TÜRKİYE CUMHURİYET MERKEZ BANKASI

WORKING PAPER NO: 15/32

Has the Forecasting Performance of the Federal Reserve's Greenbooks Changed over Time?

November 2015

Ozan EKŞİ
Cüneyt ORMAN
Bedri Kamil Onur TAŞ

Address:
Central Bank of the Republic of Turkey
Head Office
Research and Monetary Policy Department
İstiklal Caddesi No: 10
Ulus, 06100 Ankara, Turkey

Phone:
+90 312 507 54 02

Facsimile:
+90 312 507 57 33

Has the Forecasting Performance of the Federal Reserve's Greenbooks Changed over Time?[*]

Ozan Ekşi[**], Cüneyt Orman[***], and Bedri Kamil Onur Taş[****]

November 2015

"The Federal Reserve slashed its benchmark interest rate [...] at an unscheduled policy meeting. That Ben Bernanke was moved to act just a week before (the scheduled meeting) will raise the suspicion that ***the Fed knows something that markets don't****."*
(Desperate Measures, the Economist, 22 January 2008)

Abstract: We investigate how the forecasting performance of the Federal Reserve Greenbooks has changed relative to commercial forecasters between 1974 and 2009. To this end, we analyze time-variation in the Greenbook coefficients in forecast encompassing regressions. Assuming that model coefficients change continuously, we estimate unobserved components models using Bayesian inference techniques. To verify that our results do not depend on the specific way change is modeled, we also allow the coefficients to change discretely rather than continuously and test for structural breaks using classical inference techniques. We find that the Greenbook forecasts have been consistently superior to the commercial forecasts at all horizons throughout our sample period. Although the forecasting performance gap has narrowed at more distant horizons after the early-to-mid 1980s, it remains significant.

JEL Codes: C11, E52, E43

Keywords: Greenbook inflation forecasts, SPF inflation forecasts, Evaluating forecasts, Time-variation in coefficients

---

[*] The views expressed herein are solely of the authors and do not represent those of the Central Bank of the Republic of Turkey or its staff.

[**] TOBB-ETU, Department of Economics, Sogutozu Cad., No: 43, Sogutozu, 06560, Ankara, Turkey. E-mail: oeksi@etu.edu.tr

[***] Central Bank of the Republic of Turkey, Istiklal Cad. 10, 06100 Ulus, Ankara, Turkey. E-mail: cuneyt.orman@tcmb.gov.tr

[****] TOBB-ETU, Department of Economics, Sogutozu Cad., No: 43, Sogutozu, 06560, Ankara, Turkey. E-mail: onurtas@etu.edu.tr

1 Introduction

Expectations about the future state of the economy play a crucial role in the setting of monetary policy at modern central banks. This is why the staff at most central banks prepares forecasts of key macroeconomic variables such as inflation, output, and unemployment. Those forecasts form the basis of discussions among decision-makers that ultimately shape monetary policy. For instance, the so-called "Greenbook" which contains the Federal Reserve (Fed hereafter) staff forecasts is distributed to Federal Open Market Committee (FOMC hereafter) members a week before the meeting. The transcripts from the FOMC meetings clearly indicate that the committee members do indeed pay close attention to these forecasts during policy deliberations. As such, the quality of Greenbook forecasts is of paramount interest to both policymakers and academics.[1]

The importance of Greenbook forecasts in the conduct of U.S. monetary policy has prompted many researchers to analyze these forecasts in detail, giving rise to a substantial literature. In their seminal paper, Romer and Romer (2000) showed that the Greenbook forecasts of inflation are superior to the forecasts produced by commercial forecasters during the 1968-1991 period. They interpreted this finding as suggesting that the Fed has substantial information about the economy beyond what is known to commercial forecasters. Subsequent studies such as Ellingsen and Söderström (2001), Sims (2002), Gavin and Mandal (2003), and Peek, Rosengren, and Tootell (2003) reached the same conclusion using data extending to the mid-1990s. However, the findings of studies focusing on the more recent period have been more diverse. Reifschneider and Tulip (2007) found no difference between the forecasting accuracy of the Greenbook and private forecasts between 1986 and 2006, whereas D'Agostino and Whelan (2008) and Gamber and Smith (2009) found that the superiority of the Greenbook forecasts disappeared after the early 1990s. Hubert (2014) also reported a narrowing of the gap between the accuracy of Greenbook and private forecasts, but he argued that the gap did not disappear.

[1] Beginning in the late 1970s, the FOMC members have also produced forecasts of their own. In this paper, the Fed forecasts refer to those produced by the staff and contained in the Greenbooks.

From a methodological perspective, the literature on the temporal change in the relative performance of the Greenbook forecasts can be broken down into two strands. In the first strand, the values of model parameters proxying for the accuracy of different forecasts are assumed to be subject to discontinuous jumps (breaks) over time. Moreover, these break dates are selected by the researcher, usually with reference to certain events that take place around this date. For instance, Gamber and Smith (2009) conduct their analyses by assuming the existence of structural breaks in the data in 1984 and 1994 (see, also, Hubert, 2014). The authors motivate their chosen break dates with reference to the start of the Great Moderation and the changes in the Fed's operating procedures that led to improvements in external communications and transparency. In the second strand, the values of model parameters are assumed to evolve in a continuous fashion rather than being subject to potential discrete jumps. Two studies that make this assumption are D'Agostino and Whelan (2008) and El-Shagi, Giesen and Jung (2013), who then carry out their analyses using non-parametric rolling-window-based estimation methods.

Each of these two approaches, however, suffers from different limitations. First, using predetermined break dates is unsatisfactory because it is generally not an easy task for the researcher to argue convincingly that the events are selected exogenously (Hansen, 1992). In addition, assuming the existence of a break at a certain date amounts to data mining and biases the results toward finding evidence of instability at the specified date (Christiano, 1992; Zivot and Andrews, 1992; Tulip, 2009). Second, using non-parametric rolling-window-based estimation methods to capture parameter instability also has a number of drawbacks. For one, the method is not based on an explicit economic or statistical theory. Moreover, as argued by Guidolin, Ravazzolo and Tortora (2013), the lack of specific parametric forms makes testing for time-variation difficult and dependent on hard-to-justify choices of the rolling-window length and the updating rules applied to choose whether constant or decaying weights should be applied.

In this paper, we investigate how the relative performance of the Fed's Greenbook forecasts changes over time using methods that are not subject to the abovementioned limitations. First, assuming that the parameters of the model change

continuously over time, we investigate time-variation in the coefficients on the Greenbook and commercial forecasts by estimating unobserved components models, where we use Bayesian inference along the lines suggested by Harvey (1990) and Durbin and Koopman (2012). Second, we allow the model parameters to change discretely rather than continuously and then test for structural breaks in the coefficients on the Greenbook and commercial forecasts using classical inference. Rather than using candidate structural break dates, we test for potential breaks by identifying the unknown break date endogenously from the data. To this end, we carry out the Quandt Likelihood Ratio (QLR) test, also known as the supremum test, which was originally proposed by Quandt (1958, 1960) and further developed by Andrews (1993). Considering that there may be more than one break in the data, we also test for multiple breaks using the approach proposed by Perron and Qu (2007).

In our analyses, we focus on inflation forecasts as in Romer and Romer (2000) because the Fed's forecasting advantage appears to be more pronounced for inflation than for output (see, for instance, Hubert, 2014). To set the stage for our analyses, we first estimate Fair and Shiller (1989, 1990) forecast encompassing regressions to find out whether the Romer and Romer (2000) finding that the Fed's Greenbook forecasts are superior to commercial forecasts continues to hold in our data sample extending to 2009:Q4.[2] Our proxy for commercial forecasts is the set of forecasts contained in the Survey of Professional Forecasters (SPF hereafter) currently conducted by the Federal Reserve Bank of Philadelphia (FRBP hereafter). This method enables us to compare the forecasting powers of the Greenbook and SPF forecasts under the assumption that the values of model parameters are constant over time. We then allow the values of the model parameters to change over time, which allows us to investigate the temporal evolution of this comparison.

[2] Fair and Shiller regressions take the form

$$\pi_{h,t} = \beta_0 + \beta_1 \pi_{h,t}^{GB} + \beta_2 \pi_{h,t}^{PF} + \varepsilon_{h,t},$$

where $\pi_{h,t}$ denotes the actual inflation in quarter h after quarter t and $\pi_{h,t}^{GB}$ and $\pi_{h,t}^{PF}$ denote, respectively, the Greenbook and commercial forecasts of $\pi_{h,t}$ in quarter t, $\varepsilon_{h,t}$ denotes the error term, and β_0 and β_1 measure the relative information contents of the forecasts about h-quarter-ahead inflation rate.

Our analyses yield three sets of results. First, our results from estimating the forecast encompassing regressions show that the coefficient estimates on the Greenbook forecasts are large (i.e. close to 1) and highly significant whereas those on the SPF forecasts are small (i.e. close to 0) and insignificant at all horizons up to four quarters during 1974:Q4-2009:Q4. Moreover, the difference in the magnitudes of the estimated coefficients on the Greenbook and SPF forecasts is higher at more distant horizons. These results imply that the Greenbook forecasts are superior to the SPF forecasts at all horizons and that the magnitude of the difference in forecast accuracy increases with the length of the horizon. We, therefore, confirm that the findings of Romer and Romer (2000) also hold for the 1974:Q4-2009:Q4 time period.

Second, our analysis which permits continuous time-variation in the model coefficients reveals that the estimated coefficients on the Greenbook forecasts are consistently large and highly significant over time while those on the SPF forecasts are small and almost always insignificant at all forecasting horizons throughout our sample period. Although the magnitude of estimated coefficients on the Greenbook forecasts decline somewhat at the three- and four-quarter-ahead horizons after 1984:Q1, they are always statistically significantly above zero whereas the coefficients on the SPF forecasts are almost always insignificantly different from zero.

Third, our results from the estimation of structural breaks largely confirm our findings from our analysis of continuous time-variation in model parameters. Like before, the estimated coefficients on the Greenbook forecasts are consistently large and highly significant over time while those on the SPF forecasts are small and insignificant at all horizons throughout our sample period. In addition, our structural break tests provide no evidence of a change in the magnitude of the estimated coefficients on the Greenbook forecasts at the current and one-quarter-ahead horizons. However, we do find evidence that the estimated coefficients on the Greenbook forecasts become smaller after 1980:Q3 at the two-, three- and four-quarter-ahead horizons, but that they are still statistically significantly larger than the estimated coefficients on the SPF forecasts. Finally, we find no evidence of additional breaks during our sample period.

Overall, our results indicate that the Fed's advantage in forecasting inflation has been consistently higher at relatively short horizons but has deteriorated somewhat at

longer horizons after the early-to-mid 1980s, albeit not to the point of disappearance. Thus, the deterioration in the Fed's forecasting advantage appears to have occurred precisely at those horizons for which their advantage was the most pronounced, i.e. at longer horizons. This implies that the forecasting accuracy differentials between the Fed and commercial forecasters have become more uniform across forecast horizons since the early-to-mid 1980s. These results hold regardless of whether the change in model parameters is modeled as continuous or discrete or whether classical or Bayesian inference methods are used. We, therefore, conclude that the Greenbook forecasts continue to possess information that is not contained in the SPF forecasts, which in turn suggests that an individual with access to both forecasts would be well-served by discarding the SPF forecast and just using the Greenbook forecast. However, as also noted by Gamber and Smith (2009), the Fed appears to be getting less bang for its buck in terms of improved forecasting accuracy relative to the private sector, particularly at a horizon of about a year which central bankers have in mind when making policy decisions.

Our findings are broadly consistent with Atkeson and Ohanian (2001) and Stock and Watson (2007) who show that the variability of the predictable component of inflation has dropped considerably after the early-to-mid 1980s, which in turn led to a dramatic fall in the forecasting power of forecasters, especially at more distant horizons. Our finding is also generally consistent with D'Agostino and Whelan (2008) and Gamber and Smith (2009) who report a decline in the Fed's advantage at all horizons (but particularly at longer horizons) under the assumption that there are breaks in the data in both the early-to-mid 1980s and 1990s. However, unlike these studies, we find that even though the advantage falls, the difference is still statistically significantly positive in all regimes separated by the break date(s). In this respect, our finding is more in line with Hubert (2014) who also finds that the advantage declined but did not disappear after both the early-to-mid 1980s and 1990s. In contrast with D'Agostino and Whelan (2008), Gamber and Smith (2009), and Hubert (2014), however, we find no evidence of a break in the data in the early-to-mid 1990s.

Our results also shed light on the possible sources of the Fed's forecasting advantage. Specifically, while our results appear to be consistent with the fall in the

variability of the predictable component of inflation after the Great Moderation as argued by Atkeson and Ohanion (2001) and Stock and Watson (2007), this is apparently not the complete story. We believe that the existence of a large Fed staff (as suggested by Romer and Romer, 2000) and their ability to better assess the current and recent past state of the economy (as suggested by Sims, 2002) as well as the confidential knowledge about troubled non-publicly-traded banks (as suggested by Peek, Rosengren, and Tootell, 2003) provide complementary explanations of the source of the Fed's overall superior forecasting performance. The decline in the Fed's forecasting advantage at more distant horizons after the early-to-mid 1980s, on the other hand, most likely reflects the improved longer-term forecasting capabilities of commercial forecasters. This, in turn, might be explained by a combination of factors including the increased credibility of the Fed regarding inflation (as suggested by D'Agostino and Whelan, 2008), the increased transparency and openness of the Fed (as suggested by Gamber and Smith, 2009), or the increased amount of resources commercial firms spend on improving their forecasting techniques (as suggested by El-Shagi et al., 2013).

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 presents the empirical methods and results. Section 4 summarizes the main results and deliberates on the underlying explanations. Section 5 provides concluding remarks.

2 Data

In our empirical work that follows, we compare the performance of the Fed's Greenbook to that of the SPF in forecasting inflation. Beginning in 1965, the Greenbooks have been prepared by the staff of the Board of Governors before each meeting of the FOMC. The SPF, on the other hand, was launched in 1968 and is currently conducted by the FRBP in the second month of each quarter. While the SPF forecasts are available on a real-time basis, the Greenbook forecasts are released to the public with a lag of five years.[3] Given that median forecasts outperform all other survey-based forecasts (Ang et al. 2007),[4] we use the median of individual commercial forecasts contained in the SPF in our analyses. Since the Greenbook and the SPF forecasts are available at different

[3] Both the Greenbook and the SPF forecasts can be obtained from the FRBP's webpage.
[4] The authors show that adjustments to take into account both linear and non-linear bias yield worse out of-sample forecasting performance.

frequencies, these two sets of forecasts are not directly comparable. Fortunately, the FRBP webpage provides the Greenbook forecasts published from the four FOMC meetings each year that corresponds (as closely as possible) to the publication dates of the SPF forecasts. For both sets of forecasts, our data sample runs from 1974:Q4 to 2009:Q4 for a total of 141 observations at each forecast horizon.[5] Because we want to compare our results to those of Romer and Romer (2000), we consider forecast horizons of up to four quarters. Forecasts are the annualized quarterly growth rates of the GNP/GDP price deflator.[6]

Our measure of actual inflation is the annualized quarterly growth rate of the GNP/GDP price deflator. We use real-time data on the GNP/GDP deflator taken from the "Real-Time Data Set for Macroeconomists" compiled by the FRBP. These estimates correspond to the "first final" estimate (also known as the "second revision") of the Bureau of Economic Analysis. We believe the second revision data are the appropriate series to use because it is closest in definition to the variable being forecast (see, for example, Romer and Romer, 2000, Tulip, 2009, and Hubert, 2014).

3 Methods and results

In this section, we have two main objectives. First, in Subsection 3.1, we estimate Fair and Shiller (1989, 1990) regressions using classical inference to investigate whether the Romer and Romer (2000) finding that the Fed's Greenbook forecasts are superior to commercial forecasts continues to hold in our data sample that contains the most recent period. Although this method helps us to capture the overall relationship between the forecasting powers of the Greenbook and commercial forecasts, it does not inform us about how this relationship changes over time. Therefore, our second objective is to use state-of-the-art methods to examine the temporal evolution of the performance of the Greenbook forecasts relative that of commercial forecasts. To this end, we allow parameters to change both discretely and continuously; we analyze discrete changes using classical inference (i.e. by testing structural breaks) and continuous changes by

[5] We start the sample at 1974 rather than 1968 because Greenbook inflation forecasts are missing for several of the quarters between 1968 and 1974.

[6] Note that prior to 1992 real GNP was the measure of real aggregate output instead of real GDP in the national accounts. Thus, our measure of inflation is the GNP deflator prior to 1992 and GDP deflator thereafter.

estimating unobserved components models where we use Bayesian inference. We take on these issues in subsections 3.2 and 3.3.

3.1 Fair and Shiller (1989, 1990) regressions

We first investigate if the main Romer and Romer (2000) findings carry over to the period after the early 1990s. Using data for the 1968:Q4-1991:Q1 period, Romer and Romer (2000) argued that the Fed has considerable information about inflation beyond what is known to commercial forecasters. They argued, in particular, that the optimal forecasting strategy of someone with access to both the Greenbook forecasts and the commercial forecasts would be to put essentially no weight on the commercial forecasts. Their argument was based on a finding of a coefficient close to one on the Greenbook forecasts and a coefficient of close to zero on the commercial forecasts in the Fair and Shiller (1989, 1990) regressions. We want to find out if this pattern on the Greenbook and the SPF forecast coefficients also holds during the period from 1974:Q4 to 2009:Q4.[7] Toward this end, we estimate Fair and Shiller regression equations of the form

$$\pi_{h,t} = \beta_0 + \beta_1 \pi_{h,t}^{GB} + \beta_2 \pi_{h,t}^{SPF} + \varepsilon_{h,t}, \tag{1}$$

where $\pi_{h,t}$ denotes the actual inflation in quarter h after quarter t and $\pi_{h,t}^{GB}$ and $\pi_{h,t}^{SPF}$ denote, respectively, the Greenbook and SPF forecasts of $\pi_{h,t}$ in quarter t, $\varepsilon_{h,t}$ denotes the error term, and h = 0, 1, 2, 3, and 4 is the forecast horizon. The coefficients β_1 and β_2 measure the relative (conditional) information contents of the forecasts about h-quarter-ahead inflation rate. If the forecasts contain independent information, then both β_1 and β_2 should be different from zero. We want to find out if β_1 is significantly different from zero and close to one and at the same time β_2 is not significantly different from zero at each horizon during our sample period. In order to control for heteroskedasticity and autocorrelation, we calculate robust standard errors using the Newey and West (1987) methodology.

The results of estimating Equation (1) are displayed in Table I. This table replicates Table 2 of Romer and Romer (2000) and compares the Greenbook and SPF inflation forecasts up to four quarters ahead. The estimated coefficient on the Greenbook inflation forecast, β_1, is large and significant at the 1 percent level at all horizons. For the two-,

[7] Our data starts at 1974:Q4 because four-quarter-ahead inflation forecasts of the Greenbook are not available before this date.

three- and four-quarter-ahead horizons, the point estimates of β_1 are above 1. For other horizons, they are smaller, but still close to 1. In addition, the estimated coefficient on the SPF inflation forecast, β_2, is small and insignificant at the 1 percent level at all horizons. These results suggest that the Greenbook forecasts do indeed possess information that is not contained in the SPF forecasts. Thus, an individual with access to both forecasts would lose nothing by just using the Greenbook forecast. We, therefore, confirm the findings of Romer and Romer (2000) and others for the time period between 1974:Q4 and 2009:Q4.

[Insert Table I about here]

3.2 Estimation of continuous changes in coefficients

In this section, we allow the value of the coefficients β_1 and β_2 in Equation (1) to continuously change over time. Allowing the coefficients to change over time enables us to examine the temporal evolution of the forecasting performances of the Greenbook and commercial forecasts. Differently from studies such as D'Agostino and Whelan (2008) and El-Shagi, Giesen and Jung (2013) who use non-parametric rolling-window-based estimation methods, we follow Harvey (1990) and Durbin and Koopman (2012) and estimate the time-varying coefficients in Equation (1) using Bayesian methods. Basically, the method estimates the likelihoods of the coefficients via the Kalman filter. These likelihoods are then convolved with the priors on the coefficients to obtain the posterior distributions of the time-varying coefficients of the model. Further details of this estimation are given in the Appendix. Our results use the median values of the posterior distributions of the coefficients, together with their 16th and 84th percentiles. Figures I and II below display the time-varying Greenbook and SPF coefficients, respectively.[8]

[Insert Figure I and Figure II about here]

[8] In line with common practice, we report the 68 percent confidence bands in both figures. We also estimated 95 percent confidence bands. Predictably, these bands are wider, but the coefficients are still significantly above zero. These results are available from the authors upon request.

Figure I shows that the value of estimated β_1 is not only significantly different from zero but also generally close to one at all horizons throughout our sample period. While the value of estimated β_1 remains fairly stable over time at the current and one-quarter-ahead horizons, it starts to fall at the two-, three-, and four-quarter-ahead horizons after 1984:Q1. The downward movement is much more apparent at the three- and four-quarter-ahead horizons than at the two-quarter-ahead horizon. In addition, Figure II shows that zero is almost always within the confidence bands, implying that the value of estimated β_2 is generally insignificantly different from zero at all horizons. These results indicate that the Greenbook forecasts are consistently superior to the SPF forecasts at all horizons and at (almost) all points in time throughout our sample period. Despite the erosion of the Greenbook forecasts' superiority particularly at the three- and four-quarter-ahead horizons after the early-to-mid 1980s, the gap remains statistically significantly positive.

3.3 Estimation of discrete changes in coefficients

We next suppose that the coefficients on the Greenbook and SPF forecasts, β_1 and β_2, in Equation (1) change discretely rather than continuously. Such discontinuous jumps are usually referred to as structural breaks. Previous researchers test for the presence of structural breaks in the explanatory power of the Greenbook forecasts at possible break dates. For instance, Gamber and Smith (2009) conduct their analyses by assuming at the outset that there might be structural breaks in the data in 1984 and 1994. D'Agostino and Whelan (2008) and Hubert (2014) also follow a similar approach. The approach adopted by Tulip (2009) is much less restrictive in this sense, as he allows for a larger number of possible break dates. Nevertheless, using predetermined dates to define a structural change is unsatisfactory since it amounts to data mining and it biases the results toward finding evidence of instability at the specified date (Christiano, 1992; Zivot and Andrews, 1992; Tulip, 2009). To see this note that the relevant tests (such as the well-known Chow test) typically split the data from the possible break point and estimate the two sub-samples separately by ordinary least-squares (OLS hereafter). If the new results provide a significantly better fit to the data than the results obtained from estimating the whole sample at once, the null hypothesis of "no structural break in the data" is rejected. However, splitting the data into two sub-samples at any point near

the break date would almost certainly improve the fit of the model to the data even if the break is not at that point. This would necessarily create a bias in estimated coefficients, the magnitude of which depends on the gap between the selected and true break dates. Therefore, it is essential that the break date be determined as accurately as possible.

Unlike previous papers, we do not assume possible breaks at certain dates. Rather, in Subsection 3.3.1, we carry out the QLR test to identify the unknown break date endogenously from the data. The QLR test was originally proposed by Quandt (1958, 1960). The distribution of this test, together with the tabulated critical values computed by simulation, is given in Andrews (1993). The QLR test estimates individual Chow statistics for each date in the data, except for some trimmed portion from both ends. We trim the first and last 10 percent of the data.[9] The date that maximizes the estimated Chow statistics is the most likely break point in the data.[10] Having determined the breaks in the data in Subsection 3.3.1, we next investigate in Subsection 3.3.2 the direction and magnitude of these changes.

3.3.1 Structural break tests for changes in forecasting power

In order to investigate the existence of breaks in the data, we estimate the following model,

$$\pi_{h,t} = \beta_0 + \beta_1 \pi_{h,t}^{GB} + \beta_1^D (\pi_{h,t}^{GB} \times D_i) + \beta_2 \pi_{h,t}^{SPF} + \varepsilon_{h,t} \tag{2}$$

where D_i is defined as

$$D_i = \begin{cases} 1 & if\ t \geq i \\ 0 & if\ t < i \end{cases}$$

and β_1^D is the coefficient on the interaction term. Here, our main interest is β_1^D. The QLR test calculates the Chow statistic for each date *i* and then finds the date at which the Chow statistic attains a maximum. If, at this date, the value of the Chow statistic exceeds a certain critical value, then we conclude that the estimated β_1^D is significantly different

[9] The common practice in the literature is to trim 15% of the data from both ends. However, we trimmed only 10% as it allowed us to observe the pattern of the break statistics around the break date. The critical values associated with the 5% significance level at 15% and 10% trimming are 8.68 and 9.11, respectively. As any break that is significant at 10% trimming is already significant at 15% trimming, the break points remain the same in both cases.

[10] Note that the break point determined by the QLR is only the most likely break point because of small sample properties.

from zero. A significant estimated β_1^D would, in turn, indicate that the relative explanatory power of the Greenbook forecasts is subject to a structural change at this date.

Figure III below displays the results from estimating Equation (2). Specifically, the figure shows the QLR statistics for estimated β_1^D at different horizons and the respective dates at which the QLR statistic achieves a maximum. Like before, a significant estimated β_1^D at the date where the QLR statistic reaches a maximum would indicate a structural change in the Greenbook forecasts' relative explanatory power at that date. The figure indicates that the QLR statistic is below the 5 percent critical value for the current- and one-quarter-ahead inflation forecasts. Hence, we do not reject the null hypothesis of no structural break in the value of estimated β_1^D at these horizons. On the other hand, the QLR statistics for the two-, three-, and four-quarter-ahead forecasts do exceed the critical value at 1980:Q3, which suggest breaks in the value of estimated β_1^D around this date.[11]

[Insert Figure III about here]

3.3.2 Analysis of changes in the coefficients at the determined break date

The results in the previous subsection suggest that the value of the coefficients on the two-, three-, and four-quarter-ahead Greenbook forecasts in the Fair and Shiller regressions change significantly after 1980:Q3. We next explore the direction and the magnitude of these changes. To this end, we estimate the following equation:

$$\pi_{h,t} = \beta_0 + \beta_1 \pi_{h,t}^{GB} + \beta_1^D (\pi_{h,t}^{GB} \times D_{i^*}) + \beta_2 \pi_{h,t}^{SPF} + \varepsilon_{h,t} \tag{3}$$

where D_{i^*} is defined as

$$D_{i^*} = \begin{cases} 1 & if\ t \geq i^* \\ 0 & if\ t < i^* \end{cases}$$

where i^*= 1980:Q3. Thus, the coefficient on the Greenbook forecast, $\pi_{h,t}^{GB}$, is equal to β_1 before the structural break date and equal to $\beta_1 + \beta_1^D$ afterwards. Our main interest is the coefficient on the interaction variable, β_1^D, which measures the change in the

[11] We also tested for the presence of multiple breaks in the data by using Perron and Qu (2007)'s approach but found no evidence of additional breaks.

relative forecasting power of the Greenbook after the break date. Table II displays the results from estimating Equation (3).

[Insert Table II about here]

Table II shows that the estimated coefficient on the interaction variable, β_1^D, is significant at all three horizons, thus confirming our finding in Subsection 3.3.1 that structural breaks occur in 1980:Q3 at the two-, three- and four-quarter-ahead horizons. In addition, the fact that the estimated β_1^D is negative implies that the relative accuracy of the Greenbook forecasts decreases at the two-, three- and four-quarter-ahead horizons after 1980:Q3. However, the value of the estimated $\beta_1 + \beta_1^D$ is close to 1 in all three cases, suggesting that the Greenbook forecasts are superior to the commercial forecasts even after this break date.

These results are largely consistent with our results from the estimation of continuous changes in coefficients using Bayesian methods in Subsection 3.2. The main difference between the two sets of findings is that the fall in the magnitude of the coefficients starts earlier in the latter case than in the former case (1980:Q3 versus 1984:Q1). Given the differences between the two approaches, it is difficult to pinpoint the exact reason(s) behind this difference in results. Nevertheless, the findings from both approaches agree on the existence of a nontrivial decline in the value of the coefficient on the Greenbook forecasts in the early-to-mid 1980s.

The narrowing of the Fed staff's forecasting advantage at the two-, three- and four-quarter-ahead horizons is also broadly consistent with D'Agostino and Whelan (2008), Gamber and Smith (2009), and Hubert (2014) who report a decline in the Fed's forecasting superiority over time, particularly at longer horizons.[12] Despite this similarity, our results differ from their results in an important way. In particular, we do not find that the Fed's advantage disappears entirely in the more recent period as argued by D'Agostino and Whelan (2008) and Gamber and Smith (2009).[13] We do find

[12] Tulip (2009) also finds that the Fed staff's forecasting power has disappeared in the recent period. However, he compares the forecasting power of the Fed's Greenbook to a random walk, whereas we (as well as the three studies mentioned above) compare it to the SPF.

[13] Using conditional regressions, Hubert (2014) also finds that the Fed's advantage disappears entirely in his data subsamples that start in the early 1990s. His unconditional regressions, however, provide some evidence that the advantage declined but did not vanish.

evidence of a reduction in the Fed's advantage at the two-, three- and four-quarter-ahead horizons (but not for other horizons), but even then the advantage is still statistically significant. Our results are not directly comparable to those of the above studies, however, as we do not consider separately the data subsamples starting in 1992 or 1994 as these authors respectively do. As noted by Gamber and Smith (2009), the results obtained from these subsamples might be complicated by a small sample problem. In addition, these subsamples cover a very stable period for which conditional comparisons lack variability (Hubert, 2014).

4. The source of the Fed's forecasting advantage

Our results in Section 3 suggest that the Fed staff has an edge in forecasting inflation over private forecasters at all horizons during 1974:Q4-2009:Q4. Results also indicate that there has been an erosion of this advantage at the two-, three- and four-quarter-ahead horizons after the early-to-mid 1980s, but it did not disappear.

What do these results imply about the source of the Fed staff's apparent forecasting advantage? As we noted earlier, one explanation may be the drop in the variability of the predictable component of inflation after the Great Moderation, particularly at longer horizons, as documented by Stock and Watson (2007) and Tulip (2009). Another explanation could be the possibility put forward by Sims (2002) who argued that the Fed's forecasting advantage might be due to its ability to collect detailed information about the current and recent past state of the economy. The Fed might be able to do so because it commits far more resources to forecasting than even the largest commercial forecasters (Romer and Romer, 2000). A better estimate of the current state in turn might provide the Fed staff an advantage over private forecasters in predicting inflation over moderate horizons. We believe that the Stock and Watson (2007), Tulip (2009), Romer and Romer (2000), and Sims (2002) explanations are complementary rather than being mutually exclusive. In particular, while the drop in the variability of the predictable component of inflation may have contributed to a fall in the forecasting accuracy of both the Fed and private forecasters, the fact that the Fed commits far more resources to forecasting implies that they can assess the current state of the economy better than private forecasters, which in turn enables them to produce better forecasts,

particularly at relatively short horizons. Another reason behind the decline in the Fed's forecasting advantage at more distant horizons after the early-to-mid 1980s could be the improved longer-term forecasting capabilities of commercial forecasters. This, in turn, might reflect the impact of the increased amount of resources commercial firms spend on improving their forecasting techniques as suggested by El-Shagi et al. (2013).

Our findings are also partly consistent with explanations based on private information possessed by the Fed. Peek, Rosengren, and Tootell (2003) argued that confidential supervisory information about troubled non-publicly-traded banks might provide the Fed with a durable advantage when forecasting the macroeconomy. Our findings would suggest that this advantage is particularly strong at horizons up to two quarters into the future rather than up to four quarters as suggested by Peek, Rosengren, and Tootell (2003). Although there are differences in the timing of events, our findings can also be viewed as providing some evidence in favor of explanations based on the increased credibility of the Fed regarding inflation as suggested by D'Agostino and Whelan (2008) or increased transparency and openness as suggested by Gamber and Smith (2009).

5. Conclusion

We investigated the relative inflation forecasting performance of the Fed staff and commercial forecasters for the period 1974:Q4-2009:Q4. We have a number of important findings.

First, our results from the Fair and Shiller regressions suggest that on average the Fed staff has an advantage in forecasting inflation over commercial forecasters at all horizons during the entire sample period. Moreover, this advantage appears to increase with the length of the horizon.

Second, our analysis of time-variation in the model parameters provides no evidence of a decline in this advantage at the current and one-quarter-ahead horizons during our sample period. Results do indicate, however, that there has been an erosion of this advantage at the two-, three- and four-quarter-ahead horizons after the early-to-mid 1980s, but it did not disappear. Therefore, the Fed's advantage in forecasting inflation seems to have deteriorated precisely at those horizons for which their

advantage was the most pronounced, as a result of which the forecasting performance differences between the Fed and commercial forecasters have become more uniform across forecast horizons since the early-to-mid 1980s.

Taken together, our results suggest that the Fed's Greenbook forecasts continue to possess information that is not contained in the commercial forecasts, albeit to a smaller degree at longer horizons. Thus, an individual with access to both forecasts would still be well-served by discarding the commercial forecasts and just using the Greenbook forecasts.

APPENDIX: Gibbs sampling algorithm of Bayesian estimation for a time-varying parameter model

The following state-space representation provides the model with time-varying coefficients:

$$Y_t = \theta_t X_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, R),$$

$$\theta_t = \theta_{t-1} + \upsilon_t, \qquad \upsilon_t \sim N(0, Q),$$

R and Q are the unknown parameters in the observation equation and in the transition equation, respectively. The state variable θ_t is also unknown and to be estimated. Since the distributions of the unknown parameters and state variable depend on each other, we employ the Gibbs algorithm which iteratively draws parameters and unobserved states conditional on each other. We run the algorithm for 12,000 replications, with 10,000 burn-in replications discarded and 2,000 replications retained. The specific steps of the algorithm are briefly explained below. For further details, we refer the readers to Blake and Mumtaz (2012).

Step 1: Initialization.

To start Gibbs sampling, we assign initialization values to unknown parameters (R and Q) of the state space model. These are set using the OLS estimates of the observation equation by using the pre-sample data from 1974:Q4 to 1978:Q2. The initial value of Q is set by using the variance of θ_t from this estimation.

Step 2: Sample θ conditional on R and Q using the Carter and Kohn (1994) algorithm.

Given Y_t, X_t, and the initial values of R and Q, we run the Kalman filter to obtain the likelihood of the state vector. To initiate this filter, we use the OLS estimate of θ from the same pre-sample data in Step 1. To initiate the covariance matrix of θ, we multiply its estimate from the pre-sample data by T_0*3.5*10^{-4}, where T_0 is the length of the pre-sample data, and 3.5*10^{-4} is a small number that is used to adjust for the fact that the training sample is typically short and the resulting estimates may be imprecise.

The model is a linear Gaussian state space model. Assuming that the prior distribution for θ_0, denoted by $p(\theta_0)$, is Gaussian, the conditional posterior distribution of $p(\theta_t / y_t, R, Q)$ is also Gaussian. A forward recursion using the Kalman filter provides expressions for posterior means and the covariance matrix:

$$p(\theta_t/y_t, R, Q) = N(\theta_{t/t}, P_{t/t}),$$

$$P_{t/t-1} = P_{t-1/t-1} + Q,$$

$$K_t = P_{t/t-1}X_t(X_t'P_{t/t-1}X_t + R)^{-1},$$

$$\theta_{t/t} = \theta_{t/t-1} + K_t(y_t - X_t'\theta_{t/t-1}),$$

$$P_{t/t} = P_{t/t-1} - K_tX_t'P_{t/t-1} \ .$$

Starting from $\theta_{T/T}$ and $P_{T/T}$, we use the Kalman filter updating equations to characterize posterior distributions of $p(\theta^T/y^T, R, Q)$:

$$p(\theta_t/\theta_{t+1}, y^T, R, Q) = N(\theta_{t/t+1}, P_{t/t+1}),$$

$$\theta_{t/t+1} = \theta_{t/t} + P_{t/t}(P_{t/t} + Q)^{-1}(\theta_{t+1} - \theta_{t/t}) \ ,$$

$$P_{t/t+1} = P_{t/t} - P_{t/t}(P_{t/t} + Q)^{-1}P_{t/t}.$$

We use these equations to obtain draws of the state variable from its conditional distribution on the rest of the model parameters. That is, we generate a random trajectory for $\theta^T(= [\theta_1, \dots, \theta_T])$ using the backward recursion starting with a draw of θ_T from $\mathcal{N}(\theta_{T/T}, P_{T/T})$ as suggested by Carter and Kohn (1994).

Step 3: Sample Q from the inverse Wishart distribution.

Conditional on a realization for θ^T, innovations in the coefficients, v_t, are observable. Assuming an inverse-Wishart for the prior distribution of Q, with scale matrix Q_0 and degree of freedom T_0, its posterior distribution is also inverse-Wishart:

$$p(Q \ / \ y^T, \theta^T) = IW(Q_1^{-1}, T_1),$$

$$Q_1 = Q_0 + \textstyle\sum_{t=1}^{T} v_t v_t' \ , \quad T_1 = T_0 + T.$$

Given that Q_0 and T_0 are defined as the parameters governing the prior distribution of the variance of the transition equation, these equations suggest that this prior distribution is convolved with likelihood to obtain posterior distribution of Q.

Step 4: Sample R from the inverse Gamma distribution.

Conditional on a realization for θ^T, residuals from the time-varying regression are observable. Assuming the inverse-Gamma for prior distribution of R, with scale parameter R_0 and degree of freedom T_0, the posterior distribution is also inverse-Gamma:

$$p(R \,/\, y^T, \theta^T) = IG(T_1/2, 1/(2R_1)),$$

$$1/R_1 = (1/R_0 + \textstyle\sum_{t=1}^{T} \varepsilon_t \varepsilon_t')/2 \ ,$$

$$T_1 = (T_0 + T)/2.$$

where R_0 is the OLS estimate of R from the pre-sample data. Hence, the parameters of the prior distribution of R are also taken from the OLS estimate of pre-sample data.

Step 5: Posterior Inference.

Go back to step 1 and generate new draws of θ^T, R, and Q. Repeat this $M_0 + M_1$ times and discard the initial M_0 draws. Use the remaining M_1 draws for posterior inference.

References

Andrews, Donald W. K., 1993, "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61: 821-856.

Ang, Andrew, Geert Bekaert, and Min Wei, 2007, "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?", *Journal of Monetary Economics*, 54:1163-1212.

Atkeson, Andrew, and Lee E. Ohanian, 2001, "Are Phillips Curves Useful for Forecasting Inflation?", *FRB Minneapolis Quarterly Review*, Winter, 2–11.

Blake, Andrew P. and Haroon Mumtaz, 2012, "Applied Bayesian Econometrics for Central Bankers", Technical Books, Centre for Central Banking Studies, Bank of England, Edition 1, Number 4.

Carter, C. K. and R. Kohn, 1994, "On Gibbs Sampling for State Space Models", *Biometrika*, 81: 541-553.

Christiano, Lawrence J., 1992, "Searching for a Break in GNP", *Journal of Business and Economic Statistics*, 10: 237-249.

D'Agostino, Antonello and Karl Whelan, 2008, "Federal Reserve Information during the Great Moderation", Journal of the European Economic Association, 6: 609-620.

Durbin, J. and S. Koopman, 2012, "Time Series Analysis by State Space Methods", Second Edition, Oxford Statistical Science Series, OUP Oxford.

Ellingsen, Tore and Ulf Söderström, 2001, "Classifying Monetary Policy", Sveriges Riksbank Working Paper 56.

El-Shagi, Makram, Sebastian Giesen, and Alexander Jung, 2013, "Does Central Bank Staff Beat Private Forecasters?", Annual Conference 2013 (Duesseldorf): Competition Policy and Regulation in a Global Economic Order 79925, Verein für Socialpolitik / German Economic Association.

Fair, Ray C. and Robert J. Shiller, 1989, "The Informational Content of Ex Ante Forecasts", *Review of Economics and Statistics*, 71: 325–331.

Fair, Ray C. and Robert J. Shiller, 1990, "Comparing Information in Forecasts from Econometric Models", *American Economic Review*, 80: 375-38.

Gamber, Edward N. and Julie K. Smith, 2009, "Are the Fed's Inflation Forecasts Still Superior to the Private Sector's?", *Journal of Macroeconomics*, 31: 240-251.

Guidolin, Massimo, Francesco Ravazzolo, and Andrea Donato Tortora, 2013, "Alternative Econometric Implementations of Multi-factor Models of the U.S. Financial Markets", *Quarterly Review of Economics and Finance*, 53: 87-111.

Hansen, Bruce E., 1992, "Testing for Parameter Stability in Linear Models", *Journal of Policy Modelling*, 14: 517-533.

Harvey, A. C., 1990, "Forecasting, Structural Time Series Models and the Kalman Filter", Cambridge University Press, Cambridge.

Hubert, Paul, 2014, "Revisiting the Greenbook's Relative Forecasting Performance", *Revue de l'OFCE - Débats et politiques*, 137.

Peek, Joe, Eric Rosengren, and Geoffrey Tootell, 2003, "Does the Federal Reserve Possess an Exploitable Informational Advantage?", *Journal of Monetary Economics*, 50: 817-839.

Perron, Pierre, and Zhongjun Qu, 2007, "Estimating and Testing Structural Changes in Multivariate Regressions", *Econometrica*, 75: 459-502.

Quandt, Richard E., 1958, "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes", *Journal of the American Statistical Association*, 53: 873-880.

Quandt, Richard E., 1960, "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, 55: 324-330.

Romer, Christina D. and David H. Romer, 2000, "Federal Reserve Private Information and the Behavior of Interest Rates", *American Economic Review*, 90: 429-457.

Sims, Christopher, 2002, "The Role of Models and Probabilities in the Monetary Policy Process", *Brookings Papers on Economic Activity*, 2, 1-62.

Stock, James, and Mark Watson, 2007, "Why Has U.S. Inflation Become Harder to Forecast?", *Journal of Money, Credit, and Banking*, 39: 3-34.

Zivot, Eric, and Donald W. K. Andrews, 1992, "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Journal of Business and Economic Statistics*, 10: 251-27.

Table I

Forecasting performances of the Greenbook and the SPF

$$\pi_{h,t} = \beta_0 + \beta_1 \pi_{h,t}^{GB} + \beta_2 \pi_{h,t}^{SPF} + \varepsilon_{h,t}$$

	Forecast horizon				
	0	1	2	3	4
β_0	0.247	0.335	0.494	0.533	0.446
	(0.85)	(0.94)	(1.15)	(1.13)	(0.96)
β_1	0.612	0.823	1.062	1.175	1.168
	(3.11)***	(3.69)***	(4.09)***	(4.24)***	(4.41)***
β_2	0.319	0.0934	-0.169	-0.276	-0.252
	(1.37)	(0.41)	(-0.63)	(-0.99)	(-1.00)
N	141	141	141	141	141
R^2	0.89	0.86	0.85	0.84	0.82

Notes: $\pi_{h,t}$ denotes the actual inflation rate and $\pi_{h,t}^{GB}$ and $\pi_{h,t}^{SPF}$ denote, respectively, the Greenbook and SPF inflation forecasts. *t*-statistics are in parentheses. ** and *** denote significance at the 5 and 1 percent levels, respectively.

Table II

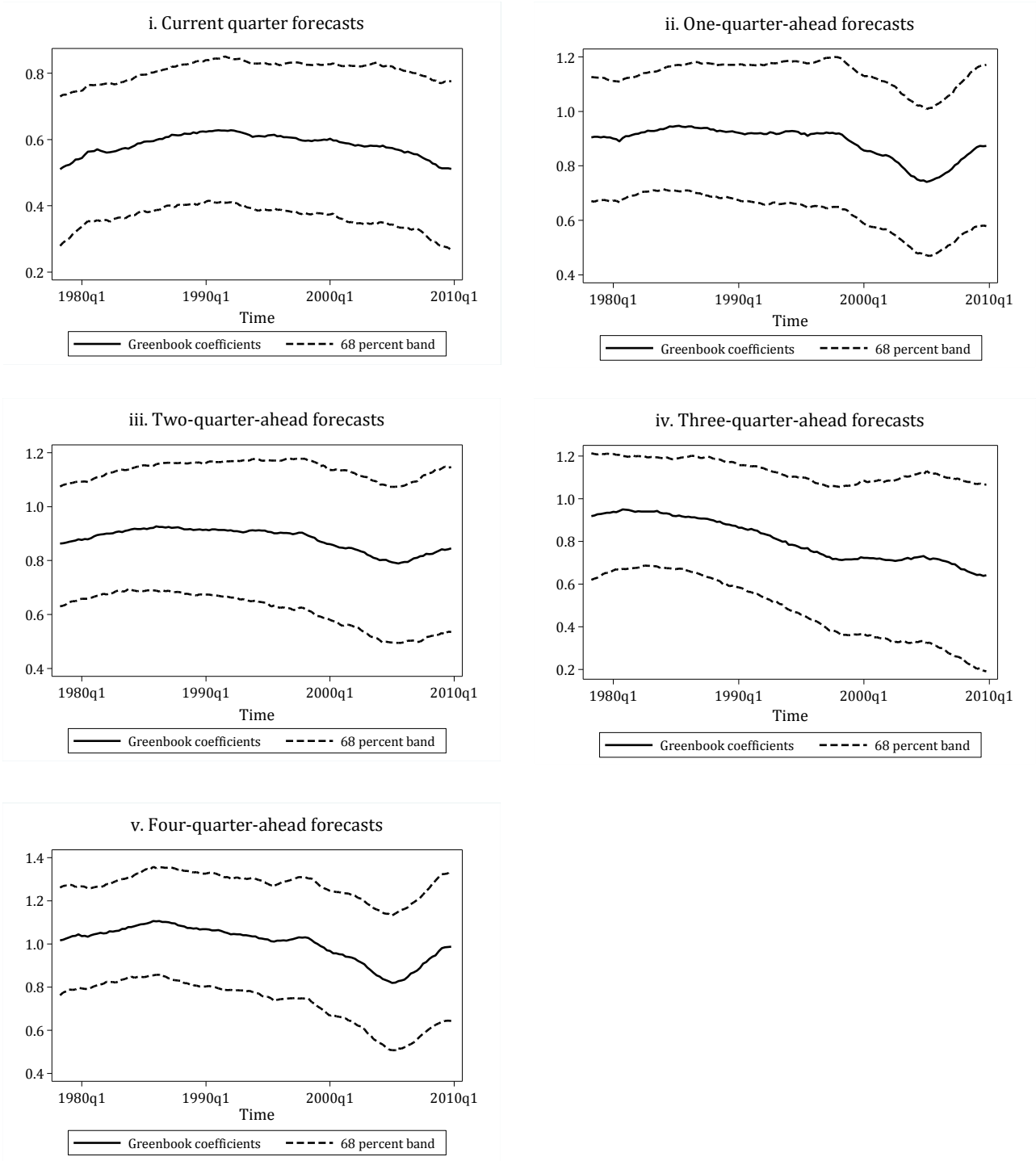
Structural changes in the forecast powers of the Greenbook and the SPF

$$\pi_{h,t} = \beta_0 + \beta_1 \pi_{h,t}^{GB} + \beta_1^D (\pi_{h,t}^{GB} \times D_{i^*}) + \beta_2 \pi_{h,t}^{SPF} + \varepsilon_{h,t}$$

	Forecast horizon		
	2	3	4
	post-1980:Q3	post-1980:Q3	post-1980:Q3
β_0	0.835	0.942	0.804
	(1.69)*	(1.86)*	(1.73)*
β_1	1.037	1.260	1.146
	(4.01)***	(4.93)***	(5.90)***
β_1^D	-0.271	-0.325	-0.380
	(-2.62)***	(-3.13)***	(-3.39)***
β_2	-0.0544	-0.243	-0.0640
	(-0.19)	(-0.90)	(-0.32)
N	141	141	141
R^2	0.86	0.85	0.85

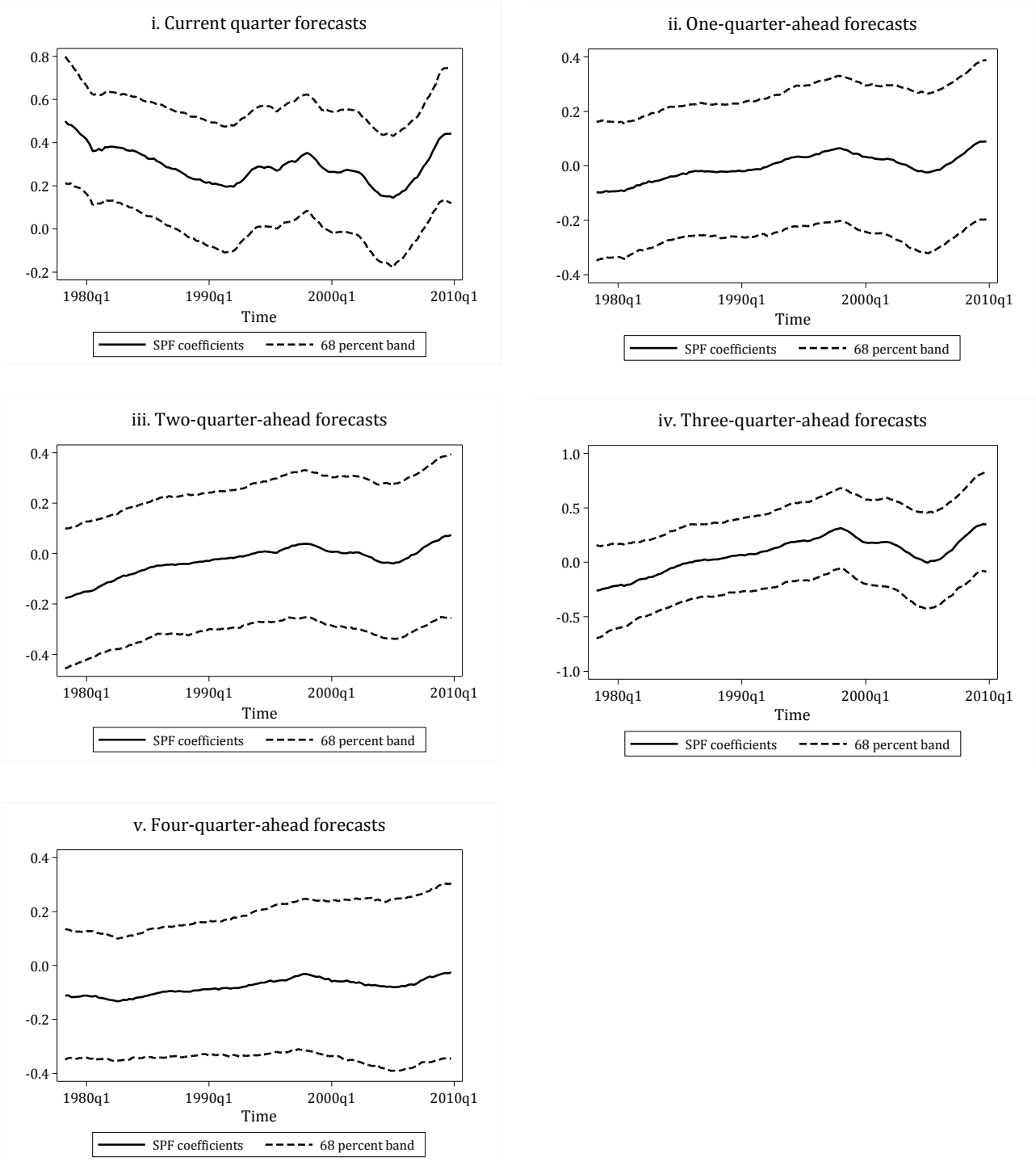
Notes: $\pi_{h.t}$ denotes the actual inflation rate and $\pi_{h.t}^{GB}$ and $\pi_{h.t}^{SPF}$ denote, respectively, the Greenbook and SPF inflation forecasts. D_{i^*} is equal to zero before the structural break date i^* and one afterwards, where i^* = 1980:Q3. *t*-statistics are in parentheses. ** and *** denote significance at the 5 and 1 percent levels, respectively.

Figure I: Time-varying coefficients of Greenbook forecasts



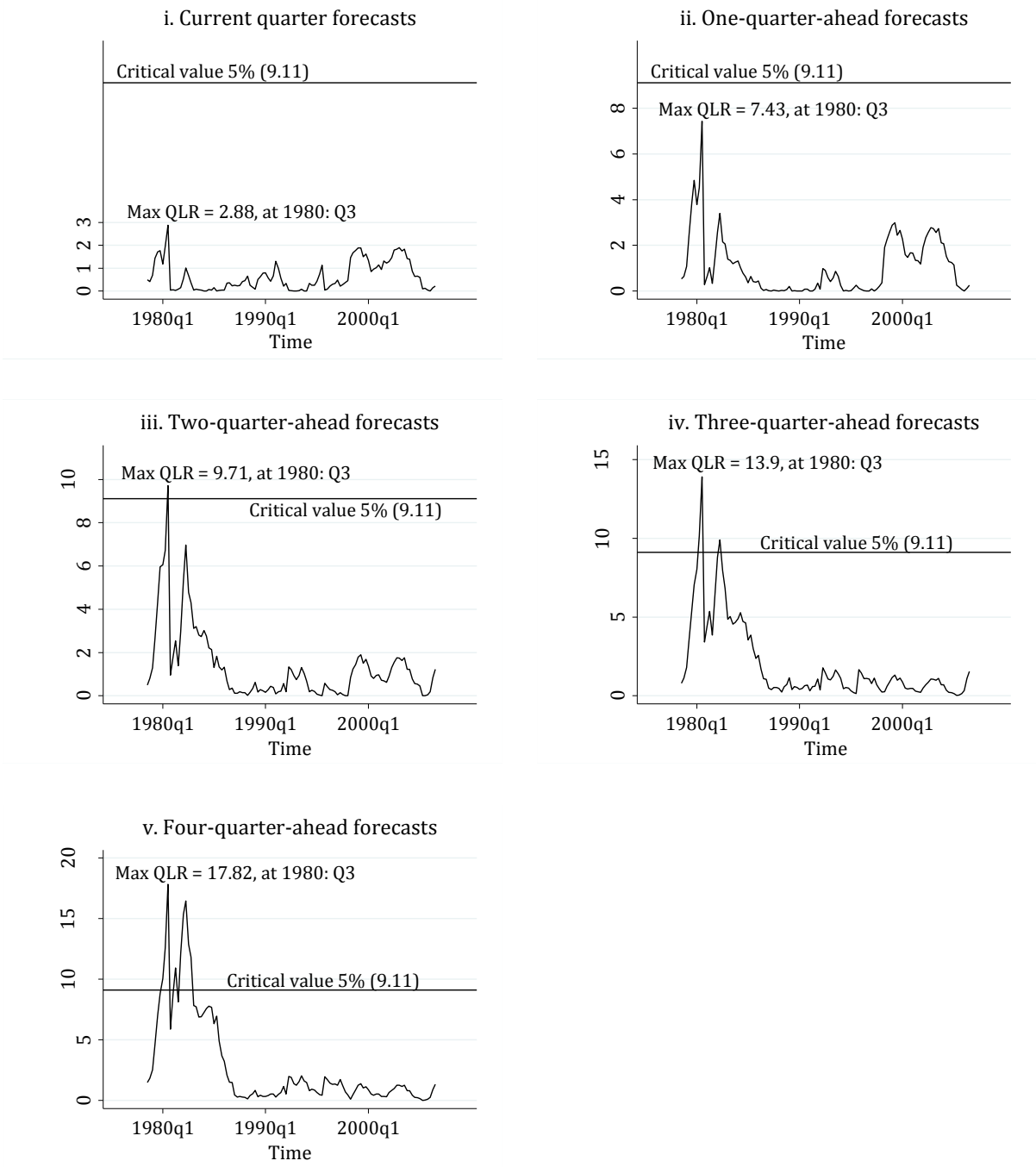
Note: Time-varying Greenbook coefficients are computed by Bayesian estimation with Kalman filtering and Gibbs sampling algorithm. 68 percent confidence bands are displayed. The procedure uses the 1974:Q4-1978:Q2 data to set priors and initial values. As a result, the sample runs from 1978:Q3 to 2009:Q4.

Figure II: Time-varying coefficients of SPF forecasts



Note: Time-varying SPF coefficients are computed by Bayesian estimation with Kalman filtering and Gibbs sampling algorithm. 68 percent confidence bands are displayed. The procedure uses the 1974:Q4-1978:Q2 data to set priors and initial values. As a result, the sample runs from 1978:Q3 to 2009:Q4.

Figure III: QLR statistics of the structural break test at different horizons



Note: The QLR statistics are based on Andrews (1993) as discussed in the text. Most likely breakpoints are dated by the time of the event, not the time of the forecast. The break tests do not use the first and last 10 percent of the data. As a result, the sample runs from 1978:Q3 to 2006:Q3.

Central Bank of the Republic of Turkey
Recent Working Papers
The complete list of Working Paper series can be found at Bank's website
(http://www.tcmb.gov.tr).

Importance of Foreign Ownership and Staggered Adjustment of Capital Outflows
(Özgür Özel ,M. Utku Özmen,Erdal Yılmaz Working Paper No. 15/31 November 2015)

Sources of Asymmetry and Non-linearity in Pass-Through of Exchange Rate and Import Price to Consumer Price Inflation for the Turkish Economy during Inflation Targeting Regime
(Süleyman Hilmi Kal, Ferhat Arslaner, Nuran Arslaner Working Paper No. 15/30 November 2015)

Selective Immigration Policy and Its Impacts on Natives: A General Equilibrium Analysis
(Şerife Genç İleri Working Paper No. 15/29 November 2015)

How Does a Shorter Supply Chain Affect Pricing of Fresh Food? Evidence from a Natural Experiment
(Cevriye Aysoy, Duygu Halim Kırlı, Semih Tümen Working Paper No. 15/28 October 2015)

Decomposition of Labor Productivity Growth: Middle Income Trap and Graduated Countries
(Gökhan Yılmaz Working Paper No. 15/27 October 2015)

Estimating Income and Price Elasticity of Turkish Exports with Heterogeneous Panel Time-Series Methods
(İhsan Bozok, Bahar Şen Doğan, Çağlar Yüncüler Working Paper No. 15/26 October 2015)

External Shocks, Banks and Monetary Policy in an Open Economy: Loss Function Approach
(Yasin Mimir, Enes Sunel Working Paper No. 15/25 September 2015)

Tüm Yeni Açılan Krediler Eşit Mi? Türkiye'de Konut Kredisi ve Konut Kredisi Dışı Borç ile Özel Kesim Tasarruf Oranı
(Cengiz Tunç, Abdullah Yavaş Working Paper No. 15/24 September 2015)

A Computable General Equilibrium Analysis of Transatlantic Trade and Investment Partnership and Trans-Pacific Partnership on Chinese Economy
(Buhara Aslan, Merve Mavuş Kütük, Arif Oduncu Working Paper No. 15/23 September 2015)

Export Behavior of the Turkish Manufacturing Firms
(Aslıhan Atabek Demirhan Working Paper No. 15/22 August 2015)

Structure of Debt Maturity across Firm Types
(Cüneyt Orman, Bülent Köksal Working Paper No. 15/21 August 2015)

Government Subsidized Individual Retirement System
(Okan Eren, Şerife Genç İleri Working Paper No. 15/20 July 2015)

The Explanatory Power and the Forecast Performance of Consumer Confidence Indices for Private Consumption Growth in Turkey
(Hatice Gökçe Karasoy, Çağlar Yüncüler Working Paper No. 15/19 June 2015)

Firm Strategy, Consumer Behavior and Taxation in Turkish Tobacco Market
(Oğuz Atuk, Mustafa Utku Özmen Working Paper No. 15/18 June 2015)

International Risk Sharing and Portfolio Choice with Non-separable Preferences
(Hande Küçük, Alan Sutherland Working Paper No. 15/17 June 2015)

A Theory of Intra-Firm Group Design
(Semih Tümen Working Paper No. 15/16 June 2015)

Government Spending Multiplier in Turkey
(Cem Çebi Working Paper No. 15/15 June 2015)