JARI HÄNNIKÄINEN

Essays on Real-Time Macroeconomic Forecasting

ACADEMIC DISSERTATION
To be presented, with the permission of
the Board of the School of Management of the University of Tampere,
for public discussion in the Lecture Room Linna K 103,
Kalevantie 5, Tampere,
on May 22nd, 2015, at 12 o'clock.

UNIVERSITY OF TAMPERE

JARI HÄNNIKÄINEN

Essays on Real-Time Macroeconomic Forecasting

Acta Universitatis Tamperensis 2054
Tampere University Press
Tampere 2015

TAMPEREEN
YLIOPISTO

ACADEMIC DISSERTATION
University of Tampere, School of Management
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Cover design by
Mikko Reinikka

Layout by
Sirpa Randell

Distributor:
verkkokauppa@juvenesprint.fi
https://verkkokauppa.juvenes.fi/

Acta Universitatis Tamperensis 2054
ISBN 978-951-44-9800-8 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1546
ISBN 978-951-44-9801-5 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

YMPÄRISTÖMERKKI MILJÖMÄRKT 441 729 Painotuote

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2015

Acknowledgements

The completion of this thesis represents an important turning point in my life. During different stages of this study, I received help from many people to whom I would now like to express my sincerest gratitude.

First, I would like to thank my supervisor, Professor Jari Vainiomäki, for patiently reading and commenting on my papers. In addition, his unconditional commitment and confidence in the success of this project was an invaluable source of motivation for me. I would also like to thank the two pre-examiners of this thesis, Docent Henri Nyberg from the University of Helsinki and Docent Juuso Vataja from the University of Vaasa, for their constructive and thorough comments on the original manuscript. Henri Nyberg has also given me detailed ideas, comments, and suggestions over the last few years, and these have been of great importance for improving the thesis as well as my understanding of macroeconomic forecasting.

Furthermore, my discussions with the seminar participants of the ALLECON seminar, the Annual Meeting of the Finnish Economic Association, the summer seminar for economic researchers in Jyväskylä, and the econometrics workshop organized by the Finnish Doctoral Programme in Economics (FDPE) have all been extremely helpful. In these seminars, I received valuable comments and suggestions from Professor Kari Heimonen, Professor Heikki Kauppi, Professor Markku Lanne, Assistant Professor Jani Luoto, Dr. Petri Kuosmanen, Dr. Helinä Laakkonen, Dr. Jukka Lahtonen, Sami Oinonen, and Juhani Raatikainen.

I owe many thanks to my colleagues at the University of Tampere as well as my colleagues at other institutions. I want to separately thank Professor Kaisa Kotakorpi, Professor Jukka Pirttilä, Dr. Sinikka Hämäläinen, Dr. Jani-Petri Laamanen, Matti Hovi, Jukka Ilomäki, Juho Jokinen, Petteri Juvonen, Markku Konttinen, Lauri Lepistö, Pietu Mänttäri, Mika Nieminen, Risto Rönkkö, Allan Seuri, and Elina Tuominen for their support. I would also like to gratefully acknowledge the financial support that I received from the FDPE.

Finally, I would like to express my deepest gratitude to those closest to me: my parents, Jouko and Tuulikki, and my brother, Juha. They have always encouraged me and have never lost their faith in this project.

Nokia, March 2015

Jari Hännikäinen

Copyright notice

Abstract

The future prospects of the economy are important for consumers, investors, and policymakers. As a result, economists provide forecasts of key macroeconomic time series for different time horizons. Economists at universities, central banks, and other forecasting institutions often find it difficult to produce accurate forecasts of uncertain future values of macroeconomic series.

There are several reasons why macroeconomic forecasting is such a challenging task. First, economic theory rarely specifies the functional form of the forecasting model or even which predictor variables should be included in the forecasting model. Thus, there is uncertainty concerning which forecasting model should be used. Second, many macroeconomic time series are subject to structural breaks. For instance, changes in tastes, technology, or institutional arrangements can cause changes in the dynamics of a series. It is well known that structural breaks matter for forecasting performance. Third, key macroeconomic data, such as real GDP and inflation series, are published with a lag and are subject to revisions. These data revisions can be quite large. Hence, it is difficult to produce accurate forecasts in real-time.

This thesis consists of an introductory chapter and four empirical essays on macroeconomic forecasting. The introductory chapter provides a review of how uncertainty about the forecasting model, structural instability, and data revisions affect macroeconomic forecasting. Model uncertainty, structural breaks, and the real-time nature of macroeconomic time series play a central role in each of the essays.

The first two essays analyze how one should generate autoregressive forecasts in the presence of structural instability and real-time data. In particular, the first essay considers the choice of the estimation window in the presence of data revisions and recent structural breaks. The Monte Carlo and empirical results for U.S. real GDP and inflation show that the expanding window estimator typically yields the most accurate forecasts after a recent structural break. The expanding window estimator performs well regardless of whether data revisions add news or reduce noise or whether we forecast first-release or final values.

The second essay compares the forecasting accuracy of alternative multi-step forecasting methods in an unstable environment. The Monte Carlo simulations indicate that the type and the timing of the break affect the relative accuracy of the multi-step forecasting methods. The iterated method typically performs the best in unstable environments, especially if the parameters are subject to small breaks. Empirical analysis of real-time U.S.

output and inflation series shows that the alternative multi-step methods only episodically improve upon the iterated method.

The other two essays investigate the real-time predictive power of interest rate spreads, which have been frequently used in the forecasting literature. The third essay studies the predictive ability of the term spread and a set of credit spreads when the short-term nominal rates have been stuck at the zero lower bound (ZLB) and the Federal Reserve has used unconventional monetary policy. The results of this essay suggest that the predictive content of the term spread has changed since the onset of the ZLB and unconventional monetary policy period. Thus, our results provide further evidence supporting the view that changes in monetary policy affect the ability of the term spread to forecast subsequent real activity. The results also indicate that the predictive power of credit spreads fluctuates over time. However, the ability of credit spreads to signal future output growth seems to be unaffected by the beginning of the ZLB and unconventional monetary policy era.

The fourth essay examines whether the mortgage spread (i.e., the difference between the 30-year mortgage rate and 10-year Treasury bond rate) is a useful leading indicator for U.S. real activity. The main finding from this study is that the mortgage spread contains predictive power for U.S. real GDP and industrial production growth. Importantly, the mortgage spread produces more accurate real-time forecasts than the widely used term spread and Gilchrist-Zakrajšek credit spread. However, the predictive power of the mortgage spread fluctuates over time. The mortgage spread has been a particularly informative leading indicator since the early 2000s.

Keywords: forecasting, structural breaks, real-time data, term spread, credit spread, zero lower bound.

Tiivistelmä

Talousennusteilla on keskeinen vaikutus kuluttajien, sijoittajien sekä raha- ja finanssipolitiikan harjoittajien päätöksiin. Talousennusteiden tärkeästä yhteiskunnallisesta asemasta johtuen ekonomistit laativat ennusteita keskeisistä makrotaloudellisista muuttujista, kuten bruttokansantuotteen (BKT) kasvusta ja kuluttajahintainflaatiosta. Ekonomistit kokevat usein tarkkojen makrotaloudellisten ennusteiden laatimisen vaikeaksi.

On olemassa useita syitä sille, miksi makrotaloudellinen ennustaminen on niin haastavaa. Ensinnäkin talousteoria on harvoin niin täsmällinen, että se määrittelisi ennustemallin funktiomuodon tai edes sen, mitä ennakoivia muuttujia tulisi sisällyttää ennustemalliin. Tästä syystä käytettävän ennustemallin valintaan liittyy huomattavaa epävarmuutta. Toiseksi useat makrotaloudelliset aikasarjat ovat kokeneet rakennemuutoksia. Nämä rakennemuutokset voivat johtua esimerkiksi kuluttajien kulutustottumusten, tuotantoteknologian tai institutionaalisten rakenteiden muutoksista. Aikaisemman empiirisen ennustekirjallisuuden perusteella rakennemuutokset vaikuttavat merkittävästi ennustemallien ennustetarkkuuteen. Lisäksi useiden makrotaloudellisten muuttujien arvot julkaistaan pitkällä viiveellä ja julkaistuja arvoja päivitetään yli ajan. Nämä päivitykset voivat olla hyvin suuria, mikä vaikeuttaa entisestään tarkkojen reaaliaikaisten ennusteiden laatimista.

Tämä väitöskirja koostuu johdantoluvusta ja neljästä makrotaloudellista ennustamista käsittelevästä esseestä. Johdantoluvussa keskustellaan siitä, miten olennaisesti ennustemalliin liittyvä epävarmuus, rakennemuutokset ja aineiston päivittäminen vaikuttavat makrotaloudelliseen ennustamiseen.

Väitöskirjan kaksi ensimmäistä esseetä analysoivat sitä, miten autoregressiiviset ennusteet tulisi laatia silloin, kun ennustettavassa aikasarjassa on tapahtunut rakennemuutos ja sarjan havaintoarvoja päivitetään yli ajan. Ensimmäisessä esseessä tarkastellaan ennustemallien parametrien estimointi-ikkunan valintaa tilanteessa, jossa aikasarjan havaintoja päivitetään yli ajan ja sarjassa on tapahtunut rakennemuutos juuri ennen ennusteen laatimishetkeä. Esseen Monte Carlo simulaatiotulokset ja empiiriset tulokset Yhdysvaltojen reaaliselle BKT:lle ja inflaatiolle osoittavat, että laajeneva estimointi-ikkuna tuottaa tällaisessa tilanteessa tyypillisesti kaikkein tarkimmat ennusteet.

Toinen essee vertailee erilaisten useiden askelien ennustemenetelmien ennustetarkkuutta silloin, kun ennustettavassa aikasarjassa on tapahtunut rakennemuutos. Monte Carlo simulaatioiden perusteella rakennemuutoksen tyyppi ja ajankohta vaikuttavat ennustemenetelmien suhteelliseen ennustetarkkuuteen. Iteratiivinen menetelmä tuottaa

tyypillisesti tarkimmat ennusteet, erityisesti silloin kun parametreihin kohdistuu vain pieniä muutoksia. Yhdysvaltojen reaaliaikaisten tuotanto- ja inflaatiosarjojen empiirinen analysointi osoittaa, että vaihtoehtoiset usean askeleen ennustemenetelmät tuottavat vain harvoin tarkempia ennusteita kuin iteratiivinen ennustemenetelmä.

Väitöskirjan kaksi viimeistä esseetä analysoivat korkoeron eli pitkän ja lyhyen valtionlainan koron erotuksen sekä luottoeron eli luottoriskin aiheuttaman korkoeron ennustekykyä. Kolmas essee tutkii korkoeron ja erilaisten luottoerojen ennustekykyä silloin, kun nimellinen lyhyt korko on nollatasolla ja Yhdysvaltojen keskuspankki on harjoittanut epätavallista rahapolitiikkaa. Esseen tulokset viittaavat siihen, että korkoeron ennustekyky on muuttunut sen jälkeen kun lyhyt korko asetettiin nollatasolle ja keskuspankki aloitti epätavallisen rahapolitiikan harjoittamisen. Tutkimustulokset tukevat siis aikaisemmassa kirjallisuudessa esitettyä näkökantaa, jonka mukaan muutokset keskuspankin tavassa harjoittaa rahapolitiikkaa muuttavat korkoeron ennustekykyä. Esseen tulokset osoittavat myös, että luottoerojen ennustekyky vaihtelee merkittävästi yli ajan. Nollakorkorajoitteen ja epätavallisen rahapolitiikan ei kuitenkaan havaita muuttavan luottoerojen ennustekykyä.

Neljäs essee tutkii sitä, onko niin sanottu asuntolainaspredi eli asuntolainan koron ja valtionlainan koron erotus hyödyllinen ennakoiva muuttuja. Esseen päätulos on se, että asuntolainaspredin avulla voidaan ennustaa Yhdysvaltojen reaalisen BKT:n ja teollisuustuotannon kasvua. Asuntolainaspredi osoittautuu tarkemmaksi ennakoivaksi muuttujaksi kuin paljon huomiota aikaisemmassa ennustekirjallisuudessa saaneet korkoero ja Gilchrist–Zakrajšek luottoero. Tulosten perusteella asuntolainaspredin ennustekyky vaihtelee yli ajan. Asuntolainaspredi on ollut erityisen hyödyllinen ennakoiva muuttuja vuodesta 2000 lähtien.

Avainsanat: ennustaminen, rakennemuutos, reaaliaikainen aineisto, korkoero, luottoero, nollakorkorajoite.

Contents

Chapter 1: Introduction 13

1.1 Model uncertainty 15

1.2 Real-time data 17

1.3 Structural breaks 21

1.4 Summaries of the essays 24

1.4.1. Chapter 2: Selection of an estimation window in the presence of data revisions and recent structural breaks 24

1.4.2 Chapter 3: Multi-step forecasting in the presence of breaks 25

1.4.3. Chapter 4: Zero lower bound, unconventional monetary policy and indicator properties of interest rate spreads 26

1.4.4 Chapter 5: The mortgage spread as a predictor of real-time economic activity 26

References........ 27

Chapter 2: Selection of an estimation window in the presence of data revisions and recent structural breaks 31

2.1 Introduction 32

2.2 Statistical framework 34

2.3 Forecasting methods 36

2.4 Monte Carlo simulations 39

2.5 Empirical application 48

2.6 Conclusions 51

References........ 60

Chapter 3: Multi-step forecasting in the presence of breaks 63

3.1 Introduction 64

3.2 Statistical framework 66

3.3 Methods for multi-step forecasting 68

3.4 Monte Carlo simulations 69

3.5 Empirical results 80

3.6 Conclusions 90

References........ 90

Chapter 4: Zero lower bound, unconventional monetary policy and indicator properties of interest rate spreads 95
4.1 Introduction 96
4.2 Methodology 99
4.3 Empirical results 102
4.4 Conclusions 112
References 113
Chapter 5: The mortgage spread as a predictor of real-time economic activity 115
5.1 Introduction 116
5.2 Methods 116
5.3 Forecasting results 117
5.4 Conclusion 123
References 123

Chapter 1
Introduction

The future prospects of the economy are important for many economic decision makers. For example, because monetary policy affects the economy with a long lag, central banks conduct forward-looking monetary policy. As a consequence, central banks' interest rate decisions are based on their forecasts of future output growth, unemployment, and inflation. Fixed-income investors are interested in real interest rate, which depends on the future inflation rate. Hence, inflation forecasts play a central role when fixed-income investors make their investment decisions. Similarly, households can benefit from wage and unemployment forecasts when deciding how much labor to supply and how much to consume.

Given the importance of the future economic outlook, economists provide forecasts of macroeconomic time series for different time horizons. Economists typically focus on forecasting key variables, such as real GDP growth, industrial production growth, price inflation, unemployment rate, wages, interest rates, stock prices, exchange rates, and commodity prices. These forecasts receive a lot of attention in the media. Economists at universities, central banks, and other forecasting institutions agree that providing accurate forecasts of uncertain future values is a difficult task. The Financial Crisis of 2007–2009 is a good example of a period when forecasting was particularly difficult. During this period, the forecasts often deviated from the true values, sometimes by a substantial margin.

There exist at least three reasons why macroeconomic forecasting is such a challenging task. First, economic theory does not usually specify the functional form of the forecasting model or even which predictor variables should be included in the forecasting model. Therefore, there is uncertainty concerning which forecasting model should be used. A variety of models have been proposed in the literature. For instance, autoregressive (AR) models, vector autoregressive models, and dynamic factor models have been used in numerous forecasting applications. Second, many macroeconomic time series are subject to structural breaks. For example, changes in tastes, technology, legislation, institutional arrangements, or government policy can cause changes in the dynamics of a series. It is well known that structural breaks matter for forecasting performance (see, e.g., Clements and Hendry, 1998, 2006; Elliott and Timmermann, 2008; Rossi, 2013). Forecast errors are typically large after structural breaks. Moreover, a forecasting model that performed well before the break might yield very inaccurate forecasts after the break. Third, key macroeconomic data, such as real GDP and inflation series, are published with a lag and are subject to revisions. The data revisions can be quite large. Hence, it is difficult to produce accurate forecasts in real-time.

Macroeconomic time series are usually serially correlated. Serial correlation between observations imply that the past values of a series are themselves useful predictors for future values. Hence, AR models are frequently used in forecasting applications. In this thesis, we focus on AR models. There are several reasons for our choice. First, despite their parsimonious form, AR models are found to perform well empirically. It appears to be relatively difficult to outperform AR models with alternative forecasting models in practice (see, e.g., Clements, 2014; Elliott and Timmermann, 2008; Stock and Watson, 1999a). Second, because of their good forecasting performance, it is standard practice to use AR model as a benchmark in forecast competitions. Third, the predictive power of a candidate predictor, say the term spread, is typically analyzed by comparing the forecasting accuracy of the AR model augmented with the candidate predictor to that of the pure AR model. If inclusion of the candidate predictor improves forecast accuracy, the candidate predictor contains marginal predictive power over and above that of the own history of the series. Ng and Wright (2013), Rossi (2013), and Stock and Watson (2003, 2007), among others, have considered the marginal predictive power of financial and macro variables for output growth and inflation.

This thesis consists of four empirical essays on macroeconomic forecasting. Two of the essays analyze how one should generate autoregressive forecasts in the presence of structural instability and real-time data. In particular, the first essay examines the forecasting performance of a set of widely used window selection methods in the presence of recent structural breaks. The second essay compares the accuracy of alternative multi-step forecasting methods in an unstable environment. The results of these essays are important not only for those who try to generate accurate real-time forecasts using AR models, but also for those who use an AR model as a benchmark in forecast competitions. The other two essays investigate the real-time marginal predictive power of interest rate spreads. The third essay studies the predictive ability of the term spread and a set of credit spreads in the 2008–2014 period when the short-term nominal rates have been stuck at the zero lower bound and the Federal Reserve has used unconventional monetary policy. This essay adds to the understanding of whether changes in the way the Federal Reserve conducts monetary policy are important for the predictive power of term and credit spreads. The fourth essay examines whether the mortgage spread is a useful real-time leading indicator for real economic activity.

Model uncertainty, structural instability, and the real-time nature of macroeconomic time series play a central role in each of the essays. In this thesis, we focus exclusively on out-of-sample forecasting. Because out-of-sample forecasting closely mimics the actual forecasting process, it provides a natural framework for analyzing the performance of alternative forecasting methods and the information content of a candidate predictor variable for subsequent economic activity. In principle, the forecasting methods discussed in this thesis can be applied to any real-time macroeconomic time series. However, in what follows, we concentrate on forecasting U.S. output growth and inflation.

The rest of this introductory chapter is organized as follows. In Section 1.1, we explain how model uncertainty complicates economic forecasting. Section 1.2 discusses the real-time nature of macroeconomic time series used in many applications. This section describes the real-time dataset used throughout this thesis, the key properties of data revisions, and how real-time data affects forecasting. Section 1.3 demonstrates that structural breaks are important for forecast accuracy. This section also emphasizes that the relative forecasting performance of two methods can fluctuate over time in an unstable environment. Finally, Section 1.4 gives a short summary of the essays, linking the main contributions of the essays to the three themes discussed in Sections 1.1–1.3.

1.1 Model uncertainty

When constructing a forecast, a forecaster has to decide which variables to use as predictors. This can be a difficult task because there are hundreds of possible predictors, representing different facets of the macroeconomy (e.g., production, employment, inflation, interest rates). Although economic theory gives guidance for variable selection, theory rarely specifies which particular variable should be included in the forecasting model. For instance, the Phillips curve indicates that real activity measures should help forecast future inflation. However, the theory does not clearly state whether unemployment, output gap, or output growth should be used as a measure of real activity.[1] Economic theory also suggests that credit spreads, which measure financial frictions, are potentially useful leading indicators for business cycle fluctuations. Given that there are a lot of alternative credit spreads, it is hard to say *a priori* which of them should be used for forecasting purposes. The inability of economic theory to pinpoint which credit spread is the most informative has generated a vast amount of literature analyzing the forecasting performance of alternative credit spreads (see, e.g., Bernanke,

1 For further discussion and empirical evaluation of Phillips curve forecasts, see, *inter alia*, Atkeson and Ohanian (2001), Faust and Wright (2013), and Stock and Watson (1999b, 2009).

1990; Faust *et al.*, 2013; Friedman and Kuttner, 1998; Gertler and Lown, 1999; Gilchrist *et al.*, 2009; Gilchrist and Zakrajšek, 2012; Mody and Taylor, 2003).

In most situations, economic theory is uninformative about the appropriate functional form of the forecasting model relating the predictors and the future value of the variable to be forecast. Hence, it is uncertain which functional form should be used. This form of model uncertainty has received a lot of attention in the literature. A variety of linear and non-linear models have been proposed. Among the alternative models, AR models (see, e.g., Marcellino *et al.*, 2006; Pesaran and Timmermann, 2005), vector autoregressive models (see, e.g., Stock and Watson, 2001), and dynamic factor models (see, e.g., Luciani, 2014; Stock and Watson, 2002a, 2002b, 2011) are probably the most commonly used in forecasting studies. The performance of the alternative models seems to depend on the forecasting problem at hand. However, parsimonious models, such as low order AR models, typically perform well in macroeconomic applications. It is particularly difficult to outperform a simple AR model when inflation is forecasted (Stock and Watson, 2007).

Policymakers and other economic agents are often interested in the medium- and long-term prospects of the economy. Hence, economists provide forecasts of key macroeconomic time series several periods ahead in time. When generating these forecasts, a forecaster encounters a multi-step forecasting problem. A forecaster has to decide whether to use the iterated or direct multi-step forecasting strategy. The iterated forecasts are made using a one-period ahead model, iterated forward for the desired number of periods. By contrast, direct forecasts are made using a horizon-specific model, and thus a forecaster has to estimate a different model for each forecast horizon. Papers that consider the relative merits of the iterated versus the direct forecast methods from a theoretical perspective include, for instance, Bao (2007), Brown and Mariano (1989), Clements and Hendry (1996b, 1998), Hoque *et al.* (1988), Ing (2003), Schorfheide (2005), and Weiss (1991). This theoretical literature emphasizes that the choice between iterated and direct multi-step forecasts involves a trade-off between bias and estimation variance. Because the iterated method uses a larger data sample in the estimation than the direct method, it produces more efficient parameter estimates. On the other hand, direct forecasts are more robust to possible model misspecification because they relate the multi-step ahead value directly to the current and past values of the predictors. The relative importance of the bias and the estimation variance in the composition of the mean squared forecast error (MSFE) values, which are used to evaluate the accuracy of the forecasts, depends on the sample size, the forecast horizon, and the (unknown) underlying data generating process (DGP). Therefore, the question of which multi-step method to use cannot be decided *ex ante* on theoretical grounds alone. Rather, which multi-step approach is the most accurate is an empirical matter.

The choice of the estimation window can substantially affect the accuracy of the forecasts. As a consequence, questions of how to weight old versus recent data and how much data to use when estimating the parameters of the forecasting model have become an essential part of the forecasting literature. In most applications, either an expanding window estimator or a rolling window estimator is used. The expanding window estimator uses the whole data sample available at the forecast origin, whereas the rolling window estimator uses only the most recent observations. When the rolling window estimator is used, the forecaster has to decide the length of the rolling window. Alternative strategies for estimating the parameters of the forecasting model include, for example, the exponentially weighted moving average (EWMA) method (see, e.g., Pesaran and Pick, 2011) and the average window method (AveW) proposed by Pesaran and Timmermann (2007). The choice of the estimation window is particularly important for forecast accuracy in the presence of structural breaks. For this reason, we discuss the window selection problem in greater detail in Section 1.3.

1.2 Real-time data

Forecasters typically use the latest available data in out-of-sample forecasting exercises. This approach is problematic when the purpose is to forecast macroeconomic variables or macroeconomic variables are used as predictors in the forecasting model. It is well known that key macroeconomic time series, such as real GDP and inflation series, are subject to important revisions. Because data are revised over time, macroeconomic forecasts based on latest available data may differ substantially from those based on real-time data. Croushore (2006, 2011) emphasizes that practical forecasting is inherently a real-time exercise. Therefore, it is important to use data values actually available at each forecast origin when simulating the out-of-sample forecasting process.

Over the past 15 years, the number of forecasting studies employing real-time data has expanded rapidly. The primary reason for this is that real-time datasets are nowadays publicly available for the U.S. and other countries.[2] For example, the Real-Time Data Set for Macroeconomists (RTDSM), compiled and maintained by the Federal Reserve Bank of Philadelphia, has been publicly available since 1999.[3] This dataset contains real-time data for several U.S. macroeconomic time series. Because the RTDSM is used in each of the four essays, we next explain how the data are organized in this dataset.

2 See the list of all publicly available real-time datasets at https://facultystaff.richmond.edu/˜dcrousho/data.htm

3 www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/

Moreover, we illustrate the key features of the real-time data. A more detailed discussion is presented in Croushore and Stark (2001).

Table 1 shows the typical structure of a real-time dataset. This table reports quarterly growth rates of real GDP (at an annualized rate). Each column of Table 1 represents a different vintage of data, i.e., the time series that a forecaster observed at the date shown in column header. Table 1 demonstrates the two key features of real-time data. First, data are published with a lag. For instance, the first release of real GDP for a given quarter is published at the end of the month following the end of that quarter. Thus, a forecaster at, say, quarter 2008:Q4 has access to the 2008:Q4 vintage values of real GDP growth up to quarter 2008:Q3. Second, and more importantly, data are revised over time. As explained in Croushore (2011), real GDP data are revised one and two months after the initial release. Real GDP series are further revised in July of each of the following three years (so called annual revisions) and approximately every five years after that (benchmark revisions). These data revisions can be very large in practice. As an example, Figure 1 plots real GDP growth in 2008:Q4 as recorded in the 2009:Q1–2014:Q3 data vintages. The figure reveals that the growth rate changes substantially over time, from -3.878% in the first available vintage (2009:Q1) to -8.541% in the 2014:Q3 vintage.

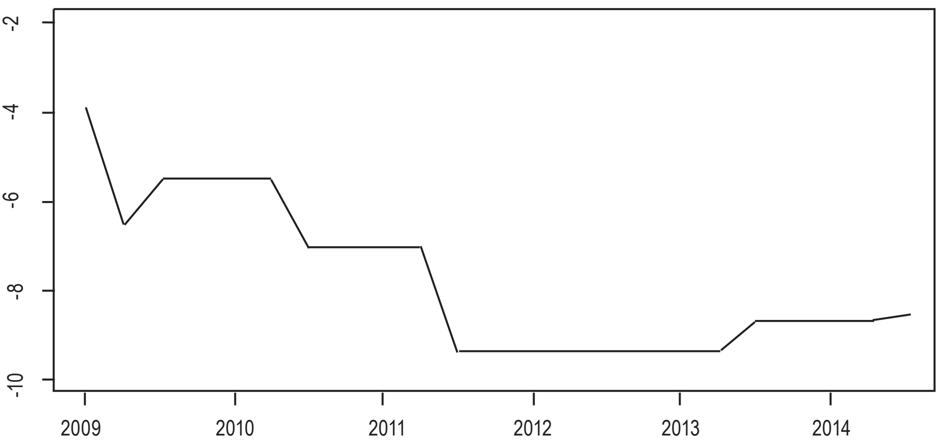
Table 1. Structure of real-time data

	Vintage					
Date	2008:Q4	2009:Q1	2009:Q2	…	2014:Q2	2014:Q3
2000:Q1	1.012	1.012	1.012	…	1.147	1.160
2000:Q2	6.234	6.234	6.234	…	7.483	7.484
2000:Q3	-0.459	-0.459	-0.459	…	0.511	0.483
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2008:Q3	-0.252	-0.512	-0.512	…	-1.985	-1.924
2008:Q4	NA	-3.878	-6.552	…	-8.701	-8.541
2009:Q1	NA	NA	-6.341	…	-5.597	-5.582
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2014:Q1	NA	NA	NA	…	0.108	-2.129
2014:Q2	NA	NA	NA	…	NA	3.872
2014:Q3	NA	NA	NA	…	NA	NA

Notes: The table shows quarterly growth rates of real GDP (at an annualized rate) for different data vintages. Data source: Real-Time Data Set for Macroeconomists (RTDSM).

The properties of data revisions have been extensively analyzed in the previous literature (see, e.g., Aruoba, 2008; Faust *et al.*, 2005; Mankiw *et al.*, 1984; Mankiw and Shapiro, 1986). In particular, the question of whether data revisions can be characterized as

Figure 1. Real GDP growth for 2008:Q4



Notes: The figure plots real GDP growth for 2008:Q4 using different data vintages from 2009:Q1 to 2014:Q3. Data source: Real-Time Data Set for Macroeconomists (RTDSM).

adding news or reducing noise has received a lot of attention. Because news and noise revisions are important concepts in the literature, we next demonstrate the difference between these two alternatives.

Under the news characterization, a government data agency optimally uses all available information in constructing the preliminary estimate, and hence data revisions reflect new information, or "news," that arrives after the announcement (Faust *et al.*, 2005). Let

$$\tilde{y}_t = y_t^{t+s} + v_t^{t+s} ,$$

where $\tilde{y}_t$ denotes the true value of y in period t, y_t^{t+s} (s ≥ 1) denotes the period $t + s$ vintage estimate of the value of y in period t, and v_t^{t+s} is the error term for that data release. Data revisions are said to add news if they have the properties of rational forecast errors. This requires that revisions $r_t^{t+s+1,t+s} = y_t^{t+s+1} - y_t^{t+s} = v_t^{t+s} - v_t^{t+s+1}$ are unpredictable given the information available at time $t + s$. The information set at time $t + s$ contains all previously published data vintages, so error terms v_t^{t+s} must be uncorrelated with the previously published vintages, i.e., $cov(y_t^{t+k}, v_t^{t+s}) = 0 \ \forall k \leq s$. Note, however, that news revisions are correlated with the true value because $cov(\tilde{y}_t, v_t^{t+s}) \neq 0$. It is straightforward to show that $var(y_t^{t+s}) < var(\tilde{y}_t)$, i.e., the variance of the preliminary estimate is smaller than that of the true value. More generally, the structure of news revisions implies that the variance increases as data are revised over time.

Under the noise characterization, the initial estimate is an observation on the final value measured with error (Mankiw and Shapiro, 1986). Measurement errors could arise, for instance, if the preliminary estimates are based on unrepresentative data or on data samples that are too small. Subsequently released estimates reduce or eliminate this measurement error, or "noise," by utilizing a more representative or larger data sample. If data revisions reduce noise, each vintage release y_t^{t+s} can be expressed as a sum of the true value $\tilde{y}_t$ and an error term ε_t^{t+s}

$$y_t^{t+s} = \tilde{y}_t + \varepsilon_t^{t+s},$$

where the error term ε_t^{t+s} is uncorrelated with the true value (i.e., $cov(\tilde{y}_t, \varepsilon_t^{t+s}) = 0$), but correlated with y_t^{t+s} (i.e., $cov(y_t^{t+s}, \varepsilon_t^{t+s}) \neq 0$). Noise revisions $r_t^{t+s+1,t+s} = y_t^{t+s+1} - y_t^{t+s} = \varepsilon_t^{t+s+1} - \varepsilon_t^{t+s}$ are correlated with data known at time $t + s$. In particular, they are correlated with the vintage $t + s$ estimate y_t^{t+s}. Hence, noise revisions are predictable. Noise revisions imply that $var(y_t^{t+s}) > var(\tilde{y}_t)$. Otherwise stated, the variance of the preliminary estimate is larger than that of the true value. More generally, the variance of the series decreases as more updated estimates become available. For a more detailed discussion of the properties of news and noise revisions, see Croushore (2011).

The results in the previous studies indicate that revisions to different macroeconomic series have different characteristics. Most importantly, at least since the mid-1980s, data revisions to U.S. output growth appear to be mainly news, whereas those to inflation mainly reduce noise (Clements and Galvão, 2013). It is important to note that although data revisions have been subject to much research, the literature on how the properties of the revision process (i.e., whether revisions add news or reduce noise) affect forecasting is scant.

There exist several reasons why using real-time data in forecasting experiments may lead to very different forecasts than using the latest available data. First, because data revisions can be very large, the parameters of the forecasting model estimated on the latest available data may differ substantially from those estimated on real-time data. Second, data revisions are also potentially important for the lag structure of the forecasting model (Stark and Croushore, 2002). Finally, real-time forecasts are conditioned on the first-release or lightly revised data actually available at each forecast origin, whereas forecasts based on the latest available data are conditioned on the latest available observations of each forecast origin.

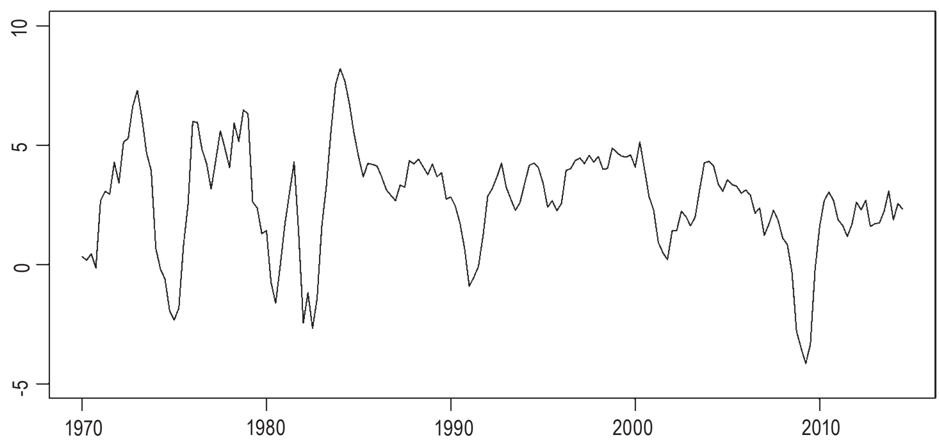
Professional forecasters construct their macroeconomic forecasts in a real-time environment using data values actually available at each forecast origin. Given the real-time nature of practical forecasting, it is important to use real-time data in forecasting experiments. Indeed, when real-time data are used, out-of-sample forecasting exercises very closely simulate actual forecasting process. Therefore, the results of such forecasting experiments should give a realistic picture of, say, the predictive power of a candidate leading indicator or the relative forecasting performance of different window selection methods.

1.3 Structural breaks

The empirical literature has found widespread evidence of instability in U.S. macroeconomic time series (Stock and Watson, 1996). The results indicate that different series have undergone different types of structural breaks. For example, the volatility of output growth has declined since the mid-1980s (McConnell and Perez-Quiros, 2000). Figure 2 demonstrates this phenomenon, called the Great Moderation, by plotting annual real GDP growth over the 1970:Q1–2014:Q3 period. On the other hand, due to changes in monetary policy in the early 1980s (Sims and Zha, 2006), both the mean and variance of inflation have decreased substantially. This fundamental change in the dynamics of inflation in the early 1980s can be seen in Figure 3, which depicts the annual CPI inflation rate from 1970:Q1 to 2014:Q3. There are several possible reasons for structural instability. For instance, changes in tastes, technology, legislation, institutional arrangements, or government policy can cause changes in the way the economy evolves.

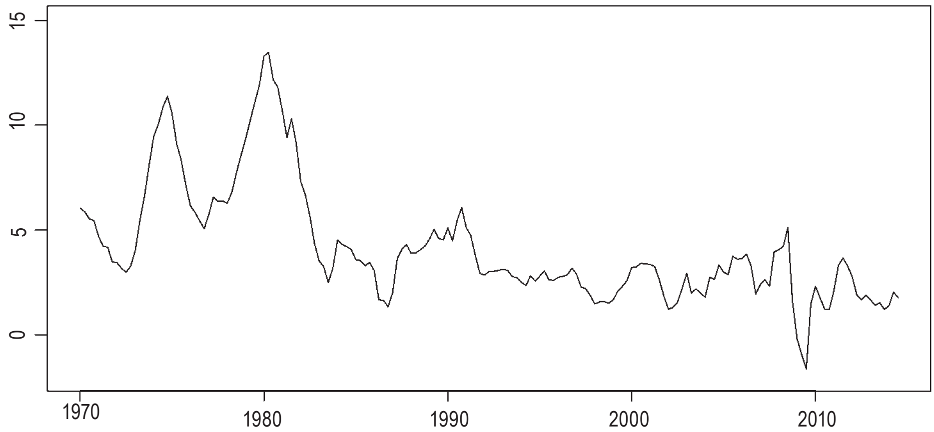
Structural breaks play a central role in economic forecasting (see, e.g., Clements and Hendry, 2006; Elliott and Timmermann, 2008; Rossi, 2013). Forecasters often find it difficult to generate accurate forecasts in the presence of structural instability. Indeed, forecast errors are typically very large after structural breaks. Breaks are also important

Figure 2. Real GDP growth



Notes: The figure depicts annual GDP growth rate from 1970:Q1 to 2014:Q3. Data source: Federal Reserve Economic Data (FRED).

Figure 3. Inflation rate



Notes: The figure shows annual CPI inflation rate from 1970:Q1 to 2014:Q3. Data source: Federal Reserve Economic Data (FRED).

because they might change the performance of a particular forecasting model or the predictive power of a candidate leading indicator. Furthermore, it is possible that a forecasting model or a candidate predictor that performed well before the break performs poorly after the break.

The choice of the estimation window can have a major impact on forecast accuracy in an unstable environment (see, e.g., Pesaran and Timmermann, 2005, 2007). As a result, a key question in the presence of structural instability is how much data to use to estimate the parameters of the forecasting model. One solution to the window selection problem is to test for breaks and use only observations after the most recent break in the estimation. This so called post-break window strategy is problematic for at least two reasons. First, it is difficult to pin down the exact break date, especially if the break has occurred close to the forecast origin and when the data are subject to revisions. There is therefore considerable uncertainty surrounding the estimate of the timing of the break and hence the length of the estimation window. Second, the parameters of the forecasting model are estimated with adequate accuracy only if the data sample is sufficiently long. Thus, the post-break window approach can be used in practice only when the (last) break has occurred sufficiently long ago.

An alternative solution to the window selection problem is to use robust estimation strategies. An estimation strategy is said to be robust if no information about the structural break is needed for its implementation. An expanding window estimator, a rolling window estimator, an exponentially weighted moving average (EWMA) method (see Pesaran and Pick, 2011), and an average window (AveW) method (Pesaran and

Timmermann, 2007) are examples of robust estimation strategies. Robust estimation strategies are more popular than the post-break window strategy in empirical studies. In particular, either the expanding window estimator or the rolling window estimator is used in a clear majority of forecasting exercises. The results in Pesaran and Timmermann (2005) indicate that the post-break window strategy may lead to inaccurate forecasts. Moreover, their Monte Carlo simulations reveal that the post-break window method usually performs poorly relative to robust window selection methods. In this thesis, we restrict ourselves to robust estimation strategies.

Structural breaks are also important for forecast evaluation. Researchers typically evaluate the accuracy of alternative forecasting models by computing relative MSFE values over the whole out-of-sample period. This approach implicitly assumes that the relative performance of the models remains constant over time. Giacomini and Rossi (2010) point out that the relative forecasting performance may change over time in the presence of structural instability. In such a case, average relative performance over the whole out-of-sample period may hide important information or even lead to incorrect conclusions. As an example, consider a situation where a forecaster compares the performance of an AR model to that of a model including an autoregressive lag and a candidate predictor. Assume that in the first half of the sample the parsimonious AR model produces more accurate forecasts, whereas in the latter half of the sample the model with the candidate predictor dominates the AR model. If the performance of the models is evaluated by the relative MSFE over the whole out-of-sample period, the forecaster might well conclude that the two models produce equally accurate forecasts. Thus, the forecaster might miss the fact that the AR model is more accurate than the model with the candidate predictor in the early part of the sample, whereas the opposite is true in the latter part.

In the applied literature, researchers traditionally analyze time variations in the relative forecasting performance by dividing the sample period into subsamples (e.g., pre- and post-1985 periods) and computing the relative MSFE values for each subperiod (see, e.g., Bordo and Haubrich, 2008a, 2008b; Stock and Watson, 2003, 2007). Although popular in practice, this approach is problematic because the subperiods are, more or less, chosen in an arbitrary fashion. It is usually difficult to say when exactly the relative forecasting performance might have changed. Different subsample choices may lead to different empirical results. A more formal way to examine time variations in the relative forecasting performance is to use the fluctuation test developed by Giacomini and Rossi (2010). This fluctuation test is designed such that it can detect changes in the relative performance at any point in the out-of-sample period. The 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) test of equal (unconditional) predictive ability computed over a rolling out-of-sample window. To be more specific, the researcher computes the Giacomini and White (2006) test statistic for each rolling window. If the maximum test statistic exceeds the critical value calculated by Giacomini and Rossi (2010), the null of equal accuracy between the two forecasting methods at each point in time is rejected.

1.4 Summaries of the essays

1.4.1. Chapter 2: Selection of an estimation window in the presence of data revisions and recent structural breaks

The first essay considers the choice of the estimation window. In particular, we analyze the forecasting performance of alternative window selection methods in the presence of data revisions and recent structural breaks. To the best of our knowledge, there are no other papers analyzing the window selection problem in a real-time environment. We focus on a set of widely used robust estimation methods. These methods include an expanding window estimator, rolling window estimators, exponentially weighted moving average methods, and the average window method. The relative accuracy of these methods is evaluated using both Monte Carlo simulations and empirical forecasting experiments.

The statistical framework used in the Monte Carlo simulations closely follows that adopted in Clements and Galvão (2013). A novelty of this framework is that it allows data revisions to be characterized either as adding news or reducing noise. Thus, we are able to analyze whether the properties of the revision process matter for the relative accuracy of the alternative window selection methods. We consider several break processes, including changes in the intercept, autoregressive parameter, and error variance. The Monte Carlo results show that the expanding window estimator often yields the most accurate forecasts after a recent break. It performs well regardless of whether revisions add news or reduce noise, or whether we forecast first-release or final values. Interestingly, our numerical results suggest that whether data revisions add news or reduce noise does not matter much for the relative ranking of the alternative window selection methods.

In the empirical application of the essay, we compare the forecasting performance of the alternative window selection methods using actual U.S. GDP growth and inflation data. Our empirical results also indicate that the expanding window estimator usually outperforms the alternatives when forecasts are generated shortly after a structural break. In particular, the expanding window estimator is clearly the best estimation strategy when we forecast GDP deflator growth after the break in the early 1980s.

1.4.2 Chapter 3: Multi-step forecasting in the presence of breaks

The second essay contributes to the existing literature by analyzing multi-step forecasting in the presence of structural breaks and data revisions. We evaluate the accuracy of the multi-step methods in a real-time, unstable environment through Monte Carlo simulations. The statistical framework used in this essay closely follows that developed in Clements and Galvão (2013). A key feature of this framework is that data revisions either add news or reduce noise. The distinction between news and noise revisions allows us to study whether the properties of the revision process matter for the multi-step forecasting problem. We compare the forecasting performance of the iterated and direct AR models and various forms of intercept corrections suggested by Clements and Hendry (1996a, 1998). We consider several break processes, including changes in the intercept, autoregressive parameter, and error variance. Furthermore, we examine how the timing of the break affects the accuracy of the methods.

Our Monte Carlo results indicate that the type and the timing of the break affect the relative performance of the multi-step methods. We find that the iterated method usually provides the most accurate multi-step forecasts in the presence of structural instability. The iterated method performs particularly well when the parameters are subject to small breaks and the break occurs early during the estimation sample. The simulation results also suggest that the relative performance of the multi-step methods is qualitatively similar regardless of whether data revisions add news or reduce noise.

In the empirical application, we explore the ability of the multi-step methods to forecast four key U.S. macroeconomic time series, namely, real GDP, industrial production, GDP deflator, and personal consumption expenditures inflation. We generate real-time multi-step out-of-sample forecasts for the 1977:Q2–2013:Q2 period. The results of this forecasting exercise lend support to the view that the iterated method typically outperforms the alternatives in an unstable environment. Indeed, we find that the alternative multi-step methods only episodically improve upon the iterated method.

1.4.3. Chapter 4: Zero lower bound, unconventional monetary policy and indicator properties of interest rate spreads

In the third essay, we investigate the real-time predictive power of interest rate spreads for U.S. real economic activity when the short-term nominal rates have been stuck at the zero lower bound (ZLB) and the Federal Reserve has used unconventional monetary policy. The results in the previous literature indicate that the predictive content of interest rate spreads fluctuates over time (Stock and Watson, 2003 and the references cited therein). Both theoretical and empirical studies highlight that regime shifts in monetary policy have a major impact on the predictive ability of interest rate spreads (see, e.g., Bordo and Haubrich, 2008a, 2008b; Estrella *et al.*, 2003; Estrella, 2005; Giacomini and Rossi, 2006). Therefore, the beginning of the ZLB and unconventional monetary policy era in December 2008, which represents a fundamental change in monetary policy, is potentially important for the leading indicator properties of interest rate spreads.

We examine the predictive power of a term spread (i.e., the difference between the yields on long-term and short-term Treasury securities) and a set of credit spreads for U.S. industrial production growth in a real-time out-of-sample forecasting exercise running from June 2003 to March 2014. The results of this exercise suggest that the predictive content of the term spread has changed since the onset of the ZLB and unconventional monetary policy period. Thus, our results provide further evidence supporting the view that changes in monetary policy affect the ability of the term spread to forecast subsequent real activity. We also find that the predictive power of credit spreads varies over time. However, the ability of credit spreads to signal future industrial production growth seems to be unaffected by the beginning of the ZLB and unconventional monetary policy period. Finally, our results show that the mortgage spread (i.e., the difference between the 30-year mortgage rate and 10-year Treasury bond rate) is a particularly useful leading indicator for U.S. industrial production growth over the whole 2003:M6–2014:M3 period.

1.4.4 Chapter 5: The mortgage spread as a predictor of real-time economic activity

The fourth essay examines the real-time predictive power of the mortgage spread for U.S. real GDP and industrial production growth. Our out-of-sample forecasting period spans from 1992:Q1 to 2012:Q4. We compare the forecasting performance of the mortgage spread to that of two widely used leading indicators, namely, the term spread and the credit spread discussed in Gilchrist and Zakrajšek (2012). Finally, we analyze whether the predictive power of the mortgage spread remains stable over time.

The main finding from this study is that the mortgage spread is a useful leading indicator for real GDP and industrial production growth. Importantly, the mortgage spread produces more accurate real-time forecasts than the term spread or the Gilchrist–Zakrajšek (2012) spread. However, the predictive power of the mortgage spread fluctuates over time. We find that the mortgage spread has been particularly informative since the early 2000s.

References

Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, 40, 319–340.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2–11.

Bao, Y. (2007). Finite-sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution. *Econometric Theory*, 23, 767–773.

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

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

Bordo, M. D. and 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.

Brown, B. W. and Mariano, R. S. (1989). Measures of deterministic prediction bias in nonlinear models. *International Economic Review*, 30, 667–684.

Clements, M. P. (2014). Real-time factor model forecasting and the effects of instability. Working Paper, University of Reading.

Clements, M. P. and Galvão, A. B. (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28, 458–477.

Clements, M. P. and Hendry, D. F. (1996a). Intercept corrections and structural change. *Journal of Applied Econometrics*, 11, 475–494.

Clements, M. P. and Hendry, D. F. (1996b). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, 58, 657–684.

Clements, M. P. and Hendry, D. F. (1998). Forecasting economic time series. Cambridge University Press, Cambridge.

Clements, M. P. and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C. W. J. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 1. Amsterdam: Elsevier, 605–657.

Croushore, D. (2006). Forecasting with real-time data. In Elliott, G., Granger, C. W. J. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 1. Amsterdam: Elsevier, 961–982.

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

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105, 111–130.

Elliott, G. and Timmermann, A. (2008). Economic forecasting. *Journal of Economic Literature*, 46, 3–56.
Estrella, A. (2005). Why does the yield curve predict output and inflation? *Economic Journal*, 115, 722–744.
Estrella, A., Rodrigues, A. P. and 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. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier, 2–56.
Faust, J., Gilchrist, S., Wright, J. H. and 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.
Faust, J., Rogers, J. H. and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, 37, 403–419.
Friedman, B. M. and Kuttner, K. N. (1998). Indicator properties of the paper-bill spread: lessons from recent experience. *Review of Economics and Statistics*, 80, 34–44.
Gertler, M. and 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. and 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. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25, 595–620.
Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.
Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102, 1692–1720.
Gilchrist, S., Yankov, V. and Zakrajšek, E. (2009). Credit market shocks and economic fluctuations: evidence from corporate bond and stock markets. *Journal of Monetary Economics*, 56, 471–493.
Hoque, A., Magnus, J. R. and Pesaran, B. (1988). The exact multi-period mean square forecast error for the first-order autoregressive model. *Journal of Econometrics,* 39, 327–346.
Ing, C-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory*, 19, 254–279.
Luciani, M. (2014). Large-dimensional dynamic factor models in real-time: a survey. Working Paper, Université libre de Bruxelles.
McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980's? *American Economic Review*, 90, 1464–1476.
Mankiw, N. G. and Shapiro, M. D. (1986). News or noise: an analysis of GNP revisions. *Survey of Current Business*, 66, 20–25.
Mankiw, N. G., Runkle, D. E. and Shapiro, M. D. (1984). Are preliminary announcements of the money stock rational forecasts? *Journal of Monetary Economics*, 14, 15–27.
Marcellino, M., Stock, J. H. and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135, 499–526.
Mody, A. and 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.
Ng, S. and Wright, J. H. (2013). Facts and challenges from the Great Recession for forecasting and macroeconomic modeling. *Journal of Economic Literature*, 51, 1120–1154.

Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business and Economic Statistics*, 29, 307–318.

Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129, 183–217.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137, 134–161.

Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier, 1203–1324.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U. S. monetary policy? *American Economic Review*, 96, 54–81.

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, 128, 99–136.

Stark, T. and Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, 24, 507–531.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14, 11–30.

Stock, J. H. and Watson, M. W. (1999a). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In Engle, R. F. and White, H. (Eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*. Oxford: Oxford University Press, 1–44.

Stock, J. H. and Watson, M. W. (1999b). Forecasting inflation. *Journal of Monetary Economics*, 44, 293–335.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15, 101–115.

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

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

Stock, J. H. and Watson, M. W. (2007). Why has U. S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39, 3–33.

Stock, J. H. and Watson, M. W. (2009). Phillips curve inflation forecasts. In Fuhrer, J., Kodrzycki, Y., Little, J. and Olivei, G. (Eds.), *Understanding Inflation and the Implications for Monetary Policy*. Cambridge: MIT Press, 99–202.

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

Weiss, A. A. (1991). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics*, 48, 135–149.

Chapter 2
Selection of an estimation window in the presence of data revisions and recent structural breaks*

Jari Hännikäinen

Abstract

In this paper, we analyze the forecasting performance of a set of widely used window selection methods in the presence of data revisions and recent structural breaks. Our Monte Carlo and empirical results for U.S. real GDP and inflation show that the expanding window estimator often yields the most accurate forecasts after a recent break. It performs well regardless of whether the revisions are news or noise, or whether we forecast first-release or final values. We find that the differences in the forecasting accuracy are large in practice, especially when we forecast inflation after the break of the early 1980s.

Keywords: Recent structural break, choice of estimation window, forecasting, real-time data

JEL codes: C22, C53, C82

* I would like to thank Henri Nyberg, Jari Vainiomäki, Helinä Laakkonen, Markku Lanne, Juhani Raatikainen and the seminar participants in the FDPE Econometrics Workshop and in the XXXV Annual Meeting of the Finnish Economic Association for helpful comments and discussion. Financial support from the FDPE is gratefully acknowledged.

2.1 Introduction

Macroeconomic time series are often serially correlated. This implies that their own past values are themselves useful predictors. Therefore, it is not surprising that autoregressive (AR) models are used extensively in economic forecasting. The previous literature has found that it is difficult to outperform AR models in practice. For example, Rossi (2013) and Stock and Watson (2003) find that only a few macroeconomic predictors systematically improve upon the AR benchmark when forecasting inflation and output growth.

However, the parameters of AR models fitted to many macroeconomic time series are unstable over time (see, e.g., Stock and Watson, 1996). This observed parameter instability can arise as a result of several reasons. For instance, changes in tastes, technology, legislation, institutional arrangements, or government policy can cause changes in the dynamics of the economy. Structural breaks are crucial because they often have a major impact on forecasting performance: a forecasting model that performed well before the break might perform extremely poorly after the break (see, e.g., Clements and Hendry, 1998; Rossi, 2013). Because tastes, technology, legislation, institutional arrangements, and government policy are likely to change in the future, structural breaks are also likely to happen in the future. Therefore, information about the forecasting performance of AR models when these models undergo structural breaks is needed. Given the empirical success of AR models and their widespread use in practice, we believe that this is an important area to investigate.

A key question in the presence of structural instability is how many observations to use to estimate the parameters of a model so that, when used to generate a forecast, a loss function such as the root mean squared forecast error (RMSFE) will be minimized. This issue has been analyzed by Eklund *et al.* (2013), Giraitis *et al.* (2013), Pesaran and Pick (2011), Pesaran and Timmermann (2005, 2007), and Pesaran *et al.* (2013). This literature typically assumes that the break has occurred in the distant past. In such a case, the standard solution to the window selection problem is to test for breaks and use only observations after the most recent break. The estimates of the timing of the break(s) can be obtained, for example, using methods developed by Altissimo and Corradi (2003), Andrews (1993), Andrews *et al.* (1996), and Bai and Perron (1998, 2003). In the presence of recent breaks, this so called post-break window strategy is not feasible. As noted by Eklund *et al.* (2013), structural break tests are not designed for detecting recent breaks. Instead, the breaks are observed with a long lag. Even if real-time detection were possible, the post-break window strategy would not be useful. The parameters of the forecasting model are estimated with adequate accuracy only if the number of observations is at least two to three times the number of parameters (see, e.g., the discussion in Pesaran and Timmermann, 2005). Hence, the post-break window strategy is applicable only when the (last) break has occurred sufficiently long ago.

Forecasting after a recent break has received very little attention in the literature. However, in practice, forecast errors are often very large after structural breaks (Clements and Hendry, 2006). This suggests that improving forecast accuracy after a recent break is a central issue in economic forecasting.

Another issue that has often been overlooked in the literature is the real-time nature of the data used in many applications. For example, GDP and inflation series are published with a lag and are subject to revisions. These revisions are usually quite large and hence forecasts based on final revised data may differ considerably from those based on real-time data. Practical forecasting is inherently a real-time exercise, and therefore ingnoring the real-time nature of the data leads to a wide discrepancy between theory and practice.

We introduce two innovations on the existing literature. First, we focus on forecasting in the presence of recent breaks. To this end, several break processes are considered, including changes in the intercept, autoregressive parameter, and error variance. Second, we take into account that most macroeconomic time series are subject to data revisions. We follow the standard practice in the literature and allow revisions to be characterized either as news or noise, in the sense of Mankiw and Shapiro (1986). To the best of our knowledge, there are no other papers analyzing the window selection problem when the data are subject to revision.

The end of the Great Moderation and the Financial Crisis of 2008 provide an excellent motivation for our exercise. It is well-known that the volatility of many U.S. macroeconomic series has declined since the mid-1980s (see, e.g., McConnell and Perez-Quiros, 2000). Recent data suggest that this phenomenon, called the Great Moderation, came to an end with the Financial Crisis. Furthermore, monetary policy has changed fundamentally since the beginning of the crisis. The nominal short-term interest rate has been stuck at the zero lower bound and the Federal Reserve has used unconventional monetary policy, both of which should change the dynamics of key macro variables. So, forecasting these days, one would certainly run into the aforementioned too-few-data-after-the-break problem and the results of this paper will be relevant.

We consider a set of widely used methods for forecasting in the presence of structural instability. These methods include rolling windows, exponentially weighted moving average models, and the average window method advocated by Pesaran and Pick (2011) and Pesaran and Timmermann (2007). The potential gains in forecasting performance from using these methods compared to the expanding window method are demonstrated through Monte Carlo simulations and empirical examples.

The main finding from this study is that, at least for macroeconomic time series such as U.S. real GDP and inflation (defined as the growth rate of the GDP deflator), the expanding window estimator tends to produce more accurate forecasts than the alternative window selection methods considered here. Our simulation results indicate that the expanding window method performs particularly well when the parameters remain constant over time or when the innovation variance changes. Our empirical results suggest that the expanding window estimator is overwhelmingly the best estimation strategy when we forecast inflation after the break in the early 1980s. In this case, the alternative methods produce 7.5–52.9 percent larger forecast errors than the expanding window estimator. The expanding window method also performs well when we make real-time GDP growth forecasts for the period 2008:Q4–2011:Q1. However, we find that, in this case, the differences in relative performances are more modest.

The remainder of the paper is organized as follows. Section 2 introduces the notation and the statistical framework. Section 3 provides a brief overview of the window selection methods. Section 4 presents the Monte Carlo simulation results and Section 5 presents the empirical results. Section 6 concludes. The appendices at the end of the paper provide the technical details.

2.2 Statistical framework

An important feature of real-time data is that the data for a period are not released until some time has passed after the end of that period. Therefore, for instance, a forecaster at period $T+1$ has access to the vintage $T+1$ values of real GDP and inflation up to time period T. Furthermore, the data are revised over time, so the first-released values and the final values may differ considerably. Although the real-time nature of macroeconomic time series clearly matters for forecasting, data revisions are rarely incorporated into the theoretical models. One exception is the statistical framework suggested by Jacobs and van Norden (2011) and further developed by Clements and Galvão (2013). This framework for modeling data revisions, which we will closely follow, relates a data vintage estimate to the true value plus an error or errors. In particular, the period $t+s$ vintage estimate of the value of y in period t, denoted by y_t^{t+s} [1], where $s = 1,\dots,l$ [2], can be expressed as a sum of the true value $\tilde{y}_t$, a news component v_t^{t+s}, and a noise component ε_t^{t+s}, so that $y_t^{t+s} = \tilde{y}_t + v_t^{t+s} + \varepsilon_t^{t+s}$.

1 Throughout this paper, superscripts refer to vintages and subscripts to time periods. This notation has become standard in the literature.

2 For simplicity, we assume that we observe l different estimates of y_t before the true value, $\tilde{y}_t$, is observed. In practice, however, the true value may never be observed.

This framework follows the standard practice in the literature and assumes that revisions either add news or reduce noise. Data revisions are said to be news if revisions are uncorrelated with the previously published vintages, $cov(y_t^{t+k}, v_t^{t+s}) = 0\ \forall k \leq s$. This implies that the initially released data are optimal forecasts of the later data. On the other hand, data revisions reduce noise if each vintage release is equal to the true value plus a noise, so that noise revisions are uncorrelated with the truth, $cov(\tilde{y}_t, \varepsilon_t^{t+s}) = 0$. For further discussion of the properties of news and noise revisions, see Croushore (2011) and Jacobs and van Norden (2011). The distinction between news and noise revisions is important in practice because revisions to different macroeconomic time series have different characteristics. For example, Clements and Galvão (2013) find that, at least since the mid-1980s, data revisions to output growth appear to be mainly news whereas those to inflation are mainly noise.

Following Clements and Galvão (2013) and Jacobs and van Norden (2011), we stack the l different vintage estimates of y_t, v_t and ε_t into vectors $\boldsymbol{y}_t = (y_t^{t+1}, ..., y_t^{t+l})'$, $\boldsymbol{v}_t = (v_t^{t+1}, ..., v_t^{t+l})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{t+1}, ..., \varepsilon_t^{t+l})'$, respectively. Now we can express each vintage of y_t as follows

$$\boldsymbol{y}_t = \boldsymbol{i}\tilde{y}_t + \boldsymbol{v}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{i}$ is an $l \times 1$ vector of ones. For the true values we consider the following AR(1) process subject to a single structural break at time T_1

$$\tilde{y}_t = \begin{cases} \rho_1 + \sum_{i=1}^{l} \mu_{v1_i} + \beta_1 \tilde{y}_{t-1} + \sigma_1 \eta_{1t} + \sum_{i=1}^{l} \sigma_{v1_i} \eta_{2t,i}, & \text{for } t < T_1, \\ \rho_2 + \sum_{i=1}^{l} \mu_{v2_i} + \beta_2 \tilde{y}_{t-1} + \sigma_2 \eta_{1t} + \sum_{i=1}^{l} \sigma_{v2_i} \eta_{2t,i}, & \text{for } t \geq T_1, \end{cases} \tag{2}$$

where η_{1t} and $\eta_{2t,i}$ $(i = 1,...,l)$ are $NIID\,(0,1)$ disturbances.[3]

The news and noise processes of each vintage are specified by

$$\boldsymbol{v}_{1t} = \begin{bmatrix} v_{1t}^{t+1} \\ v_{1t}^{t+2} \\ \vdots \\ v_{1t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^{l} \mu_{v1_i} \\ \sum_{i=2}^{l} \mu_{v1_i} \\ \vdots \\ \mu_{v1_l} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{l} \sigma_{v1_i} \eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v1_i} \eta_{2t,i} \\ \vdots \\ \sigma_{v1_l} \eta_{2t,l} \end{bmatrix}, \boldsymbol{\varepsilon}_{1t} = \begin{bmatrix} \varepsilon_{1t}^{t+1} \\ \varepsilon_{1t}^{t+2} \\ \vdots \\ \varepsilon_{1t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \mu_{\varepsilon 1_1} \\ \mu_{\varepsilon 1_2} \\ \vdots \\ \mu_{\varepsilon 1_l} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon 1_1} \eta_{3t,1} \\ \sigma_{\varepsilon 1_2} \eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon 1_l} \eta_{3t,l} \end{bmatrix} \tag{3}$$

3 We focus on the shortest possible lag length, because we want to minimize the number of possible breaks in the autoregressive structure. Furthermore, it is easier to calibrate the parameters (see the discussion below) when the lag order is one. Eklund *et al.* (2013) and Pesaran and Timmermann (2005) also consider an AR(1) specification in the presence of breaks.

for $t < T_1$ and

$$\boldsymbol{v}_{2t} = \begin{bmatrix} v_{2t}^{t+1} \\ v_{2t}^{t+2} \\ \vdots \\ v_{2t}^{t+l} \end{bmatrix} = -\begin{bmatrix} \sum_{i=1}^{l} \mu_{v2_i} \\ \sum_{i=2}^{l} \mu_{v2_i} \\ \vdots \\ \mu_{v2_l} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{l} \sigma_{v2_i}\eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v2_i}\eta_{2t,i} \\ \vdots \\ \sigma_{v2_l}\eta_{2t,l} \end{bmatrix}, \boldsymbol{\varepsilon}_{2t} = \begin{bmatrix} \varepsilon_{2t}^{t+1} \\ \varepsilon_{2t}^{t+2} \\ \vdots \\ \varepsilon_{2t}^{t+l} \end{bmatrix} = -\begin{bmatrix} \mu_{\varepsilon 2_1} \\ \mu_{\varepsilon 2_2} \\ \vdots \\ \mu_{\varepsilon 2_l} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon 2_1}\eta_{3t,1} \\ \sigma_{\varepsilon 2_2}\eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon 2_l}\eta_{3t,l} \end{bmatrix} \quad (4)$$

for $t \geq T_1$.

The shocks are assumed to be mutually independent, i.e., if $\boldsymbol{\eta}_t = [\eta_{1t}, \boldsymbol{\eta}'_{2t}, \boldsymbol{\eta}'_{3t}]$, then $E(\boldsymbol{\eta}_t) = \mathbf{0}$ and $E(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t) = I$. We assume that $\tilde{y}_t$ is a stationary process, so that $|\beta_j| < 1$ (for j = 1,2). Because $\tilde{y}_t$ is a stationary process and both the news and noise terms are stationary, (1) implies that y_t is also a stationary process. Note that the means of the news and noise terms, denoted by μ_{vji} and $\mu_{\varepsilon j_i}$ (for j = 1,2 and i = 1,...,l), are allowed to be non-zero. This is an important feature because in practice revisions to macroeconomic data have non-zero means (see, e.g., Aruoba, 2008; Clements and Galvão, 2013; Croushore, 2011).

As discussed earlier, this framework is similar to that adopted in Clements and Galvão (2013) and Jacobs and van Norden (2011). The main point of departure from their framework is that we allow the process of the true values to be subject to a recent structural break. Our setup is quite general and allows for changes in intercept, slope, and error variance immediately after the break. Another novelty of our framework is that the means and variances of the news and noise revisions are also allowed to change.

2.3 Forecasting methods

In the presence of data revisions and structural breaks, a forecaster faces two key questions. First, a forecaster has to decide how to take into account the real-time nature of the data when estimating the parameters of the forecasting model. The most commonly used approach, called the end-of-sample vintage approach (EOS), uses observations from the latest available (T +1) vintage

$$y_t^{T+1} = \alpha_0 + \alpha_1 y_{t-1}^{T+1} + e_{t,EOS}, \qquad \text{for } t = 2, ..., T. \quad (5)$$

The forecast of y_{T+1} is conditioned on the latest available vintage value of the forecast origin data, so that $\hat{y}_{T+1,EOS} = \hat{\alpha}_0 + \hat{\alpha}_1 y_T^{T+1}$. Although popular in practice, the EOS approach has a fundamental shortcoming: a large part of the data used in model estimation has been revised many times (early in the sample), while the forecast is conditioned on first-release data (the latest observation).

An alternative estimation strategy is the real-time vintage approach (RTV) suggested by Koenig *et al.* (2003). The central idea in the RTV approach is that the data used in estimation and the data on which the forecast is conditioned should be of a similar maturity. Therefore, the forecasting model is estimated on first-release data

$$y_t^{t+1} = \beta_0 + \beta_1 y_{t-1}^t + e_{t,RTV}, \qquad \text{for } t = 2, \ldots, T, \tag{6}$$

and the corresponding forecast is $\hat{y}_{T+1,RTV} = \hat{\beta}_0 + \hat{\beta}_1 y_T^{T+1}$. Note that the two forecasts are conditioned on exactly the same data. The only difference between the two approaches is the data used in the estimation.

The results in Clements and Galvão (2013) and Koenig *et al.* (2003) indicate that the RTV approach produces more accurate forecasts than the EOS approach. However, it is not known whether this result holds in the presence of structural instability. Thus, our plan is to shed light on the relative accuracy of these two methods in the presence of recent breaks.

The second question a forecaster faces is how much data to use to estimate the parameters of the forecasting model. One solution to this window selection problem is to test for breaks and use only observations over a post-break window. If the structural break has occurred recently, this post-break window strategy is infeasible for two reasons. First, it is difficult or even impossible to estimate accurately the timing of a recent break. Second, even if an accurate detection of a recent break were possible, the post-break window strategy is infeasible because a sufficient number of post-break observations, say at least two to three times the number of parameters, is required for accurate estimation. Once the real-time nature of the data is taken into account, the problems associated with the post-break window strategy get compounded since the break may not be as apparent in real-time. Moreover, post-break observations are less 'mature', which will cause problems with accuracy.

An alternative solution is to use robust estimation strategies. An estimation strategy is said to be robust if no information about the structural break is needed for its implementation. Therefore, robust methods are also valid in the presence of recent breaks. In this paper, we focus exclusively on robust methods. We compare the forecasting performance of a set of widely used estimation strategies when the underlying time series process has undergone a recent structural break. Common to all of these strategies is that the estimation window should exceed a minimum length, denoted by $\underline{\omega}$.

The first strategy is the expanding window estimator

$$\hat{\boldsymbol{\beta}}_{T,EXP} = \left(\sum_{t=1}^{T} \boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}^{'} \right)^{-1} \sum_{t=1}^{T} \boldsymbol{x}_{t-1}^{'} y_t,$$

where $\boldsymbol{x}_t = (1, y_t)'$. The expanding window estimator uses the whole data sample available at the forecast origin. The expanding window forecast for period T +1 is computed by $\hat{y}_{T+1,EXP} = \hat{\beta}'_{T,EXP} \boldsymbol{x}_T$.

The second strategy is the rolling window estimator

$$\hat{\boldsymbol{\beta}}_{T,ROLL}(m) = \left(\sum_{t=T-m+1}^{T} \boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}^{'} \right)^{-1} \sum_{t=T-m+1}^{T} \boldsymbol{x}_{t-1}^{'} y_t,$$

where $m \in \underline{\omega}, ..., T$ is the length of the rolling window. The parameters are estimated using the m most recent observations. The resulting forecast for period T +1 is computed by $\hat{y}_{T+1,ROLL}(m) = \hat{\beta}'_{T,ROLL}(m)\boldsymbol{x}_T$. Giacomini and White (2006) argue that when the forecasting model is misspecified (due to inadequately modeled dynamics, inadequately modeled heterogeneity, incorrect functional form, or any combination of these), the rolling window estimator often provides more reliable forecasts than the expanding window estimator.

The third alternative is the exponentially weighted moving average (EWMA) method. This method, unlike rolling regressions, gives a positive weight to each observation. The central idea is that if the relation of interest has changed over time, the most recent observations are more informative than the earlier ones. Thus, the most recent observations receive the highest weight in the estimation:

$$\hat{\boldsymbol{\beta}}_{T,EWMA} = \left(\lambda \sum_{t=1}^{T} (1-\lambda)^{T-t} \boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}^{'} \right)^{-1} \lambda \sum_{t=1}^{T} (1-\lambda)^{T-t} \boldsymbol{x}_{t-1}^{'} y_t,$$

where $0 < \lambda < 1$ is the down-weighting parameter. The forecast for period T +1 is computed by $\hat{y}_{T+1,EWMA} = \hat{\beta}'_{T,EWMA} \boldsymbol{x}_T$. Pesaran and Pick (2011) find that the choice of the down-weighting parameter greatly affects the forecasting performance of the EWMA method.

A final alternative is the average window (AveW) method suggested by Pesaran and Timmermann (2007). This method builds on the common finding that forecast combinations often reduce forecast errors (see, e.g., Timmermann, 2006). Therefore, rather than selecting a single estimation window, the AveW method combines forecasts from models estimated on different observation windows. The AveW method gives an equal weight to each forecast,

$$\hat{y}_{T+1,AveW} = (T - \underline{\omega} + 1)^{-1} \sum_{m=\underline{\omega}}^{T} \hat{y}_{T+1,ROLL}(m),$$

where $\hat{y}_{T+1,ROLL}(m)$ denotes the forecast generated by a rolling window of size m.

2.4 Monte Carlo simulations

In this section, we evaluate the forecasting performance of alternative window selection methods in a set of Monte Carlo experiments. These experiments are based on the statistical framework introduced in Section 2. Our interest in this paper lies in the point forecasts shortly after a structural break. Therefore, we assume that a single break has occurred at time $T_1 = T$. One-step ahead forecasts are made recursively for the next ten periods, i.e., for the periods $T+1,\ldots,T+10$. We assume that no breaks occur during the forecasting period.

To ensure that our simulation results are empirically relevant, we calibrate the parameters on actual U.S. output and inflation data. We start by considering the case where the parameters remain stable over time (experiment 1 in Table 1). In this case the mean of the true process lies between 2.0 and 2.5, which corresponds roughly to the average annual inflation and real GDP growth since the mid-1980s. The parameters of this model are used as benchmarks in the rest of the experiments. We consider both moderate (0.25) and large (0.5) changes in the autoregressive parameter in either direction (experiments 2–5). We also consider changes in the error variance. We allow σ to increase from 1.5 to 4.5 (experiment 6) and decrease from 1.5 to 0.5 (experiment 7). Finally, we study the effects of breaks in the constant term (experiments 8–9).

We assume that the revisions are either pure news ($\sigma_{v_i} \neq 0$, $\sigma_{\varepsilon_i} = 0$ for $i = 1, \ldots, l$) or pure noise ($\sigma_{v_i} = 0$, $\sigma_{\varepsilon_i} \neq 0$ for $i = 1, \ldots, l$). This allows us to analyze whether the properties of the revision process matters for the window selection problem. We set $l =$ 14, so that we observe 14 different estimates of y_t before the true value, $\tilde{y}_t$, is observed. Following Clements and Galvão (2013), we assume that the first and the fifth revisions are non-zero mean. The means of these revisions are set to four and two percent of the mean of the first-release data, y_t^{t+1}, both before and after the break. Similarly, the standard deviation of the first revision is set to 40 percent of the standard deviation of the first-release data. The standard deviations of revisions 2–13 and 14 are set to 20 and 10 percent of the standard deviation of the first-release data, respectively. For convenience, the parameter values used in the Monte Carlo experiments are reported in Table 1.[4]

We examine the ability of various window selection methods to forecast both the first-release values (y_{t+1}^{t+2}) and the final values (y_{t+1}^{t+16}). Because we assume that revisions have non-zero mean, the final values differ systematically from the first-release values. As a consequence, the forecasting models in (5) and (6) produce unbiased forecasts for the first-release values, but biased forecasts for the final values. In order to produce unbiased forecasts for the final values, we use the bias correction method suggested by Clements and Galvão (2013). The bias correction is the sample estimate of the difference between the final value and the first-release value calculated using data up to the forecast origin. To be more specific, the forecast for the final value is computed using the formula $\hat{y}_{t+1}^{t+16} = \hat{y}_{t+1}^{t+2} + (t-14)^{-1}\sum_{i=1}^{t-14}(y_i^{i+15} - y_i^{i+1})$. An alternative approach, of course, would be to use the fully revised data as the left hand side variable in (5) and (6). As discussed in Clements and Galvão (2013), these two approaches are asymptotically equivalent. However, the bias correction method yields more accurate forecasts in small samples.

We focus on a set of widely used robust estimation strategies, including the rolling window, the exponentially weighted moving average (EWMA), and the average window (AveW) method. We analyze the forecasting performance of a short rolling window using the most recent 20 observations and a long rolling window using the most recent 40 observations. These rolling windows correspond to five and 10 years of quarterly data, respectively. The down-weighting parameter, λ, in the EWMA method is set to 0.05 (henceforth EWMAS). In addition, we follow Eklund *et al.* (2013) and consider a method that combines different down-weighting parameters. More specifically, we calculate an equally weighted forecast using down-weighting parameters of 0.1, 0.2 and 0.3 (henceforth EWMAA). We assume that the minimum estimation window length, $\underline{\omega}$, in the AveW method is 10 observations.

The expanding window estimator is the most efficient estimation method when the underlying time series process is stable over time. Therefore, it is used as a benchmark in our Monte Carlo simulations. For each robust estimation strategy we compute RMSFE values relative to those produced by the expanding window benchmark. Values below (above) unity indicate that the candidate method produces more (less) accurate forecasts

4 Formulas for the means and standard deviations of the first-release and final data are presented in Appendix A. This appendix also presents the formulas for the means and standard deviations of data revisions when the revisions are either pure news or pure noise. Appendix B gives the means and standard deviations of the first-release and final data for each experiment.

Table 1. Simulation setup

True process Experiments	ρ_1	ρ_2	β_1	β_2	σ_1	σ_2				
1: No break	1	1	0.5	0.5	1.5	1.5				
2: Moderate break in β (increase)	1	1	0.5	0.75	1.5	1.5				
3: Moderate break in β (decrease)	1	1	0.5	0.25	1.5	1.5				
4: Large break in β (increase)	1	1	0.25	0.75	1.5	1.5				
5: Large break in β (decrease)	1	1	0.75	0.25	1.5	1.5				
6: Increase in post-break variance	1	1	0.5	0.5	1.5	4.5				
7: Decrease in post-break variance	1	1	0.5	0.5	1.5	0.5				
8: Break in mean (increase)	1	1.5	0.5	0.5	1.5	1.5				
9: Break in mean (decrease)	1	0.5	0.5	0.5	1.5	1.5				
News Experiments	μ_{v1_1}	μ_{v2_1}	μ_{v1_5}	μ_{v2_5}	σ_{v1_1}	σ_{v2_1}	$\sigma_{v1_{2,..,13}}$	$\sigma_{v2_{2,...,13}}$	$\sigma_{v1_{14}}$	$\sigma_{v2_{14}}$
1: No break	0.085	0.085	0.043	0.043	0.783	0.783	0.391	0.391	0.196	0.196
2: Moderate break in β (increase)	0.085	0.195	0.043	0.098	0.783	2.238	0.391	1.119	0.196	0.560
3: Moderate break in β (decrease)	0.085	0.054	0.043	0.027	0.783	0.634	0.391	0.317	0.196	0.158
4: Large break in β (increase)	0.054	0.195	0.027	0.098	0.634	2.238	0.317	1.119	0.158	0.560
5: Large break in β (decrease)	0.195	0.054	0.098	0.027	2.238	0.634	1.119	0.317	0.560	0.158
6: Increase in post-break variance	0.085	0.085	0.043	0.043	0.783	2.348	0.391	1.174	0.196	0.587
7: Decrease in post-break variance	0.085	0.085	0.043	0.043	0.783	0.261	0.391	0.130	0.196	0.065
8: Break in mean (increase)	0.085	0.128	0.043	0.064	0.783	0.783	0.391	0.391	0.196	0.196
9: Break in mean (decrease)	0.085	0.043	0.043	0.021	0.783	0.783	0.391	0.391	0.196	0.196
Noise Experiments	$\mu_{\varepsilon1_1}$	$\mu_{\varepsilon2_1}$	$\mu_{\varepsilon1_{2,...,5}}$	$\mu_{\varepsilon2_{2,...,5}}$	$\sigma_{\varepsilon1_1}$	$\sigma_{\varepsilon2_1}$	$\sigma_{\varepsilon1_{2,4,...,14}}$	$\sigma_{\varepsilon2_{2,4,...,14}}$	$\sigma_{\varepsilon1_{3,5,...,13}}$	$\sigma_{\varepsilon2_{3,5,...,13}}$
1: No break	0.113	0.113	0.038	0.038	0.728	0.728	0.188	0.188	0.325	0.325
2: Moderate break in β (increase)	0.113	0.226	0.038	0.075	0.728	0.953	0.188	0.246	0.325	0.426
3: Moderate break in β (decrease)	0.113	0.075	0.038	0.025	0.728	0.651	0.188	0.168	0.325	0.291
4: Large break in β (increase)	0.075	0.226	0.025	0.075	0.651	0.953	0.168	0.246	0.291	0.426
5: Large break in β (decrease)	0.226	0.075	0.075	0.025	0.953	0.651	0.246	0.168	0.426	0.291
6: Increase in post-break variance	0.113	0.113	0.038	0.038	0.728	2.183	0.188	0.564	0.325	0.976
7: Decrease in post-break variance	0.113	0.113	0.038	0.038	0.728	0.243	0.188	0.063	0.325	0.108
8: Break in mean (increase)	0.113	0.170	0.038	0.057	0.728	0.728	0.188	0.188	0.325	0.325
9: Break in mean (decrease)	0.113	0.057	0.038	0.019	0.728	0.728	0.188	0.188	0.325	0.325

than the benchmark. Relative RMSFE values are computed with sample sizes $T = 50$, 100, and 150. The results are based on 10,000 replications and are shown in Tables 2–5.

First, we compare the forecasting performance of alternative window selection methods when the revisions are pure news. The results, presented in Tables 2 and 3, reveal that the forecasting methods that generate the lowest RMSFE values in most of the experiments are the expanding window benchmark and the EWMAS method. Indeed, the expanding window estimator produces the most accurate forecasts in 52 of the 108 dependent variable/experiment/vintage approach/sample size combinations considered here. It performs particularly well when the parameters remain stable over time (experiment 1) or when the variance of the time series changes (experiments 6 and 7). There is a simple explanation for these findings. When the parameters remain fixed over time, it is optimal to use as many observations as possible in the estimation. Similarly, when a break only affects the volatility of the series, the variance of the parameter estimation error can be reduced by using a longer estimation window. The expanding window estimator performs poorly only when the autoregressive parameter is subject to large changes (experiments 4 and 5). Such breaks imply huge changes in the mean of the process, and are thus unlikely to occur in practice. Therefore, the weak performance of the expanding window estimator in the presence of large slope shifts should not be overemphasized. Interestingly, we find that it is more difficult to outperform the benchmark when the RTV approach is used. Similarly, the expanding window method performs better when the first-release values are the ones to be forecast.

Another prominent window selection method is the EWMAS approach. This approach performs well when the autoregressive parameter increases substantially after the break (experiment 4). In this case, the improvements over the expanding window benchmark are quite large, ranging from 0.6 to 8 percent. The EWMAS method also does particularly well in experiment 5 when we forecast first-release values, and in experiment 9 when we forecast the final values. Note that the EWMAS method improves upon the benchmark more often when the EOS approach is used.

In the few cases where the expanding window or EWMAS approach do not dominate, the EWMAA and AveW methods generate the best forecasts. The AveW method produces forecasts that are very close to those produced by the expanding window estimator. Therefore both the gains and losses in relative accuracy are more modest than with the other methods. The EWMAA method, on the other hand, performs well when the slope parameter decreases substantially after the break (experiment 5) and we forecast the final values, but extremely poorly in the vast majority of the experiments. The rolling windows fare no better: they rarely improve upon the benchmark and never produce the most accurate forecasts.

Table 2. Relative RMSFE values when revisions are news

	First-release												
	EOS							RTV					
Exp.	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW	*T* = 50	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW
1	1.743	1.042	1.007	1.007	1.078	1.011		1.732	1.036	1.006	1.010	1.096	1.011
2	3.906	1.016	1.001	0.999	1.059	1.002		3.925	1.030	1.004	1.014	1.126	1.010
3	1.673	1.029	1.005	0.999	1.054	1.001		1.646	1.023	1.003	1.003	1.071	1.001
4	4.159	0.984	0.990	0.971	1.010	0.980		4.160	1.004	0.995	0.992	1.087	0.992
5	2.208	1.061	1.014	0.954	0.954	0.998		2.114	1.043	1.008	0.968	0.982	0.997
6	5.213	1.046	1.009	1.034	1.174	1.022		5.248	1.053	1.010	1.044	1.213	1.027
7	0.654	1.171	1.038	0.997	1.030	1.035		0.621	1.137	1.028	1.007	1.063	1.033
8	1.789	1.033	1.006	0.998	1.056	1.005		1.794	1.025	1.004	1.000	1.070	1.004
9	1.822	1.027	1.005	0.993	1.044	1.002		1.794	1.025	1.004	0.999	1.068	1.004
							T = 100						
1	1.714	1.056	1.020	1.015	1.093	1.008		1.707	1.047	1.016	1.017	1.110	1.007
2	3.917	1.011	0.996	0.991	1.051	0.992		3.920	1.028	1.003	1.008	1.121	0.997
3	1.665	1.040	1.016	1.006	1.061	1.001		1.642	1.030	1.011	1.007	1.075	1.000
4	4.302	0.955	0.962	0.941	0.980	0.962		4.275	0.980	0.972	0.966	1.062	0.970
5	2.116	1.111	1.057	0.982	0.993	1.009		2.053	1.079	1.037	0.987	1.011	1.005
6	5.149	1.057	1.021	1.041	1.187	1.015		5.166	1.072	1.027	1.057	1.243	1.020
7	0.612	1.240	1.109	1.043	1.098	1.034		0.592	1.185	1.080	1.040	1.112	1.028
8	1.782	1.041	1.015	1.004	1.066	1.002		1.793	1.032	1.010	1.003	1.077	1.001
9	1.809	1.037	1.013	0.999	1.055	1.001		1.784	1.034	1.011	1.004	1.079	1.002
							T = 150						
1	1.711	1.061	1.025	1.020	1.099	1.006		1.707	1.052	1.021	1.020	1.113	1.005
2	3.944	1.008	0.993	0.988	1.050	0.988		3.939	1.026	1.000	1.007	1.120	0.992
3	1.661	1.042	1.018	1.008	1.066	1.000		1.641	1.031	1.012	1.009	1.078	0.999
4	4.414	0.933	0.940	0.920	0.956	0.952		4.383	0.959	0.951	0.945	1.042	0.957
5	2.085	1.122	1.071	0.994	1.004	1.009		2.030	1.085	1.048	0.995	1.019	1.005
6	5.128	1.061	1.024	1.044	1.188	1.011		5.138	1.076	1.032	1.061	1.242	1.015
7	0.598	1.268	1.136	1.067	1.122	1.030		0.583	1.197	1.095	1.053	1.127	1.021
8	1.766	1.046	1.019	1.008	1.072	1.002		1.780	1.036	1.013	1.006	1.082	1.001
9	1.811	1.040	1.017	1.001	1.056	1.001		1.788	1.035	1.014	1.005	1.079	1.001

Notes: The experiments are as defined in Table 1. Method m = 20 denotes rolling window of size 20, whereas m = 40 denotes rolling window of size 40. EWMAS denotes EWMA method with $\lambda = 0.05$ and EWMAA denotes EWMA method with $\lambda = 0.1$, 0.2 and 0.3. AveW denotes the average window method. The sample size is T. The break occurs at period $T_1 = T$. One-step ahead forecasts are generated recursively for periods $T+1,\dots,T+10$. The first column in each panel shows the RMSFE for the expanding window estimator. In subsequent columns, RMSFE values are computed relative to those produced by the expanding window estimator.

Table 3. Relative RMSFE values when revisions are news

	Final value												
	EOS							RTV					
Exp.	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW	*T* = 50	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW
1	2.383	1.018	1.003	1.000	1.035	1.003		2.366	1.017	1.003	1.004	1.049	1.005
2	5.991	1.006	1.000	0.998	1.024	1.000		6.001	1.012	1.001	1.005	1.054	1.004
3	2.146	1.012	1.002	0.995	1.023	0.997		2.114	1.010	1.001	1.000	1.038	0.999
4	6.155	0.990	0.995	0.984	0.999	0.989		6.151	0.999	0.997	0.994	1.035	0.995
5	2.895	1.004	1.002	0.945	0.909	0.979		2.777	1.000	1.000	0.958	0.937	0.983
6	7.044	1.025	1.005	1.018	1.099	1.012		7.066	1.029	1.006	1.025	1.122	1.015
7	0.932	1.059	1.013	0.978	0.967	0.999		0.881	1.057	1.012	0.993	1.007	1.008
8	2.428	1.012	1.002	0.994	1.021	0.999		2.424	1.011	1.002	0.997	1.032	1.000
9	2.460	1.010	1.001	0.992	1.015	0.998		2.428	1.010	1.001	0.997	1.031	1.000
							T = 100						
1	2.345	1.027	1.010	1.006	1.048	1.003		2.335	1.024	1.009	1.008	1.059	1.003
2	6.008	1.002	0.997	0.994	1.018	0.996		6.008	1.010	1.000	1.001	1.049	0.998
3	2.129	1.021	1.009	1.001	1.032	0.999		2.103	1.017	1.006	1.003	1.043	0.999
4	6.277	0.977	0.981	0.971	0.986	0.981		6.256	0.989	0.986	0.983	1.026	0.985
5	2.688	1.051	1.030	0.970	0.952	0.999		2.613	1.035	1.019	0.976	0.970	0.998
6	7.012	1.031	1.011	1.022	1.105	1.008		7.023	1.040	1.015	1.031	1.138	1.011
7	0.852	1.111	1.050	1.008	1.023	1.010		0.823	1.090	1.039	1.014	1.044	1.011
8	2.398	1.020	1.007	0.999	1.031	1.000		2.403	1.017	1.005	1.000	1.039	1.000
9	2.431	1.017	1.006	0.996	1.024	0.999		2.405	1.017	1.006	1.000	1.040	1.000
							T = 150						
1	2.340	1.031	1.013	1.009	1.052	1.003		2.333	1.027	1.010	1.010	1.062	1.003
2	6.026	1.000	0.996	0.993	1.018	0.994		6.022	1.009	0.999	1.001	1.050	0.996
3	2.113	1.023	1.010	1.003	1.036	0.999		2.092	1.018	1.007	1.004	1.045	0.999
4	6.371	0.963	0.968	0.957	0.971	0.975		6.348	0.975	0.973	0.968	1.011	0.978
5	2.623	1.064	1.041	0.980	0.966	1.002		2.559	1.042	1.026	0.983	0.980	1.000
6	6.957	1.033	1.013	1.024	1.105	1.006		6.964	1.042	1.017	1.033	1.138	1.008
7	0.827	1.135	1.069	1.026	1.045	1.012		0.806	1.102	1.049	1.023	1.058	1.010
8	2.393	1.022	1.009	1.001	1.033	1.000		2.401	1.017	1.006	1.001	1.041	1.000
9	2.430	1.019	1.008	0.997	1.025	0.999		2.408	1.017	1.007	1.000	1.038	1.000

See the notes to Table 2.

Our results indicate that the choice between the EOS and RTV approaches is not clear-cut: the RTV approach yields more accurate forecasts in experiments 1, 3, 5, 7, and 9, whereas the EOS approach yields more accurate forecasts in experiment 6. The evidence for experiments 2, 4, and 8 is mixed. Hence, the EOS approach can be recommended only when the volatility of the series increases after a break. Another point worth noticing is that the sample size also matters for forecasting accuracy. We find that the selection of the estimation window becomes more important when the sample size increases.

The results for noise revisions are reported in Tables 4 and 5. These results are qualitatively similar to those presented in Tables 2 and 3, suggesting that the news versus noise issue does not matter much for the relative ranking of the alternative window selection methods. If anything, the view that emerges from Tables 2–5 is that the expanding window estimator performs slightly better when the revisions reduce noise. In such cases, it produces the best forecasts in 55 of the 108 cases. The EWMAS method also performs quite well when the revisions reduce noise. However, the evidence for its predictive ability is not as convincing as it is when the revisions are news. The results for noise revisions imply that it is more difficult to improve upon the benchmark when the EOS approach is used. This is a surprising result because the opposite was the case when the revisions were news. Again, the expanding window estimator performs better when the first-release values are the ones to be forecast.

When the revisions reduce noise, the ranking between the EOS and RTV approaches is very different. The EOS approach produces more reliable forecasts in experiments 2, 4, 5, and 8, whereas the RTV approach produces more accurate forecasts in experiments 3, 6, and 9. The evidence for experiments 1 and 7 is mixed. Once again, the choice of the estimation window matters more when the sample size is large. The results reported in Tables 2–5 reveal that the differences in the forecasting accuracy are larger when the revisions reduce noise. This result suggests that the choice of correct estimation window is more important when the revisions reduce noise.

To sum up, our results are consistent with the view that the news versus noise issue does not matter much for the relative ranking of alternative window selection methods. We find that the expanding window estimator often produces the best forecasts after a recent break—regardless of whether the revisions add news or reduce noise. However, the news versus noise issue matters for the relative accuracy of the EOS and RTV approaches. In general, our results suggest that the RTV approach yields more accurate forecasts when the revisions add news, whereas the EOS approach generates more reliable forecasts when the revisions reduce noise. This result is consistent with the findings in Clements and Galvão (2013).

Table 4. Relative RMSFE values when revisions are noise

	First-release												
	EOS							RTV					
Exp.	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW	*T* = 50	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW
1	1.733	1.031	1.006	1.012	1.102	1.009		1.734	1.036	1.007	1.009	1.091	1.010
2	2.023	1.006	0.998	0.988	1.049	0.993		2.101	0.998	0.996	0.972	1.013	0.986
3	1.769	1.017	1.003	1.005	1.080	1.000		1.746	1.024	1.004	1.005	1.080	1.003
4	2.260	0.954	0.982	0.936	0.953	0.957		2.368	0.944	0.980	0.925	0.922	0.952
5	2.038	1.010	1.001	0.999	1.047	0.995		2.063	1.004	0.999	0.980	1.007	0.988
6	5.297	1.056	1.011	1.050	1.246	1.029		5.263	1.054	1.010	1.045	1.224	1.027
7	0.615	1.108	1.025	1.007	1.056	1.026		0.624	1.130	1.031	1.004	1.048	1.030
8	1.780	1.025	1.004	1.007	1.085	1.005		1.810	1.025	1.003	0.999	1.067	1.003
9	1.797	1.023	1.004	1.004	1.081	1.004		1.797	1.025	1.005	0.999	1.068	1.004
							T = 100						
1	1.717	1.039	1.013	1.017	1.110	1.005		1.714	1.046	1.016	1.015	1.102	1.006
2	2.027	1.004	0.995	0.983	1.045	0.988		2.112	0.993	0.990	0.964	1.007	0.982
3	1.755	1.022	1.007	1.007	1.085	0.999		1.727	1.033	1.011	1.009	1.088	1.001
4	2.352	0.919	0.945	0.902	0.919	0.945		2.464	0.911	0.943	0.892	0.888	0.944
5	2.040	1.017	1.007	1.003	1.050	0.997		2.058	1.014	1.008	0.987	1.017	0.993
6	5.191	1.072	1.027	1.060	1.264	1.019		5.164	1.070	1.026	1.055	1.243	1.019
7	0.594	1.149	1.059	1.031	1.091	1.020		0.597	1.181	1.073	1.033	1.091	1.023
8	1.774	1.024	1.005	1.006	1.084	1.005		1.802	1.023	1.005	0.998	1.065	1.003
9	1.798	1.028	1.009	1.006	1.084	1.001		1.797	1.031	1.011	1.002	1.073	1.001
							T = 150						
1	1.723	1.042	1.016	1.020	1.117	1.004		1.720	1.050	1.019	1.019	1.110	1.005
2	2.031	0.999	0.991	0.978	1.042	0.986		2.121	0.986	0.983	0.957	0.999	0.981
3	1.750	1.022	1.007	1.008	1.084	0.998		1.721	1.032	1.012	1.010	1.086	0.999
4	2.411	0.894	0.921	0.878	0.891	0.941		2.526	0.887	0.919	0.869	0.862	0.941
5	2.030	1.018	1.010	1.006	1.053	0.998		2.046	1.016	1.012	0.990	1.019	0.996
6	5.172	1.078	1.033	1.066	1.271	1.015		5.145	1.075	1.032	1.060	1.250	1.014
7	0.584	1.157	1.071	1.041	1.103	1.014		0.586	1.194	1.092	1.049	1.109	1.019
8	1.767	1.031	1.011	1.010	1.092	1.000		1.795	1.032	1.012	1.003	1.073	0.999
9	1.788	1.029	1.011	1.008	1.087	1.000		1.786	1.034	1.012	1.004	1.075	0.999

See the notes to Table 2.

Table 5. Relative RMSFE values when revisions are noise

	Final value												
	EOS							RTV					
Exp.	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW	*T* = 50	RMSFE	m = 20	m = 40	EWMAS	EWMAA	AveW
1	1.576	1.037	1.007	1.014	1.121	1.010		1.574	1.044	1.009	1.011	1.110	1.013
2	1.829	1.003	0.997	0.980	1.047	0.988		1.922	0.994	0.994	0.961	1.003	0.981
3	1.656	1.019	1.003	1.005	1.089	1.000		1.628	1.028	1.005	1.006	1.091	1.004
4	2.130	0.937	0.977	0.917	0.923	0.945		2.251	0.929	0.976	0.907	0.891	0.942
5	1.998	1.007	1.000	0.993	1.033	0.993		2.037	1.000	0.998	0.973	0.989	0.985
6	4.823	1.067	1.013	1.060	1.291	1.035		4.785	1.065	1.012	1.054	1.266	1.033
7	0.576	1.122	1.028	1.008	1.064	1.029		0.576	1.160	1.038	1.011	1.071	1.041
8	1.637	1.028	1.004	1.006	1.095	1.005		1.670	1.029	1.004	0.997	1.075	1.003
9	1.665	1.025	1.004	1.003	1.088	1.004		1.663	1.028	1.005	0.997	1.074	1.004
							T = 100						
1	1.560	1.048	1.017	1.021	1.133	1.007		1.555	1.057	1.021	1.019	1.125	1.008
2	1.834	0.997	0.990	0.971	1.039	0.982		1.935	0.984	0.984	0.950	0.994	0.976
3	1.640	1.024	1.008	1.007	1.094	0.998		1.608	1.037	1.013	1.010	1.099	1.001
4	2.238	0.895	0.931	0.876	0.882	0.933		2.363	0.890	0.932	0.870	0.852	0.934
5	1.995	1.014	1.007	0.997	1.037	0.996		2.028	1.009	1.007	0.979	0.999	0.992
6	4.704	1.087	1.032	1.073	1.313	1.024		4.675	1.085	1.032	1.067	1.290	1.023
7	0.547	1.174	1.070	1.036	1.107	1.023		0.545	1.220	1.090	1.044	1.117	1.030
8	1.635	1.027	1.005	1.006	1.095	1.005		1.667	1.027	1.005	0.997	1.073	1.003
9	1.665	1.030	1.010	1.005	1.092	1.000		1.663	1.035	1.013	1.001	1.080	1.000
							T = 150						
1	1.563	1.050	1.018	1.023	1.139	1.004		1.557	1.060	1.023	1.022	1.133	1.005
2	1.843	0.991	0.985	0.965	1.035	0.980		1.950	0.976	0.977	0.942	0.983	0.975
3	1.636	1.024	1.008	1.008	1.093	0.997		1.603	1.037	1.014	1.011	1.096	0.999
4	2.308	0.866	0.903	0.849	0.848	0.930		2.437	0.862	0.905	0.843	0.822	0.932
5	1.986	1.015	1.010	0.999	1.039	0.998		2.018	1.011	1.011	0.981	0.999	0.995
6	4.685	1.094	1.040	1.080	1.323	1.018		4.656	1.091	1.038	1.073	1.299	1.017
7	0.536	1.185	1.084	1.049	1.121	1.017		0.534	1.233	1.111	1.061	1.135	1.023
8	1.626	1.034	1.012	1.010	1.102	1.000		1.659	1.036	1.013	1.001	1.080	0.998
9	1.651	1.032	1.012	1.006	1.094	0.999		1.649	1.037	1.014	1.002	1.082	0.999

See the notes to Table 2.

2.5 Empirical application

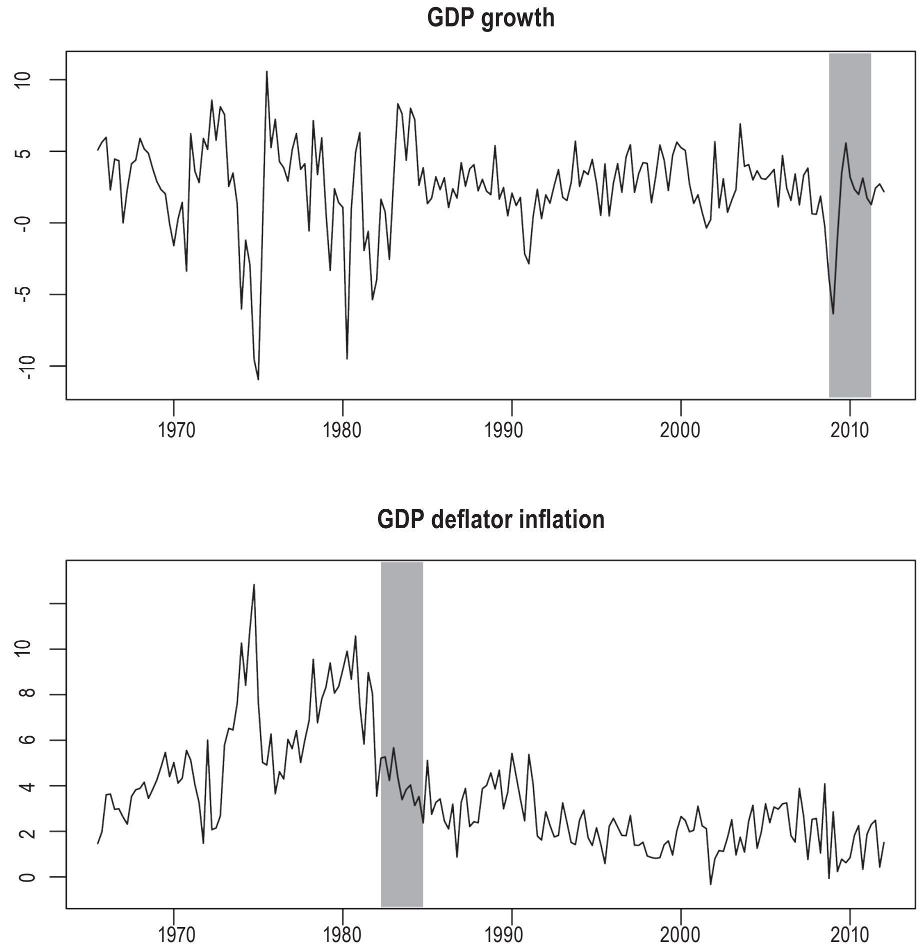
In this section, we compare the forecasting performance of the alternative window selection methods discussed above using actual U.S. data. We consider one-step ahead forecasts of real GDP and GDP deflator inflation (at an annualized rate). All forecasts are out-of-sample. In other words, at each forecast origin $t+1$, the $t+1$ vintage estimates of data up to period t are used to estimate the parameters of a forecasting model that is then used to generate a forecast for period $t+1$. All real-time data is quarterly and the sample period runs from 1965:Q4 to 2012:Q2. Different vintages of real GDP and GDP deflator series are obtained from the Federal Reserve Bank of Philadelphia's real-time database.

The goal of our application is to compare the different forecasting performances in the presence of a recent break. As discussed in the Introduction, structural break tests provide inaccurate estimates of the timing of the break(s). Therefore, a problem that arises in this analysis is how to select the relevant forecasting periods. To this end we consider the following strategy. In Figure 1, we plot the first-release quarterly growth rates of real GDP and GDP deflator over the 1965:Q4–2012:Q2 period. A time period is considered as a starting point of a forecasting period if the latest available observation differs considerably from the earlier ones. Our approach suggests that 2008:Q3 is a potential break point in the dynamics of the GDP growth. As a result, the GDP forecasts are made for the period 2008:Q4–2011:Q1. On the other hand, we find that the behavior of the inflation series changed after 1982:Q1 and hence the GDP deflator inflation forecasts are made for the period 1982:Q2–1984:Q3.

The performance of the various window selection methods compared to the expanding window benchmark is summarized in Table 6. Panel A shows the results for the real GDP forecasts, whereas Panel B has the inflation forecasts. The first row in both Panels provides the RMSFE value of the benchmark expanding window estimator. The subsequent rows show the RMSFE of a candidate window selection method relative to the RMSFE of the benchmark. Forecasts of final values are bias corrected first-release forecasts. The correction is based on the sample mean of the difference between the final values, y_t^{t+15}, and the first-release values, y_t^{t+1}, calculated with data up to the forecast origin. We use y_{t+1}^{t+16} as true values for inflation and the vintage 2012:Q2 values as true values for real GDP growth. To ensure that our empirical results are comparable to our Monte Carlo results, we consider an AR(1) specification.[5]

5 We also considered AR(2) and AR(4) models. The results for these specifications are qualitatively similar to those presented in Table 6.

Figure 1. First-release growth rates



Notes: The figure depicts the quarterly growth rates of real GDP and GDP deflator (annualized) over the 1965:Q4–2012:Q2 period. The shaded areas denote out-of-sample forecasting periods.

Table 6. Out-of-sample relative RMSFE values

A. GDP growth

	First-release		Final value	
	EOS	RTV	EOS	RTV
Expanding window	3.219	2.776	4.419	3.967
m = 20	1.081	1.210	1.031	1.128
m = 40	1.047	1.120	1.003	1.077
EWMAS	1.033	1.166	1.002	1.103
EWMAA	1.052	1.273	1.019	1.157
AveW	0.998	1.084	0.991	1.056

B. GDP deflator inflation

	First-release		Final value	
	EOS	RTV	EOS	RTV
Expanding window	1.003	0.961	1.559	1.484
m = 20	1.529	1.502	1.460	1.488
m = 40	1.347	1.328	1.293	1.315
EWMAS	1.104	1.115	1.123	1.149
EWMAA	1.426	1.075	1.280	1.143
AveW	1.217	1.202	1.212	1.227

Notes: Forecasting periods for real GDP growth and GDP deflator inflation are 2008:Q4–2011:Q1 and 1982:Q2–1984:Q3, respectively. The first row in each panel shows the root mean squared forecast error for the expanding window estimator. Subsequent rows show the ratio of the RMSFE of a candidate window selection method to the RMSFE of the benchmark expanding window estimator. Forecasts of final values are bias corrected first-release forecasts. The correction is based on the sample mean of the difference between the final values, y_t^{t+15}, and the first-release values, y_t^{t+1}, calculated with data up to the forecast origin. We use y_{t+1}^{t+16} as true values for GDP deflator inflation and the vintage 2012:Q2 values as true values for real GDP growth.

The AveW and the expanding window estimator produce the most accurate real GDP forecasts. When we use the EOS approach, the AveW method does marginally better than the expanding window estimator. By contrast, the expanding window estimator turns out to be the best method when the RTV approach is used. For the GDP deflator inflation, the expanding window estimator is overwhelmingly the best estimation window method. It produces the most accurate forecasts in each of the four dependent variable/vintage approach combinations considered here. The differences in the forecasting abilities are very large. The relative RMSFE values range between 1.075 and 1.529, indicating that the alternative window selection methods produce 7.5–52.9 percent larger forecast errors than the expanding window benchmark.

Our simulation results are useful in explaining why it is difficult to outperform the expanding window estimator after a recent break. For example, if the break only affects the innovation variance, σ^2, our simulation results indicate that the expanding window estimator produces the most accurate forecasts. The two breaks considered here most likely caused changes in the innovation variance. In particular, the results in the literature indicate that the variance of the inflation series has reduced substantially since the early 1980s. This would explain why none of the alternative methods systematically improve upon the expanding window benchmark. Another reason for the good performance of the expanding window estimator lies in the fact that the means of the series have declined after the breaks (at least temporarily). The simulation results show that when the mean declines after the break, the expanding window estimator performs well relative to the alternatives (see experiments 3 and 9). Note also that the differences in the relative predictive abilities are larger for the GDP deflator inflation. As discussed in Section 2, revisions to the GDP deflator inflation are mainly noise, whereas those to the GDP are mainly news (see, e.g., Clements and Galvão, 2013). Thus, our results suggest that the differences in the relative predictive abilities are larger when the revisions reduce noise. In addition, our results indicate that, in general, the expanding window method performs better when the first-release values are the ones to be forecast. These two findings are consistent with our simulation results in Tables 2–5.

The rolling window methods and the EWMAA method perform poorly in our empirical applications. In particular, a short rolling window typically produces forecasts that are substantially worse than those produced by the expanding window benchmark. These empirical findings are in line with our Monte Carlo simulations. Indeed, our simulation results suggest that these methods rarely outperform the expanding window estimator.

The results in Table 6 also indicate that one key determinant of the forecasting performance is the choice of how to use the real-time data to estimate the parameters of the forecasting model. A substantial amount of the literature on real-time forecasting uses the EOS approach. In our empirical examples, the RTV approach produces more accurate forecasts after a recent break regardless of whether we consider forecasting the real GDP or the GDP deflator inflation. We find that the RTV approach yields improvements of 4.8%–13.8% over the EOS approach.

2.6 Conclusions

This paper analyzes the forecasting performance of various window selection methods after a recent break when the data are subject to revision. Several practical recommendations for choosing the estimation window emerge from our analysis. First, our Monte Carlo and empirical results suggest that the expanding window method usually provides the most accurate forecasts after a recent break. It performs well regardless of whether the revisions add news or reduce noise, or whether we forecast the first-release or the final values. Thus, the evidence in favor of the expanding window estimator seems well established. Second, we find that rolling windows perform the worst of all the methods. They never produce the most accurate forecasts in any of the cases considered here. Furthermore, they rarely improve upon the expanding window estimator. This is an important result because rolling windows are used extensively in the literature. In short, our results suggest that the use of rolling windows should be rethought, at least when making forecasts after a recent break. Third, our results imply that whether the revisions add news or reduce noise does not matter much for the relative ranking of the alternative window selection methods. Finally, no clear ranking between the EOS and RTV vintage approaches emerges. In general, our Monte Carlo results suggest that the RTV approach produces more accurate forecasts when the revisions add news, whereas the EOS approach yields more reliable forecasts when the revisions reduce noise. The RTV approach performs particularly well in our empirical examples.

Our results could be extended in several ways. We have considered only cases where the autoregressive process has been subject to a single, recent break. In practice, however, autoregressive processes are likely to be subject to multiple breaks. Therefore, analyzing the forecasting performance in the presence of multiple breaks might be a fruitful area for future research. In addition, our statistical framework neglects some important features of the actual data revision process, including time variations in the revision mean and variance. Incorporating these features into the statistical framework may lead to a better understanding of the relative forecasting accuracy of alternative window selection methods in the presence of data revisions.

Appendix A

In this Appendix, we derive formulas for the means and variances of the first-release data, y_t^{t+1}, and final data, $\tilde{y}_t$. Recall that $\tilde{y}_t = \rho + \sum_{i=1}^{l} \mu_{v_i} + \beta\tilde{y}_{t-1} + \sigma\eta_{1t} + \sum_{i=1}^{l} \sigma_{v_i}\eta_{2t,i}$ and $y_t^{t+1} = \tilde{y}_t - \sum_{i=1}^{l} \mu_{v_i} - \sum_{i=1}^{l} \sigma_{v_i}\eta_{2t,i} - \mu_{\varepsilon_1} + \sigma_{\varepsilon_1}\eta_{3t,1}$. Both y_t^{t+1} and $\tilde{y}_t$ are (covariance) stationary processes. We set $l = 14$, so that we observe 14 different estimates of y_t before the true value, $\tilde{y}_t$, is observed. The expected value of $\tilde{y}_t$ is

$$E(\tilde{y}_t) = \mu_{\tilde{y}} = \frac{\rho + \sum_{i=1}^{l} \mu_{v_i}}{1-\beta}.$$

Therefore, the expected value of y_t^{t+1} is

$$\begin{aligned} E(y_t^{t+1}) &= E(\tilde{y}_t) - \sum_{i=1}^{l} \mu_{v_i} - \mu_{\varepsilon_1} \\ &= \frac{\rho + \beta\sum_{i=1}^{l} \mu_{v_i}}{1-\beta} - \mu_{\varepsilon_1}. \end{aligned}$$

If the revisions are pure news, the expected values of the first-release and final data are

$$E(\tilde{y}_t) = \frac{\rho + \sum_{i=1}^{l} \mu_{v_i}}{1-\beta} \quad \text{and} \quad E(y_t^{t+1}) = \frac{\rho + \beta\sum_{i=1}^{l} \mu_{v_i}}{1-\beta}.$$

If the revisions are pure noise, the expected values of the first-release and final data are

$$E(\tilde{y}_t) = \frac{\rho}{1-\beta} \quad \text{and} \quad E(y_t^{t+1}) = \frac{\rho}{1-\beta} - \mu_{\varepsilon_1}.$$

The revisions are defined by $r_t^i = y_t^{t+1+i} - y_t^{t+i}$, for $i = 1,...,l$. For example, the first revision at time t is equal to $r_t^1 = y_t^{t+2} - y_t^{t+1}$, i.e., the difference between the second-release value and the first-release value. Equations (1), (2), and (3) imply that $y_t^{t+1} = \rho + \beta\tilde{y}_{t-1} + \sigma\eta_{1t} - \mu_{\varepsilon_1} + \sigma_{\varepsilon_1}\eta_{3t,1}$ and $y_t^{t+2} = \rho + \mu_{v_1} + \beta\tilde{y}_{t-1} + \sigma\eta_{1t} + \sigma_{v_1}\eta_{2t,1} - \mu_{\varepsilon_2} + \sigma_{\varepsilon_2}\eta_{3t,2}$.

Hence,

$$r_t^1 = y_t^{t+2} - y_t^{t+1} = \mu_{v_1} + \sigma_{v_1}\eta_{2t,1} - \mu_{\varepsilon_2} + \sigma_{\varepsilon_2}\eta_{3t,2} + \mu_{\varepsilon_1} - \sigma_{\varepsilon_1}\eta_{3t,1}.$$

Following Clements and Galvão (2013), we assume that the first and the fifth revisions have non-zero mean. To be more specific, we assume that the means of the first and fifth revisions are, respectively, δ and $\delta/2$ times the mean of the first-release data. In what follows, we set $\delta = 0.04$. Our assumptions imply that for news revisions,

$$\begin{aligned} E(r_t^1) &= \mu_{v_1} \\ E(r_t^2) &= \mu_{v_2} \\ &\vdots \\ E(r_t^{14}) &= \mu_{v_{14}}, \end{aligned}$$

so that $\mu_{v_2} = \mu_{v_3} = \mu_{v_4} = \mu_{v_6} = ... = \mu_{v_{14}} = 0$, $E(r_t^1) = \mu_{v_1}$ and $E(r_t^5) = \mu_{v_5}$. Setting $E(r_t^1) = \delta E(y_t^{t+1})$ yields

$$\mu_{v_1} = \delta \frac{\rho + \beta \sum_{i=1}^{l} \mu_{v_i}}{1 - \beta}.$$

Using the fact that $E(r_t^1) = 2E(r_t^5)$, i.e., $\mu_{v_1} = 2\mu_{v_5}$, we can express μ_{v_1} and μ_{v_5} as

$$\mu_{v_1} = \frac{\delta\rho}{1 - (1 + 1.5\delta)\beta} \quad \text{and} \quad \mu_{v_5} = \frac{\mu_{v_1}}{2}.$$

The situation is more complicated if the revisions are pure noise. The structure of the DGP and our assumptions imply that

$$
\begin{aligned}
E(r_t^1) &= -\mu_{\varepsilon_2} + \mu_{\varepsilon_1} = \delta E(y_t^{t+1}) \\
E(r_t^2) &= -\mu_{\varepsilon_3} + \mu_{\varepsilon_2} = 0 \\
E(r_t^3) &= -\mu_{\varepsilon_4} + \mu_{\varepsilon_3} = 0 \\
E(r_t^4) &= -\mu_{\varepsilon_5} + \mu_{\varepsilon_4} = 0 \\
E(r_t^5) &= -\mu_{\varepsilon_6} + \mu_{\varepsilon_5} = \frac{\delta}{2} E(y_t^{t+1}) \\
E(r_t^6) &= -\mu_{\varepsilon_7} + \mu_{\varepsilon_6} = 0 \\
&\vdots \\
E(r_t^{13}) &= -\mu_{\varepsilon_{14}} + \mu_{\varepsilon_{13}} = 0 \\
E(r_t^{14}) &= \mu_{\varepsilon_{14}} = 0.
\end{aligned}
$$

Revisions 6–14 have zero mean, which implies that $\mu_{\varepsilon_6} = \mu_{\varepsilon_7} = ... = \mu_{\varepsilon_{13}} = \mu_{\varepsilon_{14}} = 0$. Because $\mu_{\varepsilon_6} = 0$, μ_{ε_5} equals $\frac{\delta}{2} E(y_t^{t+1})$. This finding implies that $\mu_{\varepsilon_2} = \mu_{\varepsilon_3} = \mu_{\varepsilon_4} = \mu_{\varepsilon_5} = \frac{\delta}{2} E(y_t^{t+1})$. Finally, we find that $\mu_{\varepsilon_1} = \frac{3\delta}{2} E(y_t^{t+1})$. So, if the revisions are pure noise,

$$
\mu_{\varepsilon_1} = \frac{1.5}{(1+1.5\delta)} \frac{\delta\rho}{(1-\beta)}, \quad \mu_{\varepsilon_2} = ... = \mu_{\varepsilon_5} = \frac{\delta}{2(1+1.5\delta)} \frac{\rho}{(1-\beta)}, \quad \mu_{\varepsilon_6} = ... = \mu_{\varepsilon_{14}} = 0.
$$

Next, we derive the variance of $\tilde{y}_t$. The true values can be expressed as follows

$$
(\tilde{y}_t - \mu_{\tilde{y}}) = \beta(\tilde{y}_{t-1} - \mu_{\tilde{y}}) + \sigma\eta_{1,t} + \sum_{i=1}^{l} \sigma_{v_i} \eta_{2t,i}, \tag{7}
$$

where $\mu_{\tilde{y}}$ denotes the expected value of $\tilde{y}_t$. The variance of $\tilde{y}$ can be found by multiplying (7) by $(\tilde{y}_t - \mu_{\tilde{y}})$ and taking expectations:

$$
E(\tilde{y}_t - \mu_{\tilde{y}})^2 = \beta E\left[(\tilde{y}_t - \mu_{\tilde{y}})(\tilde{y}_{t-1} - \mu_{\tilde{y}})\right] + E\left[(\tilde{y}_t - \mu_{\tilde{y}})\sigma\eta_{1t}\right] + E\left[(\tilde{y}_t - \mu_{\tilde{y}}) \sum_{i=1}^{l} \sigma_{v_i} \eta_{2t,i}\right]. \tag{8}
$$

Note that

$$E\left[(\tilde{y}_t - \mu_{\tilde{y}})\sigma\eta_{1t}\right] = \sigma^2 E(\eta_{1t}^2) = \sigma^2 \qquad \text{and}$$

$$E\left[(\tilde{y}_t - \mu_{\tilde{y}})\sum_{i=1}^{l}\sigma_{v_i}\eta_{2t,i}\right] = \sum_{i=1}^{l}\sigma_{v_i}^2 E(\eta_{2t,i}^2) = \sum_{i=1}^{l}\sigma_{v_i}^2.$$

Thus, (8) can be rewritten as

$$\gamma_0 = \beta\phi_1\gamma_0 + \sigma^2 + \sum_{i=1}^{l}\sigma_{v_i}^2, \tag{9}$$

where γ_0 denotes the variance and ϕ_1 the first autocorrelation coefficient. Using the fact that for an AR(1) process, $\phi_1 = \beta$, we have

$$\gamma_0 = \frac{\sigma^2 + \sum_{i=1}^{l}\sigma_{v_i}^2}{1-\beta^2}.$$

The variance of y_t^{t+1} can be derived as follows

$$var(y_t^{t+1}) = var(\tilde{y}_t - \sum_{i=1}^{l}\mu_{v_i} - \sum_{i=1}^{l}\sigma_{v_i}\eta_{2t,i} - \mu_{\varepsilon_1} + \sigma_{\varepsilon_1}\eta_{3t,1})$$

$$var(y_t^{t+1}) = var(\tilde{y}_t) + \sum_{i=1}^{l}\sigma_{v_i}^2 var(\eta_{2t,i}) + \sigma_{\varepsilon_1}^2 var(\eta_{3t,1}) - 2\sum_{i=1}^{l}\sigma_{v_i} cov(\tilde{y}_t, \eta_{2t,i})$$
$$+2\sigma_{\varepsilon_1} cov(\tilde{y}_t, \eta_{3t,1}) - 2\sum_{i=1}^{l}\sigma_{v_i}\sigma_{\varepsilon_1} cov(\eta_{2t,i}, \eta_{3t,1}).$$

Because $cov(\tilde{y}_t, \eta_{2t,i}) = \sum_{i=1}^{l}\sigma_{v_i}$, $cov(\tilde{y}_t, \eta_{3t,1}) = 0$, and $cov(\eta_{2t,i}, \eta_{3t,1}) = 0$, we have

$$var(y_t^{t+1}) = var(\tilde{y}_t) + \sum_{i=1}^{l}\sigma_{v_i}^2 + \sigma_{\varepsilon_1}^2 - 2\sum_{i=1}^{l}\sigma_{v_i}^2$$

$$= \frac{\sigma^2 + \sum_{i=1}^{l} \sigma_{v_i}^2}{1-\beta^2} - \sum_{i=1}^{l} \sigma_{v_i}^2 + \sigma_{\varepsilon_1}^2$$

$$= \frac{\sigma^2 + \beta^2 \sum_{i=1}^{l} \sigma_{v_i}^2}{1-\beta^2} + \sigma_{\varepsilon_1}^2.$$

Therefore, when the revisions are pure news, we have

$$\sigma_{y_t^{t+1}}^2 = \frac{\sigma^2 + \beta^2 \sum_{i=1}^{l} \sigma_{v_i}^2}{1-\beta^2}.$$

When the revisions are pure noise, the variance is

$$\sigma_{y_t^{t+1}}^2 = \frac{\sigma^2}{1-\beta^2} + \sigma_{\varepsilon_1}^2.$$

Next, we derive the variances of the data revisions. Let $\sigma_{r_i}^2$ (for i = 1,...,l) denote the variance of the ith revision. The variance of the first revision is

$$var(r_t^1) = var(y_t^{t+2} - y_t^{t+1}) = var(\mu_{v_1} + \sigma_{v_1}\eta_{2t,1} - \mu_{\varepsilon_2} + \sigma_{\varepsilon_2}\eta_{3t,2} + \mu_{\varepsilon_1} - \sigma_{\varepsilon_1}\eta_{3t,1}).$$

If the revisions are pure news, $var(r_t^1) = \sigma_{r_1}^2 = var(\mu_{v_1} + \sigma_{v_1}\eta_{2t,1}) = \sigma_{v_1}^2 var(\eta_{2t,1}) = \sigma_{v_1}^2$. If the revisions are pure noise, $var(r_t^1) = \sigma_{r_1}^2 = var(-\mu_{\varepsilon_2} + \sigma_{\varepsilon_2}\eta_{3t,2} + \mu_{\varepsilon_1} - \sigma_{\varepsilon_1}\eta_{3t,1}) = \sigma_{\varepsilon_2}^2 var(\eta_{3t,2}) + \sigma_{\varepsilon_1}^2 var(\eta_{3t,1}) = \sigma_{\varepsilon_2}^2 + \sigma_{\varepsilon_1}^2$.

We set $\sigma_{r_1} = \alpha\sigma_{y_t^{t+1}}$, where α denotes the ratio of the standard deviation of the first revision to the standard deviation of the first-release data. Furthermore, we assume that $\sigma_{r_2,\ldots,r_{13}} = \frac{\alpha}{2}\sigma_{y_t^{t+1}}$ and that $\sigma_{r_{14}} = \frac{\alpha}{4}\sigma_{y_t^{t+1}}$. In what follows, we set $\alpha = 0.4$. Thus, the variance of the first revision, when the revisions are pure news, can be found by solving the equation

$$\sigma_{v_1}^2 = \alpha^2 \frac{\sigma^2 + \beta^2 \sum_{i=1}^{l} \sigma_{v_i}^2}{1-\beta^2}.$$

Note that $1/4\sigma^2_{v_1} = \sigma^2_{v_2} = ... = \sigma^2_{v_{13}}$ and $1/16\sigma^2_{v_1} = \sigma^2_{v_{14}}$. This implies that $\sum_{i=1}^{l} \sigma^2_{v_i} = 4.0625\sigma^2_{v_1}$. Using this fact we can express the variance of the first revision as

$$\sigma^2_{v_1} = \alpha^2 \frac{\sigma^2 + 4.0625\beta^2\sigma^2_{v_1}}{1-\beta^2}.$$

After some algebra, we find that

$$\sigma^2_{v_1} = \frac{\alpha^2\sigma^2}{1-(1+4.0625\alpha^2)\beta^2}.$$

So, the formulas for the standard deviations are

$$\sigma_{v_1} = \sqrt{\frac{\alpha^2\sigma^2}{1-(1+4.0625\alpha^2)\beta^2}}, \quad \sigma_{v_2} = ... = \sigma_{v_{13}} = \sigma_{v_1}/2, \quad \sigma_{v_{14}} = \sigma_{v_1}/4.$$

Next, we consider noise revisions. We have

$$\begin{aligned}
\sigma^2_{r_1} &= \sigma^2_{\varepsilon_2} + \sigma^2_{\varepsilon_1} \\
\sigma^2_{r_2} &= \sigma^2_{\varepsilon_3} + \sigma^2_{\varepsilon_2} \\
&\vdots \\
\sigma^2_{r_{13}} &= \sigma^2_{\varepsilon_{14}} + \sigma^2_{\varepsilon_{13}} \\
\sigma^2_{r_{14}} &= \sigma^2_{\varepsilon_{14}}.
\end{aligned}$$

Using the fact that revisions 2–13 have equal variance, we find that

$$\sigma^2_{\varepsilon_2} = \sigma^2_{\varepsilon_4} = ... = \sigma^2_{\varepsilon_{12}} = \sigma^2_{\varepsilon_{14}} \quad \text{and} \quad \sigma^2_{\varepsilon_3} = \sigma^2_{\varepsilon_5} = ... = \sigma^2_{\varepsilon_{13}}.$$

Note that $\sigma_{r_{13}} = \alpha/2\sigma_{y_t^{t+1}}$ and $\sigma_{r_{14}} = \alpha/4\sigma_{y_t^{t+1}}$, implying that $4\sigma^2_{r_{14}} = \sigma^2_{r_{13}}$. Therefore,

$$4\sigma^2_{\varepsilon_{14}} = \sigma^2_{\varepsilon_{14}} + \sigma^2_{\varepsilon_{13}},$$

which in turn implies that $\sigma^2_{\varepsilon_{13}} = 3\sigma^2_{\varepsilon_{14}}$. Plugging $\sigma^2_{\varepsilon_2} = ... = \sigma^2_{\varepsilon_{14}} = (\frac{\alpha}{4})^2 \left[\frac{\sigma^2}{1-\beta^2} + \sigma^2_{\varepsilon_1}\right]$ into

$\sigma^2_{\varepsilon_2} + \sigma^2_{\varepsilon_1} = \alpha^2 \sigma^2_{y_t^{t+1}}$ yields

$$\frac{\alpha^2}{16}\left[\frac{\sigma^2}{1-\beta^2} + \sigma^2_{\varepsilon_1}\right] + \sigma^2_{\varepsilon_1} = \alpha^2\left[\frac{\sigma^2}{1-\beta^2} + \sigma^2_{\varepsilon_1}\right].$$

After some algebra, we find that

$$\sigma^2_{\varepsilon_1} = \frac{15\alpha^2}{16-15\alpha^2}\frac{\sigma^2}{1-\beta^2}.$$

So, the formulas for the standard deviations are

$$\sigma_{\varepsilon_1} = \sqrt{\frac{15\alpha^2}{16-15\alpha^2}\frac{\sigma^2}{1-\beta^2}},$$

$$\sigma_{\varepsilon_2} = \sigma_{\varepsilon_4} = ... = \sigma_{\varepsilon_{14}} = \sqrt{\left(\frac{\alpha}{4}\right)^2 \frac{16}{16-15\alpha^2}\frac{\sigma^2}{1-\beta^2}},$$

$$\sigma_{\varepsilon_3} = \sigma_{\varepsilon_5} = ... = \sigma_{\varepsilon_{13}} = \sqrt{3\left(\frac{\alpha}{4}\right)^2 \frac{16}{16-15\alpha^2}\frac{\sigma^2}{1-\beta^2}}.$$

Appendix B

Means and standard deviations

News Experiment	$E(\tilde{y}_{1t})$	$E(\tilde{y}_{2t})$	$E(y_{1t}^{t+1})$	$E(y_{2t}^{t+1})$	$\sigma_{\tilde{y}_{1t}}$	$\sigma_{\tilde{y}_{2t}}$	$\sigma_{y_{1t}^{t+1}}$	$\sigma_{y_{2t}^{t+1}}$
1	2.255	2.255	2.128	2.128	2.514	2.514	1.957	1.957
2	2.255	5.171	2.128	4.878	2.514	7.187	1.957	5.595
3	2.255	1.442	2.128	1.361	2.514	2.035	1.957	1.584
4	1.442	5.171	1.361	4.878	2.035	7.187	1.584	5.595
5	5.171	1.442	4.878	1.361	7.187	2.035	5.595	1.584
6	2.255	2.255	2.128	2.128	2.514	7.541	1.957	5.871
7	2.255	2.255	2.128	2.128	2.514	0.838	1.957	0.652
8	2.255	3.383	2.128	3.191	2.514	2.514	1.957	1.957
9	2.255	1.128	2.128	1.064	2.514	2.514	1.957	1.957
Noise Experiment	$E(\tilde{y}_{1t})$	$E(\tilde{y}_{2t})$	$E(y_{1t}^{t+1})$	$E(y_{2t}^{t+1})$	$\sigma_{\tilde{y}_{1t}}$	$\sigma_{\tilde{y}_{2t}}$	$\sigma_{y_{1t}^{t+1}}$	$\sigma_{y_{2t}^{t+1}}$
1	2.000	2.000	1.887	1.887	1.732	1.732	1.879	1.879
2	2.000	4.000	1.887	3.774	1.732	2.268	1.879	2.460
3	2.000	1.333	1.887	1.258	1.732	1.549	1.879	1.680
4	1.333	4.000	1.258	3.774	1.549	2.268	1.680	2.460
5	4.000	1.333	3.774	1.258	2.268	1.549	2.460	1.680
6	2.000	2.000	1.887	1.887	1.732	5.196	1.879	5.636
7	2.000	2.000	1.887	1.887	1.732	0.577	1.879	0.626
8	2.000	3.000	1.887	2.830	1.732	1.732	1.879	1.879
9	2.000	1.000	1.887	0.943	1.732	1.732	1.879	1.879

References

Altissimo, F. and Corradi, V. (2003). Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics*, 117, 207–244.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61, 821–856.

Andrews, D. W. K., Lee, I. and Ploberger, W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, 70, 9–38.

Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, 40, 319–340.

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

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1–22.

Clements, M. P. and Galvão, A. B. (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28, 458–477.

Clements, M. P. and Hendry, D. F. (1998). Forecasting economic time series. Cambridge University Press: Cambridge.

Clements, M. P. and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C. W. J. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 1. Amsterdam: Elsevier, 605–657.

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

Eklund, J., Kapetanios, G. and Price, S. (2013). Robust forecast methods and monitoring during structural change. *Manchester School*, 81, 3–27.

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

Giraitis, L., Kapetanios, G. and Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics*, 177, 153–170.

Jacobs, J. P. A. M. and van Norden, S. (2011). Modeling data revisions: measurement error and dynamics of true values. *Journal of Econometrics*, 161, 101–109.

Koenig, E. F., Dolmas, S. and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics*, 85, 618–628.

Mankiw, N. G. and Shapiro, M. D. (1986). News or noise: an analysis of GNP revisions. *Survey of Current Business*, 66, 20–25.

McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980's? *American Economic Review*, 90, 1464–1476.

Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business and Economic Statistics*, 29, 307–318.

Pesaran, M. H., Pick, A. and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177, 134–152.

Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129, 183–217.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137, 134–161.

Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier, 1203–1324.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14, 11–30.

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

Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C. W. J. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 1. Amsterdam: Elsevier, 135–196.

Chapter 3
Multi-step forecasting in the presence of breaks*

Jari Hännikäinen

Abstract

This paper analyzes the relative performance of multi-step forecasting methods in the presence of breaks and data revisions. Our Monte Carlo simulations indicate that the type and the timing of the break affect the relative accuracy of the methods. The iterated method typically performs the best in unstable environments, especially if the parameters are subject to small breaks. This result holds regardless of whether data revisions add news or reduce noise. Empirical analysis of real-time U.S. output and inflation series shows that the alternative multi-step methods only episodically improve upon the iterated method.

Keywords: Structural breaks, multi-step forecasting, intercept correction, real-time data

JEL codes: C22, C53, C82

* I would like to thank Henri Nyberg, Jukka Lahtonen, Sami Oinonen, Jari Vainiomäki and participants of the FDPE Econometrics Workshop in Helsinki, Hecer Time Series Econometrics seminar in Helsinki, Allecon seminar in Jyväskylä, and the XXXVI Annual Meeting of the Finnish Economic Association in Kuopio for helpful comments and discussions. Financial support from the FDPE is gratefully acknowledged.

3.1 Introduction

The medium- and long-term prospects of the economy are important for consumers, investors, and policymakers. For example, it is well known that monetary policy affects the economy with a long lag. As a result, central banks conduct forward-looking monetary policy, i.e., central banks' interest rate decisions are based on their forecasts of future output growth, unemployment, and inflation. Given the importance of the medium- and long-term economic outlook, economists provide forecasts of key macroeconomic time series several periods ahead in time. These macroeconomic series are often serially correlated, implying that their own past values are themselves useful predictors. Therefore, autoregressive (AR) models are used extensively in economic forecasting. Despite their parsimonious form, it appears to be difficult to outperform AR models in practice (see, e.g., Elliott and Timmermann, 2008; Rossi, 2013; Stock and Watson, 2003).

When generating a multi-step forecast, a forecaster has to decide whether to use the iterated or direct forecasting strategy. In the iterated approach, forecasts are made using a one-period ahead model, iterated forward for the desired number of periods. A central feature of the iterated approach is that the model specification is the same regardless of the forecast horizon. Direct forecasts, on the other hand, are made using a horizon-specific model. Thus, a forecaster estimates a separate model for each forecast horizon. The theoretical literature analyzing the relative merits of the iterated versus the direct forecast methods includes, e.g., Bao (2007), Brown and Mariano (1989), Chevillon and Hendry (2005), Clements and Hendry (1996b, 1998), Findley (1985), Hoque *et al.* (1988), Ing (2003), Schorfheide (2005), and Weiss (1991). This literature emphasizes that the choice between iterated and direct multi-step forecasts is not clear cut, but rather involves a trade-off between bias and estimation variance. The iterated method uses the largest available data sample in the estimation and thus produces more efficient parameter estimates than the direct method. In contrast, direct forecasts are more robust to model misspecification. Which element, the bias or the estimation variance, dominates in the composition of the mean squared forecast error (MSFE) values in practice depends on the sample size, the forecast horizon, and the (unknown) underlying DGP, and therefore the question of which method to use cannot be decided ex ante on theoretical grounds alone. Hence, the question of which multi-step forecasting method to use is an empirical one. In their empirical analysis of 170 U.S. monthly macroeconomic time series, Marcellino *et al.* (2006) and Pesaran *et al.* (2011) find that the iterated approach typically outperforms the direct approach, especially if the sample size is small, if the forecast horizon is long, and if long lags of the variables are included in the forecasting model.

Although the parameters in many of the macroeconomic time series are unstable over time (Stock and Watson, 1996), work on multi-step forecasting in the presence of breaks has been virtually absent from the literature. However, it is widely accepted that structural breaks play a central role in economic forecasting (see, e.g., Clements and Hendry, 2006; Elliott and Timmermann, 2008; Rossi, 2013). Forecast errors are typically very large after structural breaks. Furthermore, it is possible that a forecasting model that performed well before the break performs poorly after the break. Forecasting models often systematically under- or over-predict in the presence of structural instability. Therefore, one way to improve their forecast accuracy in an unstable environment is to use intercept corrections, advocated by Clements and Hendry (1996a, 1998). Intercept corrections are based on the idea that if the forecasts systematically differ from the true values, i.e., if the forecast errors are systematically either positive or negative, then adjusting the mechanistic, model-based forecast by the previous forecast error (or an average of the most recent errors) should reduce the forecast bias and hence improve forecast performance.

Another issue that has been overlooked in the multi-step forecasting literature is the fact that key macroeconomic data, such as GDP and inflation series, are subject to revisions. The real-time nature of macroeconomic time series is potentially important for the relative performance of multi-step forecasting methods for at least three reasons. First, because data revisions are usually quite large, the parameters estimated on the final revised data may differ considerably from those estimated on the real-time data. Second, data revisions can also affect the dynamic lag structure of the forecasting model. Finally, real-time forecasts are conditioned on the first-release or lightly revised data actually available at each forecast origin, whereas forecasts based on the final revised data are conditioned on the latest available observations of each forecast origin. Practical forecasting is inherently a real-time exercise and thus the relative accuracy of multi-step forecasting methods should be evaluated using real-time data.

The main contributions of this paper are as follows. First, we analyze the relative performance of multi-step forecasting methods in the presence of breaks through Monte Carlo simulations. Our comparison includes the iterated and direct AR models and various forms of intercept correction. We consider several break processes, including changes in the intercept, autoregressive parameter, and error variance. We also examine how the timing of the break affects the accuracy of the methods. Second, we take into account in our simulations that most macroeconomic time series are subject to data revisions. A novelty of our simulation framework is that data revisions can either add news or reduce noise (see, e.g., Mankiw and Shapiro, 1986). The distinction between news and noise revisions allows us to study whether the properties of the revision process matter for the multi-period forecasting problem. Finally, the real-time accuracy of the multi-step forecasting methods for four key U.S. macroeconomic time series, namely, real GDP, industrial production, GDP deflator, and personal consumption expenditures (PCE) inflation, is compared.

The remainder of this paper is organized as follows. Section 2 introduces the notation and the statistical framework. Section 3 provides a brief overview of the multi-step forecasting methods. Section 4 presents the Monte Carlo simulation results and Section 5 presents the empirical results. Section 6 concludes.

3.2 Statistical framework

Key macroeconomic time series are published with a lag and are subject to revisions. For instance, a forecaster at period $T+1$ has access to the vintage $T+1$ values of GDP up to time period T. In addition, because of data revisions, the first-released value and the final value for a period may differ substantially. These two features of real-time data clearly matter for forecasting. As a result, we incorporate the publication lag and data revisions into our statistical framework. The statistical framework used in this paper follows that adopted in Clements and Galvão (2013), Hännikäinen (2014), and Jacobs and van Norden (2011). It relates a data vintage estimate to the true value plus an error or errors. More specifically, the period $t+s$ vintage estimate of the value of y in period t, denoted by y_t^{t+s} [1], where $s = 1, \ldots, l$ [2], can be expressed as the sum of the true value $\tilde{y}_t$, a news component v_t^{t+s}, and a noise component ε_t^{t+s}, i.e., $y_t^{t+s} = \tilde{y}_t + v_t^{t+s} + \varepsilon_t^{t+s}$.

In this framework, revisions either add news or reduce noise. Data revisions are news if they are uncorrelated with the previously published vintages, $cov(y_t^{t+k}, v_t^{t+s}) = 0 \; \forall k \leq s$. On the other hand, data revisions reduce noise if each vintage release is equal to the true value plus a noise. Noise revisions are uncorrelated with the true values, $cov(\tilde{y}_t, \varepsilon_t^{t+s}) = 0$. For further discussion of the properties of news and noise revisions, see Croushore (2011) and Jacobs and van Norden (2011).

We stack the l different vintage estimates of y_t, v_t and ε_t into vectors $\boldsymbol{y}_t = (y_t^{t+1}, \ldots, y_t^{t+l})'$, $\boldsymbol{v}_t = (v_t^{t+1}, \ldots, v_t^{t+l})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{t+1}, \ldots, \varepsilon_t^{t+l})'$, respectively. Using these vectors we can express each vintage of y_t as follows

$$\boldsymbol{y}_t = \boldsymbol{i}\tilde{y}_t + \boldsymbol{v}_t + \boldsymbol{\varepsilon}_t, \qquad (1)$$

1 Throughout this section, superscripts refer to vintages and subscripts to time periods.

2 Following Clements and Galvão (2013), we assume that we observe l different estimates of y_t before the true value, $\tilde{y}_t$, is observed. In practice, however, data may continue to be revised forever, so the true value may never be observed.

where $\boldsymbol{i}$ is an $l \times 1$ vector of ones. For simplicity, we consider an AR(1) process for the true values and assume that a single break has occurred at time T_1[3]

$$\tilde{y}_t = \begin{cases} \rho_1 + \sum_{i=1}^{l} \mu_{v1_i} + \beta_1 \tilde{y}_{t-1} + \sigma_1 \eta_{1t} + \sum_{i=1}^{l} \sigma_{v1_i} \eta_{2t,i}, & \text{for } t < T_1, \\ \rho_2 + \sum_{i=1}^{l} \mu_{v2_i} + \beta_2 \tilde{y}_{t-1} + \sigma_2 \eta_{1t} + \sum_{i=1}^{l} \sigma_{v2_i} \eta_{2t,i}, & \text{for } t \geq T_1, \end{cases} \tag{2}$$

where η_{1t} and $\eta_{2t,i}$ ($i = 1, \ldots, l$) are $NIID$ (0,1) disturbances. This setup allows for changes in the error variance, the intercept, and the slope immediately after the break.

The news and noise components in (1) before and after the break are specified by

$$\boldsymbol{v}_{1t} = \begin{bmatrix} v_{1t}^{t+1} \\ v_{1t}^{t+2} \\ \vdots \\ v_{1t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^{l} \mu_{v1_i} \\ \sum_{i=2}^{l} \mu_{v1_i} \\ \vdots \\ \mu_{v1_l} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{l} \sigma_{v1_i} \eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v1_i} \eta_{2t,i} \\ \vdots \\ \sigma_{v1_l} \eta_{2t,l} \end{bmatrix}, \boldsymbol{\varepsilon}_{1t} = \begin{bmatrix} \varepsilon_{1t}^{t+1} \\ \varepsilon_{1t}^{t+2} \\ \vdots \\ \varepsilon_{1t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \mu_{\varepsilon 1_1} \\ \mu_{\varepsilon 1_2} \\ \vdots \\ \mu_{\varepsilon 1_l} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon 1_1} \eta_{3t,1} \\ \sigma_{\varepsilon 1_2} \eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon 1_l} \eta_{3t,l} \end{bmatrix} \tag{3}$$

for $t < T_1$ and

$$\boldsymbol{v}_{2t} = \begin{bmatrix} v_{2t}^{t+1} \\ v_{2t}^{t+2} \\ \vdots \\ v_{2t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^{l} \mu_{v2_i} \\ \sum_{i=2}^{l} \mu_{v2_i} \\ \vdots \\ \mu_{v2_l} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{l} \sigma_{v2_i} \eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v2_i} \eta_{2t,i} \\ \vdots \\ \sigma_{v2_l} \eta_{2t,l} \end{bmatrix}, \boldsymbol{\varepsilon}_{2t} = \begin{bmatrix} \varepsilon_{2t}^{t+1} \\ \varepsilon_{2t}^{t+2} \\ \vdots \\ \varepsilon_{2t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \mu_{\varepsilon 2_1} \\ \mu_{\varepsilon 2_2} \\ \vdots \\ \mu_{\varepsilon 2_l} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon 2_1} \eta_{3t,1} \\ \sigma_{\varepsilon 2_2} \eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon 2_l} \eta_{3t,l} \end{bmatrix} \tag{4}$$

for $t \geq T_1$.

The shocks are assumed to be mutually independent. Otherwise stated, if $\boldsymbol{\eta}_t = [\eta_{1t}, \boldsymbol{\eta}'_{2t}, \boldsymbol{\eta}'_{3t}]'$, then $E(\boldsymbol{\eta}_t) = \mathbf{0}$ and $E(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t) = I$. We assume that $\tilde{y}_t$ is a stationary process, so that $|\beta_j| < 1$ (for $j = 1,2$). Because $\tilde{y}_t$ is a stationary process and both the news and noise terms are stationary, $\boldsymbol{y}_t$ is also a stationary process. The means of the news and noise terms, denoted by μ_{vj_i} and $\mu_{\varepsilon j_i}$ (for $j = 1,2$ and $i = 1,\ldots,l$), are allowed to be non-zero. This is an important feature because the previous literature has found that revisions to

3 We consider an AR(1) model rather than, say, an AR(4) model because it is easier to calibrate the parameters (see Section 4 below) when the model contains only one lag. Eklund *et al.* (2013), Hännikäinen (2014), and Pesaran and Timmermann (2005) also focus on an AR(1) model in the presence of breaks.

macroeconomic data typically have non-zero means (see, e.g., Aruoba, 2008; Clements and Galvão, 2013; Croushore, 2011).

3.3 Methods for multi-step forecasting

In this section, we explain how the multi-step forecasts are computed in the iterated and direct approaches. We assume that the variable of interest, y_t, is a stationary process. For simplicity, we focus on an AR(1) model. The generalization to AR(p) models is straightforward.

Iterated forecasts are made using a one-period ahead model, iterated forward for the required number of periods. The one-step ahead AR model for y_t, ignoring data revisions, is

$$y_{t+1} = \alpha + \beta y_t + \varepsilon_t. \tag{5}$$

The parameters in (5) are estimated by OLS and the iterated forecast of y_{t+h} is then calculated as follows:

$$\hat{y}^I_{t+h|t} = \hat{\alpha} + \hat{\beta}\hat{y}^I_{t+h-1|t},$$

where $\hat{y}_{t|t} = y_t$. Note that the same model specification is used for all forecast horizons.

Under the direct approach, the dependent variable in the forecasting model is the multi-step ahead value being forecasted. Thus, a forecaster selects a separate model for each forecast horizon. The direct forecasting model, ignoring data revisions, is

$$y_{t+h} = \phi + \rho y_t + \varepsilon_{t+h}. \tag{6}$$

The parameters in (6) are estimated by OLS using data through period t (i.e., y_t is the last observation on the left-hand side of the multi-step regression). Then, the direct forecast of y_{t+h} is constructed as

$$\hat{y}^D_{t+h|t} = \hat{\phi} + \hat{\rho} y_t.$$

As discussed in the Introduction, intercept corrections offer some protection against structural instability. If the forecasting model systematically either under- or over-predicts after a break, intercept corrections based on the previous forecast errors reduce forecast bias. On the other hand, intercept corrections increase forecast error variance.

Following Clements and Hendry (1996a, 1998), we consider three alternative intercept corrections to the iterated approach. The first strategy is a so called constant adjustment method, where the adjustment over the forecast period is held constant at the average of the m most recent forecast errors, denoted by $e_t^* = \frac{1}{m}\sum_{j=1}^{m} e_{t+1-j}$:

$$\tilde{y}^I_{t+h|t} = \hat{\alpha} + \hat{\beta}\tilde{y}^I_{t+h-1|t} + e^*_t,$$

which implies that

$$\tilde{y}^I_{t+h|t} = \hat{y}^I_{t+h|t} + \sum_{i=0}^{h-1} \hat{\beta}^i e^*_t.$$

The second strategy only adjusts the one-step ahead forecast. The iterated forecast generated by this one-off adjustment method is

$$\vec{y}^I_{t+h|t} = \hat{\alpha} + \hat{\beta}\vec{y}^I_{t+h-1|t}, \quad \vec{y}^I_{t+1|t} = \tilde{y}^I_{t+1|t} = \hat{\alpha} + \hat{\beta}y_t + e^*_t,$$

so that

$$\vec{y}^I_{t+h|t} = \hat{y}^I_{t+h|t} + \hat{\beta}^{h-1}e^*_t.$$

The third strategy, called the full-adjustment method, adjusts the model-based forecast by the full amount of the average of the m most recent forecast errors:

$$\bar{y}^I_{t+h|t} = \hat{y}^I_{t+h|t} + e^*_t.$$

In addition, we consider a full-adjustment to the direct forecasting method. In this case, the average of the m most recent forecast errors from the direct model, denoted by $e^*_{t,D} = \frac{1}{m}\sum_{j=1}^{m} e^D_{t+1-j}$, is used to adjust the model-based forecast:

$$\bar{y}^D_{t+h|t} = \hat{y}^D_{t+h|t} + e^*_{t,D}.$$

3.4 Monte Carlo simulations

In this section, we perform a number of Monte Carlo simulation experiments to evaluate the performance of the multi-step forecasting methods in the presence of breaks. These experiments are based on the statistical framework introduced in Section 2. A sample size of 100 observations, which corresponds to 25 years of quarterly data, is used in the experiments. We assume that a single break has occurred prior to the forecast origin. Because the timing of the break might affect the relative accuracy of the multi-step methods, we consider three different break points: $T_1 = 25, 50$, and 99.

We calibrate the parameter values on actual U.S. data following Hännikäinen (2014). The parameters remain constant over time in experiment 1 (see Table 1). In this case, the selected parameter values imply that the mean of the true process lies between 2.0 and 2.5, which corresponds roughly to the average U.S. annual inflation and real GDP growth over the past 25 years. The parameters in experiment 1 are used as pre-break parameters in the rest of the experiments (with the exceptions of experiments 4–5). We consider several break processes. First, we analyze how moderate (0.25) and large (0.5) changes in the autoregressive parameter in either direction affect the relative performance of the multi-step methods (experiments 2–5). Second, we consider breaks in the error variance. We allow σ to increase from 1.5 to 4.5 (experiment 6) and decrease from 1.5 to 0.5 (experiment 7). Finally, we examine how changes in the constant term affect the accuracy of the methods (experiments 8–9).

We assume that the data revisions are either pure news ($\sigma_{v_i} \neq 0$, $\sigma_{\varepsilon_i} = 0$ for $i = 1, \ldots, l$) or pure noise ($\sigma_{v_i} = 0$, $\sigma_{\varepsilon_i} \neq 0$ for $i = 1, \ldots l$). This allows us to analyze whether the properties of the revision process matter for the relative performance of the multi-step forecasting methods. We set $l = 14$, so that we observe 14 different estimates of y_t before the true value, $\tilde{y}_t$, is observed.[4] Consistent with the previous work in Clements and Galvão (2013) and Hännikäinen (2014), only the first and fifth revisions are assumed to have non-zero means. The means of these revisions are set to four and two percent of the mean of the first-release data, y_t^{t+1}, both before and after the break. Similarly, the standard deviation of the first revision is set to 40 percent of the standard deviation of the first-release data. The standard deviations of revisions 2–13 and 14 are set to 20 and 10 percent of the standard deviation of the first-release data, respectively. For convenience, the parameter values used in the Monte Carlo experiments are shown in Table 1.[5]

4 As discussed in Croushore (2011), GDP and inflation data for period t are subject to annual revisions at the end of July of each of the following three years. Our choice $l = 14$ is motivated by the fact that y_t^{t+15} will have undergone all the regular revisions irrespectively of which quarter of the year t falls in. For a similar approach, see Clements and Galvão (2013).

5 Appendix A summarizes the means and standard deviations of the first-release and final data for each experiment. The details of the calibration process are presented in Hännikäinen (2014).

Table 1. Simulation setup

True process Experiments	ρ_1	ρ_2	β_1	β_2	σ_1	σ_2
1: No break	1	1	0.5	0.5	1.5	1.5
2: Moderate break in β (increase)	1	1	0.5	0.75	1.5	1.5
3: Moderate break in β (decrease)	1	1	0.5	0.25	1.5	1.5
4: Large break in β (increase)	1	1	0.25	0.75	1.5	1.5
5: Large break in β (decrease)	1	1	0.75	0.25	1.5	1.5
6: Increase in post-break variance	1	1	0.5	0.5	1.5	4.5
7: Decrease in post-break variance	1	1	0.5	0.5	1.5	0.5
8: Break in mean (increase)	1	1.5	0.5	0.5	1.5	1.5
9: Break in mean (decrease)	1	0.5	0.5	0.5	1.5	1.5

News Experiments	μ_{v1_1}	μ_{v2_1}	μ_{v1_5}	μ_{v2_5}	σ_{v1_1}	σ_{v2_1}	$\sigma_{v1_{2,\dots,13}}$	$\sigma_{v2_{2,\dots,13}}$	$\sigma_{v1_{14}}$	$\sigma_{v2_{14}}$
1: No break	0.085	0.085	0.043	0.043	0.783	0.783	0.391	0.391	0.196	0.196
2: Moderate break in β (increase)	0.085	0.195	0.043	0.098	0.783	2.238	0.391	1.119	0.196	0.560
3: Moderate break in β (decrease)	0.085	0.054	0.043	0.027	0.783	0.634	0.391	0.317	0.196	0.158
4: Large break in β (increase)	0.054	0.195	0.027	0.098	0.634	2.238	0.317	1.119	0.158	0.560
5: Large break in β (decrease)	0.195	0.054	0.098	0.027	2.238	0.634	1.119	0.317	0.560	0.158
6: Increase in post-break variance	0.085	0.085	0.043	0.043	0.783	2.348	0.391	1.174	0.196	0.587
7: Decrease in post-break variance	0.085	0.085	0.043	0.043	0.783	0.261	0.391	0.130	0.196	0.065
8: Break in mean (increase)	0.085	0.128	0.043	0.064	0.783	0.783	0.391	0.391	0.196	0.196
9: Break in mean (decrease)	0.085	0.043	0.043	0.021	0.783	0.783	0.391	0.391	0.196	0.196

Noise Experiments	$\mu_{\varepsilon1_1}$	$\mu_{\varepsilon2_1}$	$\mu_{\varepsilon1_{2,\dots,5}}$	$\mu_{\varepsilon2_{2,\dots,5}}$	$\sigma_{\varepsilon1_1}$	$\sigma_{\varepsilon2_1}$	$\sigma_{\varepsilon1_{2,4,\dots,14}}$	$\sigma_{\varepsilon2_{2,4,\dots,14}}$	$\sigma_{\varepsilon1_{3,5,\dots,13}}$	$\sigma_{\varepsilon2_{3,5,\dots,13}}$
1: No break	0.113	0.113	0.038	0.038	0.728	0.728	0.188	0.188	0.325	0.325
2: Moderate break in β (increase)	0.113	0.226	0.038	0.075	0.728	0.953	0.188	0.246	0.325	0.426
3: Moderate break in β (decrease)	0.113	0.075	0.038	0.025	0.728	0.651	0.188	0.168	0.325	0.291
4: Large break in β (increase)	0.075	0.226	0.025	0.075	0.651	0.953	0.168	0.246	0.291	0.426
5: Large break in β (decrease)	0.226	0.075	0.075	0.025	0.953	0.651	0.246	0.168	0.426	0.291
6: Increase in post-break variance	0.113	0.113	0.038	0.038	0.728	2.183	0.188	0.564	0.325	0.976
7: Decrease in post-break variance	0.113	0.113	0.038	0.038	0.728	0.243	0.188	0.063	0.325	0.108
8: Break in mean (increase)	0.113	0.170	0.038	0.057	0.728	0.728	0.188	0.188	0.325	0.325
9: Break in mean (decrease)	0.113	0.057	0.038	0.019	0.728	0.728	0.188	0.188	0.325	0.325

Table 2. Relative MSFE values when revisions add news

Break date		T_1 = 25				T_1 = 50			
Forecast horizon		2	4	8	12	2	4	8	12
Exp.1	Constant	1.419	1.623	1.713	1.729	–	–	–	–
	One-off	1.040	1.003	1.000	1.000	–	–	–	–
	Full	1.182	1.175	1.178	1.183	–	–	–	–
	Direct	1.009	1.017	1.024	1.027	–	–	–	–
	Full direct	1.380	1.481	1.504	1.514	–	–	–	–
Exp.2	Constant	1.436	1.821	2.360	2.528	1.427	1.787	2.271	2.483
	One-off	1.064	1.016	1.004	1.001	1.058	1.015	1.004	1.001
	Full	1.129	1.097	1.107	1.105	1.121	1.090	1.096	1.104
	Direct	1.007	1.026	1.047	1.052	1.009	1.024	1.048	1.068
	Full direct	1.398	1.601	1.723	1.713	1.389	1.592	1.740	1.786
Exp.3	Constant	1.369	1.406	1.412	1.447	1.344	1.339	1.334	1.341
	One-off	1.024	0.999	1.000	1.000	1.022	0.993	0.999	1.000
	Full	1.205	1.167	1.161	1.176	1.166	1.082	1.060	1.070
	Direct	1.006	1.011	1.011	1.006	1.003	1.004	1.006	0.997
	Full direct	1.332	1.351	1.346	1.370	1.296	1.261	1.235	1.248
Exp.4	Constant	1.450	1.838	2.351	2.503	1.396	1.797	2.330	2.544
	One-off	1.069	1.018	1.006	1.001	1.045	1.015	1.007	1.002
	Full	1.135	1.101	1.103	1.092	1.102	1.085	1.092	1.094
	Direct	1.003	1.014	1.032	1.052	1.003	1.031	1.068	1.073
	Full direct	1.408	1.612	1.722	1.728	1.368	1.634	1.814	1.877
Exp.5	Constant	1.390	1.471	1.435	1.476	1.235	1.160	1.013	0.997
	One-off	1.060	0.983	0.984	0.993	0.981	0.919	0.962	0.985
	Full	1.144	1.023	0.934	0.924	1.019	0.849	0.764	0.752
	Direct	0.994	0.952	0.875	0.832	0.986	0.944	0.886	0.845
	Full direct	1.275	1.189	0.981	0.964	1.106	0.861	0.609	0.575
Exp.6	Constant	1.457	1.657	1.768	1.775	1.450	1.627	1.808	1.795
	One-off	1.054	1.007	1.000	1.000	1.053	1.004	1.001	1.000
	Full	1.210	1.194	1.203	1.193	1.200	1.176	1.211	1.200
	Direct	1.008	1.022	1.024	1.033	1.015	1.026	1.032	1.033
	Full direct	1.421	1.513	1.552	1.564	1.411	1.496	1.576	1.568
Exp.7	Constant	1.314	1.477	1.531	1.541	1.264	1.382	1.397	1.403
	One-off	0.997	0.992	0.998	1.000	0.974	0.984	0.998	1.000
	Full	1.106	1.096	1.084	1.088	1.063	1.032	1.018	1.014
	Direct	1.013	1.018	1.021	0.997	1.017	1.025	1.023	1.027
	Full direct	1.288	1.348	1.355	1.358	1.232	1.263	1.240	1.252
Exp.8	Costant	1.438	1.663	1.773	1.775	1.412	1.625	1.720	1.704
	One-off	1.048	1.004	1.000	1.000	1.038	1.001	1.000	1.000
	Full	1.188	1.174	1.180	1.174	1.168	1.157	1.149	1.140
	Direct	1.007	1.014	1.021	1.022	1.006	1.013	1.013	1.003
	Full direct	1.398	1.520	1.526	1.547	1.376	1.485	1.491	1.495
Exp.9	Costant	1.406	1.652	1.682	1.716	1.361	1.551	1.589	1.601
	One-off	1.038	1.004	1.000	1.000	1.022	0.994	0.999	1.000
	Full	1.168	1.174	1.143	1.157	1.134	1.106	1.082	1.076
	Direct	1.007	1.015	1.014	1.013	1.007	1.017	1.011	1.005
	Full direct	1.370	1.526	1.477	1.499	1.328	1.423	1.392	1.396

Notes: The experiments are as defined in Table 1. 'Constant' denotes the method of constant adjustment to the iterated model; 'One-off ' denotes the one-off adjustment to the iterated method. 'Full' and 'Full direct' denote full adjustment to the iterated and direct methods, respectively. Intercept corrections are based on the average of the latest 4 forecast errors. The sample size is T = 100. The break occurs at T_1 = 25, 50, or 99. MSFE values are computed relative to those produced by the iterated forecasting method.

T_1 = 99			
2	4	8	12
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
0.999	1.024	1.058	1.070
0.981	0.995	1.000	1.000
0.979	0.987	1.004	1.009
1.006	1.010	1.008	1.008
1.007	1.029	1.047	1.060
1.416	1.511	1.464	1.474
1.044	0.995	0.999	1.000
1.186	1.114	1.059	1.060
1.009	1.028	1.032	1.038
1.406	1.475	1.406	1.413
0.961	0.989	1.004	1.014
0.984	0.999	1.000	1.000
0.960	0.983	0.995	1.003
1.006	1.005	1.004	1.003
0.973	0.997	1.010	1.018
1.271	1.272	1.267	1.285
0.949	0.907	0.964	0.987
0.990	0.838	0.770	0.761
1.028	1.091	1.159	1.186
1.617	1.875	1.792	1.750
1.121	1.196	1.240	1.220
1.013	1.003	1.000	1.000
1.050	1.054	1.062	1.051
1.003	1.012	1.007	1.014
1.068	1.090	1.097	1.097
3.431	4.419	4.663	4.721
1.212	1.007	0.999	1.000
2.057	1.966	1.905	1.911
1.032	1.070	1.094	1.109
3.494	4.074	4.044	4.091
1.338	1.493	1.535	1.563
1.026	1.000	1.000	1.000
1.139	1.130	1.116	1.127
1.010	1.020	1.026	1.031
1.327	1.420	1.415	1.440
1.269	1.391	1.400	1.429
1.004	0.995	0.999	1.000
1.091	1.078	1.057	1.069
1.011	1.023	1.029	1.031
1.260	1.316	1.292	1.326

Table 3. Relative MSFE values when revisions reduce noise

Break date		$T_1 = 25$				$T_1 = 50$			
Forecast horizon		2	4	8	12	2	4	8	12
Exp.1	Constant	1.413	1.639	1.680	1.701	–	–	–	–
	One-off	1.046	1.006	1.000	1.000	–	–	–	–
	Full	1.186	1.193	1.170	1.176	–	–	–	–
	Direct	1.008	1.015	1.015	1.022	–	–	–	–
	Full direct	1.354	1.470	1.451	1.482	–	–	–	–
Exp.2	Constant	1.492	1.877	2.328	2.633	1.431	1.750	2.120	2.234
	One-off	1.104	1.029	1.005	1.002	1.063	1.006	1.001	1.000
	Full	1.177	1.124	1.107	1.122	1.132	1.078	1.067	1.065
	Direct	1.004	1.013	1.028	1.029	1.004	1.019	1.024	1.017
	Full direct	1.403	1.608	1.718	1.806	1.368	1.583	1.716	1.723
Exp.3	Constant	1.383	1.424	1.377	1.419	1.350	1.407	1.429	1.423
	One-off	1.034	1.000	1.000	1.000	1.030	0.999	1.000	1.000
	Full	1.222	1.189	1.146	1.174	1.177	1.135	1.126	1.123
	Direct	1.008	1.013	1.008	1.011	1.007	1.012	1.008	1.005
	Full direct	1.347	1.360	1.299	1.347	1.302	1.314	1.314	1.319
Exp.4	Constant	1.517	1.902	2.385	2.668	1.345	1.636	1.971	2.095
	One-off	1.113	1.033	1.006	1.002	1.029	0.995	0.998	0.999
	Full	1.188	1.132	1.109	1.103	1.080	1.030	1.025	1.016
	Direct	1.004	1.015	1.028	1.024	0.992	0.995	1.009	1.003
	Full direct	1.440	1.667	1.778	1.809	1.321	1.582	1.734	1.767
Exp.5	Constant	1.455	1.569	1.576	1.617	1.430	1.505	1.541	1.489
	One-off	1.089	1.004	0.996	0.999	1.089	0.983	0.982	0.993
	Full	1.206	1.105	1.038	1.029	1.161	1.008	0.913	0.883
	Direct	1.027	1.012	0.952	0.932	1.009	0.962	0.867	0.830
	Full direct	1.407	1.445	1.327	1.275	1.328	1.263	1.088	0.985
Exp.6	Constant	1.473	1.644	1.746	1.730	1.484	1.688	1.740	1.811
	One-off	1.071	1.008	1.000	1.000	1.072	1.010	1.000	1.000
	Full	1.229	1.196	1.195	1.187	1.232	1.204	1.180	1.205
	Direct	1.004	1.023	1.027	1.031	1.008	1.023	1.035	1.039
	Full direct	1.413	1.490	1.503	1.510	1.429	1.510	1.494	1.546
Exp.7	Constant	1.372	1.532	1.587	1.573	1.329	1.479	1.530	1.543
	One-off	1.028	0.999	0.999	1.000	1.011	0.993	0.999	1.000
	Full	1.155	1.135	1.119	1.107	1.122	1.105	1.091	1.086
	Direct	1.011	1.025	1.021	1.016	1.008	1.025	1.017	1.025
	Full direct	1.326	1.382	1.373	1.354	1.275	1.348	1.338	1.350
Exp.8	Costant	1.450	1.691	1.780	1.785	1.426	1.655	1.730	1.757
	One-off	1.062	1.008	1.000	1.000	1.058	1.008	1.000	1.000
	Full	1.203	1.192	1.186	1.181	1.185	1.165	1.151	1.149
	Direct	1.009	1.015	1.018	1.012	1.006	1.010	1.013	1.011
	Full direct	1.404	1.521	1.529	1.524	1.378	1.492	1.506	1.530
Exp.9	Costant	1.472	1.686	1.738	1.759	1.417	1.599	1.674	1.655
	One-off	1.067	1.008	1.000	1.000	1.048	1.000	0.999	1.000
	Full	1.214	1.187	1.159	1.163	1.171	1.131	1.112	1.101
	Direct	1.004	1.014	1.013	1.007	1.006	1.005	1.002	1.004
	Full direct	1.401	1.507	1.478	1.486	1.362	1.429	1.446	1.433

See the notes to Table 2.

T_1 = 99			
2	4	8	12
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
1.118	1.196	1.296	1.325
0.970	0.990	1.000	1.000
1.007	1.011	1.055	1.072
1.018	1.029	1.016	1.012
1.112	1.159	1.208	1.227
1.431	1.521	1.498	1.490
1.060	1.002	0.999	1.000
1.207	1.145	1.099	1.096
1.006	1.016	1.028	1.031
1.393	1.447	1.391	1.402
0.947	1.014	1.056	1.074
0.958	0.997	1.000	1.000
0.923	0.975	1.011	1.025
1.026	1.017	1.012	1.009
0.975	1.025	1.054	1.066
1.089	1.048	1.005	0.993
0.956	0.932	0.972	0.990
0.966	0.860	0.806	0.793
1.012	1.058	1.120	1.137
1.133	1.174	1.128	1.109
1.204	1.290	1.295	1.297
1.036	1.006	1.000	1.000
1.100	1.091	1.073	1.073
1.006	1.011	1.018	1.016
1.118	1.121	1.111	1.108
3.371	4.432	4.723	4.886
1.232	1.019	0.999	1.000
2.051	2.008	1.948	1.992
1.020	1.043	1.069	1.087
3.510	4.229	4.182	4.288
1.383	1.524	1.602	1.591
1.045	1.004	1.000	1.000
1.173	1.149	1.153	1.140
1.005	1.014	1.025	1.028
1.351	1.429	1.454	1.440
1.284	1.375	1.416	1.447
1.016	0.995	0.999	1.000
1.110	1.074	1.070	1.080
1.010	1.021	1.027	1.026
1.264	1.280	1.285	1.309

Table 4. Squared bias relative to the MSFE of the iterated benchmark when revisions add news

Break date		T_1 = 25				T_1 = 50			
Forecast horizon		2	4	8	12	2	4	8	12
Exp.1	Iterated	0.003	0.003	0.004	0.004	–	–	–	–
	Constant	0.000	0.000	0.000	0.000	–	–	–	–
	One-off	0.002	0.003	0.004	0.004	–	–	–	–
	Full	0.001	0.001	0.001	0.001	–	–	–	–
	Direct	0.004	0.003	0.004	0.004	–	–	–	–
	Full direct	0.000	0.000	0.000	0.000	–	–	–	–
Exp.2	Iterated	0.002	0.004	0.004	0.006	0.014	0.025	0.033	0.036
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	One-off	0.000	0.003	0.004	0.005	0.005	0.019	0.031	0.035
	Full	0.000	0.001	0.002	0.003	0.002	0.009	0.015	0.017
	Direct	0.002	0.004	0.004	0.003	0.014	0.026	0.033	0.030
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Exp.3	Iterated	0.020	0.027	0.022	0.027	0.058	0.071	0.087	0.078
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	One-off	0.011	0.026	0.022	0.027	0.030	0.064	0.087	0.078
	Full	0.001	0.003	0.002	0.003	0.005	0.009	0.016	0.013
	Direct	0.019	0.023	0.014	0.014	0.056	0.064	0.073	0.061
	Full direct	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001
Exp.4	Iterated	0.005	0.010	0.017	0.016	0.024	0.037	0.051	0.053
	Constant	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
	One-off	0.002	0.008	0.017	0.016	0.009	0.028	0.048	0.052
	Full	0.001	0.004	0.010	0.009	0.004	0.012	0.022	0.024
	Direct	0.005	0.010	0.013	0.008	0.023	0.035	0.045	0.039
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003
Exp.5	Iterated	0.066	0.136	0.185	0.190	0.163	0.324	0.421	0.439
	Constant	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001
	One-off	0.024	0.105	0.175	0.186	0.050	0.236	0.392	0.429
	Full	0.010	0.047	0.082	0.086	0.021	0.113	0.196	0.216
	Direct	0.055	0.091	0.089	0.061	0.150	0.275	0.325	0.299
	Full direct	0.000	0.000	0.000	0.002	0.001	0.001	0.006	0.016
Exp.6	Iterated	0.001	0.000	0.000	0.001	0.000	0.001	0.001	0.000
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	One-off	0.000	0.000	0.000	0.001	0.000	0.001	0.001	0.000
	Full	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Direct	0.001	0.000	0.000	0.001	0.000	0.002	0.001	0.000
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Exp.7	Iterated	0.023	0.031	0.033	0.027	0.023	0.025	0.024	0.030
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	One-off	0.012	0.028	0.033	0.027	0.011	0.022	0.023	0.030
	Full	0.003	0.008	0.009	0.006	0.002	0.004	0.004	0.007
	Direct	0.024	0.034	0.036	0.028	0.024	0.028	0.025	0.032
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Exp.8	Iterated	0.001	0.003	0.002	0.002	0.020	0.027	0.033	0.040
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	One-off	0.000	0.003	0.002	0.002	0.008	0.023	0.033	0.040
	Full	0.000	0.001	0.000	0.001	0.002	0.005	0.008	0.011
	Direct	0.001	0.002	0.000	0.000	0.019	0.025	0.026	0.026
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001
Exp.9	Iterated	0.018	0.022	0.021	0.020	0.058	0.075	0.083	0.089
	Constant	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	One-off	0.008	0.019	0.021	0.020	0.027	0.064	0.082	0.089
	Full	0.002	0.004	0.004	0.004	0.008	0.015	0.020	0.023
	Direct	0.017	0.019	0.014	0.009	0.058	0.072	0.071	0.069
	Full direct	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001

Notes: The table reports the squared bias of the different methods as a ratio of the MSFE of the iterated benchmark model.

$T_1 = 99$			
2	4	8	12
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
0.072	0.118	0.158	0.174
0.057	0.098	0.135	0.150
0.067	0.116	0.157	0.174
0.062	0.107	0.146	0.162
0.073	0.121	0.160	0.177
0.067	0.117	0.156	0.173
0.155	0.210	0.236	0.222
0.052	0.063	0.074	0.066
0.116	0.198	0.235	0.222
0.079	0.119	0.142	0.131
0.161	0.225	0.251	0.235
0.074	0.111	0.129	0.118
0.127	0.191	0.240	0.253
0.104	0.166	0.213	0.226
0.122	0.191	0.240	0.253
0.109	0.172	0.220	0.232
0.129	0.194	0.242	0.254
0.111	0.176	0.224	0.235
0.494	0.623	0.687	0.693
0.091	0.076	0.059	0.051
0.287	0.512	0.654	0.682
0.220	0.351	0.428	0.441
0.515	0.686	0.790	0.804
0.215	0.355	0.448	0.455
0.000	0.000	0.001	0.000
0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.000
0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.000
0.000	0.000	0.000	0.000
0.015	0.018	0.020	0.025
0.001	0.001	0.002	0.003
0.009	0.016	0.020	0.025
0.004	0.006	0.008	0.011
0.015	0.019	0.023	0.027
0.001	0.001	0.002	0.003
0.119	0.162	0.177	0.185
0.091	0.123	0.133	0.140
0.109	0.159	0.177	0.185
0.100	0.141	0.155	0.163
0.121	0.169	0.183	0.190
0.112	0.168	0.189	0.194
0.173	0.231	0.260	0.257
0.081	0.109	0.123	0.121
0.140	0.222	0.259	0.257
0.107	0.159	0.183	0.181
0.179	0.243	0.273	0.268
0.101	0.150	0.172	0.168

For simplicity, we focus on forecasting the first-release values and assume that the lag structure of the forecasting model is correctly specified, i.e., the forecasts are generated using an AR(1) model.[6] We estimate the parameters of the forecasting models using the entire data sample from the latest available vintage. Following Clements and Hendry (1996a), the intercept corrections are based on the average of the latest four forecast errors.[7] The iterated multi-step forecasting method is used as a benchmark in our Monte Carlo simulations. For each alternative method we compute MSFE values relative to those produced by the iterated benchmark. Values below (above) unity imply that the candidate method produces more (less) accurate forecasts than the benchmark. Multi-step forecasts are computed for horizons of 2, 4, 8, and 12 periods. The results are based on 10,000 replications and are shown in Tables 2 and 3.

Table 2 shows the relative performance of the multi-step forecasting methods when the data revisions are pure news. The results indicate that the iterated method generates the best forecasts in most of the experiments. In particular, the iterated method dominates the other methods when the parameters remain constant over time (experiment 1), or the variance increases (experiment 6), or the intercept increases (experiment 8). The iterated method also performs particularly well when the autoregressive parameter decreases moderately (experiment 3), or when the constant term decreases (experiment 9), although it does not always deliver the most accurate forecasts. In these few cases, however, the best performing alternative makes only a very slight improvement over the iterated approach. By contrast, the iterated method performs poorly when the autoregressive parameter decreases substantially after the break (experiment 5).

The timing of the structural break ($T_1 = 25, 50, 99$) has an impact on the performance of the various approaches. The iterated method appears to be the superior method when the break occurs early ($T_1 = 25$) during the sample, but its performance deteriorates when the break occurs closer to the forecast origin. There is a simple explanation for this finding. Table 4 reports the (squared) forecast bias of each method relative to the MSFE of the benchmark iterated model. As the timing of the break gets closer to the forecast origin, forecasts become more biased, because fewer post-break values are available for estimation. This implies that the importance of the bias component in determining the accuracy of the forecasts increases. The iterated method is more prone to bias than the other methods. Therefore, it is less successful when the break date T_1 gets close to the end of the sample.

6 The results are qualitatively similar if we use the bias correction method suggested by Clements and Galvão (2013) to forecast the final values or if we consider an AR(2) forecasting model. A full set of results is available upon request.

7 The general conclusions are the same if the intercept corrections are based on the most recent forecast error or the average of the latest two or three forecast errors.

Moreover, the relative performance of the iterated method improves as the forecast horizon increases. This happens for a subtle reason. As the forecast horizon increases, the parameters of the direct model are estimated with fewer observations. The parameters of the iterated model, on the other hand, are estimated with the largest possible sample size regardless of the forecast horizon. Thus, for a fixed sample size, it becomes less desirable to use an inefficient direct method as the forecast horizon lengthens. Intercept corrections reduce the forecast bias at the cost of increased forecast error variance. The additional uncertainty induced by intercept corrections grows with the forecast horizon. Hence, the bias–variance trade-off is less favorable to intercept corrections at long horizons.

The results in Table 2 suggest that various forms of intercept correction yield relatively poor forecasts in the presence of structural instability. The only exception is the case where the slope parameter decreases substantially after the break (experiment 5). In this case, the improvements over the iterated benchmark are very large at longer forecast horizons (i.e., $h = 8$ and 12). Hence it is mainly in situations where a break is believed to decrease substantially the AR parameter (i.e., when both the mean and variance decrease substantially) that intercept corrections can be recommended. In the rest of the experiments, intercept corrections have the most potential when the break has occurred close to the forecast origin (i.e., $T_1 = 99$) and the forecast horizon is short (i.e., $h = 2$ and 4). The one-off adjustment to the iterated method is generally more successful at reducing the MSFE values than the other forms of intercept correction. The constant adjustment to the iterated method and the full adjustment to the direct method perform worst among all the methods. They produce significantly higher MSFE values than the iterated benchmark in most of the experiments.

A comparison of the iterated and direct methods reveals that the iterated method typically delivers more accurate forecasts in the presence of breaks. The direct forecasts only dominate the iterated ones when the autoregressive parameter decreases substantially (experiment 5) and the timing of the break is either $T_1 = 25$ or $T_1 = 50$. Thus, there is only very limited evidence that the direct method helps reduce MSFE values in an unstable environment. The explanation for this finding is again related to the bias–variance trade-off. It appears that in an unstable environment, the reduction in bias obtained from the direct model is less important than the reduction in estimation variance arising from estimating the iterated model.

The results for noise revisions are summarized in Table 3. These results are qualitatively similar to those presented in Table 2. Thus, whether the data revisions add news or reduce noise does not matter much for the relative performance of the multi-period forecasting methods. If anything, the iterated method performs slightly better in relative terms when data revisions reduce noise.

3.5 Empirical results

Next, we compare the relative performance of the multi-step forecasting methods using actual U.S. real-time data. We consider h-step ahead forecasts of real GDP and industrial production growth, the GDP deflator, and the PCE inflation rate (annualized). All forecasts are out-of-sample. At each forecast origin $t + 1$, the $t + 1$ vintage estimates of data up to period t are used to estimate the parameters of a forecasting model that is then used to generate a forecast for period $t + h$. Forecasts are generated for horizons of $h = 2, 4, 8$, and 12 quarters. The parameters of the forecasting models are re-estimated at each forecast origin using a rolling window of 100 observations.[8] We consider two fixed lag lengths, namely $p = 1$ and $p = 4$. In addition, we determine the lag length by the Bayes Information Criterion (BIC) and the Akaike Information Criterion (AIC). The possible lag lengths are $p = 1, \ldots, 4$. At each forecast origin the model with the lowest information criteria is chosen. Because the BIC and AIC values are recomputed at each forecast origin, the order of the forecasting model can change from one period to the next.[9] Intercept corrections are based on the average of the four most recent forecast errors.[10] For simplicity, we focus on forecasting the first-release values. The general conclusions are the same if we forecast the final, 2013:Q3 vintage values. All real-time data is quarterly and the sample period runs from 1947:Q2 to 2013:Q2. Different vintages are obtained from the Federal Reserve Bank of Philadelphia's real-time database.

We start our analysis by considering the whole out-of-sample period spanning from 1977:Q2 to 2013:Q2. The performance of the various multi-step forecasting methods relative to the iterated benchmark over this period is summarized in Table 5. Panels A and B report the results for the real GDP and industrial production, whereas Panels C and D contain the results for the GDP deflator and PCE inflation. The first row in each Panel provides the root MSFE value of the benchmark iterated estimator. The subsequent rows show the MSFE values of the candidate methods relative to the

8 As discussed in Rossi (2013), different estimation window sizes may lead to different results. We check the robustness of our results by considering four different rolling window sizes, namely 40, 60, 80, and 100. The results are similar for the four rolling windows, and therefore we report the results for the rolling window of 100 observations only.

9 Iterated models selected by the AIC on average include two lags for real activity measures and three lags for inflation series. The BIC selects iterated models with only one lag for the real output series and models with two or three lags for the inflation series. For the direct models, the AIC recommends on average one or two lags, whereas the BIC recommends an optimal lag length of one.

10 The results are qualitatively similar if intercept corrections are based on the most recent forecast error or the average of the latest two or three forecast errors.

MSFE value of the benchmark model. The statistical significance is evaluated using the Giacomini and White (2006) test of equal unconditional predictive ability.

The results in Panels A and B indicate that the iterated method typically produces the lowest, or nearly the lowest, MSFE values for both real GDP and industrial production irrespective of which lag method or forecast horizon is employed. Even in the few cases where at least one of the other methods generates more accurate multi-step forecasts, even the best performing alternative provides only modest improvements over the iterated benchmark. For real GDP, the one-off adjustment method systematically dominates the benchmark at $h = 2$. Similarly, when short-lag selection methods ($p = 1$ and BIC) are used, the direct forecast is preferable to the iterated one at the shortest forecast horizon. However, the p-values indicate that these differences in the predictive ability are not statistically significant. When industrial production is forecasted, only the direct estimator outperforms the iterated benchmark in a few cases. Again, the difference in the predictive accuracy in these cases is so small that the null cannot be rejected, suggesting that the improvement from the direct estimator is too small to be of practical forecasting value. For both measures of economic activity, the constant adjustment to the iterated method and the full-adjustment to both the iterated and direct methods perform very poorly and they never improve upon the benchmark. Indeed, the iterated method produces statistically significantly more accurate forecasts than these three forms of intercept correction in the clear majority of cases.

Inspection of Panels C and D reveal that the conclusions are substantially different for the price series. Most importantly, the iterated method performs worse in relative terms when future inflation is forecasted. For the GDP deflator, the one-off and full-adjustment to the iterated model dominate the iterated benchmark, with one exception, regardless of the forecast horizon and lag selection method. These improvements are large and generally statistically significant. In particular, the relative MSFE value at $h = 4$ for the full-adjustment method when an AR(1) specification is used is 0.691, indicating a 30.9% improvement relative to the benchmark. The results also show that the performance of the constant adjustment to the iterated method, the direct method and the full-adjustment to the direct method relative to the iterated benchmark depends on the method of lag selection. The ability of these methods to forecast the future GDP deflator is superior to the iterated benchmark in the majority of cases when the AR(1) model is used. On the other hand, if the results for the AR(1) specification are excluded, the iterated method is almost universally preferred to these three alternative methods. The good performance of these three methods when the AR(1) model is considered is probably due to the fact that low order AR models do not capture the true dynamics of the GDP deflator and are hence misspecified. At least the AR(1) model yields less accurate forecasts than the other lag methods.

Table 5. MSFE values relative to the iterated benchmark based on the same lag selection method

		AR(1)				AR(4)				BIC				AIC			
Forecast horizon		2	4	8	12	2	4	8	12	2	4	8	12	2	4	8	12
(A) GDP	Iterated	2.664	2.719	2.723	2.688	2.716	2.786	2.720	2.682	2.673	2.724	2.720	2.685	2.663	2.724	2.722	2.686
	Constant	1.227	1.343	1.518**	1.742***	1.257	1.366	1.660**	1.908***	1.279	1.461*	1.654**	1.923***	1.275	1.426	1.603**	1.873***
	One-off	0.978	0.995	1.000	1.000	0.995	0.993	1.003	1.001	0.993	1.004	0.999	1.000	0.995	1.002	0.999	1.000
	Full	1.075	1.098	1.205	1.338***	1.114	1.082	1.239	1.385***	1.127	1.142	1.219	1.355***	1.127	1.126	1.199	1.337***
	Direct	0.978	1.019	1.015	0.999	1.022	1.010	1.040	1.015	0.971	1.023	1.017	1.001	1.006	1.038	1.021	1.005
	Full direct	1.217	1.387	1.375*	1.483***	1.236	1.347	1.491**	1.466***	1.208	1.384*	1.379*	1.486***	1.270*	1.405*	1.424**	1.481***
(B) Industrial production	Iterated	6.639	6.744	6.748	6.859	6.705	6.798	6.767	6.832	6.618	6.744	6.749	6.860	6.781	6.716	6.752	6.831
	Constant	1.504**	1.787***	1.575**	1.872***	1.424*	1.714**	1.653**	1.938***	1.505**	1.786***	1.571**	1.865***	1.395*	1.693**	1.619**	1.880***
	One-off	1.057	1.009	1.000	1.000	1.033	1.028	1.002	1.001	1.057	1.009	1.000	1.000	1.025	1.015	1.001	1.001*
	Full	1.272*	1.359**	1.195	1.388***	1.208	1.297*	1.253*	1.429***	1.271*	1.359**	1.195	1.386***	1.191	1.301*	1.247*	1.436***
	Direct	1.004	1.004	1.020	0.978	0.997	0.982	1.065	1.004	1.012	1.000	1.019	0.981	0.976	1.019	1.015	0.997
	Full direct	1.472**	1.596**	1.408**	1.657***	1.380*	1.497**	1.569***	1.593***	1.486**	1.584**	1.400**	1.655***	1.438*	1.594**	1.405**	1.684***
(C) GDP deflator	Iterated	1.479	1.779	2.279	2.444	1.231	1.396	1.851	2.122	1.289	1.436	1.910	2.186	1.246	1.412	1.861	2.124
	Constant	0.971	0.725	0.811	1.114	1.151	1.158	1.471*	2.040***	1.123	1.122	1.384	1.907**	1.123	1.142	1.474*	2.063***
	One-off	0.872**	0.838***	0.948***	0.983***	0.967	0.891***	0.937**	0.968*	0.955	0.878***	0.926***	0.964**	0.957	0.889***	0.935**	0.967*
	Full	0.857*	0.691***	0.718***	0.764***	1.015	0.866*	0.853**	0.900*	0.992	0.831**	0.813***	0.862**	0.995	0.856**	0.847**	0.896*
	Direct	0.886***	0.675***	0.756***	0.895	1.019	1.076***	1.196***	1.292**	0.965	1.037	1.080**	1.231*	1.005	1.059**	1.160***	1.318**
	Full direct	0.869	0.665*	0.857	1.127	1.096	1.155	1.487	1.757*	0.998	1.023	1.225	1.504	1.062	1.094	1.381	1.720*
(D) PCE inflation	Iterated	1.898	2.051	2.482	2.678	1.830	1.955	2.323	2.555	1.847	2.016	2.389	2.609	1.823	1.960	2.337	2.568
	Constant	1.364***	1.345*	1.397	1.629*	1.485***	1.658***	1.985***	2.404***	1.385***	1.473**	1.669**	2.078**	1.446***	1.604***	1.892***	2.334***
	One-off	1.061	0.940*	0.954***	0.980**	1.080**	1.012	0.969	0.982	1.061	0.990	0.962*	0.983	1.067*	1.005	0.969	0.984
	Full	1.129**	0.949	0.879**	0.867***	1.247***	1.085	0.986	0.958	1.168**	1.025	0.936	0.925	1.221**	1.078	0.976	0.953
	Direct	1.009	0.947	0.959	0.983	0.994	1.018	1.162**	1.146	1.003	0.967	1.036	1.070	0.997	1.016	1.121*	1.125
	Full direct	1.241**	1.079	1.127	1.446	1.333***	1.244	1.553*	1.785*	1.305**	1.166	1.220	1.552	1.330**	1.230	1.419	1.630

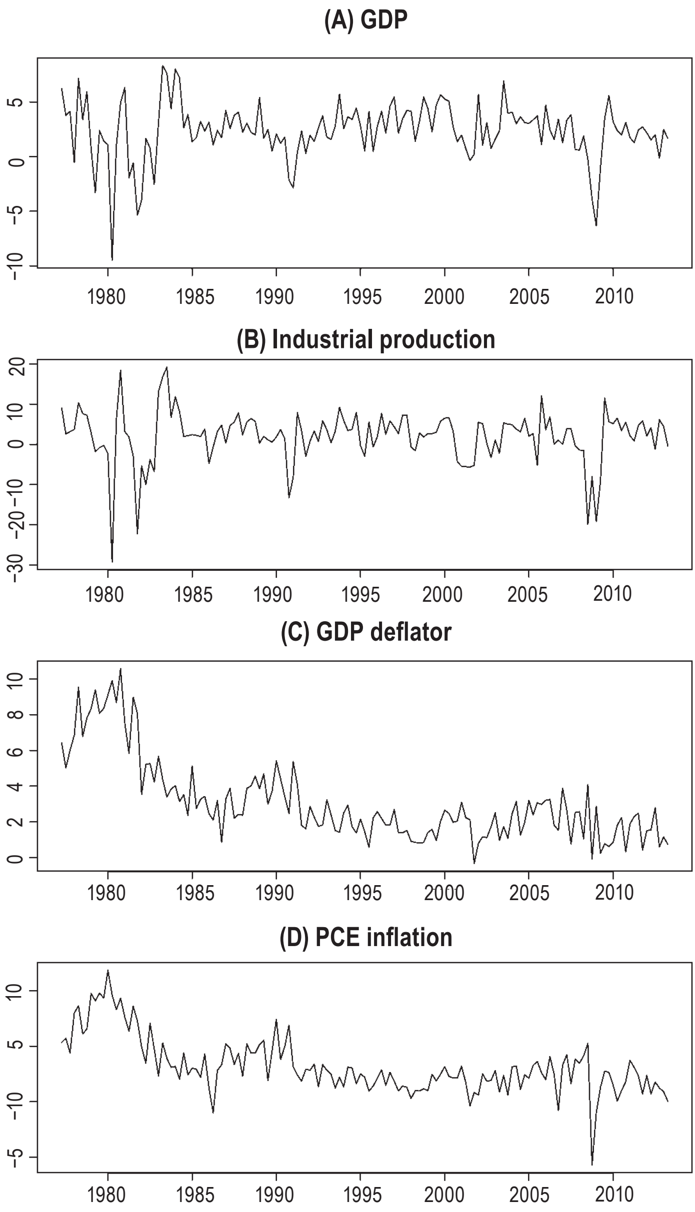
Notes: Forecast period spans from 1977:Q2 to 2013:Q2. The first row in each panel shows the root mean squared forecast error for the iterated benchmark. Subsequent rows report the ratio of the MSFE of each candidate multi-step method relative to the MSFE of the iterated benchmark. Intercept corrections are based on the average of the latest 4 forecast errors. Asterisks mark rejection of the two-sided Giacomini and White (2006) test at the 1%(***), 5%(**), and 10%(*) significance levels, respectively.

The evidence for the one-off and full-adjustment to the iterated method is less convincing when changes in PCE inflation are forecasted. These methods generate smaller forecast errors than the iterated benchmark at $h = 8$ and $h = 12$. Although the improvements are quite large, the null of equal accuracy is rejected at conventional significance levels only for the AR(1) model. In contrast, the one-off and full-adjustment to the iterated method produce higher MSFE values than the benchmark at $h = 2$, sometimes by quite a substantial margin. According to the p-values, the null is rejected in favor of the iterated benchmark at this horizon in six of eight cases. The direct estimator beats the iterated one when the forecasts are computed using an AR(1) model, but using longer lags in the forecasting model eliminates the advantage of the direct estimator, particularly at long horizons ($h = 8$ and $h = 12$). In contrast with the GDP deflator results, the constant-adjustment to the iterated method and the full-adjustment to the direct method never produce better PCE inflation forecasts than the iterated benchmark. Indeed, at the longest horizon $h = 12$, these methods are markedly worse than the benchmark.

All in all, the results in Table 5 indicate that the iterated method provides the most accurate real-time output forecasts, whereas the one-off and full-adjustment to the iterated method help improve the accuracy of the inflation forecast. Thus, there seems to be no single dominant multi-step forecasting method (cf. Marcellino *et al.*, 2006; Pesaran *et al.*, 2011). Figure 1 plots the quarterly growth rates of the four macroeconomic time series (at an annualized rate) over the out-of-sample period. The figure demonstrates that the series have undergone different types of structural breaks. In particular, it is well documented that the volatility of the real GDP and industrial production growth have decreased since the mid-1980s (see, e.g., McConnell and Perez-Quiros, 2000). The simulation results in Section 4 show that when the volatility changes, the iterated method performs well relative to the other multi-step methods. On the other hand, due to changes in monetary policy, both the mean and variance of the two inflation variables have decreased substantially since the early 1980s (Sims and Zha, 2006). The Monte Carlo results show that when both the mean and variance decrease substantially, e.g., when the autoregressive parameter of an AR(1) model decreases substantially (see Appendix A), the iterated method yields rather poor forecasts. Hence, the Monte Carlo results are very helpful in understanding why it is difficult to find a single multi-step method that dominates across all variables.

The results in Section 4 also suggest that the timing of the break affects the accuracy of the multi-step methods, implying that the relative forecasting performance might be time-varying in an unstable environment. To examine this possibility, Figure 2 plots the Giacomini and Rossi (2010) fluctuation test as well as the two-sided critical values at the 5% significance level (dashed horizontal lines) for an AR(4) model at $h = 4$. The

Figure 1. Quarterly growth rates



Notes: Sample period 1977:Q2–2013:Q2. The figure plots the first-release growth rates, annualized.

fluctuation test is implemented by using a centered rolling window of 40 observations. The truncation parameter is set to $P^{1/5} \approx 3$, where P denotes the number of out-of-sample observations.[11] Positive (negative) values of the test indicate that the candidate multi-step forecasting method has produced more (less) accurate forecasts than the iterated benchmark. If the fluctuation test statistic crosses either the upper or the lower critical value, the null of equal local predictive ability at each point in time is rejected.

Several results stand out. First, despite the large differences in the relative predictive ability reported in Table 5, the fluctuation test rejects the null of equal accuracy at each point in time only in three cases. Interestingly, the fluctuation test reveals that the one-off and full-adjustment to the iterated method contain substantial incremental real-time predictive information for the GDP deflator in the early 1980s. However, later in the sample, these two forms of intercept correction give less accurate forecasts than the iterated benchmark. Broadly speaking, these findings are consistent with the aforementioned observation that both the mean and variance of the GDP deflator have decreased substantially in the early 1980s. The simulation results in Tables 2–3 suggest that in the presence of large and recent decrease in both the mean and variance of a series only the one-off and full-adjustment to the iterated method of the five alternatives should dominate the benchmark (see the results for $T_1 = 99$). Furthermore, as time passes after the break, the gains from these two intercept corrections should diminish.

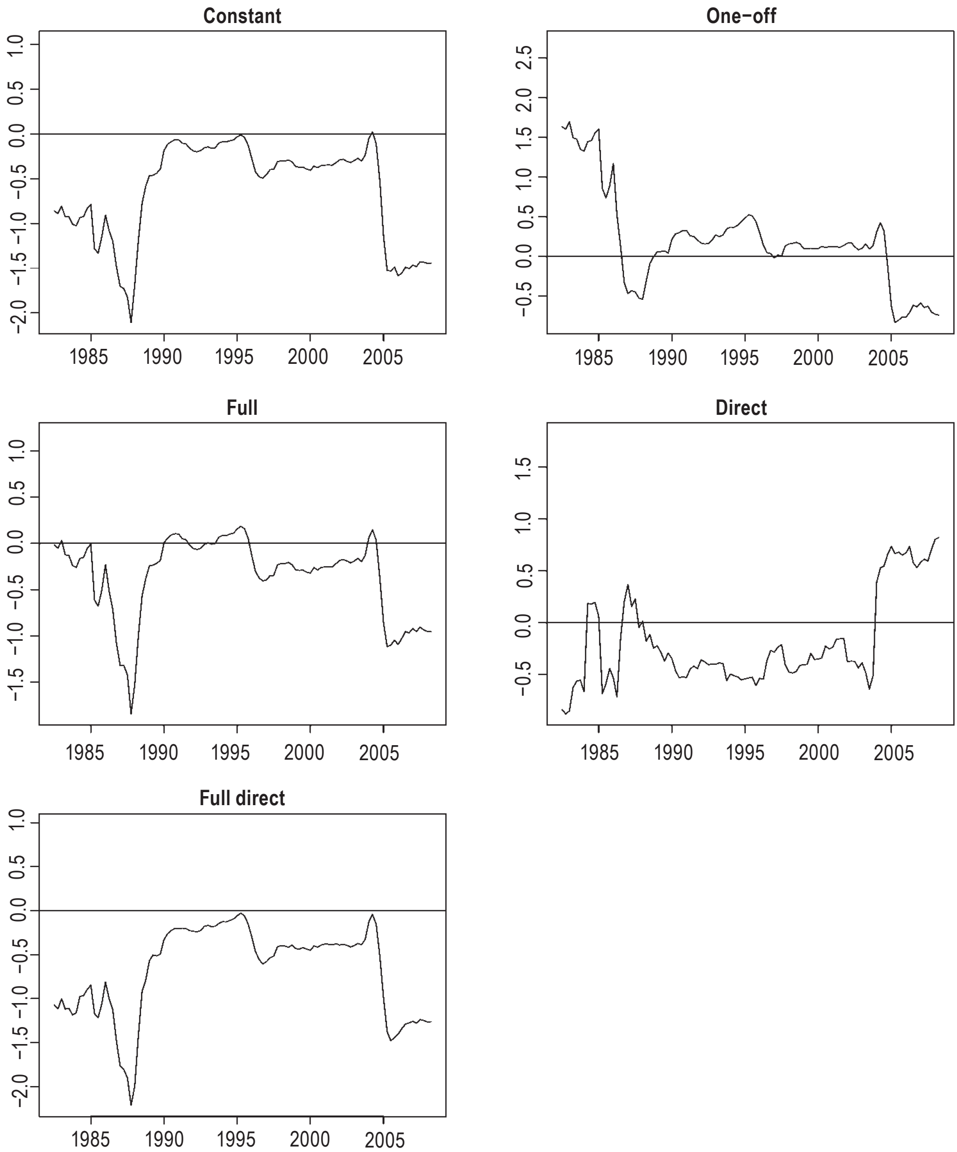
The fluctuation test for the two output variables show that the track record of the constant adjustment to the iterated method and the full-adjustment to both the iterated and direct method is not good. In fact, the fluctuation test implies that these methods yield systematically worse forecasts than the iterated benchmark over the whole out-of-sample period (the value of the test statistic is always negative), although the null of equal accuracy at each point in time cannot be rejected. Similarly, the direct estimator almost universally produces larger forecast errors for the price series than the iterated estimator.

Overall, the fluctuation test indicates that the alternative multi-step methods only episodically improve upon the iterated benchmark. Therefore, the results over the whole out-of-sample period might give a somewhat misleading picture of their predictive ability. Most notably, the one-off and full-adjustment to the iterated method do not systematically beat the iterated benchmark when GDP deflator is forecasted, but rather they perform particularly well only in the early 1980s. The empirical results, as well as the simulation results, support the view that the iterated method typically produces the most accurate real-time forecasts in unstable environment. However, the results also highlight that if both the mean and variance of the series decrease substantially and

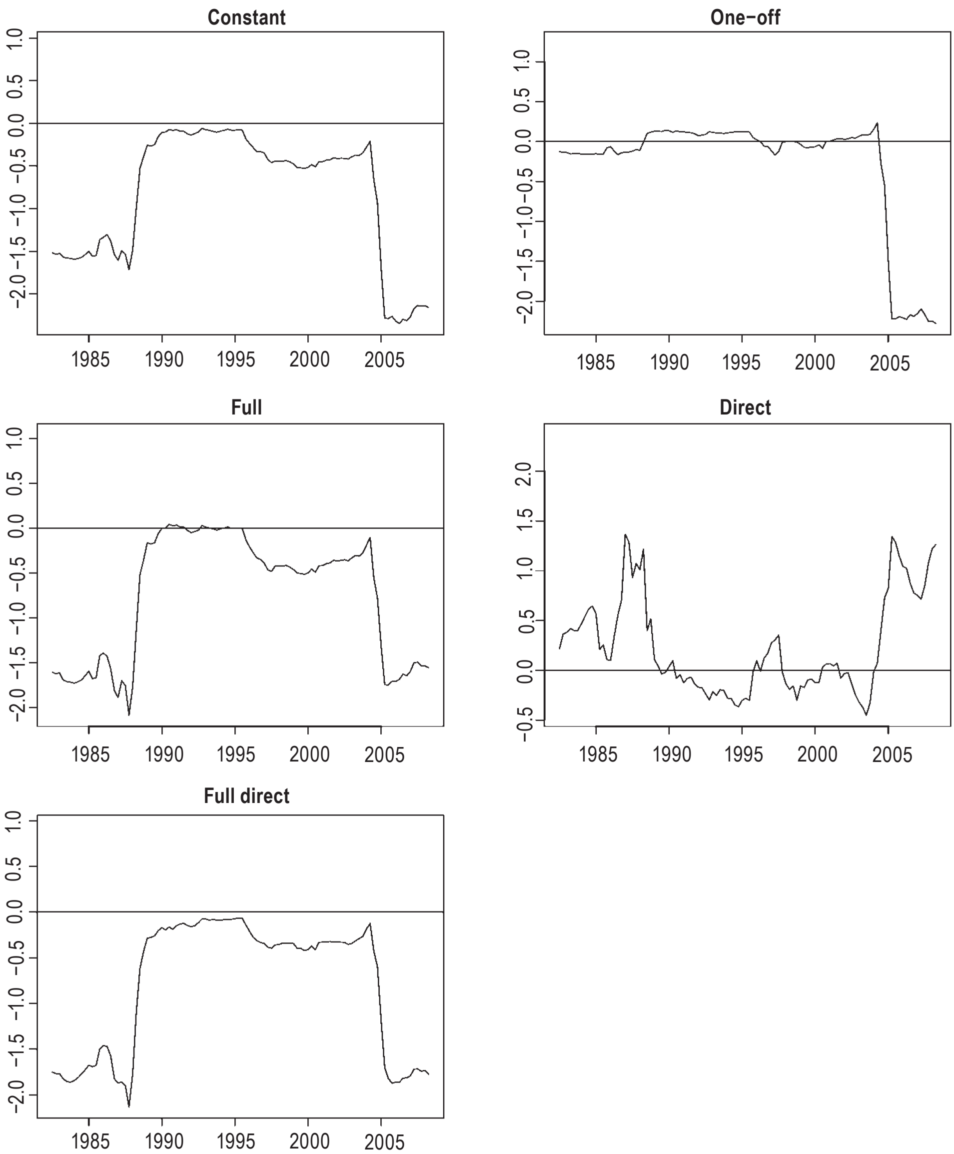
11 The length of the out-of-sample period is 145 observations when $h = 2$ and 135 when $h = 12$.

Figure 2. Fluctuation test for equal out-of-sample predictability at h = 4

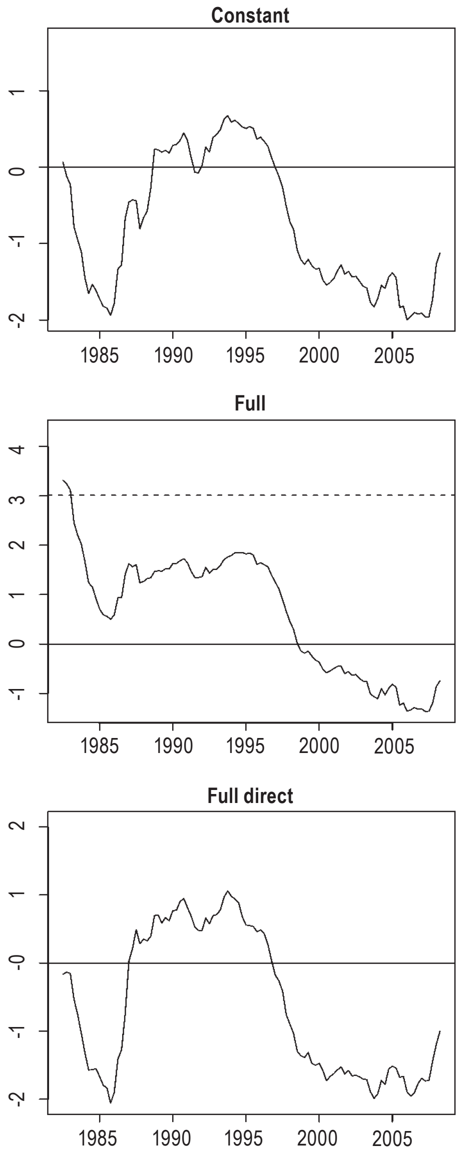
(A) GDP



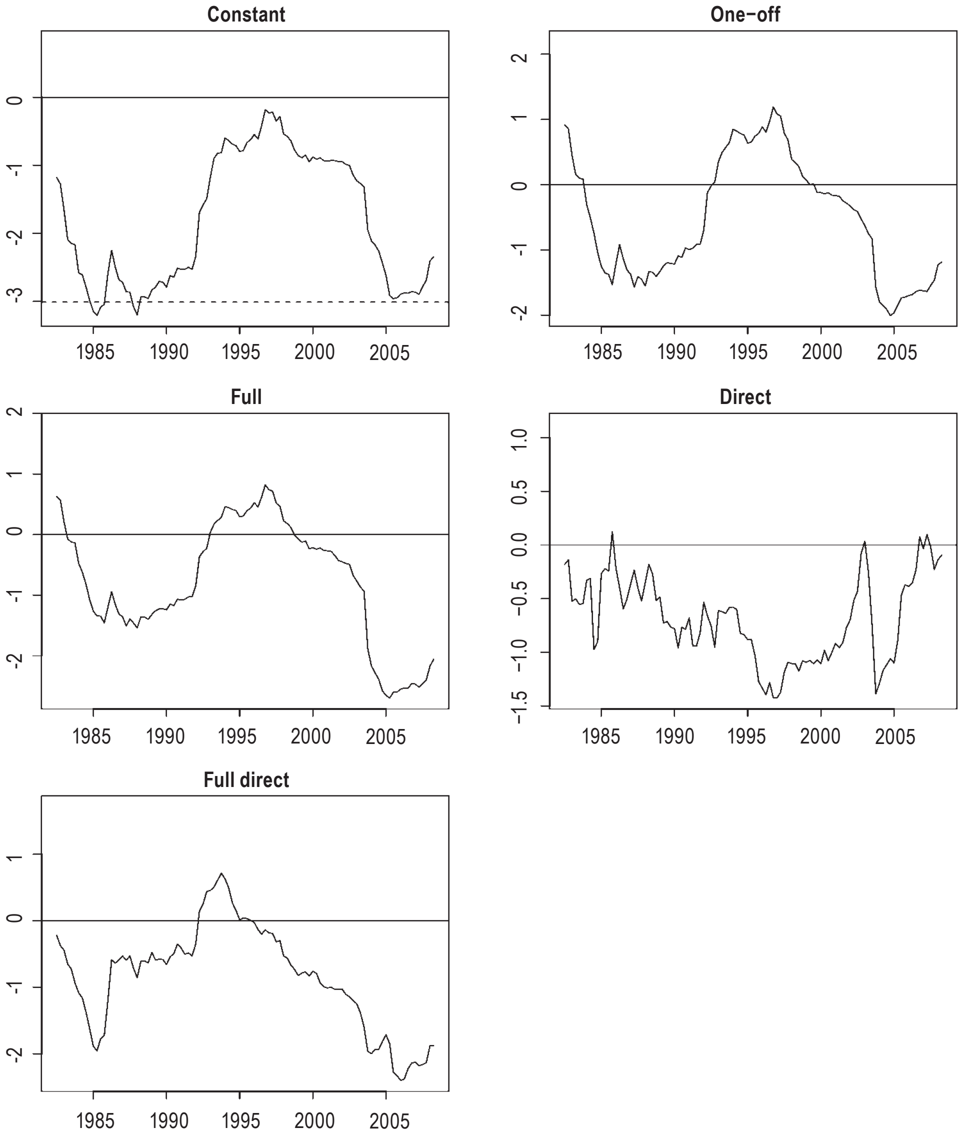
(B) Industrial production



(C) GDP deflator



(D) PCE inflation



Notes: The figure plots the two-sided Giacomini and Rossi (2010) fluctuation test based on sequences of the Giacomini and White (2006) unconditional test statistic for AR(4) specification. The test is implemented by using a centered rolling window of 40 observations. The sample period spans from 1977:Q4 to 2013:Q2. Positive (negative) values indicate that the candidate method has produced more (less) accurate forecasts than the benchmark. The dashed lines represent critical values at the 5% level. If the absolute value of the fluctuation test exceeds the critical value, the null that the two multi-step methods have equal predictive ability at each point in time is rejected.

the multi-step forecasts are made shortly after the break, the iterated method produces inaccurate forecasts and performs poorly in relative terms. In such a case, an alternative multi-step method, perhaps a one-off adjustment to the iterated method, should be used.

3.6 Conclusions

This paper analyzes the real-time performance of various multi-step forecasting methods in the presence of structural breaks. Our Monte Carlo and empirical analysis leads us to three main conclusions. First, our results suggest that the iterated method provides the most accurate multi-step forecasts in the presence of structural instability, especially if the parameters are subject to small or medium-size breaks. The good performance of the iterated method suggests that the error component dominates the bias component in the composition of MSFE values in an unstable environment. Second, the alternative multi-step methods, which are less prone to bias, have the most potential when the parameters are subject to large breaks and forecasts are made shortly after the break. Third, in the presence of breaks, the relative performance of the multi-step methods might be time-varying. For instance, it is only in the early 1980s that the one-off and full-adjustment to the iterated method provide more accurate GDP deflator forecasts than the iterated method.

The finding that the type as well as the timing of the break affects the relative merit of the multi-step methods is an intriguing one. The previous literature has found strong evidence for parameter instability in U.S. macroeconomic time series. These series have been subject to different types of breaks at different dates. This observation together with our findings might help explain why it is so difficult to find a single multi-step method that performs well across all variables at all time periods. Clearly, it would be interesting to analyze the time-variations further using the dataset of 170 U.S. monthly macroeconomic time series studied in Marcellino *et al.* (2006) and Pesaran *et al.* (2011).

References

Aruoba, S. B. (2008). Data revisions are not well-behaved. *Journal of Money, Credit and Banking*, 40, 319–340.

Bao, Y. (2007). Finite-sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution. *Econometric Theory*, 23, 767–773.

Brown, B. W. and Mariano, R. S. (1989). Measures of deterministic prediction bias in nonlinear models. *International Economic Review*, 30, 667–684.

Chevillon, G. and Hendry, D. F. (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting*, 21, 201–218.

Clements, M. P. and Galvão, A. B. (2013). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, 28, 458–477.
Clements, M. P. and Hendry, D. F. (1996a). Intercept corrections and structural change. *Journal of Applied Econometrics*, 11, 475–494.
Clements, M. P. and Hendry, D. F. (1996b). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, 58, 657–684.
Clements, M. P. and Hendry, D. F. (1998). Forecasting economic time series. Cambridge University Press: Cambridge.
Clements, M. P. and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C. W. J. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 1. Amsterdam: Elsevier, 605–657.
Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, 49, 72–100.
Eklund, J., Kapetanios, G. and Price, S. (2013). Robust forecast methods and monitoring during structural change. *Manchester School*, 81, 3–27.
Elliott, G. and Timmermann, A. (2008). Economic forecasting. *Journal of Economic Literature*, 46, 3–56.
Findley, D. F. (1985). Model selection for multi-step-ahead forecasting. In Baker, H. A. and Young, P. C. (Eds.), *Proceedings of the Seventh Symposium on Identification and System Parameter Estimation*. Oxford: Pergamon, 1039–1044.
Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25, 595–620.
Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.
Hoque, A., Magnus, J. R. and Pesaran, B. (1988). The exact multi-period mean square forecast error for the first-order autoregressive model. *Journal of Econometrics*, 39, 327–346.
Hännikäinen, J. (2014). Selection of an estimation window in the presence of data revisions and recent structural breaks. Tampere Economic Working Paper, No. 92, University of Tampere.
Ing, C-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory*, 19, 254–279.
Jacobs, J. P. A. M. and van Norden, S. (2011). Modeling data revisions: measurement error and dynamics of "true" values. *Journal of Econometrics*, 161, 101–109.
Mankiw, N. G. and Shapiro, M. D. (1986). News or noise: an analysis of GNP revisions. *Survey of Current Business*, 66, 20–25.
Marcellino, M., Stock, J. H. and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135, 499–526.
McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the United States: what has changed since the early 1980's? *American Economic Review*, 90, 1464–1476.
Pesaran, H. M., Pick, A. and Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, 164, 173–187.
Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129, 183–217.
Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier, 1203–1324.

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, 128, 99–136.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U. S. monetary policy? *American Economic Review*, 96, 54–81.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14, 11–30.

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

Weiss, A. A. (1991). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics*, 48, 135–149.

Appendix A

Table 6. Means and standard deviations

News Experiment	$E(\tilde{y}_{1t})$	$E(\tilde{y}_{2t})$	$E(y_{1t}^{t+1})$	$E(y_{2t}^{t+1})$	$\sigma_{\tilde{y}1t}$	$\sigma_{\tilde{y}2t}$	$\sigma_{y_{1t}^{t+1}}$	$\sigma_{y_{2t}^{t+1}}$
1	2.255	2.255	2.128	2.128	2.514	2.514	1.957	1.957
2	2.255	5.171	2.128	4.878	2.514	7.187	1.957	5.595
3	2.255	1.442	2.128	1.361	2.514	2.035	1.957	1.584
4	1.442	5.171	1.361	4.878	2.035	7.187	1.584	5.595
5	5.171	1.442	4.878	1.361	7.187	2.035	5.595	1.584
6	2.255	2.255	2.128	2.128	2.514	7.541	1.957	5.871
7	2.255	2.255	2.128	2.128	2.514	0.838	1.957	0.652
8	2.255	3.383	2.128	3.191	2.514	2.514	1.957	1.957
9	2.255	1.128	2.128	1.064	2.514	2.514	1.957	1.957
Noise Experiment	$E(\tilde{y}_{1t})$	$E(\tilde{y}_{2t})$	$E(y_{1t}^{t+1})$	$E(y_{2t}^{t+1})$	$\sigma_{\tilde{y}1t}$	$\sigma_{\tilde{y}2t}$	$\sigma_{y_{1t}^{t+1}}$	$\sigma_{y_{2t}^{t+1}}$
1	2.000	2.000	1.887	1.887	1.732	1.732	1.879	1.879
2	2.000	4.000	1.887	3.774	1.732	2.268	1.879	2.460
3	2.000	1.333	1.887	1.258	1.732	1.549	1.879	1.680
4	1.333	4.000	1.258	3.774	1.549	2.268	1.680	2.460
5	4.000	1.333	3.774	1.258	2.268	1.549	2.460	1.680
6	2.000	2.000	1.887	1.887	1.732	5.196	1.879	5.636
7	2.000	2.000	1.887	1.887	1.732	0.577	1.879	0.626
8	2.000	3.000	1.887	2.830	1.732	1.732	1.879	1.879
9	2.000	1.000	1.887	0.943	1.732	1.732	1.879	1.879

Chapter 4
Zero lower bound, unconventional monetary policy and indicator properties of interest rate spreads*

Jari Hännikäinen

Abstract

This paper re-examines the out-of-sample predictive power of interest rate spreads when the short-term nominal rates have been stuck at the zero lower bound and the Fed has used unconventional monetary policy. Our results suggest that the predictive power of some interest rate spreads have changed since the beginning of this period. In particular, the term spread has been a useful leading indicator since December 2008, but not before that. Credit spreads generally perform poorly in the zero lower bound and unconventional monetary policy period. However, the mortgage spread has been a robust predictor of economic activity over the 2003–2014 period.

Keywords: business fluctuations, forecasting, interest rate spreads, monetary policy, zero lower bound, real-time data

JEL codes: C53, E32, E44, E52, E58

---

* This paper is forthcoming in *Review of Financial Economics*. I would like to thank the editor, Gerald A. Whitney, two anonymous referees, Kari Heimonen, Heikki Kauppi, Petri Kuosmanen, Markku Lanne, Jani Luoto, Henri Nyberg, Jukka Pirttilä, Jari Vainiomäki, Juuso Vataja and seminar participants in Helsinki, Tampere and Jyväskylä for helpful comments and suggestions. Financial support from the FDPE is gratefully acknowledged.

4.1 Introduction

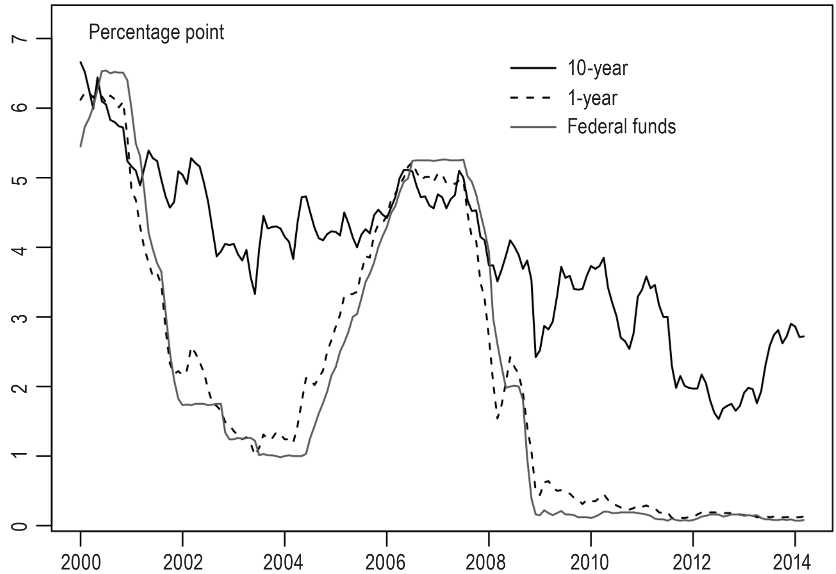
The empirical literature focusing on forecasting U.S. real macroeconomic variables has found that interest rate spreads have substantial predictive power for future economic activity. In particular, the term spread, i.e., the difference between the yields on long-term and short-term Treasury securities, has been identified as one of the most informative leading indicators (see, e.g., Stock and Watson, 2003). The term spread has predictive power because it is an indicator of the stance of monetary policy, which is an important driver of business cycles. The relationship between the term spread and future output growth is positive, i.e., higher spread indicates higher future growth.

The previous literature has also documented that various credit spreads contain significant information about subsequent real activity (see, e.g., Bernanke, 1990; Bernanke and Blinder, 1992; Faust *et al.*, 2013; Friedman and Kuttner, 1992, 1998; Gertler and Lown, 1999; Gilchrist *et al.*, 2009; Gilchrist and Zakrajšek, 2012; Mody and Taylor, 2003). Credit spread means either the difference between the yields on various corporate bonds and government bonds of comparable maturity or the difference between the yields on two private debt instruments differing with respect to their rating categories. Credit spreads are informative about future activity because they are indicators of changes in the supply of credit and market participants' expectations of default. They are also, at least to some extent, indicators of an effective monetary policy because the central bank's actions affect the supply of credit and the likelihood of defaults.

The predictive power of interest rate spreads varies over time. For example, it is a well-known fact that the ability of the term spread to forecast future economic activity has diminished since the mid-1980s (Stock and Watson, 2003 and the references cited therein). The changes in the predictive content of the term spread often correspond closely to major changes in the conduct of monetary policy (Bordo and Haubrich, 2008; Estrella *et al.*, 2003; Giacomini and Rossi, 2006). Therefore, regime shifts in monetary policy are potentially important for the predictive power of the term spread. Similarly, because credit spreads are, at least to some extent, indicators of the stance of monetary policy, changes in monetary policy may also affect their predictive ability.

The financial crisis in 2008 changed the Fed's monetary policy altogether. Prior to the crisis the federal funds rate – the Fed's traditional monetary policy instrument – was well above zero. Since December 2008, the federal funds rate has been essentially stuck at the zero lower bound (ZLB). Figure 1 demonstrates this fundamental change in monetary policy by plotting ten-year and one-year Treasury rates and the federal funds rate from 2000 through 2014. Although the federal funds rate has been at the

Figure 1. Treasury rates since 2000



Notes: Sample period 2000:M1–2014:M3. The data are extracted from the Federal Reserve Economic Data (FRED) (Federal Reserve Bank of St. Louis).

lower bound of zero[1], the recovery from the crisis has been slow. Therefore, the Fed has started to use unconventional monetary policies. The Fed has launched asset purchase programs, often referred to as quantitative easing, and used forward guidance. The aim of these two unconventional policies is to lower long-term rates and hence boost economic activity.

The fundamental change in monetary policy since December 2008 is potentially important for the predictive power of interest rate spreads for several reasons. First, in the non-ZLB environment, the term spread correlates negatively with the short-term rate and is uncorrelated with the long-term rate (see Table 2). In contrast, when the short-term rate is fixed at or near zero, the term spread fluctuates essentially one-for-one with the long-term rate. Second, related to the first reason, the possible values of the term spread are restricted when the short-term rate is fixed at the ZLB. In the non-ZLB period, when both the short-term and long-term rates fluctuate, the term spread can be negative, zero, or positive. When the short-term rate is fixed at or near zero, the term spread equals the long-term rate and can thus have only non-negative values. Third, as discussed in Krippner (2013), the term spread is a directionally misleading

1 Investors always have the option of holding cash, so interest rates cannot be reduced below zero.

measure of the stance of monetary policy in ZLB/unconventional monetary policy environments. Tight monetary policy periods in non-ZLB/conventional monetary policy environments have corresponded with low values of the term spread. However, in the ZLB/unconventional monetary policy environment since December 2008, the term spread decreases because the long-term rate falls while the short-term rate remains essentially fixed at the zero level. Hence, the decreasing spread could be misinterpreted as a tightening of monetary policy when actually the use of unconventional methods substantially eases monetary policy. Fourth, the long-term rate depends on the entire path of expected future short-term rates. Hence, if the short-term rates are assumed to be at the zero level for a sufficiently long period, the ZLB constraint on short-term rates should also affect the behavior of the long-term rates. However, Swanson and Williams (2014) find that, for instance, the ten-year Treasury rate was essentially unconstrained by the zero bound throughout 2008–2010. Since late 2011, the sensitivity of the ten-year Treasury rate to macroeconomic news has fallen, indicating that the long-term rate has been affected by the ZLB.[2] This finding suggests that the predictive ability of interest rate spreads depending on the long-term Treasury rate might have changed since the onset of the ZLB/unconventional monetary policy period.

The short-term rates in the U.S. have been effectively constrained by the ZLB only in the 1930s and since 2008. Although very low interest rates have been rare, Bernanke *et al.* (2004) and Chung *et al.* (2012) argue that the ZLB restriction is nowadays much more likely to become binding than in the past. The primary reason for this is the change in the way central banks conduct monetary policy. Modern central banks have adopted an inflation target and are thus committed to keeping inflation at a low level. Low and less volatile inflation has in turn allowed for lower interest rates. Low inflation and interest rates increase the probability that negative shocks will force the central bank to lower the short-term rate to the ZLB. As a consequence, we believe that empirical study of the leading indicator properties of interest rate spreads when the ZLB restriction is binding is highly worthwhile.

In this paper, we examine whether the ZLB and unconventional monetary policy has affected the real-time out-of-sample predictive power of the term spread and a set of credit spreads for U.S. industrial production. The main finding from this study is that the predictive content of the term spread has changed since the beginning of the ZLB/unconventional monetary policy period. We find that the term spread does not contain predictive power for future economic activity in non-ZLB/conventional monetary

2 Swanson and Williams (2014) offer two explanations for their findings. Until late 2011, market participants expected that the Fed would raise the short-term rate from zero within a few quarters, which minimized the effect of the ZLB on long-term Treasury rates. On the other hand, the unconventional monetary policy actions have helped offset the effects of the ZLB on long-term rates.

policy environments. However, the term spread is a useful leading indicator in the ZLB/unconventional monetary policy period. Thus, our results support the view that changes in monetary policy affect the predictive ability of the term spread (see Estrella, 2005). The results also indicate that the mortgage spread (i.e., the difference between the 30-year mortgage rate and ten-year Treasury bond rate) is a particularly informative leading indicator. It is a robust predictor of industrial production growth across a variety of sample periods and forecast horizons. The mortgage spread systematically contains predictive power in our real-time forecasting exercise both in the non-ZLB/conventional monetary policy and ZLB/unconventional monetary policy periods.

The remainder of the paper is organized as follows. In Section 2, we describe the econometric methodologies. Section 3 presents the empirical results, and Section 4 contains concluding remarks.

4.2 Methodology

In this section, we briefly describe the econometric methodologies used in this paper. The purpose of this study is to examine whether different spreads forecast future economic activity in the ZLB/unconventional monetary policy period.[3] In order to analyze this question, we follow Rossi (2013) and Stock and Watson (2003) and estimate the following linear, horizon-specific *h*-step ahead regression model:

$$Y_{t+h}^{h} = \beta_0 + \sum_{i=0}^{p} \beta_{1i} X_{t-3i} + \sum_{j=0}^{q} \beta_{2j} Y_{t-j} + u_{t+h}^{h}, \qquad t = 1, ..., T \tag{1}$$

where the dependent variable and the lagged dependent variable are $Y_{t+h}^{h} = (1200/h)\, ln(IP_{t+h}/IP_t)$ and $Y_{t-j} = 400 ln(IP_{t-3j-1}/IP_{t-3j-4})$, respectively, IP_t is the industrial

3 Monthly industrial production is used to gauge the state of the economy. The most frequently used measure of economic activity in the previous literature is the quarterly GDP. In our case, the number of observations is important because the ZLB/unconventional monetary policy period is relatively short (running from December 2008 to March 2014). Therefore, monthly industrial production is more appropriate for our purposes.

production at month t^4, X_t is the candidate predictor, and u^h_{t+h} is an error term.[5] The forecast horizon h is chosen such that we forecast economic activity one, two, three, and four quarters ahead (i.e., $h = 3,6,9,12$). The forecasting regression (1) is estimated by OLS.

We evaluate the forecasting performance of various interest rate spreads using a real-time out-of-sample forecasting exercise. We follow the procedure proposed by Stock and Watson (2003) and allow the lags of Y_t to vary between zero and four and the lags of X_t to vary between one and four in the forecasting model (1) (so we have 20 different models for each interest rate spread). At each forecast origin, the model with the lowest Bayesian information criteria (BIC) is chosen. Unlike Stock and Watson (2003), we use a rolling estimation scheme. This estimation scheme is more appropriate for our purposes than a recursive scheme for two reasons. First, as Giacomini and White (2006) point out, when the forecasting model is misspecified, it is often the case that a limited memory estimator provides more reliable forecasts than an expanding window estimator. Second, tests of equal predictive ability (discussed below) require limited memory estimators and thus rule out the recursive estimation scheme.

A standard way to quantify out-of-sample forecast performance is to compute the mean squared forecast error (MSFE) of a candidate forecast relative to a benchmark. Because the growth rate of industrial production is serially correlated and thus its own past values are themselves informative about future industrial production growth, it is natural to use an autoregressive (AR) model as a benchmark. The results from the literature indicate that it is relatively hard to outperform the AR benchmark (see, e.g., Elliott and Timmermann, 2008; Rossi, 2013; Stock and Watson, 2003). For the benchmark model, we consider lags between one and four and again choose the optimal lag length at each forecast origin with the BIC. If the relative MSFE is less than one, the model with the spread has produced more accurate forecasts than the AR benchmark. This implies that the spread contains marginal predictive power. However, the difference in the predictive content might not be statistically significant. The relative MSFE could be less than one simply because of sampling variability. Thus, we need more formal test procedures for deciding which spreads contain predictive power.

4 The one month publication lag in the industrial production series is taken into account. We use quarterly lags instead of monthly lags because we want to include information from the latest year to the forecasting regression and still keep the model relatively parsimonious.

5 Alternatively, we could use univariate regression equations including only current and lagged values of the candidate predictor as regressors. However, this approach has an important shortcoming: the industrial production series is serially correlated and thus its own past values are themselves useful predictors. By including the lagged values of the dependent variable, we consider the marginal predictive power of the spreads, i.e., whether they have predictive content for Y^h_{t+h} when its own past values Y_t are already taken into account.

In our setting, forecast evaluation is complicated by the fact that both the model using the spread and the benchmark model have a recursive BIC lag length selection. This implies that we might possibly use both nested and non-nested models when generating a sequence of out-of-sample forecasts. The Giacomini and White (2006) test of equal conditional predictive ability and test of equal unconditional predictive ability allow the comparison of both nested and non-nested models as well as models that change from time to time and are thus appropriate for our purposes.

The test of equal unconditional predictive ability tests the null hypothesis that the two forecasting methods are equally accurate on average over the out-of-sample period. Rejection of the null hypothesis implies that one of the two methods produces on average more accurate forecasts than the other method. On the other hand, the test of equal conditional predictive ability examines whether some available information (above and beyond past average behavior) can be used to predict which forecast will be more accurate for a specified future date. Under the null hypothesis the two methods are equally accurate and thus one cannot predict which method will be more accurate using the information in the conditioning set. Rejection of the null hypothesis indicates that the conditioning information (e.g., some feature of the economy) can be used to decide which forecasting method is preferable at each forecast origin. Because we are interested in analyzing whether the ZLB and unconventional monetary policy change the predictive ability of different spreads, we condition the relative predictive ability on an indicator taking the value of one when the ZLB restriction is binding and zero otherwise.[6] In our case, the null hypothesis states that the forecasting model using the spread and the AR benchmark have equal predictive ability regardless of whether the short-term rate is at the ZLB or not.

Giacomini and Rossi (2010) point out that the relative forecasting performance may change over time in unstable environments. In such a case, average relative performance over the whole out-of-sample period may hide important information and even lead to incorrect conclusions. We analyze time variations in the relative forecasting performance using methods developed by Giacomini and Rossi (2010). Their fluctuation test is simply the Giacomini and White (2006) test of equal unconditional predictive ability computed over a rolling out-of-sample window size of m. This fluctuation test examines whether the local relative forecasting performance of the methods is equal at each point in time. Under the null hypothesis the two methods yield equally accurate forecasts at each point in time. If the null hypothesis is rejected, one of the methods outperformed its competitor at some point in time.

6 In other words, we use the test function $h_t = (1, ZLB_t)'$, where ZLB_t is a dummy variable that takes a value of one when the ZLB restriction is binding (2008:M12–2014:M3) and zero otherwise.

4.3 Empirical results

This section describes the data and summarizes our empirical results. The sample period runs from 1987:M9 to 2014:M3. Different vintages of an industrial production series used in an out-of-sample forecasting exercise were obtained from the Philadelphia Fed's real-time database. The monthly interest rate data were obtained from the St. Louis Fed's FRED database.[7] Definitions of the alternative spreads used in this paper are given in Table 1. The first ten of these spreads have been frequently used in the literature. The inclusion of the last spread, namely the mortgage spread, is motivated by the recent work of Hall (2011) and Walentin (2014). Using a SVAR model, Walentin (2014) shows that mortgage spread shocks have sizeable effects on the macroeconomy. However, the predictive power of the mortgage spread has not been analyzed in the literature. The mortgage spread is potentially informative about future growth because it is an indicator of changes in the supply of credit in the residental mortgage markets.[8]

Table 1. Definitions of the variables

Series label	*Definition*
TS10y.3m	Treasury bond (10 years) — Treasury bill (3 months)
TS10y.1y	Treasury bond (10 years) — Treasury bill (1 year)
TS10y.Ffs	Treasury bond (10 years) — Federal funds rate (overnight)
TS1y.3m	Treasury bill (1 year) — Treasury bill (3 months)
Paper.bill	Commercial paper (3 months) — Treasury bill (3 months)
Aaa.10y	Long-term corporate bond (Aaa rating) — Treasury bond (10 years)
Baa.10y	Long-term corporate bond (Baa rating) — Treasury bond (10 years)
Baa.Aaa	Long-term corporate bond (Baa rating) — long-term corporate bond (Aaa rating)
Hy.10y	High-yield bond — Treasury bond (10 years)
Hy.Aaa	High-yield bond — long-term corporate bond (Aaa rating)
Mortgage	Mortgage rate (30 years) — Treasury bond (10 years)

We start our analysis by considering correlations between the spreads and the federal funds rate, ten-year Treasury bond rate, and 3- and 12-month-ahead industrial production growth. Table 2 shows the correlations both in the non-ZLB/conventional monetary policy period (1987:M9–2008:M11) and in the ZLB/unconventional monetary policy

7 The Merrill Lynch U.S. High-Yield Master II index for the period 1986:M9–1996:M12 is taken from Mark Watson's webpage. During this period the high-yield index is the last daily observation of the month.

8 We follow Hall (2011) and calculate the mortgage spread as the difference between 30-year mortgage rate and 10-year Treasury bond rate. Hall (2011) points out that the 10-year Treasury bond provides a close match to the actual duration of the 30-year fixed rate mortgage. Therefore, the mortgage spread does not contain term premium, but rather it is a credit spread.

period (2008:M12–2014:M3).[9] Several results stand out. First, as one might expect, the federal funds rate and the ten-year Treasury rate are positively correlated in the non-ZLB/conventional monetary policy period. Due to the fact that the federal funds rate has been fixed at or near zero since December 2008, the federal funds rate and the ten-year Treasury rate are uncorrelated in the ZLB/unconventional monetary policy period. Interestingly, the ten-year Treasury rate is positively correlated with 3- and 12-month-ahead industrial production growth both in the non-ZLB and ZLB environments. Thus, a higher long-term rate indicates higher future growth. On the other hand, the federal funds rate is generally uncorrelated with future industrial production growth. Second, and most importantly, the correlation coefficients presented in Table 2 suggest that the behavior of the term spread has changed fundamentally since the beginning of the ZLB/unconventional monetary policy period. The term spreads (with the exception being the TS1y.3m spread based on two short-term rates) are negatively correlated with the federal funds rate but uncorrelated with the ten-year Treasury rate in the non-ZLB period. Thus, changes in the term spreads mostly reflect changes in the federal funds rate during this period. By contrast, in the ZLB period when the federal funds rate has been fixed at or near zero, the term spreads vary essentially one-for-one with the ten-year Treasury rate. The results indicate that the term spreads are significantly correlated with 12-month-ahead industrial production growth in both periods. However, correlations are much stronger in the later period. The term spreads are correlated with 3-month-ahead industrial production growth only in the ZLB period, probably because in the ZLB period term spreads fluctuate one-for-one with the ten-year Treasury rate, which itself is correlated with 3-month-ahead industrial production growth. The changes in the correlations suggest that the predictive power of the term spreads might have changed since the beginning of the ZLB/unconventional monetary policy period. Third, correlations between credit spreads and the federal funds rate and the ten-year Treasury rate have in some cases changed, but these changes are less dramatic. In general, credit spreads are significantly correlated with both 3- and 12-month-ahead industrial production growth.

Next, we evaluate whether the various interest rate spreads contain predictive power in a real-time out-of-sample forecasting exercise. We consider first the whole out-of-sample period running from 2003:M6 to 2014:M3. The results for this period are summarized in Table 3. The first row provides the root MSFE of the benchmark AR

9 In December 2008, the Fed set the federal funds rate to a range of 0% to 0.25%, where the federal funds rate has remained since then. The first large-scale asset purchase program, commonly referred to as QE1, was announced on November 25, 2008. However, the program was formally launched on December 16, 2008.

model.[10] For the subsequent rows, the first line reports the MSFE of a forecasting model using both the lagged values of industrial production growth and a candidate spread relative to the MSFE of the benchmark model. Values less (more) than one indicate that the model with a candidate spread has produced more (less) accurate forecasts than the benchmark, implying that the spread contains (does not contain) marginal predictive power. The *p*-value of the one-sided Giacomini and White (2006) test of equal unconditional predictive ability is reported in parentheses.[11]

The results reported in Table 3 suggest that the mortgage spread is a particularly informative leading indicator. The mortgage spread contains statistically significant predictive power for all four forecast horizons. Furthermore, its ability to forecast future industrial production growth is superior to all other spreads, regardless of the forecast horizon. The results also show that the difference between the Aaa corporate bond rate and the ten-year Treasury bond rate (i.e., the Aaa.10y spread) is a useful predictor of industrial production growth, although the null of equal accuracy cannot be rejected at conventional significance levels. The evidence for the rest of the credit spreads is mixed, but none of these spreads contains predictive power across all horizons. Various measures of the term spread also perform relatively poorly in the real-time forecasting exercise. Indeed, only in a few cases does inclusion of the term spread increase forecast accuracy. This result is interesting because the literature has identified the term spread as one of the most informative leading indicators (see, e.g., Stock and Watson, 2003).

10 Forecast errors are calculated using the latest available data, i.e., the vintage of April 2014. The results are qualitatively similar if forecast errors are computed using the first available real-time vintages of data.

11 As discussed in Rossi (2013), different estimation window sizes may lead to different empirical results. We check the robustness of our results by considering three different rolling window sizes, namely 120, 150, and 180 observations. The results are similar for the three rolling windows, and hence we report the results for the rolling window of 150 observations only.

Table 2. Correlations between the spreads and the federal funds rate, the 10-year Treasury rate, and industrial production growth

	Fed funds rate		10-year Treasury		IP_{t+3}		IP_{t+12}	
	pre-ZLB	ZLB	pre-ZLB	ZLB	pre-ZLB	ZLB	pre-ZLB	ZLB
Fed funds rate	1.00***	1.00***	0.79***	0.19	0.10	-0.26**	0.04	-0.05
10-year Treasury	0.79***	0.19	1.00***	1.00***	0.22***	0.28**	0.24***	0.54***
TS10y.3m	-0.55***	0.12	0.06	1.00***	0.07	0.30**	0.16***	0.54***
TS10y.1y	-0.66***	0.06	-0.10*	0.98***	0.01	0.38***	0.12**	0.55***
TS10y.Ffs	-0.66***	0.13	-0.06	1.00***	0.10*	0.30**	0.22***	0.54***
TS1y.3m	0.15***	0.41***	0.54***	0.60***	0.21***	-0.25**	0.21***	0.26*
Paper.bill	0.43***	0.21*	0.33***	0.02	-0.34***	-0.72***	-0.44***	-0.52***
Aaa.10y	-0.58***	0.16	-0.59***	-0.33***	-0.39***	-0.74***	-0.32***	-0.35***
Baa.10y	-0.50***	0.36***	-0.46***	-0.10	-0.52***	-0.73***	-0.43***	-0.28**
Baa.Aaa	-0.16***	0.41***	-0.07	0.00	-0.49***	-0.67***	-0.41***	-0.23*
Hy.10y	-0.08	0.37***	-0.05	0.07	-0.60***	-0.66***	-0.46***	-0.19
Hy.Aaa	0.06	0.38***	0.08	0.10	-0.59***	-0.64***	-0.44***	-0.17
Mortgage	-0.13**	0.01	-0.34***	-0.53***	-0.47***	-0.72***	-0.53***	-0.59***

Notes: Sample period: Monthly data from 1987:M9 to 2014:M3. The pre-ZLB period runs from 1987:M9 to 2008:M11 and the ZLB period spans from 2008:M12 to 2014:M3. *, **, *** denote statistically significant at 10, 5, 1% levels, respectively.

Table 3. Out-of-sample mean squared forecast errors

Spread	*h* = 3	*h* = 6	*h* = 9	*h* = 12
Uni.	6.55	6.53	6.28	5.95
TS10y.3m	1.03	1.02	1.09	1.02
	(0.83)	(0.69)	(0.79)	(0.62)
TS10y.1y	1.00	1.00	0.97	0.99
	(0.50)	(0.47)	(0.16)	(0.47)
TS10y.Ffs	1.01	1.08	1.09	1.00
	(0.68)	(0.78)	(0.72)	(0.49)
TS1y.3m	1.18	1.16	1.09	1.02
	(0.94)	(0.84)	(0.75)	(0.57)
Paper.bill	0.96	1.07	1.01	1.01
	(0.35)	(0.64)	(0.53)	(0.54)
Aaa.10y	0.93	0.92	0.94	0.98
	(0.14)	(0.15)	(0.24)	(0.41)
Baa.10y	0.96	1.10	1.02	0.89
	(0.39)	(0.67)	(0.58)	(0.17)
Baa.Aaa	0.95	1.15	1.24	1.17
	(0.33)	(0.73)	(0.78)	(0.74)
Hy.10y	0.94	1.13	1.16	1.09
	(0.31)	(0.73)	(0.90)	(0.88)
Hy.Aaa	0.97	1.22	1.22	1.11
	(0.39)	(0.81)	(0.90)	(0.92)
Mortgage	0.69	0.61	0.62	0.67
	(0.01)	(0.02)	(0.04)	(0.05)

Notes: Out-of-sample forecasting period runs from 2003:M6 to 2014:M3. The first row shows the root mean squared forecast error for the univariate autoregression. In subsequent rows, the first line reports the ratio of the MSFE of a candidate model relative to the MSFE of the benchmark model; the *p*-value of the one-sided Giacomini and White (2006) test of equal unconditional predictive ability is reported in parentheses. The truncation lag for the Newey-West (1987) HAC estimator is *h*-1, where *h* is the forecast horizon.

The results reported in Table 3 focus on average predictive power over the whole out-of-sample period. However, the purpose of this study is to examine whether the ZLB and unconventional monetary policy affect the predictive content of different spreads. In order to analyze this question, we divide the out-of-sample period into two parts. The first period runs from 2003:M6 to 2008:M11 and it characterizes a period with normal monetary policy. The second period spans from 2008:M12 to 2014:M3. During this second period, short-term interest rates have been stuck at the ZLB and the Fed has used unconventional monetary policy. The results for these two subperiods are summarized in Table 4. The first row provides the root MSFE of the benchmark AR model in the two sample periods. In subsequent rows, the first line reports the MSFE of a forecasting model using a candidate spread relative to the MSFE of the benchmark model in the first subperiod; the second line reports the relative MSFE in the second period; and the third line reports the *p*-value of the Giacomini and White (2006) test of equal conditional predictive ability. This test is implemented by conditioning the relative predictive ability on an indicator taking the value of one when the short-term rate has been at the ZLB (2008:M12–2014:M3) and zero otherwise. Under the null hypothesis the model with the spread and the benchmark model have equal predictive ability regardless of whether the short-term rate is at the ZLB or not.

The results for the term spread models are particularly interesting. The results suggest that the predictive power of the term spread differs substantially in the two subperiods. In the first period, the relative MSFE values are above one, indicating that the term spreads do not contain predictive power in the non-ZLB/conventional monetary policy environment.[12] However, later in the sample when the short-term rate has been fixed at the ZLB and the Fed has used unconventional policies, the term spreads have predictive power for future industrial production growth (the relative MSFE values are below one). The change in the predictive power is in most cases statistically significant and especially large when the forecast horizon is long (i.e., $h = 9$ and 12). Thus, the results support the view that changes in monetary policy matter for the predictive power of the term spread (see, e.g., Estrella, 2005; Giacomini and Rossi, 2006).

On the other hand, the predictive ability of the mortgage spread seems to be unaffected by the change in monetary policy that took place in late 2008. The mortgage spread is the best leading indicator in both subperiods. It produces the most accurate real-time forecasts in each of the eight forecast horizon/sample period combinations considered. Interestingly, inclusion of the mortgage spread substantially improves forecast accuracy. For instance, the 9-month-ahead forecast based on the lagged values of industrial production growth and the mortgage spread have a relative MSFE of 0.43 in the second period, indicating a 57% improvement relative to the AR benchmark.

12 This finding is consistent with the results presented in Ng and Wright (2013) and Rossi and Sekhposyan (2010). They find that the predictive ability of the term spread has diminished since the mid-1980s. In particular, their results show that the term spread does not contain marginal predictive power in the 1990s and early 2000s.

Table 4. Tests of equal conditional predictive ability

Spread	h = 3	h = 6	h = 9	h = 12
Uni.	6.89	7.12	7.11	6.85
	6.17	5.79	5.11	4.57
TS10y.3m	1.08	1.07	1.19	1.12
	0.96	0.92	0.87	0.76
	(0.10)	(0.03)	(0.05)	(0.03)
TS10y.1y	1.03	1.03	1.02	1.06
	0.96	0.95	0.87	0.80
	(0.27)	(0.21)	(0.13)	(0.04)
TS10y.Ffs	1.08	1.21	1.23	1.11
	0.93	0.87	0.75	0.70
	(0.04)	(0.04)	(0.04)	(0.05)
TS1y.3m	1.30	1.29	1.19	1.11
	1.03	0.94	0.86	0.75
	(0.30)	(0.27)	(0.40)	(0.57)
Paper.bill	0.92	0.89	0.88	0.92
	1.02	1.37	1.30	1.27
	(0.33)	(0.38)	(0.41)	(0.42)
Aaa.10y	1.05	0.99	1.04	1.09
	0.77	0.81	0.71	0.66
	(0.11)	(0.27)	(0.32)	(0.09)
Baa.10y	0.88	0.84	0.89	0.93
	1.08	1.55	1.35	0.77
	(0.50)	(0.36)	(0.47)	(0.64)
Baa.Aaa	0.84	0.84	0.85	0.86
	1.10	1.68	2.14	2.03
	(0.23)	(0.21)	(0.22)	(0.26)
Hy.10y	0.91	0.92	1.04	1.11
	0.98	1.50	1.44	1.04
	(0.86)	(0.60)	(0.41)	(0.34)
Hy.Aaa	0.94	0.94	1.03	1.08
	1.01	1.70	1.64	1.19
	(0.92)	(0.55)	(0.39)	(0.35)
Mortgage	0.76	0.70	0.70	0.73
	0.59	0.47	0.43	0.51
	(0.04)	(0.13)	(0.22)	(0.25)

Notes: The first out-of-sample forecasting period runs from 2003:M6 to 2008:M11 and the second from 2008:M12 to 2014:M3. The first row provides the root MSFE for the univariate autoregression in the two sample periods. In subsequent rows, the first line reports the MSFE of a candidate model relative to the MSFE of the benchmark model in the first period; the second line reports the relative MSFE in the second period; the *p*-value of the Giacomini and White (2006) test of equal conditional predictive ability is reported in parentheses. The test function is $h_t = (1, ZLB_t)'$, where ZLB_t is a dummy variable taking the value of one when the ZLB restriction is binding (2008:M12–2014:M3) and zero otherwise.

The effect of the ZLB restriction/unconventional monetary policy on the predictive content of the rest of the credit spreads is somewhat mixed. The difference between the Aaa corporate bond rate and the ten-year Treasury bond rate (the Aaa.10y spread) has predictive power for future industrial production only in the ZLB/unconventional monetary policy period. In general, however, the results indicate that credit spreads perform well in the first period but perform poorly in the second period. Although the differences in the relative MSFE values are large, the null of equal conditional predictive ability cannot be rejected at conventional significance levels. Note that some credit spreads (e.g., the Baa-Aaa corporate bond spread) perform poorly, whereas some credit spreads (e.g., the Aaa.10y and Mortgage spread) perform well in the ZLB/unconventional monetary policy period. Hence, no consensus on how the ZLB restriction and unconventional monetary policy affect the real-time predictive power of credit spreads emerges. This is probably due to the fact that credit spreads do not depend directly on the short-term rate and are thus only weakly correlated with the stance of monetary policy. Changes in the structure of the credit market are potentially more important for the predictive power of credit spreads than changes in monetary policy.

So far we have assumed that the forecasting ability of the interest rate spreads either remains constant over time (Table 3) or differs in the non-ZLB/conventional monetary policy and ZLB/unconventional monetary policy periods (Table 4). However, Giacomini and Rossi (2010) point out that the forecasting performance may be time varying. In such a case, average performance (either unconditional or conditional) over the whole out-of-sample period may hide important information and even lead to incorrect conclusions. Thus, we next consider the Giacomini and Rossi (2010) fluctuation test robust to instabilities. The fluctuation test is implemented by using a centered rolling window of 45 observations. We focus on the shortest 3-month-ahead forecast horizon because we want to maximize the number of out-of-sample observations for the ZLB/unconventional monetary policy period. Figure 2 reports both the fluctuation test statistic as well as the one-sided critical value at the 5% significance level (dashed horizontal line). Positive (negative) values of the fluctuation test indicate that the interest rate spread model has produced more (less) accurate forecasts than the AR benchmark. If the value of the fluctuation test exceeds the critical value, the null of equal local predictive ability at each point in time is rejected.

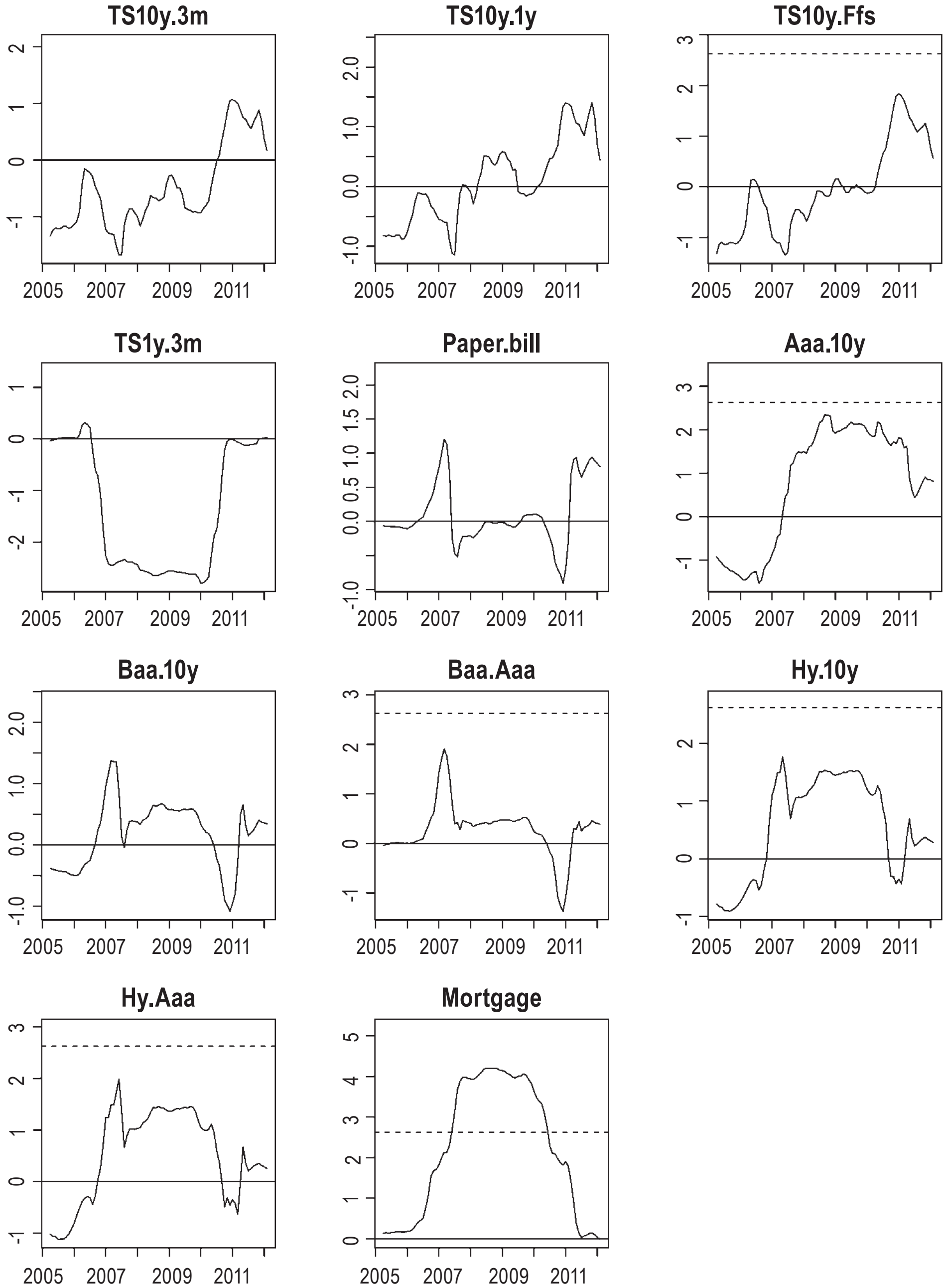
Inspection of Figure 2 reveals interesting details concerning the predictive ability of the term spread. At the beginning of the out-of-sample period, various term spread models typically produce larger MSFE values than the AR benchmark, implying that term spreads do not contain predictive power. Recently, however, the term spreads (with the exception being the TS1y.3m spread) have been informative leading indicators.

For windows centered since early 2010, inclusion of the term spread improves forecast accuracy. Therefore, the fluctuation test suggests that the predictive power of the term spread has changed. The timing of this change corresponds closely to the beginning of the ZLB/unconventional monetary policy period.

The fluctuation test shows that the good performance of the mortgage spread reported in Tables 3 and 4 is not due to some specific subperiod. The forecasting model using both the lagged values of industrial production growth and the mortgage spread systematically produces more accurate real-time industrial production forecasts than the AR benchmark in the 2003–2014 period (the value of the fluctuation test is systematically positive). The null is rejected at the 5% significance level for all windows centered at 2007:M7 through 2010:M6, indicating that for those windows the mortgage spread contains statistically significant predictive power. Because the mortgage spread performs well over the whole out-of-sample period, the beginning of the ZLB/unconventional monetary policy environment has not changed its ability to forecast future industrial production growth.

The evidence for the paper-bill spread and the Baa.10y and Baa-Aaa corporate bond spreads is mixed. In general, these spreads do not add incremental predictive information in the real-time forecasting exercise. The results also suggest that the performance of the Aaa.10y spread and high-yield spreads as predictors of industrial production growth is somewhat episodic. For all windows centered before early 2007, inclusion of these spreads reduces forecast accuracy. However, later in the sample, the Aaa.10y spread and both high-yield spreads contain predictive information. Note that the predictive power of these credit spreads changed well before the short-term rate hit the ZLB and the Fed started to use unconventional monetary policy. Generally speaking, the fluctuation test does not show systematic deterioration/improvement in the forecasting ability of credit spreads since the beginning of the ZLB/unconventional monetary policy environment. Hence, the predictive power of credit spreads seems to be unaffected by the ZLB and unconventional monetary policy.

Figure 2. Fluctuation test for equal out-of-sample predictability (h = 3 months)



Notes: The figure plots the Giacomini and Rossi (2010) fluctuation test based on sequences of the Giacomini and White (2006) (GW) unconditional test statistics. The fluctuation test is implemented by using a centered rolling window of 45 observations (i.e., $\mu = m/P$ is approximately 0.4, where m is the size of the rolling window of the GW statistics and P is the number of out-of-sample observations). The sample period spans from 2003:M6 to 2014:M3. Positive (negative) values indicate that the interest rate spread model has produced more (less) accurate forecasts than the benchmark. The dashed line represents the critical value at the 5% significance level. If the fluctuation test statistic exceeds the critical value (2.770), the null that the two models have equal predictive ability at each point in time is rejected.

All in all, the results indicate that the predictive power of the term spread is unstable over time. The term spread has no predictive power for U.S. industrial production growth at the beginning of the out-of-sample period. Recently, however, the term spread has been a useful leading indicator. The literature has indicated that changes in monetary policy regimes are important for the predictive content of the term spread (see, e.g., Giacomini and Rossi, 2006). Therefore, the onset of the ZLB/unconventional monetary policy period provides a potential explanation for the observed change in predictive ability. The ZLB on nominal interest rates and unconventional monetary policy affect the behavior of the term spread. Therefore, it is not surprising that the timing of the change in the predictive content seems to correspond closely to the beginning of the ZLB/unconventional monetary policy period.

In general, the track record of credit spreads as indicators of U.S. industrial production growth is not good. The results show that most credit spreads contain predictive power only episodically. The predictive content of credit spreads seems to be unaffected by the ZLB and unconventional monetary policy. This finding is not surprising. Credit spreads contain predictive power primarily because they indicate changes in the supply of credit and expectations of default (Ng and Wright, 2013). Therefore, it is natural to interpret changes in the predictive power as being driven by other reasons than the ZLB and unconventional monetary policy. The real-time forecasting exercise suggests that the mortgage spread is a particularly informative leading indicator. The mortgage spread is a robust predictor of future economic activity across a variety of sample periods and forecast horizons. Furthermore, the mortgage spread systematically produces more accurate forecasts than the other spreads.

4.4 Conclusions

This paper analyzed the leading indicator properties of various interest rate spreads when the short-term rate has been fixed at the ZLB and the Fed has used unconventional monetary policy. The re-examination is motivated by the fact that the ZLB on nominal interest rates and unconventional monetary policy affect the behavior of the term spread. Our results suggest that the predictive content of the term spread in the ZLB/unconventional monetary policy period differs from that in the non-ZLB/conventional monetary policy period. In normal times, the term spread is not informative about future industrial production growth. However, when the short-term rate is fixed at the zero level and the Fed uses unconventional monetary policy, the term spread contains predictive power for industrial production growth. The results are consistent with the view that changes in the monetary policy regime affect the predictive power of the term spread.

Most credit spreads contain predictive power only episodically in our real-time forecasting exercise. The instability in predictive relationships highlights the burdens associated with using credit spreads as business cycle indicators; predictors that perform well in one period may work poorly in another. Although the predictive power of credit spreads fluctuates over time, the ability of credit spreads to signal future industrial production growth seems to be unaffected by the beginning of the ZLB and unconventional monetary policy era.

Our results indicate that the mortgage spread is a particularly useful leading indicator for U.S. industrial production growth. It outperforms the term spread and a set of widely used credit spreads in our real-time forecasting exercise regardless of the forecast horizon and sample period under investigation. Importantly, we find that the mortgage spread contains substantial predictive power both in the non-ZLB/conventional monetary policy and ZLB/unconventional monetary policy periods. Thus, the results suggest that the ZLB and unconventional monetary policy do not change the predictive content of the mortgage spread.

Although the mortgage spread is a robust predictor, our sample period is relatively short, running from 2003 to 2014. It would be interesting to examine the predictive power of the mortgage spread using a longer sample period from the 1970s to the present. Furthermore, one would like to know whether the mortgage spread has predictive power for other measures of economic activity, such as GDP and consumption. We leave these issues for future research.

References

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

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

Bernanke, B. S., Reinhart, V. R. and Sack, B. P. (2004). Monetary policy alternatives at the zero bound: an empirical assessment. *Brookings Papers on Economic Activity*, 1–78.

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

Elliott, G. and Timmermann, A. (2008). Economic forecasting. *Journal of Economic Literature*, 46, 3–56.

Chung, H., Laforte, J-P., Reifschneider, D. and Williams, J. C. (2012). Have we underestimated the likelihood and severity of zero lower bound events? *Journal of Money, Credit and Banking*, 44, 47–82.

Estrella, A., Rodrigues, A. P. and 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.

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

Faust, J., Gilchrist, S., Wright, J. H. and 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.

Friedman, B. M. and Kuttner, K. N. (1992). Money, income, prices and interest rates. *American Economic Review*, 82, 472–492.

Friedman, B. M. and Kuttner, K. N. (1998). Indicator properties of the paper-bill spread: lessons from recent experience. *Review of Economics and Statistics*, 80, 34–44.

Gertler, M. and 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. and 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. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25, 595–620.

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

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102, 1692–1720.

Gilchrist, S., Yankov, V. and Zakrajšek, E. (2009). Credit market shocks and economic fluctuations: evidence from corporate bond and stock markets. *Journal of Monetary Economics*, 56, 471–493.

Hall, R. E. (2011). The long slump. *American Economic Review*, 101, 431–469.

Krippner, L. (2013). Measuring the stance of monetary policy in conventional and unconventional environments. Centre for Applied Macroeconomic Analysis Working Paper, No. 6., Australian National University, Canberra.

Mody, A. and 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.

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

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

Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G. and Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier, 1203–1324.

Rossi, B. and Sekhposyan, T. (2010). 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. and Watson, M. W. (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature*, 41, 788–829.

Swanson, E. T. and Williams, J. C. (2014). Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *American Economic Review*, 104, 3154–3185.

Walentin, K. (2014). Business cycle implications of mortgage spreads. *Journal of Monetary Economics*, 67, 62–77.

Chapter 5
The mortgage spread as a predictor of real-time economic activity*

Jari Hännikäinen

Abstract

We analyze the predictive content of the mortgage spread for U.S. economic activity. We find that the spread contains predictive power for real GDP and industrial production. Furthermore, it outperforms the term spread and Gilchrist–Zakrajšek credit spread in a real-time out-of-sample forecasting exercise. However, the predictive ability of the mortgage spread varies over time.

Keywords: mortgage spread, forecasting, real-time data

JEL codes: C53, E37, E44

* I would like to thank Karl Walentin and Jari Vainiomäki for helpful comments and suggestions. Financial support from the FDPE is gratefully acknowledged.

5.1 Introduction

In a recent paper, Walentin (2014) shows that the spread between the mortgage rate and government bond rate (the mortgage spread) affects economic activity. However, the predictive ability of the mortgage spread has received little attention in the previous literature. The mortgage spread is potentially informative about future growth because it is an indicator of change in the supply of credit in mortgage markets. To the best of our knowledge, Hännikäinen (2015) is the only study analyzing the predictive power of the mortgage spread. He finds that the mortgage spread predicts U.S. industrial production growth in a relatively short period from 2003:M6 to 2014:M3.

This paper contributes to the existing literature by analyzing the real-time out-of-sample predictive power of the mortgage spread for U.S. real activity. We forecast both real GDP and industrial production growth. Our out-of-sample forecasting period, from 1992:Q1 to 2012:Q4, is substantially longer than that of Hännikäinen (2015). We compare the forecasting performance of the mortgage spread to that of two widely used leading indicators, namely, the term spread and a credit spread discussed in Faust *et al.* (2013), Gilchrist *et al.* (2009), and Gilchrist and Zakrajšek (2012) (henceforth GZ spread). Finally, we examine whether the predictive power remains stable over time.

The main finding from this study is that the predictive ability of the mortgage spread exceeds that of the term spread and GZ spread. However, the predictive power of the mortgage spread fluctuates over time. We find that the mortgage spread has been a particularly informative leading indicator since the early 2000s.

5.2 Methods

Following Stock and Watson (2003), we analyze the predictive power using the linear, horizon-specific *h*-step ahead model:

$$Y^h_{t+h} = \beta_0 + \sum_{i=0}^{p} \beta_{1i} X_{t-i} + \sum_{j=0}^{q} \beta_{2j} Y_{t-j} + u^h_{t+h}, \qquad t = 1, ..., T \tag{1}$$

where $Y^h_{t+h} = (400/h)ln(GDP_{t+h}/GDP_t)$ is the growth over the h quarters, $Y_{t-j} = 400ln(GDP_{t-j}/GDP_{t-j-1})$, and X_t is the spread.

We estimate (1) at each forecast origin by OLS using a rolling window of 60 observations. We allow the lags of Y_t to vary between zero and four and the lags of X_t to vary between one and four. We determine the lag lengths by minimizing the Bayes Information Criterion (BIC).

We quantify out-of-sample forecast performance by computing the mean squared forecast error (MSFE) of the mortgage spread forecast relative to that obtained from an autoregressive (AR) model. For the AR model, we consider lags between one and four and choose the lag length with the BIC. If the relative MSFE is less than one, the model with the spread has produced more accurate forecasts than the AR model. This implies that the spread contains marginal predictive power. The statistical significance is evaluated using the one-sided Giacomini and White (2006) test of equal unconditional predictive ability.

If the relative forecasting performance varies over time, the average performance over the whole out-of-sample period may give a misleading picture of the predictive power. We analyze time variations in the relative forecasting performance using the Giacomini and Rossi (2010) fluctuation test, which is equal to the Giacomini and White (2006) test computed over a rolling out-of-sample window.

5.3 Forecasting results

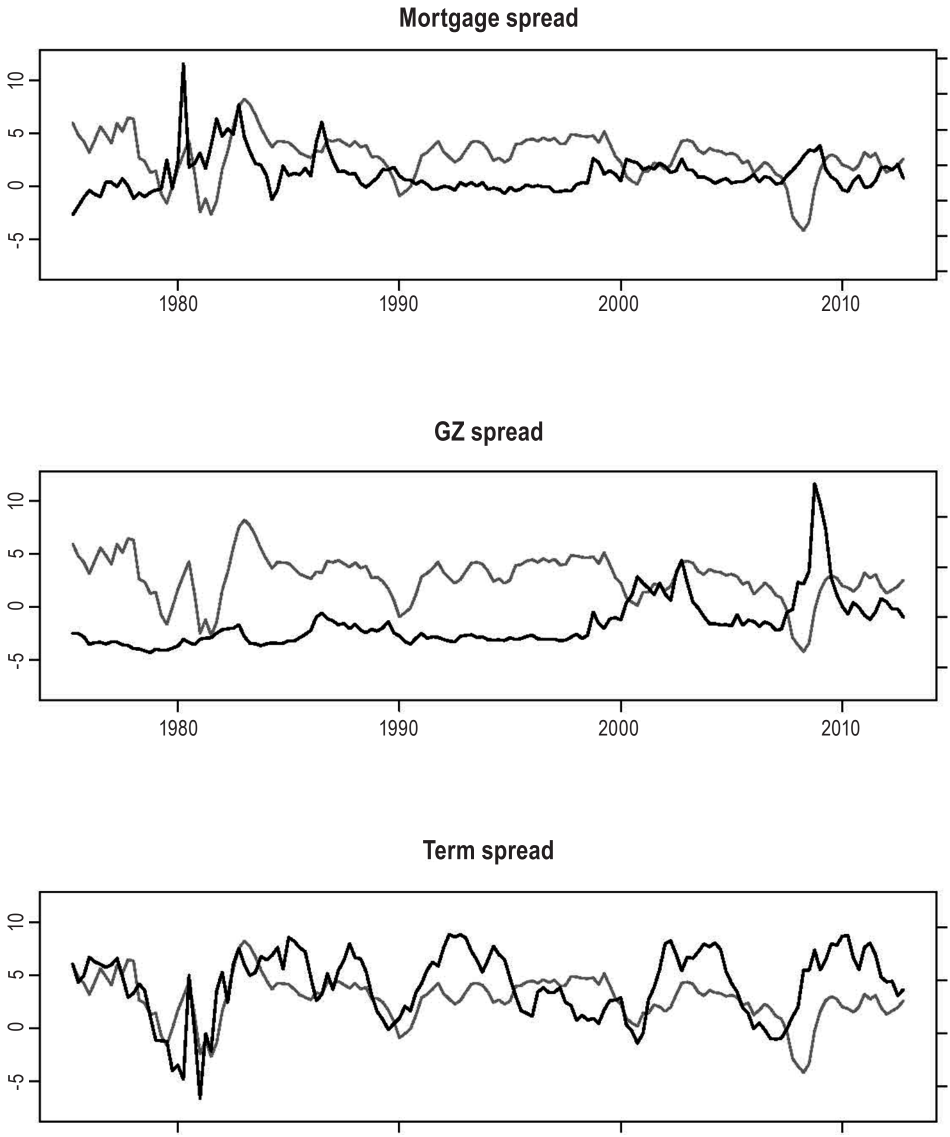
We analyze whether the mortgage spread is a useful leading indicator for real GDP and industrial production.[1] We compare the predictive power of the mortgage spread to that of the term spread (10-year Treasury bond rate – three-month Treasury bill rate) and the GZ spread. The sample period runs from 1975:Q1 to 2012:Q4. Different vintages of real GDP and industrial production are obtained from the Philadelphia Fed's real-time database. The interest rate data are from the St. Louis Fed's database and the GZ spread is downloaded from Simon Gilchrist's web page. Following Faust *et al.* (2013), the forecasts are made using data available in the middle month of each quarter. For real GDP and industrial production, we use the February, May, August, and November vintages of data. All interest rates are from the first month of each quarter. Figure 1 plots the spreads and the annual GDP growth rate from 1975:Q2 to 2012:Q4.

First, we consider the whole out-of-sample period 1992:Q1–2012:Q4. The results for real GDP are summarized in Panel A of Table 1, whereas Panel B contains the results for industrial production. Table 1 shows the MSFE of a candidate spread model relative to the MSFE of the AR benchmark.

The results show that the mortgage spread contains predictive power for real GDP growth. The mortgage spread model produces more accurate forecasts than the AR benchmark, regardless of the forecast horizon and whether we forecast the first-release

1 Following Hall (2011), the mortgage spread is defined as the difference between 30-year mortgage rate and 10-year Treasury bond rate.

Figure 1. Mortgage spread, GZ spread, and term spread



Notes: The figure depicts the spreads (black line, right scale) and annual GDP growth rate (grey line, left scale) from 1975:Q2 to 2012:Q4.

or the final values.[2] Interestingly, the ability of the mortgage spread to forecast real GDP growth is superior to that of the term spread and GZ spread in seven of the eight forecast horizon/true value combinations. The term spread and GZ spread typically perform poorly in the forecasting exercise.

Table 1. Out-of-sample MSFE values

	First-release				Final values			
A. GDP	*h*=1	*h*=2	*h*=3	*h*=4	*h*=1	*h*=2	*h*=3	*h*=4
Mortgage spread	0.88	0.92	0.94	1.02	0.91	0.92	0.94	0.99
GZ spread	0.90	1.06	1.27	1.34	0.90	1.04	1.21	1.28
Term spread	1.22	1.16	1.22	1.27	1.14	1.14	1.20	1.23
B. Industrial production								
Mortgage spread	0.87*	0.79*	0.80*	0.87	0.87*	0.79*	0.81*	0.87
GZ spread	0.92	1.19	1.32	1.28	0.93	1.19	1.32	1.29
Term spread	1.11	1.00	1.02	0.98	1.09	0.99	1.01	0.98

Notes: Asterisks mark rejection of the Giacomini and White (2006) test at the 1% (***), 5% (**), and 10% (*) significance levels, respectively.

Panel B suggests that the mortgage spread is a useful leading indicator for industrial production. The mortgage spread model produces lower MSFE values than the benchmark for all forecast horizons. In six of the eight forecast horizon/true value combinations, the mortgage spread contains statistically significant predictive power. Furthermore, it outperforms the term spread and GZ spread in each of the eight cases.

As a robustness check, Table 2 reports the results when the spread model contains the current value of the spread and one autoregressive lag and an AR(1) model is the benchmark. The results in Table 2 are similar to those presented in Table 1.

Next, we plot the relative MSFE values computed over a rolling window of 40 quarters. Figures 1 and 2 show the results for real GDP and industrial production, respectively. To save space, we report the results only for the first-release values at $h = 1$ and $h = 4$. The results for the other two horizons, and for the final values, are qualitatively similar.

The performance of the mortgage spread as a predictor of output growth is somewhat episodic. At the beginning of the out-of-sample period, the mortgage spread model produces less accurate forecasts than the benchmark. However, later in the sample, inclusion of the mortgage spread improves forecast accuracy. The fluctuation test rejects the null at $h = 1$ for windows centered at 2004:Q1–2006:Q1 for real GDP

2 We use values recorded in the real-time dataset two quarters after the target quarter as final values (cf. Faust *et al.*, 2013).

and at 2004:Q2–2007:Q2 for industrial production. Because the mortgage spread performs well in the latter part of the sample, the results imply that the frictions in the mortgage market are important in explaining recent business cycle fluctuations (see Hännikäinen, 2015; Walentin, 2014).

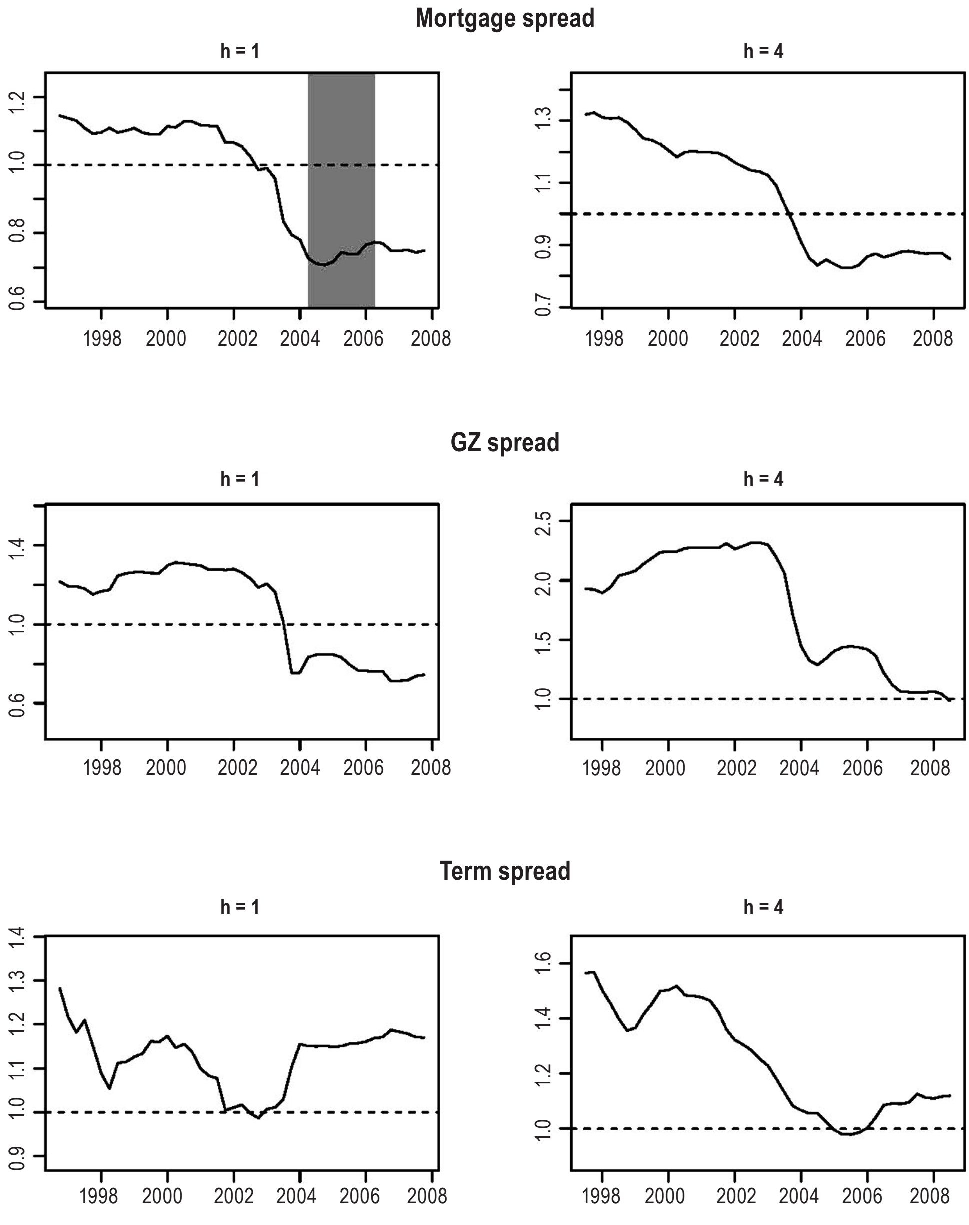
Table 2. Out-of-sample MSFE values (fixed lag lengths)

	First-release				Final values			
A. GDP	*h*=1	*h*=2	*h*=3	*h*=4	*h*=1	*h*=2	*h*=3	*h*=4
Mortgage spread	0.83**	0.89	0.96	0.98	0.87*	0.90	0.95	0.96
GZ spread	0.90	1.07	1.27	1.32	0.91	1.05	1.21	1.28
Term spread	1.12	1.11	1.11	1.16	1.06	1.08	1.10	1.14
B. Industrial production								
Mortgage spread	0.83**	0.81*	0.82	0.88	0.83**	0.81*	0.83*	0.89
GZ spread	0.95	1.19	1.31	1.34	0.97	1.20	1.32	1.35
Term spread	1.02	0.99	0.97	0.95	1.00	0.99	0.97	0.95

See the notes to Table 1.

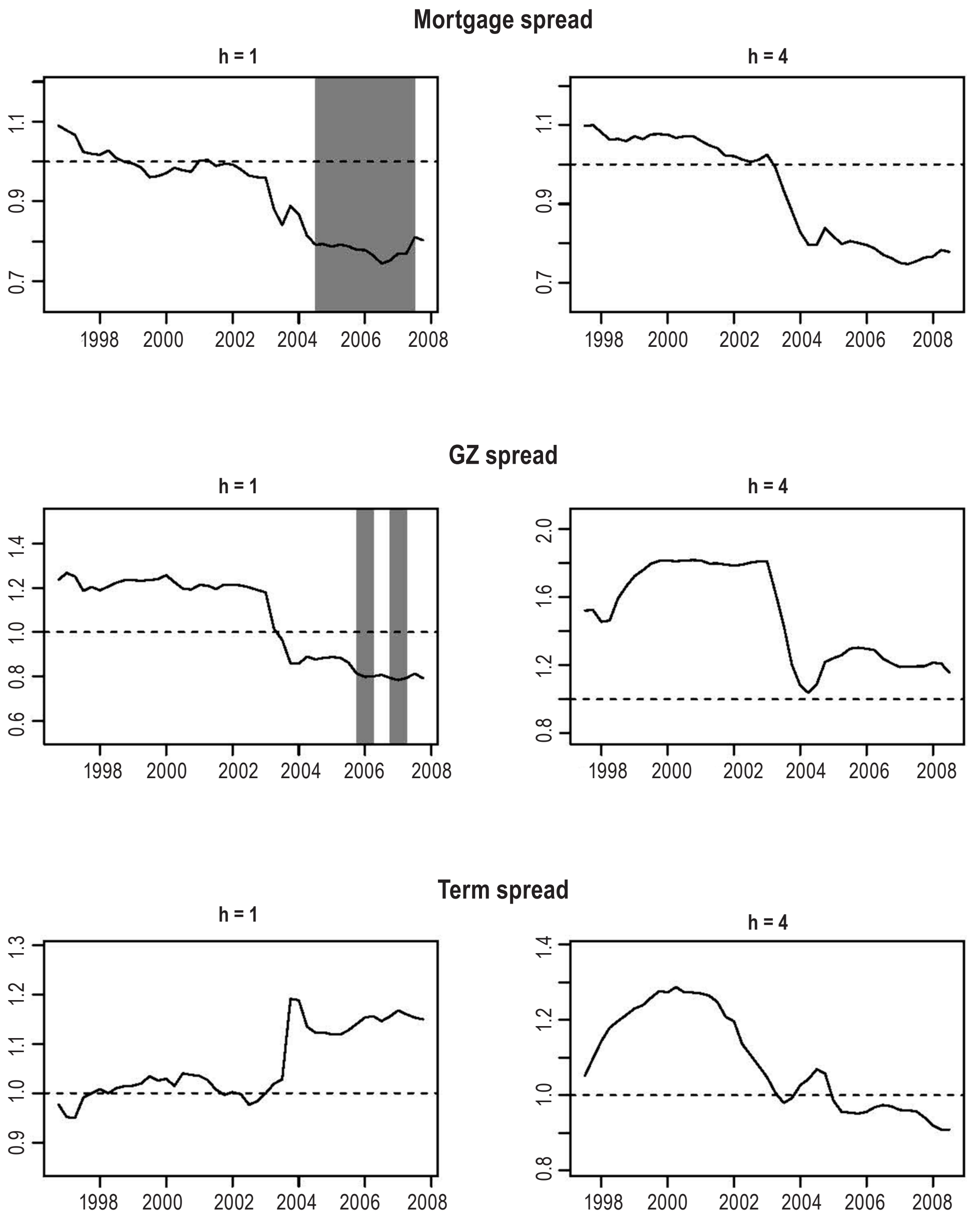
Figures 1 and 2 reveal that the term spread and GZ spread have episodically predictive power (cf. Ng and Wright, 2013). However, the rolling relative MSFE values for these two spreads are typically above one. Figures 1 and 2 confirm our previous finding that the predictive ability of the mortgage spread, in most cases, exceeds that of the term spread and GZ spread.

Figure 2. Rolling relative MSFE values for real GDP



Notes: The shaded areas denote the midpoints of windows in which the Giacomini and Rossi (2010) fluctuation test rejects the null of equal forecast accuracy at the 10% significance level.

Figure 3. Rolling relative MSFE values for industrial production



See the notes to Figure 2.

5.4 Conclusion

This paper examined whether the mortgage spread has real-time predictive power for U.S. economic activity. We find that the mortgage spread is a useful leading indicator for real GDP and industrial production growth. However, the predictive power fluctuates over time. The mortgage spread has been particularly informative since the early 2000s. Interestingly, our results show that the mortgage spread typically outperforms the widely used term spread and GZ spread.

References

Faust, J., Gilchrist, S., Wright, J. H. and 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.

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

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

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American Economic Review*, 102, 1692–1720.

Gilchrist, S., Yankov, V. and Zakrajšek, E. (2009). Credit market shocks and economic fluctuations: evidence from corporate bond and stock markets. *Journal of Monetary Economics*, 56, 471–493.

Hall, R. E. (2011). The long slump. *American Economic Review*, 101, 431–469.

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

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

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

Walentin, K. (2014). Business cycle implications of mortgage spreads. *Journal of Monetary Economics*, 67, 62–77.