window scheme respectively. The variable $f_{k-h}(\widehat{\beta_t})$ is the in-sample prediction of the real GDP growth rate $g_{k-h\rightarrow k}$, which varies across the scheme and forecast horizon: $k = t - m + h + 1, \dots, t$ for recursive regressions and $k = h + 1, \dots, t$. for rolling regressions. The actual values of $g_{k-h\rightarrow k}$ are compared to the in-sample predictions real GDP growth $f_{k-h}(\widehat{\beta_t})$. These differences are then squared so that distinctions between negative or positive values dissipate, and subsequently added together. This value is then divided by the sample size to ascertain the mean and to make the difference in sample size irrelevant. In essence, $\overline{FE}_t(\widehat{\beta_t})$ and $\overline{FE}_t^*(\widehat{\beta_t})$ compute the mean squared forecast error of each of the benchmark models used in the time-varying parameter models.

In the second stage, a series of SFEs are computed for each of the four chosen benchmark models. The computation of the out-of-sample forecast error and the average in-sample forecast error enables the production of the series of SFE:

$$SFE_{t+h}(\widehat{\beta_t}) = FE_{t+h}(\widehat{\beta_t}) - \overline{FE}_t(\widehat{\beta_t}) \qquad \text{for } t = m, m+1, \dots, T-h. \tag{13}$$

Eq. (13) defines the SFE as the difference between average in-sample forecast error and out-of-sample forecast error. At each moment in time, a mismatch between expected and actual forecast performance can be observed. If a large discrepancy between past and future performance occurs, there is an SFE. If the error is large enough, a forecast break has occurred.

In the third stage, two metrics are extracted from the distribution of SFEs. To test whether a perceived value of SFEs is significantly large or small, the mean and the asymptotic standard deviation should be computed in the following terms:

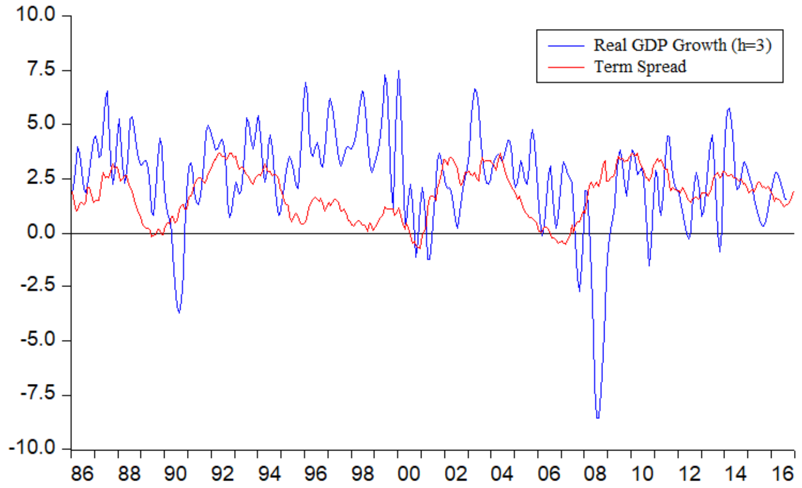
$$\overline{SFE}_{m,n} = n^{-1}\sum_{t=m}^{T-h} SFE_{t+h}\left(\hat{\beta}_t\right) \tag{14}$$

$$\hat{\sigma}_{m,n} = \sqrt{\frac{\sum(SFE_{t+h}(\widehat{\beta_t}) - \overline{SFE}_{m,n})^2}{n}} \tag{15}$$

Eq. (14) describes the average of SFEs for a specific in-sample and out-of-sample window and forecast horizon. The mean is computed by adding up the entire series of SFEs and subsequently dividing it by the amount of observations.

In the final stage, a confidence band is constructed by using the standard deviation and assuming that the SFE follows a t-distribution. In formal terms, a forecast break occurs when the SFE is significantly different from the average SFE. A confidence band of 95 percent can be construed by using the relevant t-values. Finally, the SFEs are plotted with the corresponding bandwidth. Whenever the SFE is not inside the bandwidth, a forecast break has occurred. If the SFE-line moves above (under) the bandwidth, it indicates a significant deterioration (improvement) of the predictive power of the term spread. If all

Fig. 2: real GDP growth at a forecast horizon of three months (h=3) and term spread are plotted. The y-axis is in percentages. The x-axis represents the timeline in calendar years.



forecast breaks are in line with the timing and theorized effect of monetary policy events, applying either the recursive or the rolling-window method of parameter estimation, the second hypothesis is confirmed.

VI. RESULTS

This chapter specifies which of the three types of benchmark models and across which forecast horizon the benchmark models should be used in the two-step procedure. Subsequently, this chapter tests the two main hypotheses and makes additional remarks.

Specification of the benchmark models

Before running the OLS regressions, an assessment of the main variables is in order. Fig. 2 plots the term spread, as calculated by Eq. (1), and the real GDP quarterly growth rate at an arbitrary horizon of 3 months ($h = 3$), as defined by Eq. (2). What immediately stands out are the large drops in real GDP growth rates followed by negative term spreads.[12] The large negative growth rates in 1990 and 2007-2008 are preceded by a negative term spread in 1988 and 2006 respectively. This observation is in line with ample evidence suggesting the term spread's ability to predict future recessions.[13] Furthermore, periods of convergence between the two variables are alternated by periods of divergence. For instance, the late 90's shows a large divergence compared to the early 90's and the early 00's. It is hard to say whether this has implications for the time-variation of the predictive power of the yield curve. On the one hand, one might argue that a consistent level of the forecasting relation would not produce this kind of apparent diverging and converging behavior. On the other hand, the coefficient of the forecasting relation is often found to be between zero and one. The gap between the term spread and real GDP

[12] Also referred to as an inverted yield curve.
[13] Wheelock and Wohar (2009) provide an extensive review on this topic.

Table 2: descriptive statistics for real GDP growth across different forecast horizons. The Augmented Dickey-Fuller (ADF) test statistic is based on Schwarz Information Criterion (SIC) lag selection. The corresponding t-statistics are -3.449 at a 1 percent level, denoted by a double asterisk, and -2.869 at 5 percent level, denoted by single asterisk.

Forecast horizon (months)	Mean	Std. Dev.	Min.	Max.	ADF
1	2.507	2.444	-9.099	8.767	-3.973**
3	2.511	2.745	-8.543	7.485	-3.035*
6	2.515	1.977	-7.227	5.943	-2.811
9	2.513	1.820	-5.606	5.218	-2.154
12	2.514	1.705	-4.156	5.133	-3.496**
15	2.514	1.615	-3.361	5.034	-3.300*
18	2.516	1.536	-2.918	5.011	-2.239
21	2.516	1.468	-2.406	4.846	-2.122
24	2.513	1.407	-1.827	4.839	-2.973*
27	2.510	1.354	-1.326	4.730	-2.301
30	2.506	1.311	-0.998	4.651	-2.588
33	2.500	1.274	-0.785	4.728	-2.530
36	2.492	1.236	-0.549	4.596	-2.303
Term Spread	1.799	1.118	-0.697	3.688	-3.042*

growth is larger at high values of the term spread than at low values. So even if the coefficient is constant, the lines would still be expected to diverge at periods of a steepening yield curve.

Table 2 shows the descriptive statistics of the term spread, as calculated by Eq. (1), and the real GDP quarterly growth rate across forecast horizons, as defined by Eq. (2). Several matters are worth noting. First, only a handful of variables appear to reject the null-hypothesis of having unit root. In the remainder of this paper these variables will be used in the benchmark models. The data containing a unit root are omitted, as they can hamper a correct estimation of models. Secondly, the standard deviation, minima, and maxima of $g_{t \to t+h}$ decrease as the forecast horizon becomes larger. This is presumably caused by the forecast horizon's inability to capture the intricate movements of the GDP observations. Indeed, differing growth rates are lumped together to produce a mean that does not represent the volatility of output data. Third, the difference in means and standard deviation across forecast horizons is due to differing sample sizes; the larger the forecast horizon, the more recent data points are lost.

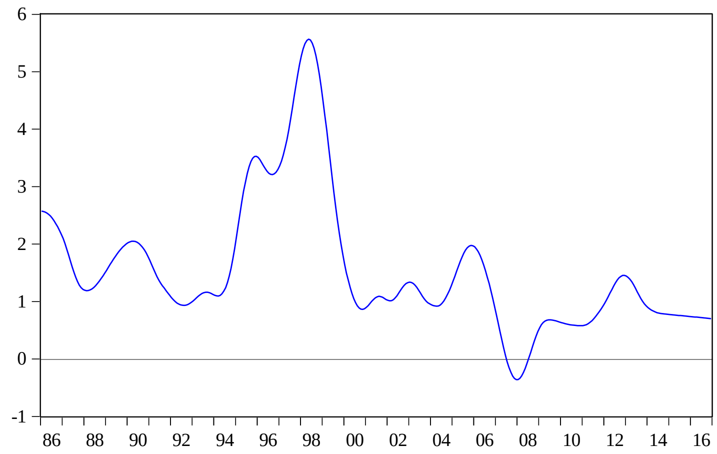
Table 3: parameter estimations for OLS regressions of Eq. (3), (4), and (5). Newey-West (1987) heteroskedasticity and autocorrelation consistent standard errors are between brackets. Single asterisk denotes statistical significance at 5 percent. Double asterisk denotes statistical significance at 1 percent.

		$h = 1$	$h = 3$	$h = 12$	$h = 15$	$h = 24$
Eq. (3)	α	2.435**	2.371**	2.068**	1.985**	1.827**
		(0.383)	(0.386)	(0.405)	(0.416)	(0.409)
	β	0.040	0.078	0.246	0.293	0.381*
		(0.180)	(0.174)	(0.153)	(0.155)	(0.160)
	Adj. R^2	-0.002	0.001	0.024	0.040	0.095
	Sample	1986M1-2016M11	1986M1-2016M9	1986M1-2015M12	1986M1-2015M9	1986M1-2014M12
Eq. (4)	α	1.525**	1.074**	0.018	-0.009	-0.009
		(0.444)	(0.401)	(0.248)	(0.211)	(0.126)
	β	0.055	0.090	0.126	0.116*	0.080*
		(0.135)	(0.119)	(0.065)	(0.058)	(0.039)
	γ	0.353**	0.506**	0.899**	0.915**	0.940**
		(0121	(0.103)	(0.070)	(0.054)	(0.040)
	Adj. R^2	0.120	0.258	0.827	0.863	0.924
	Sample	1986M2-2016M11	1986M4-2016M9	1987M1-2015M12	1987M4-2015M9	1988M1-2014M12
Eq. (5)	α	1.472**	1.039*	0.025	-0.008	0.006
		(0.457)	(0.411)	(0.250)	(0.102)	(0.127)
	β	0.085	0.110	0.125	0.119**	0.071
		(0.136)	(0.124)	(0.066)	(0.031)	(0.038)
	γ	0.353**	0.506**	0.897**	0.913**	0.941**
		(0.122)	(0.103)	(0.069)	(0.027)	(0.034)
	Adj. R^2	0.121	0.259	0.827	0.863	0.923
	Sample	1986M2-2016M11	1986M4-2016M9	1987M1-2015M12	1987M4-2015M9	1988M1-2014M12

Table 3 shows the parameter estimation output of the OLS regressions of future real GDP percentage growth on the yield spread as defined in Eq.'s (3), (4), and (5).[14] Several observations regarding these models are noteworthy. Firstly, the models containing lagged real GDP growth have a considerably higher adjusted R^2 than the model without. Generally, the adjusted R^2 of a regression shows how well a model fits the data, while penalizing for the number of parameters. Indeed, it shows to which extent variation in future real output growth is explained by variation in the current term spread. Moreover, the adjusted R^2 of Eq. (3) is quite low compared to what is generally observed in literature. For instance, Abdymomunov (2013) finds, at a forecast horizon of two years, an adjusted R^2 that is twice as a large as measured here. A reason for the adjusted R^2 to be lower than in other literature could be the mismatch between the actual observation between quarterly real GDP and monthly term spread data. Chosen models from Eq.'s (4) and (5) show a relatively high adjusted R^2, especially at horizons between one and two years. These models fit the data quite well. Secondly, the slope of the OLS regression lies between zero and one across all models and all forecast horizons. A positive value indicates that if the yield curve is steep, real GDP will increase over a period of h months from now. More precisely, in Eq. (2) at a horizon of two years, if the term spread is five hundred basis points, real GDP growth is expected to be 0.4 percent. Thirdly, there appears to be marginal predictive content in the current and lagged term spread in excess of lagged real GDP growth. The coefficient of real GDP growth, γ, is close to unity and highly significant. Despite this, there is still some minor predictive content in the term spread, especially in Eq. (4) at a two-year horizon. Fourthly, using either the lagged or current term spread makes almost

[14] The method of cubic spine interpolation poses a threat of artificially imposing serial correlation on the regressors, thereby violating one of the OLS assumptions. Newey-West (1987) standard errors are applied to circumvent this problem.

Figure 3: graph of time-variation in β_t of Eq. (21), represented by the blue line. The y-axis denotes the value of beta-t. The x-axis represents the timeline in calendar years.



no difference to the size of the forecast coefficient or adjusted R^2. Fifthly, Table 3 shows four regressions that contain a statistically significant forecast coefficient:

$$g_{t \to t+24} = \alpha + \beta * s_t + \varepsilon_{t+24,24}, \qquad \varepsilon_{t+24,24} \sim N(0, \delta_\varepsilon^2) \qquad (17)$$

$$g_{t \to t+15} = \alpha + \beta * s_t + \gamma g_{t-15 \to t} + \varepsilon_{t+15,15}, \qquad \varepsilon_{t+15,15} \sim N(0, \delta_\varepsilon^2) \qquad (18)$$

$$g_{t \to t+24} = \alpha + \beta * s_t + \gamma g_{t-24 \to t} + \varepsilon_{t+24,24}, \qquad \varepsilon_{t+24,24} \sim N(0, \delta_\varepsilon^2) \qquad (19)$$

$$g_{t \to t+15} = \alpha + \beta * s_{t-1} + \gamma g_{t-15 \to t} + \varepsilon_{t+15,15}, \qquad \varepsilon_{t+15,15} \sim N(0, \delta_\varepsilon^2) \qquad (20)$$

Eq. 17 represents Eq. (1) with a forecast horizon of 24 months, Eq. 18 is Eq. (2) with a forecast horizon of 15 months, Eq. 19 represents Eq. (2) with a forecast horizon of 24 months, and Eq. (20) is Eq. (3) with a forecast horizon of 15 months. These models are thus the benchmark models used in the two-step procedure.

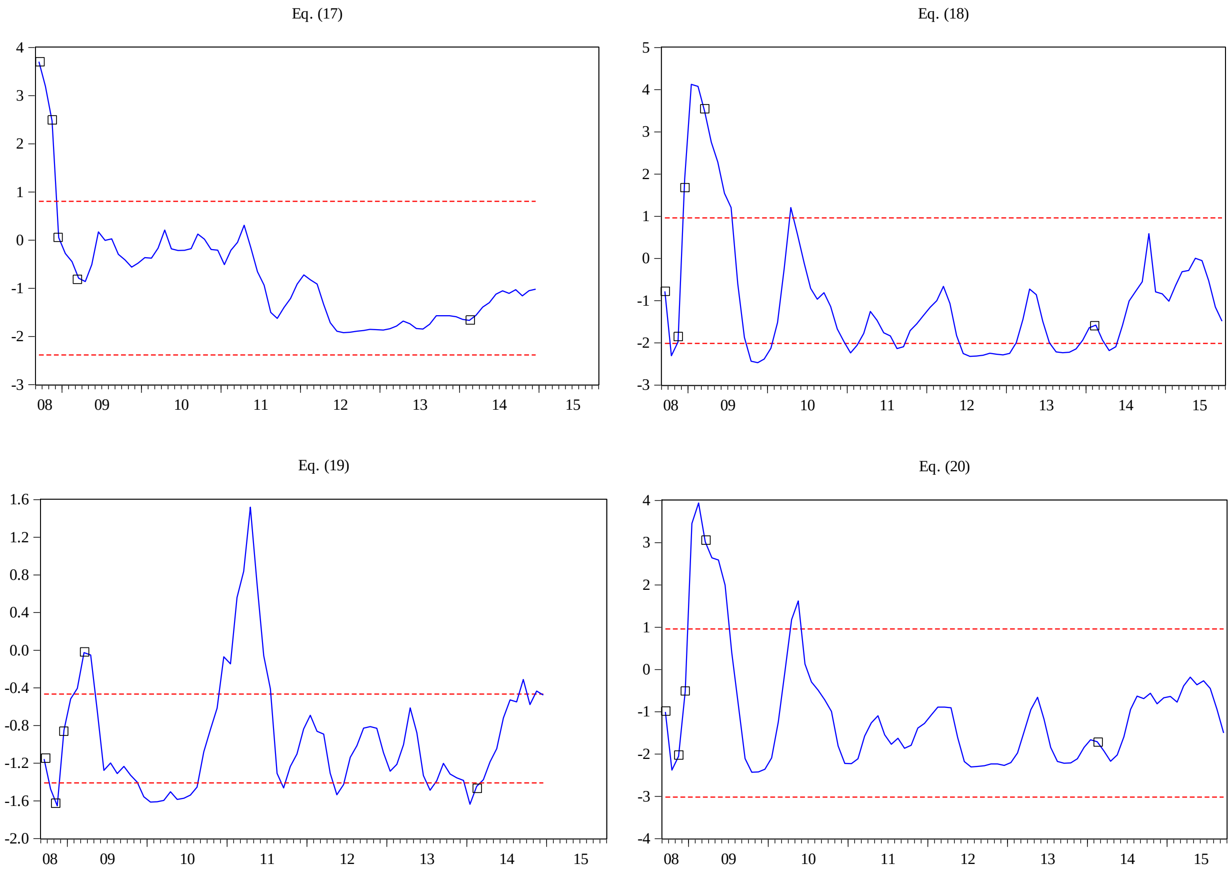
Hypothesis I

To test the first hypothesis, Eq. (17) is adjusted to contain a time-varying slope coefficient:

$$g_{t \to t+24} = \alpha + \beta_t * s_t + \varepsilon_{t+24,24}, \qquad \varepsilon_{t+24,24} \sim N(0, \delta_\varepsilon^2) \qquad (21)$$

Eq. (21) represents Eq. (17) with a time-varying slope coefficient. Only this equation is used to test whether the predictive power of the yield curve is unstable over time. De Pace (2013, p. 6) shows that adding lagged values of real GDP growth in the regression function does not alter the general intuition of the evidence posed. For brevity's sake, Eq. (18) - (20) are thus not investigated. Fig. (3) shows the graph of the evolution of β_t of Eq. (21) over time. Several things are worth noting. Firstly, the estimates of the forecast coefficient are largely positive, suggesting there is considerable predictive content in term spread during the investigated timespan. Secondly, large positive and negative values are visible. A

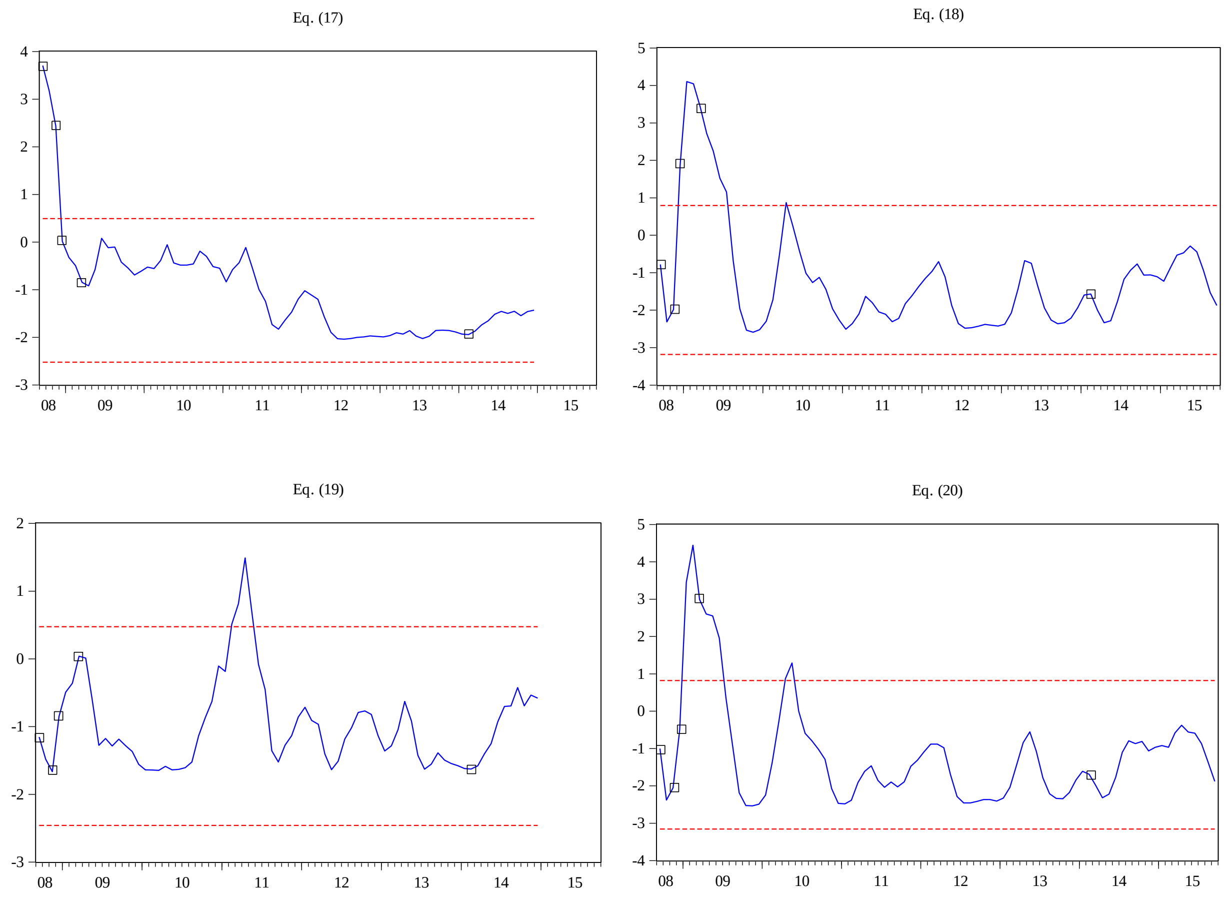
Figure 4: recursive estimates of Eq.'s (17) to (20). The blue lines represent SFE over time. The squares located on the blue lines indicate the occurrence of a major monetary policy event defined by this thesis. The red lines indicate the 95 percent confidence bandwidth by computing a two-sided t-test. For Eq.'s (17) and (19) the relevant t-value is $t(75) = 1.988$. For Eq.'s (18) and (20) the relevant t-value is $t(84) = 1.992$.



positive peak of 5.5 is visible May 1998. The observed values are higher than the positive but below unity values generally reported in other literature.[15] The largest negative values are found in the period July 2007 to June 2008, with a low point of -0.4 in December 2007. This finding is counterintuitive at first sight. In Fig. 2, an inverted yield curve is observed in the same period. Combining that observation with the estimates here would indicate that real GDP growth is expected while in reality the output level shrank as soon as the GFC occurred. These observations of very high and very low values become less counterintuitive, however, when compared to the results of De Pace (2013). He finds that the predictive content of the Treasury term spread has statistically dissolved between 1990Q2 and 2010Q1. If the predictive power of the yield actually broke down within this period, counterintuitive findings, as seen in Fig. 3, can be expected. All in all, the evidence presented here suggest that the first null-hypothesis should be confirmed as the slope-coefficient varies through time.

[15] Although De Pace (2013, p. 9) finds a peak at 1.5.

Figure 5: rolling-window estimates of Eq.'s (17) to (20). The blue lines represent SFE over time in percentage points. The squares located on the blue lines indicate the occurrence of a major monetary policy event defined by this thesis. The red lines indicate the 95 percent confidence bandwidth by computing a two-sided t-test . For Eq.'s (17) and (19) the relevant t-value is $t(75) = 1.988$. For Eq.'s (18) and (20) the relevant t-value is $t(84) = 1.992$.



Hypothesis II

To investigate whether the instability of the forecasting relation is linked to the five US major monetary policy events, the SFE of the benchmark models, defined in Eq. (17) – (20), are computed. The starting point is September 2008, which is the date on which the first major monetary policy event, the start of the GFC, occurred. Fig. 4 presents the SFE estimations of the parameters of each of the benchmark models. The panel related to Eq. (17) reports significant deteriorations of the forecast performance during the GFC and QE, though both findings contradict the theorized effect. The graph linked to Eq. (18) indicates that the entrapment of interest rates at the ZLB and forward guidance both coincide with a significant deterioration, where only the former change in predictive power conforms to this thesis' expectations. The graphical depiction of results produced by Eq. (19) demonstrates a significant improvement of the forecasting relation coinciding with QE and forward guidance, and a significant improvement at the appointment of Chairwoman Yellen. All these observations are in line with the theorized effect of these events. The graph associated with Eq. (20) shows that only the introduction of forward guidance coincides with a significant deterioration of the forecasting relation. Furthermore, Fig. 5 presents the SFE estimations, using rolling-window estimations of the parameters of each of the benchmark models. The results pertaining to Eq. (17) report a deterioration during the start of the GFC and QE, similar to the recursive model. The output produced by Eq. (18) reports a deterioration coinciding with the occurrence of interest rates stuck at the ZLB and forward guidance. The rolling-window estimation of Eq. (19) reports no major monetary policy events coinciding with forecast breaks. This is in contrast with the recursive estimation, which found three. Eq. (20) reports that the usage of forward guidance coincides with a forecast breakdown. The forecast breaks that the rolling-window estimation finds are all in contrast with the theorized effect of the major monetary policy events.

Fig.'s (4) and (5) also report untheorized break dates. In the recursive model, Eq. (18) reports a period of forecast breakdowns in the period December 2008 to July 2009, with a peak at January 2009. This period is followed by significant improvement of the forecasting relation between October 2009 until January 2010 and another significant deterioration in April 2010. Further significant improvements occur in August and September 2011, June 2012 to January 2013, August to November 2013, and April and May 2015. Eq. (19) shows a signification deterioration of the forecasting relation during December 2010 to July 2011. Also, significant improvements of the predictive content of the term spread are seen in September 2011, May and June 2012, July 2013, and January 2014. Eq. (20) reports a breakdown of the forecasting ability from January to June 2009 and in April and May 2010. In the rolling-window estimations, Eq. (18) demonstrates forecast breakdowns during December 2008 to July 2009 and in April 2010. Eq. (19) shows that forecast breakdowns occur during March to May 2011, peaking at April 2011. Eq. (20) reports significant deteriorations in the forecast performance in January to July 2009 and April to May 2010.

In conclusion, the evidence presented here suggests that the second null-hypothesis should be rejected. None of the major monetary policy events consistently coincide with a theorized forecast break date using either the recursive or the rolling-window method of parameter estimation. In only four of the total of twelve cases where the forecast break dates do, in fact, coincide with major monetary policy events, a change in forecast performance occurs conform the monetary-based view. Also, a considerable amount of untheorized forecast break dates is found, which further undermines the notion that monetary policy conduct is the confounding factor of the forecast relation. Lastly, it is of importance to place these results into the theoretical framework of the monetary-based view depicted in the top panel of Fig. 1. Keeping this framework in mind, the forecast break test results do not indicate that the bond market did not react to monetary policy events that this thesis characterizes as significant. Nor do these findings indicate that the Fed is unable to affect economic future real GDP growth. Instead, it is suggested that the monetary-based view, as it is depicted in Fig. 1, does at least not provide a complete picture of the underlying mechanism.

Recursive vs. rolling-window regressions

Some differences are visible when comparing the results distilled from the recursive and rolling-window estimation methods. At first glance, the shape of all four graphical depictions of the SFE produced by recursive estimation are similar when compared to its rolling-window counterparts. The main apparent difference is the change in the 95 percent confidence bandwidth. Rolling window estimations seem to produce a wider bandwidth. For this reason, the recursive estimates of Eq.'s (18) to (20) yield different results than its rolling-window counterparts. With Eq.'s (18) and (19) in particular, the bandwidth produced by the rolling-window regressions broadens compared to the recursive. The increased bandwidth signals that the SFEs produced by the rolling-window regressions have a larger variation. Thus, it seems that the loss in past data points, compared to the recursive method, caused the rolling-window method to have a lower forecast accuracy.

The differences in the statistical output above solicit a brief discussion on the pros and cons of recursive versus rolling-window parameter estimation in the context of this research. As the forecasting relation between the Treasury term spread and future U.S. real GDP growth is subject to structural change, forecasting agents may be inclined to use rolling-window parameter estimation. That is, using a sub-sample of the available data is especially helpful now that is has been established that the earliest observations are unrelated to more recent ones. So, rolling-window estimation excludes information distant to current observations but instead captures the more relevant short-run time-variation. In the context of the Treasury term spread's predictive power, rolling-window parameter estimations are likely to be less biased and produce lower forecast errors than its recursive counterparts. However, the decrease in sample size, in relation to recursive parameter estimation, is likely to increase the variance of the forecasting parameter estimates. In sum, forecasting agents will have to strike a balance between minimizing the bias or the variance in the parameter estimations in the forecast models.

To bypass these weaknesses, it is also possible to take the so-called model averaging approach. This method entails producing a weighted average of the slope-coefficient across different models. In economic literature, a dichotomy exists how weights are assigned to the forecast coefficient of each model. On the one hand, Bayesian model averaging (BMA) applies the Bayesian Rule and the law of total probability to produce a distribution of weighted averages of the forecasting coefficient, depending on the time-varying nature of the forecast coefficient in each separate model. It therefore considers the estimated variances of the recursive and rolling windows model as well as the weighted variance of the forecast coefficient across these two estimation models. A common critique is that the complexity and computational work of applying BMA is burdensome. On the other hand, Frequentist model averaging (FMA) produces an average of the forecast coefficient using smoothed weights. This so-called FMA estimator is asymptotically distributed. A drawback to this approach is the overly optimistic confidence intervals it produces.

VII. CONCLUSION

This paper examines the relation between the Treasury term spread and future U.S. real GDP growth. It confirms results found in past literature suggesting that the forecasting relation is unstable over time. This thesis contributes to existing work by examining the underlying theory of the predictive power of the U.S. Treasury yield curve. More precisely, this paper focuses on finding evidence of the Fed's monetary policy conduct as the confounding factor of both term spread behavior and future output changes. The results produced by the forecast break test suggest that the monetary-based view does not provide a complete explanation of the predictive power of the Treasury term spread. The recursive estimation of the forecast parameter shows a few instances where the major monetary policy events coincide with forecast break. However, almost half of these findings report a change in forecast relation that contradicts the monetary-based view. The rolling-window counterpart finds a slightly lower number of break dates occurring at the moment that a major monetary policy event occurs, though all of these show a different change in forecast performance than the theory. Taken at face value, the reported findings also do not provide evidence in favor of the consumption-based view. The selected monetary policy events cannot be considered to have fundamentally shifted the manner in which bond investors smooth their consumption. However, the validity of this competing theory can be tested by investigating whether the found forecast break dates coincide with significant changes in inflationary persistence. As is noted above, the consumption-based theory hinges on the presumption that the nominal yield curve is shaped similar to its real counterpart.

This thesis circumvents the drawbacks of methods used to discover the sources of instability of the forecasting relation, also taking into account the changing monetary policy environment where the FFR is no longer a suitable proxy for the Fed's monetary policy conduct. More precisely, this paper uses a two-sided version of Giacomini and Rossi's (2009) forecast breakdown test instead of the commonly used time-varying parameters and breakpoint analyses. The forecast break test applied here, for example, is not hampered by structural breaks occurring at higher moments. Moreover, FFR is not required as an input variable, in contrast to the commonly used methods. A recommendation for future research would be to incorporate one of the two model averaging approaches: BMA or FMA. This paper examines the recursive and rolling-window estimations separately. In order to more accurately gauge the surprise forecast error, it could be more informative to get a better grasp of the true forecast error itself. The surprise forecast error, the gap between the in-sample and out-of-sample forecast error is expected to be lower, and the corresponding bandwidth are measured with higher accuracy. Thus, an intriguing venture would be to apply a model averaging approach, which combines the recursive and rolling-window estimation of the forecast coefficient, in order to uncover the forecast break dates. Another area of future research lies in examining whether positive investment shocks coincide with negative forecast breaks. An investment boom entails higher future GDP growth and contra-inflationary monetary policy, thus a negative forecast coefficient arises. Such findings would support the monetary-based view.

VIII. REFERENCES

Abdymomunov, A., 2013. Predicting output using the entire yield curve. Journal of Macroeconomics 37, 333-344.

Andrews, D., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61(4), 821-856.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66(1), 47-78.

Benati, L., Goodhart, C., 2008. Investigating time-variation in the marginal predictive power of the yield spread. Journal of Economics Dynamics & Control 32, 1236-1272.

Borio, C., Zabai, A., 2016. Unconventional monetary policies: a re-appraisal. BIS Working Papers 570.

Chinn, M., Kucko, K., 2015. The predictive power of the yield curve across countries and time. International Finance 18(2), 129-156.

Clarida, R., Galí, J., and Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics 115, 147-1180.

Claus, E., Dungey, M., 2016. Can monetary policy surprises affect the term structure? Journal of Macroeconomics 47, 68-83.

Cukierman, A., 2013. Monetary policy and institutions before, during, and after the global financial crisis. Journal of Financial Stability 9(3), 373-384.

De Pace, P., 2013. Gross domestic product growth predictions through the yield spread: time-variation and structural breaks. International Journal of Finance and Economics 18, 1-24.

Estrella, A., 2005. Why does the yield curve predict output and inflation? The Economic Journal 115(505), 722-744.

Estrella, A., Hardouvelis, G. A., 1991. The term structure as a predictor of real economic activity. Journal of Finance 46(2), 555-576.

Estrella, A., Mishkin, F. S., 1996. Predicting U.S. recessions: financial variables as leading indicators. Federal Reserve Bank of New York Research Paper no. 9609, May.

Estrella, A., Rodrigues, A. P., Schich, S., 2003. How stable is the predictive power of the yield curve: evidence from Germany and the U.S. The Review of Economics and Statistics 85(3), 629-644.

Estrella, A., Trubin, M. R., 2006. The yield curve as a leading indicator: some practical issues. Current Issues in Economics and Finance 12(5), 1-7.

Feroli, M., 2004. Monetary policy and the information content of the yield spread. Topics in Macroeconomics 4(1), article 13.

Filardo, A., Hofmann, B., 2014. Forward guidance at the zero lower bound. BIS Quarterly Review March, 37-53.

Gagnon, J., Raskin, M., Remache, J., Sack, B., 2011. The financial market effects of the Federal Reserve's large-scale asset purchases. International Journal of Central Banking 7(1), 3-43.

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(s1), 783-795.

Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. Review of Economic Studies Limited 76, 669-705.

Hamilton, J. D., Wu, J. C., 2012. The effectiveness of alternative monetary policy tools in a zero lower bound environment. Journal of Money, Credit and Banking 44(1), 3-46.

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

Harvey, C. R., 1988. The real term structure and consumption growth. Journal of Financial Economics 22(2), 305-333.

Jarrow, R., Li, H., 2014. The impact of quantitative easing on the US term structure of interest rates. Review of Derivatives Research 17(3), 287-321.

Krishnamurthy, A., Vissing-Jorgensen, A., 2011. The effects of quantitative easing on interest rates: channels and implications for policy. NBER Working Paper (No. 17555).

Laurent, R. D., 1988. An interest rate-based indicator of monetary policy. Economic Perspectives, 3-14.

Mussa, M., 2005. Asset prices and monetary policy. In: Hunter, W., Kaufman, G., Pomerleano, M. (Ed.), Asset price bubbles: the implications for monetary, regulatory, and international policies. MIT press, Massachusetts, 41-50.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.

Plosser, C. I., Rouwenhorst, K. G., 1994. International term structures and real economic growth. Journal of Monetary Economics 33, 133-155.

Rendu de Lint, C., Stolin, D., 2003. The predictive power of the yield curve: a theoretical assessment. Journal of Monetary Economics 50(7), 1603-1622.

Sarno, L., Thornton, D. L., Valente, G., 2007. The empirical failure of the expectations hypothesis of the term structure of bond yield. Journal of Financial and Quantitative Analysis 42(1), 81-100.

Sarris, A. H., 1973. Kalman filter models: a Bayesian approach to estimation of time-varying regression coefficients. Annals of Economic and Social Measurement 2(4), 501-523.

Swanson, E. T., 2015. Measuring effects of unconventional monetary policy on asset prices. NBER Working Paper Series (December), 1-23.

Taylor, J. B., 1993. Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy 39, 195-214.

Taylor, J.B., 1999. A historical analysis of monetary policy rules. In: Taylor, J. (Ed.), Monetary policy rules. The University of Chicago Press, Chicago.

Wheelock, D.C., Wohar, M.E., 2009. Can the term spread predict output growth and recessions? A survey of literature. Federal Reserve Bank of St. Louis Review 91(5.1), 419-440.

Young, P.C., Taylor, C.J., Tych, W., Pedregal, D.J., 2007. The Captain Toolbox. Centre for Research on Environmental Systems and Statistics. Lancaster University, United Kingdom. Available at: http://www.es.lancs.ac.uk/cres/captain.