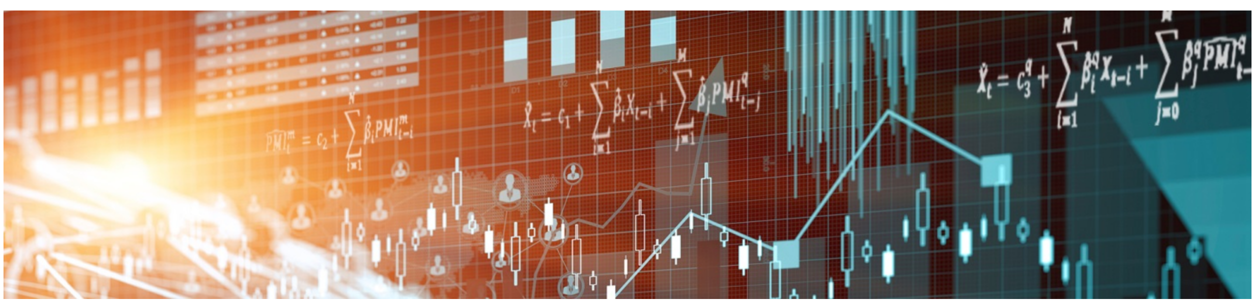
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2018-52

Evaluating the Bank of Canada Staff Economic Projections Using a New Database of Real-Time Data and Forecasts



by Julien Champagne, Guillaume Poulin-Bellisle and Rodrigo Sekkel

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2018-52

November 2018

Evaluating the Bank of Canada Staff Economic Projections Using a New Database of Real-Time Data and Forecasts

Julien Champagne,[1] Guillaume Poulin-Bellisle[2] and Rodrigo Sekkel[3]

[1] Canadian Economic Analysis Department
[2] International Economic Analysis Department
[3] Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
champagnej@bankofcanada.ca
poug@bankofcanada.ca
rsekkel@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

We thank Jason Allen, Sharon Kozicki, Lawrence Schembri, Jean-Philippe Cayen, Dalibor Stevanovic, Shaun Vahey, Simon van Norden, Dean Croushore, and seminar participants at the Bank of Canada; at CIRANO (Montréal); the 2018 Canadian Economics Association Conference (McGill University); and the 2018 International Association for Applied Econometrics (UQAM, Montréal) for useful conversations and suggestions, and Jaime Trujillo and Sarah Burkinshaw for excellent research assistance.

Abstract

We present a novel database of real-time data and forecasts from the Bank of Canada's staff economic projections. We then provide a forecast evaluation for GDP growth and CPI inflation since 1982: we compare the staff forecasts with those from commonly used time-series models estimated with real-time data and with forecasts from other professional forecasters and provide standard bias tests. Finally, we study changes in the predictability of the Canadian economy following the announcement of the inflation-targeting regime in 1991. Our database is unprecedented outside the United States, and our evidence is particularly interesting, as it includes over 30 years of staff forecasts, two severe recessions and different monetary policy regimes. The database will be made available publicly and updated annually.

Bank topics: Monetary policy; Economic models; Inflation targets; Econometric and statistical methods
JEL codes: C32, E17, E37

Résumé

Nous présentons une nouvelle base de données en temps réel et de prévisions tirées des projections économiques du personnel de la Banque du Canada. Nous évaluons ensuite les prévisions relatives à la croissance du PIB et à l'inflation mesurée par l'indice des prix à la consommation depuis 1982. Pour ce faire, nous comparons les prévisions du personnel aux prévisions de modèles économétriques courants estimés à partir de données en temps réel ainsi qu'aux prévisions d'autres professionnels en la matière. Nous effectuons ensuite des tests pour vérifier la présence de biais. Enfin, nous étudions la variation de la prévisibilité de l'économie canadienne après l'annonce du régime de ciblage de l'inflation en 1991. Notre base de données est la première de cette ampleur à l'extérieur des États-Unis, et nos résultats sont particulièrement intéressants, car ils couvrent plus de 30 années de prévisions du personnel, deux récessions majeures et différents régimes de politique monétaire. La base de données, accessible au public, sera mise à jour annuellement.

Sujets : Politique monétaire ; Modèles économiques ; *Cibles en matière d'inflation* ;
Méthodes économétriques et statistiques
Codes JEL : C32, E17, E37

Non-technical summary

Central banks use monetary policy instruments to achieve their mandates, such as inflation targeting (IT) and maximum sustainable employment. However, there are well-known lags between changes in these instruments and their ultimate effects, and these effects, as well as future economic conditions, are always uncertain. Consequently, the accuracy of central banks' staff economic forecasts plays a crucial role in the effective implementation of monetary policy.

We present a new, rich database containing both real-time historical data and forecasts collected from archived Bank of Canada staff economic projections. Our database consists of quarterly vintages of numerous Canadian macroeconomic aggregates such as real GDP, total and core CPI, GDP components, the output gap, the unemployment rate and many others, some of which go back to the early 1970s. Notably, we have a consistent set of real-time data and forecasts for GDP growth and CPI inflation since 1982Q2.

We highlight several uses of this data-set. We start by providing a thorough evaluation of the Bank of Canada staff forecasts for Canadian GDP growth and CPI inflation. We compare the accuracy of the staff forecasts with simple benchmark models and test these forecasts for systematic bias. We find that the staff forecast for the current quarter for both real GDP growth and CPI inflation are significantly more accurate than those from the econometric models. This reflects the fact that some hard and soft indicators for the current quarter are available to the staff at the time the forecasts are prepared. This advantage is much less persistent for real GDP growth than for CPI inflation: it essentially disappears after one quarter for the former but persists all through the forecasting horizons for the latter.

We also study the subsample stability of our results before and after the introduction of IT in 1991, and find that for both real GDP growth and CPI inflation, forecast errors are smaller in the latter sample. However, for CPI inflation, the decline in forecast errors for some of the models was much larger than for the staff, signaling a change in the predictability of inflation. We investigate these results further and perform a variance decomposition of GDP growth and CPI inflation into predictability and uncertainty components before and after IT. While both components have decreased for GDP growth following IT, the decline in uncertainty is more important. For inflation, we find a substantial drop in predictability, while uncertainty has stayed roughly constant. Finally, we perform standard statistical bias tests and find that staff GDP growth forecasts are biased at long horizons, whereas they are not for CPI inflation.

After comparing the staff forecasts with the benchmark models, we examine how well they contrast with those of other professional forecasters. Using a smaller subset of our sample, starting in 1994, we argue that the Bank of Canada staff forecasts are more accurate and informative than those of the average professional forecaster. Finally, we show that the forecasts prepared by the Bank of Canada's Governing Council and published in the *Monetary Policy Report* (MPR) are usually more accurate than the staff forecast for the current year, bur less so for next year, although the MPR forecasts are based on one more month of data.

1 Introduction

Central banks' staff economic forecasts play a critical role in the effective implementation of monetary policy. For instance, there are well-documented lags between changes in the stance of monetary policy and desired outcomes; moreover, there is uncertainty around future economic conditions. It is therefore crucial to evaluate the accuracy of central banks' staff forecasts.

Some central banks and multilateral institutions (such as the International Monetary Fund (IMF) and the World Bank) periodically publish evaluations of their forecasting performance. Most of these evaluations are for a recent period, usually from the mid-1990s onward (e.g., see Groen et al. (2009) for the Bank of England and Timmermann (2007) for the IMF's *World Economic Outlook*). A notable exception is the Federal Reserve Board (FRB), which publishes both its staff (the 'Greenbooks') and the Federal Open Market Committee (FOMC) forecasts.[1] Specifically, the FRB's staff Greenbook forecasts are publicly available and go back to 1967. Related, some prominent papers have exposed the importance of using real-time data when evaluating forecast performance or testing new forecasting methods (e.g., Croushore and Stark (2002, 2003), Faust and Wright (2009)). The pioneering work in this literature started at the Federal Reserve Bank of Philadelphia, where vintages of real-time data covering many U.S. macroeconomic aggregates were collected into a single database (Croushore and Stark (2001)). Other researchers followed course and more real-time databases containing vintages of macroeconomic data for different countries became available for research (e.g., Garratt and Vahey (2006) for the U.K., Giannone et al. (2012) for the euro area and Fernandez et al. (2011) for some OECD countries). The availability of real-time data and, importantly, the Greenbook forecasts have sparked a large literature in empirical macroeconomics. For instance, several papers analyze the quality of these forecasts and conclude that they are at or near the frontier of predictability;[2] as a result, they have become the *de facto* testing ground for new rationality tests (e.g., Patton and Timmermann (2012) and Rossi and Sekhposyan (2016)). Furthermore, the Greenbooks have

[1]The FRB publishes the Greenbook forecasts with a five-year lag.

[2]For examples, see Romer and Romer (2000), Sims (2002), Romer and Romer (2008) or Faust and Wright (2009).

been used to study various important questions in macroeconomics, such as the effects of monetary policy (Romer and Romer (2004), Coibion et al. (2017), Cloyne et al. (2018)), the monetary policy implications of output gap revisions (Orphanides (2003)), and fiscal and monetary policy interactions (Croushore and van Norden (2017)), among others.[3]

In this paper, we present a new database containing both real-time historical data and forecasts collected from archived Bank of Canada staff economic projections. Our database is comprised of quarterly vintages of numerous Canadian macroeconomic aggregates such as real gross domestic product (GDP) and its components, total and core Consumer Price Index (CPI), the output gap, the unemployment rate, interest rates and many others, some of which go back to the early 1970s. Notably, we have a consistent set of real-time data and forecasts for GDP growth and CPI inflation since 1982.[4] To our knowledge, no other such extensive data-set of central bank staff forecasts is available for such a long time period apart from the FRB's Greenbooks. Our data-set is interesting as it provides a nice case study of the interplay between central bank forecasting and a change of monetary policy regime (announcement of inflation targeting (IT) in 1991). During the sample covered by our database, Canada also experienced two severe recessions (1991–92 and 2008–09) and large swings in commodity prices and exchange rates.

We then use this novel data-set to provide a thorough evaluation of the Bank of Canada staff forecasts for GDP growth and CPI inflation. We compare the accuracy of the staff forecasts with benchmark econometric models and test these forecasts for systematic bias. We further use the data-set to study changes in predictability in the Canadian economy after the announcement of IT. We highlight several takeaways. First, the staff nowcasts (its forecasts for the current quarter) for both real GDP growth and CPI inflation are significantly more accurate than those from standard econometric models. This reflects the fact that some hard and soft indicators for the current quarter are available to the staff when the forecasts are prepared. Second, this nowcasting advantage is much less persistent for real GDP growth than for CPI inflation, as it essentially disappears after

[3]Another strand in the literature has stressed the importance of large data-sets for macro analysis. Such databases have been created for the U.S. by Stock and Watson (1996) and McCracken and Ng (2016) and by Fortin-Gagnon et al. (2018) for Canada.

[4]More precisely, each projection vintage starting in 1982Q1 has forecasts for the current quarter (nowcast) and up to 8 quarters ahead for both GDP growth and CPI inflation. See Section 2 and Appendix A for more details.

one quarter for the former but persists all through the forecasting horizons (8 quarters ahead) for the latter. Third, we show that when we provide the benchmark models with the staff short-term forecasts, their longer-term forecasts for GDP growth improve relative to the staff's, while for CPI inflation, the staff's advantage remains. Fourth, when studying the subsample stability of our results before and after the introduction of IT in 1991, we find that for both real GDP growth and CPI inflation, forecast errors are smaller in the latter sample. On the one hand, for real GDP growth the results are qualitatively the same across both samples: the staff is more accurate than the models for the short-term forecasts but slightly less accurate than a simple autoregressive (AR) model over longer horizons. On the other hand, results are markedly different for CPI inflation: although the forecast errors for both staff and the models have declined since IT was introduced, they have done so much more forcefully for some econometric models. Fifth, we build on these results and perform a variance decomposition of GDP growth and CPI inflation into predictability and uncertainty components before and after IT; we find that while both components have decreased for GDP growth following IT, the decline in uncertainty is relatively more important. For inflation, we find a substantial drop in predictability while uncertainty has stayed roughly constant. Our findings for inflation are consistent with the U.S. evidence presented by Tulip (2006), Stock and Watson (2007) and Faust and Wright (2013). Finally, we perform standard statistical bias tests and find that staff GDP growth forecasts are biased at long horizons, whereas they are not for CPI inflation.

Several papers show that the FRB's Greenbook forecasts compare well with other forecasters. For example, using data from 1965 to 1991, Romer and Romer (2000) find that the Greenbook inflation forecasts are more accurate than private sector forecasts. Moreover, Romer and Romer (2008) find that the Greenbook forecasts are more informative about future real GDP growth, the unemployment rate and inflation than the ones prepared by the FOMC. We conduct a similar analysis, albeit with a shorter sample, and argue that the Bank of Canada staff forecasts are more accurate and informative than the average professional forecasters' forecasts. Furthermore, we find that the forecasts prepared by the Bank of Canada's Governing Council and published in the *Monetary Policy Report* (MPR) are usually more accurate than the staff's for the current year, but less so for the next year, although the MPR forecasts are based on one more month of data.

The remainder of the paper is organized as follows. Section 2 provides a description of our new database of real-time historical data and Bank of Canada staff's forecasts. In Section 3, we provide an evaluation of the staff forecasts. Section 4 assesses the robustness of our results to different subsamples (before and since the introduction of IT in Canada in 1991) and evaluates how these changes translate into changes in predictability and uncertainty. In Section 5, we compare the staff forecasts with those of professional forecasters and from the MPR. Section 6 concludes.

2 The real-time database of Bank of Canada staff projections

We construct a novel database of real-time data and forecasts that we collect from archived Bank of Canada staff economic projections. These staff projections contain vintages of historical (real-time) data for numerous macroeconomic aggregates at a quarterly frequency. Consequently, each data vintage gives a snapshot of the staff's information set on the date of the vintage. These historical data are usually available with the release of the quarterly national income and expenditure accounts (NAC), i.e. at the beginning of the months of March, June, September and December. Moreover, a set of quarterly forecasts is associated with each data vintage for all the variables: Bank of Canada staff has been producing four exhaustive sets of forecasts each year following the release of the NAC and they are generally carried out around the end of the given quarter.[5] Each quarterly vintage of real-time data and staff forecasts is saved into what is called the "staff economic projections." We went through a large amount of (unpublished) documents to ensure that the older data vintages in our database contain the correct information. For example, we corroborated forecast values from hard copies of archived staff economic reports, as well as historical data values from archived *Bank of Canada Reviews*. This database will be publicly available on the

[5] In a few instances between 2006 and 2012, Staff produced eight projections per year, one before each of the eight monetary policy announcements (e.g., see Champagne and Sekkel (2018)). For the forecast comparison exercise in this paper, we use one projection per quarter even when we have more projections available.

Bank of Canada website and updated annually.[6]

The staff forecasts are an important part of the analysis presented to the Governing Council every quarter in the weeks leading up to the publication of the Bank's MPR.[7] They are analogous to the Greenbook forecasts prepared by the FRB staff; they are judgmental in the sense that forecasts are prepared using different sources of information and economic models. These models to nowcast, forecast and analyze the Canadian economy have evolved substantially over the years. From about 1970 to 1993, staff used econometric models such as RDX1, RDX2 and RDXF (Helliwell (2005)), while from 1993 to 2005, staff worked with the Quarterly Projection Model (QPM, e.g., Poloz et al. (1994)). Between 2005 and 2015, the staff used ToTEM, an open-economy, New Keynesian DSGE-based model (Dorich et al. (2013)), and since 2015, staff has broadened its toolkit by adding another model (called LENS, e.g., Gervais and Gosselin (2014)) to complement ToTEM, a forecasting model in the spirit of the FRB-U.S. (e.g., Brayton et al. (2014)).

Our database contains quarterly data vintages for real and nominal GDP (and its components), CPI, core CPI, the output gap, the unemployment rate, the U.S./Canada nominal exchange rate and other macroeconomic aggregates from the early 1970s onward. All variables in our database are either in levels or indexes, other than rates like the output gap and unemployment. Because there is a five-year ban on access to staff data, we can only use forecast data prepared up to the 2013Q4 vintage. For this paper, we restrict the forecast sample to data vintages starting in 1982Q2, such that all the vintages contain (at least) real-time data extending back to 1973Q1 and include a nowcast for the current quarter as well as forecasts for (at least) 8 quarters ahead. As a result, there are 128 data vintages that include both real-time historical data and forecasts, ranging from 1982Q2 to 2013Q4, while the dates which the forecasts refer to range from 1982Q2 (the first nowcast) to 2015Q4 (the last 8-quarters-ahead forecast). We hope this new database will prove useful for future research. We refer the reader to the Appendix for further details about the new database of Bank of Canada staff projections.

[6]Some parts of this database have been used in two other papers (e.g., see Champagne et al. (2018) and Champagne and Sekkel (2018)).

[7]See Murray (2013) for details about the information and analysis presented by the staff to the Governing Council. We highlight the fact that these are staff estimates and are not the same as the estimates provided in the MPR.

The baseline analysis of the paper focuses on GDP growth and CPI inflation forecasts.[8] Growth rates are computed as quarter-over-quarter, annualized rates, i.e. $400 * (y_t / y_{t-1} - 1)$, where y is the variable of interest. We compute the forecast error for horizon h (e_{t+h}) as the actual (y_{t+h}) minus the forecast (y^F_{t+h}) value:

$$e_{t+h} = y_{t+h} - y^F_{t+h} \tag{1}$$

It is not clear what vintage to treat as actual data, since data are continuously revised (especially for GDP growth; less so for CPI). Following Tulip (2006), Faust and Wright (2009) and many others, we use the data as recorded in our real-time data-set two quarters after the quarter which the data refer to.[9] This definition of actuals minimizes the effects of changes in data definitions as forecast errors and should be conceptually similar to the series staff were trying to forecast at any given point in time.

Figure 1 shows some illustrative data. The dots show the staff 4-quarter ahead forecasts for GDP growth (upper panel) and CPI inflation (lower panel) while the solid lines represent the analogous actual values between 1982 and 2013. Dates denote the time of the event, not the time the forecasts were produced. It is clear that the swings in economic activity in the 1980s and early 1990s were perceived as more persistent by the staff than those fluctuations after the mid-1990's. For example, the staff did relatively well at anticipating the slowdown in economic growth that preceded the 1991–92 recession, but greatly overestimated GDP growth during the recovery phase. On the inflation front (lower panel), CPI inflation declined from highs above 11 per cent (q/q, annualized) in 1982 (not shown) to hovering between 3 and 6 per cent throughout the 1980s until the first inflation targets set for 1992.[10] Thereafter, it decreased substantially to an average of about 2 per cent and has experienced milder fluctuations since. Looking at the forecasts, it seems that staff perceived the shocks to CPI inflation as temporary after the introduction of IT up to the 2008–09 financial crisis. From 2009 onwards, there are larger swings in the 1-year-ahead forecasts, reflecting

[8] In the Appendix, we provide analogous forecast evaluation for core CPI inflation and the GDP components.

[9] For example, the actual value for 2010Q1 is taken from the 2010Q3 vintage.

[10] It is worth pointing out that CPI inflation was actually increasing in the late 1980s, which led to the introduction of the IT regime. Further note that the spike in 1991 is due to the introduction of the federal sales tax.

the fluctuations in economic activity, commodity prices and exchange rates during this period.

3 Evaluating the Bank of Canada staff forecasts

In this section, we evaluate the Bank of Canada staff forecasts by examining their accuracy, comparing them with forecasts from econometric models and testing for the presence of bias. Although there is an untold number of benchmark models that can be used to evaluate the staff projections, we restrict the analysis to simple time series models. Notwithstanding their simplicity, outperforming them is a daunting task for more sophisticated alternatives (Faust and Wright (2009)).

3.1 Benchmark time series models

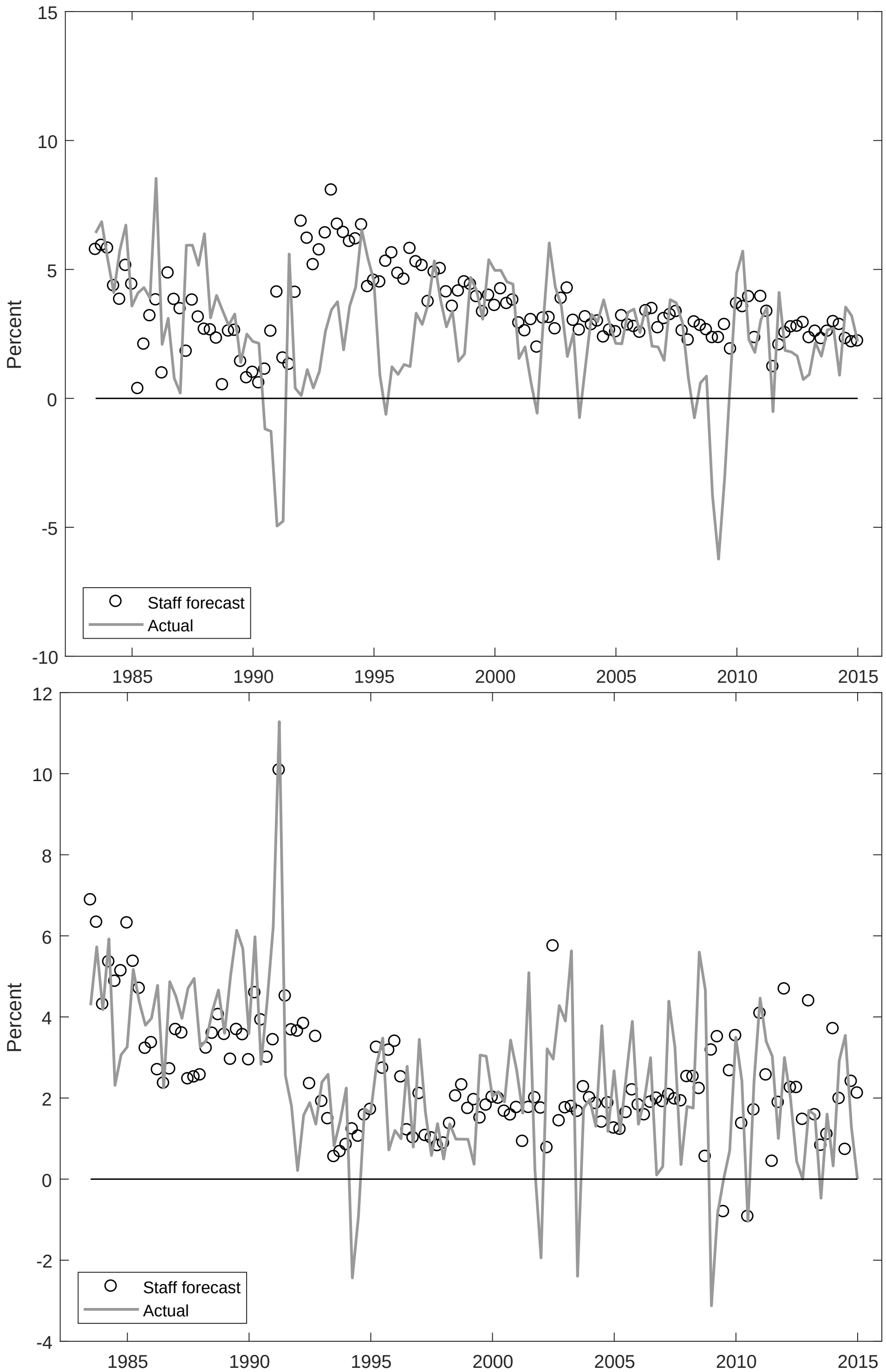
The following forecasting models are estimated recursively with real-time data, and produce forecasts for annualized quarter-over-quarter GDP growth and CPI inflation the 1982Q2 to 2015Q4 period. At quarter T, we consider forecasts for the current quarter ($h = 0$) to 8 quarters ahead.

1. *A univariate AR(4) model.* We estimate the model $y_t = \rho_0 + \sum_{i=1}^{4} y_{t-i} + \varepsilon_t$, and iterate it forward to generate forecasts for y_{T+h}, where h is the horizon of the forecast, in quarters. Faust and Wright (2009) find that the AR(4) model generates out-of-sample forecasts for U.S. inflation that are hard to improve upon.

2. *The Markov-Switching AR model (MS-AR)*: $y_t = \rho_{0,i}(s_t) + \sum_{j=1}^{p} \rho_j y_{t-j} + \varepsilon$. This model is similar to the AR(4) above, but allows the constant to switch between two states s_t.

3. *The Vector Autoregressive model (VAR)*: $Y_t = \rho_0 + \sum_{j=1}^{p} \rho_j Y_{t-j} + \varepsilon_t$, where $Y_t = (x_t, \pi_t, i_t)$, x_t is real GDP growth, π_t is CPI inflation and i_t is the 3-month trea-

Figure 1: Staff 4-quarter-ahead GDP growth and CPI forecasts *vs.* actual values



Note: This figure shows staff 4-quarter-ahead GDP growth (upper panel) and 4-quarter-ahead CPI inflation (lower panel) forecasts (dots) as well as the corresponding actual values (solid line).

sury bill rate. As with the univariate models above, we estimate the VAR model with 4 lags.

4. *The Atkenson-Ohanian Random Walk (AORW).* This model simply takes $\frac{1}{4}\sum_{j=1}^{4} y_{T-i}$ as a forecast for y_{T+h}, where, as above, h is the horizon of the forecast in quarters. Atkeson and Ohanian (2001) show that out-of-sample forecasts for U.S. inflation from this modified random walk model are more accurate than traditional Phillips Curve models.

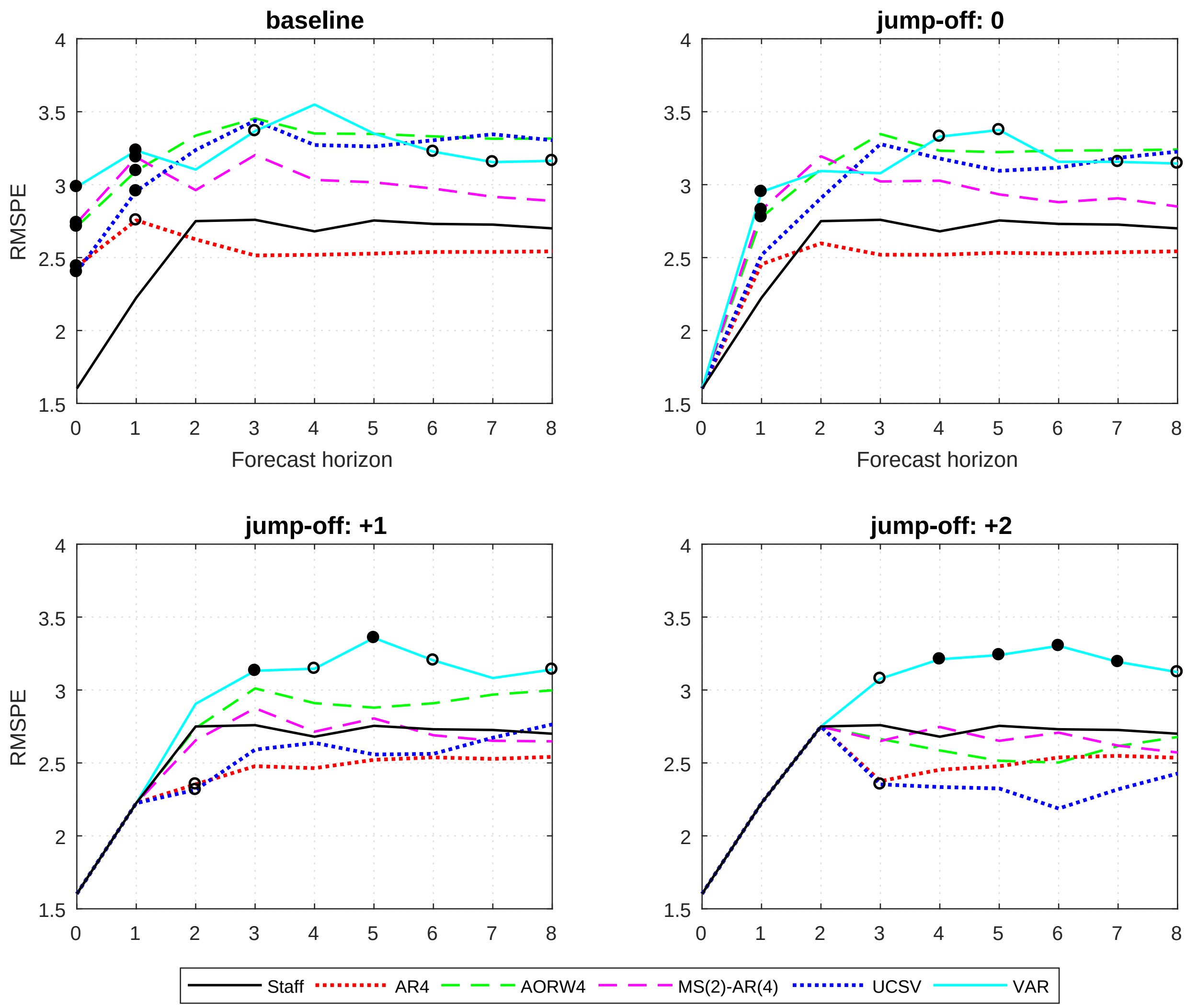
5. *The Unobserved Components Stochastic Volatility (UCSV) model of Stock and Watson (2007).* This model decomposes inflation and GDP growth into trend and cycle components and uses the filtered estimates of the trend as a forecast for y_{T+h}. Let $y_t = \tau_t + \eta_t^T$ and $\tau_t = \tau_{t-1} + \eta_t^P$, where η_t^T and η_t^P are *iid* $N(0, \sigma_{T,t}^2)$ and *iid* $N(0, \sigma_{P,t}^2)$, respectively. Both conditional variances evolve as a random walk $log(\sigma_{T,t}^2) = log(\sigma_{T,t-1}^2) + \psi_{1,t}$ and $log(\sigma_{P,t}^2) = log(\sigma_{P,t-1}^2) + \psi_{2,t}$, with $(\psi_{1,t}, \psi_{2,t})'$ both *iid* $N(0, \lambda I)$. As in Stock and Watson (2007), we fix λ at 0.04.

3.2 Forecast comparison results

Our main metric for forecast accuracy is the root mean square prediction error (RMSPE). We provide RMSPE for the nowcast (horizon 0) through 8 quarters ahead and signal the statistical significance of the difference of RMSPE between the staff's and each econometric model at the 10 and 5 per cent levels with circles ○ and ●, respectively. The upper left panels ('baseline') of Figures 2 and 3 display the main results for the full sample (1982Q2–2015Q4) for real GDP growth and CPI inflation, respectively. For both GDP growth and CPI inflation, the staff output growth nowcasts ($h = 0$) are significantly more accurate than the econometric models, by a factor of about 1.5 for GDP and 4.0 for inflation. This is consistent with the evidence from Faust and Wright (2009) and Chauvet et al. (2013) for the U.S., and speaks to the view that central banks, or professional forecasters in general, exploit a large number of hard and soft indicators when assessing the current state of the economy. The persistence of the staff's forecasting advantage over the models is very different for GDP growth relative to CPI inflation. While this advantage persists throughout the forecast horizons (up to 8-quarter-ahead) for CPI inflation, it essentially disappears for horizons longer than 2 quarters for GDP growth.

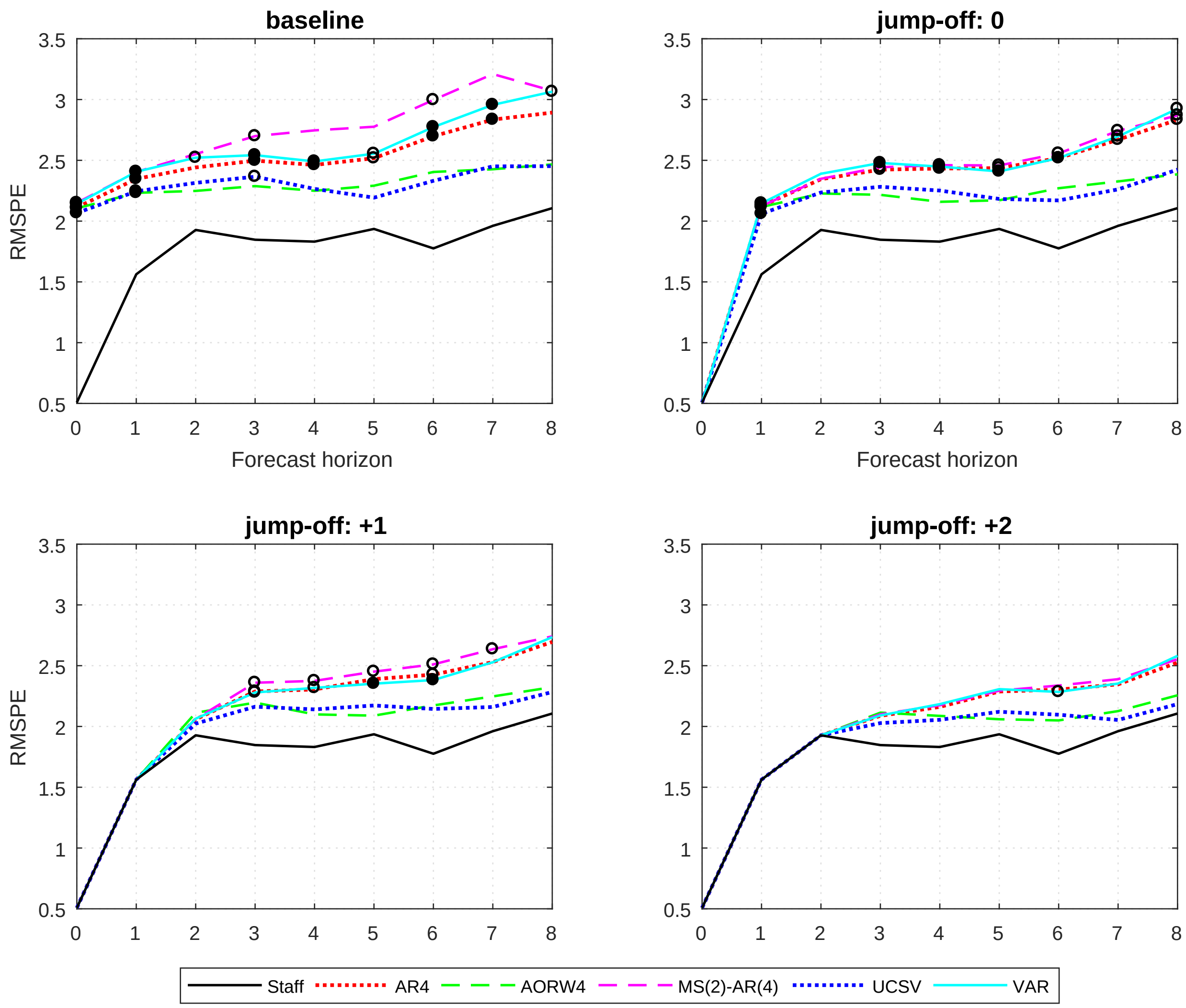
As discussed in Section 2, the staff forecasts are finalized towards the end of the quarter. Hence, the staff already has a number of monthly indicators available for the current quarter to inform its nowcasts. For example, when the staff finalizes its quarterly inflation forecasts, it already has knowledge of the inflation data for the first month of the quarter and thus has a significant informational advantage over the econometric models. Sims (2002) conjectured that the good performance of the FRB's Greenbook forecasts is mainly due to their better nowcasts. Using a large set of real-time data and forecasts from the Greenbooks, Faust and Wright (2009) test this conjecture by feeding small- and large-scale forecasting models with the Greenbook short-horizon forecasts. We conduct a similar exercise here. For a given forecast prepared by the staff for quarter t, quarterly data for quarter t-1 are available to estimate the benchmark models. We thus append the staff forecasts for quarter t and beyond to the model's information set. That is, we give the econometric models the staff view on the state of the economy, and study how the accuracy of their long-run forecasts changes in light of this information. Following Faust and Wright (2009)'s terminology, we call the quarter to which we update the data for the benchmark models the "jumping-off point." We consider jumping-off points from the current quarter t (i.e. the nowcast) to t+2.

Figure 2: Staff RMSPEs for GDP growth *vs.* Model Benchmarks at Different Jumping-off Points



Note: This figure shows the RMSPEs of GDP forecasts from the staff and econometric benchmarks estimated in real time up to horizon $t+8$. In each panel, benchmarks are conditioned on a different set of information based on the staff forecasts up to different jumping-off points. Circles (○ and ●) indicate that the difference in RMSPE between the benchmarks and the staff is statistically significant at 10 and 5 per cent, respectively. Forecasted dates range from 1982Q2 to 2015Q4.

Figure 3: Staff RMSPEs for Inflation *vs.* Model Benchmarks at Different Jumping-off Points



Note: This figure shows the RMSPEs of CPI inflation forecasts from the staff and econometric benchmarks estimated in real time up to horizon $t+8$. In each panel, benchmarks are conditioned on a different set of information based on the staff forecasts up to different jumping-off points. Circles (○ and ●) indicate that the difference in RMSPE between the benchmarks and the staff is statistically significant at 10 and 5 per cent, respectively. Forecasted dates range from 1982Q2 to 2015Q4.

The remaining panels ('jump-off': 0, 1, and 2) in Figures 2 and 3 present how the staff RMSPEs for GDP growth and CPI inflation compare with the benchmark models once they are given the staff nowcasts, the 1- and the 2-quarter-ahead forecasts, respectively. For both GDP growth and CPI inflation, the models' predictions are marginally more accurate as they are given more of the staff short-term forecasts; nonetheless, for most models the improvements are small, especially for their long-horizon forecasts. For GDP growth, when the models are given the $t+1$ and $t+2$ forecasts ('jump-off': 1 and 2), the staff RMSPEs are no better than the average model. For CPI inflation, the staff's relative forecast accuracy again differs markedly from GDP growth. As Faust and Wright (2009, 2013) find for the Greenbook forecasts, the Bank of Canada staff CPI inflation forecasts dominate the time series forecasts for all horizons and jumping-off points.

When focusing solely on the econometric models, it is interesting to note that for real GDP growth the AR model is clearly more accurate than the MS-AR and the VAR model for virtually every horizon and jumping-off point. The lower RMSPE of the AR model persists throughout the forecasting horizon. The AORW and the UCSV models are the models that benefit the most from the jumping-off data points. On the inflation front, the UCSV model and the AORW random walk models are more accurate than the remaining econometric models, a result similar to Faust and Wright (2009) for the U.S.

We provide a series of robustness tests and additional evidence in the online Appendix. First, we report an alternative evaluation metric, namely the percentage of times (quarters) the benchmark models have a more accurate forecast than the staff, and find that our results above are consistent with this alternative comparison method. Second, we restrict our sample to quarters classified as recessions by the C.D. Howe Institute; we find that the RMPSEs are markedly higher for both staff and econometric models than in our baseline results, but the qualitative conclusions remain unchanged. Third, we compare the staff forecasts for the GDP components (consumption, investment, government expenditures, exports and imports) with the simple AR(4) model. We show that the findings for GDP growth translate to all of its components, i.e. the staff has an advantage over the AR(4) model over short horizons, but this advantage dissipates after the 1-quarter-ahead forecast. Finally, we report RMSPEs for the core CPI inflation forecasts and find that they are smaller than the total CPI inflation RMPSEs reported above. Again, the qualitative results are similar to those reported for CPI inflation.

3.3 How important are real-time data when evaluating the staff forecasts?

Various papers have highlighted the importance of using real-time data when estimating and evaluating forecasting models (i.e. Croushore (2011)). To examine the importance of evaluating the Bank of Canada staff forecasts with real-time data, as well as estimating and forecasting with the benchmark time series models, we perform the following exercise. First, we calculate the forecast errors for the staff using the latest vintage of real-time historical data in our sample (i.e. 2016Q2). The first row of Table 1 shows the ratio of the RMSPEs of the staff forecasts computed using this final vintage over the ones computed with the real-time data vintages. Ratios over one imply that computing the staff forecast errors using the last vintage results in higher RMSPEs than calculating errors with real-time data. For all forecast horizons, the staff RMSPEs calculated with the last vintage are higher than the ones calculated with the real-time data. For some horizons, especially the nowcasts, the difference in RMSPEs can be quite sizable. The staff RMSPEs are 18 per cent higher for GDP growth, and 39 per cent for CPI inflation at the nowcast horizon when the staff errors are calculated with the last vintage instead of the real-time data.

The remaining row in Table 1 shows the ratio of the RSMPEs of the econometric models estimated with the last vintage to the models estimated with the real-time data. For the former, we also use the last available vintage to calculate the forecast errors, while for the latter we use the real-time data. The evidence for the econometric models is more mixed, with some presenting increases in RMSPE but others benefiting from marginal gains. Overall, the gains are larger for GDP growth than for CPI inflation, as one would expect since CPI is substantially less revised than GDP. This evidence is suggestive that forecast comparisons between benchmark models (estimated with revised data) and judgmental forecasts produced in real time that use revised data as actual values will generally tend to favour the time series models.

Table 1: RMSPEs with real-time *vs.* latest available data vintage as actuals

Model	T0	T+1	T+2	T+4	T+6	T+8
GDP growth						
Staff	1.26	1.09	1.08	1.11	1.13	1.12
AR(4)	1.12	1.10	1.13	1.15	1.14	1.13
AO RW	1.11	1.08	1.06	1.10	1.12	1.11
MSM(2)-AR(4)	0.98	0.98	1.07	0.99	1.03	1.02
UCSV	1.14	1.12	1.08	1.12	1.14	1.12
VAR	1.08	1.07	1.06	0.94	1.02	1.03
Inflation						
Staff	1.37	1.05	1.04	1.03	1.03	1.01
AR(4)	0.99	1.00	1.00	1.00	1.01	0.99
AO RW	1.02	1.02	1.02	1.01	1.00	1.00
MSM(2)-AR(4)	0.98	0.98	0.97	0.92	0.93	0.96
UCSV	0.99	1.01	1.02	1.00	1.01	0.98
VAR	0.95	0.98	0.96	0.96	0.97	0.95

Note: This table reports RMSPE ratios for forecast errors calculated using the last available vintage as the actual data relative to using the vintage two quarters after the forecasted period. Ratios over one thus imply that forecast errors computed with the last vintage available are larger. When using the last vintage as actual data, the benchmarks are provided with pseudo real-time data from the last vintage while the staff forecast is unchanged. Forecasted dates range from 1982Q2 to 2015Q4.

3.4 Are the staff forecasts biased?

Next, we determine if the Bank of Canada staff forecasts show any evidence of bias. A forecast is said to be biased if it overestimates (or underestimates) its realized values in a systematic way. A large number of papers examine the unbiasedness of Greenbook forecasts. For instance, Romer and Romer (2000) find that the Greenbook inflation forecasts are unbiased for the 1965 to 1991 period using traditional Mincer and Zarnowitz (1969) regressions. Sims (2002) extends their sample to 1995 and finds some evidence that Greenbook inflation forecasts are biased upwards; however, he corroborates Romer and Romer (2000) results that the Greenbook forecasts are of very high quality and could be useful to private forecasters if provided in a timely manner. Many papers use the Greenbook forecasts as their testing grounds for new approaches testing forecast unbiasedness. For example, Joutz and Stekler (2000) and Patton and Timmermann (2012) propose new bias tests that explore the multiple horizon nature of forecasts and apply them to the Greenbooks.

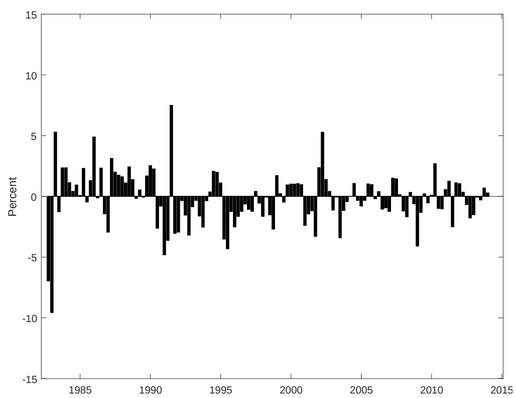
We begin by plotting the forecast errors, defined as in Equation (1), for 1- and 4-quarters-ahead GDP growth (Figure 4, panels (a) and (b)) and CPI inflation (Figure 4, panels (c) and (d)) for our full sample. There is no clear evidence of bias in the 1-quarter-ahead GDP growth forecasts, although there are localized periods of systematic underprediction (e.g., 1985 to 1990) and overprediction (e.g., 1990 to 2000). In contrast, there is clear evidence of a negative bias in the 4-quarters-ahead forecast errors, as the forecasts at this horizon are systematically above their actual values. Finally, panels (c) and (d) suggest no discernible bias in the CPI inflation forecasts for both horizons.

Next, we run Mincer and Zarnowitz (1969) regressions to examine more closely the rationality of the forecasts for both real GDP growth and CPI inflation. The test is based on the simple equation

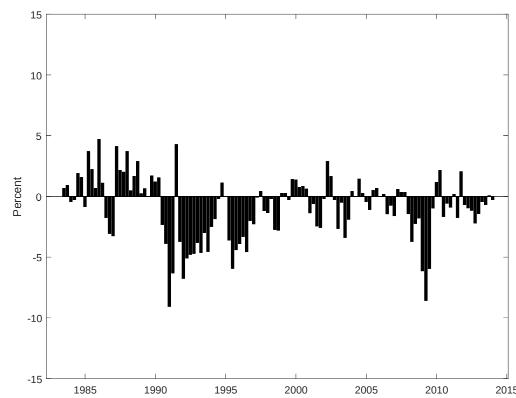
$$y_{t+h} = \alpha + \beta y^F_{t+h} + \varepsilon_{t+h}, \tag{2}$$

where y^F_{t+h} is the h-quarters-ahead forecast for y_t made in period t and y_{t+h} is the corresponding actual data. A forecast is said to be rational if $\alpha = 0$ and $\beta = 1$. As shown by

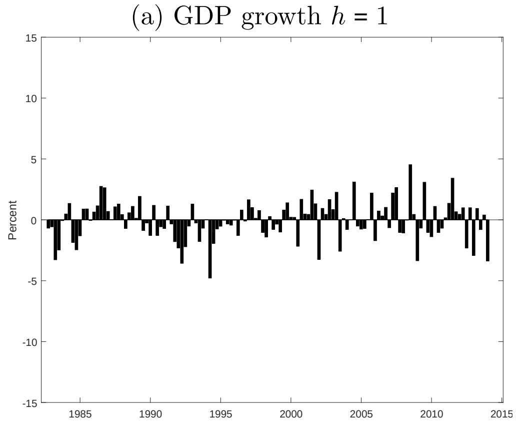
Figure 4: Staff forecast errors for GDP growth and CPI inflation



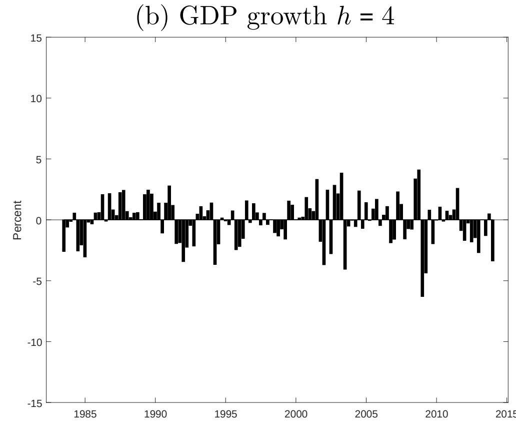
(a) GDP growth $h = 1$



(b) GDP growth $h = 4$



(c) CPI inflation $h = 1$



(d) CPI inflation $h = 4$

Note: This figure shows the forecast errors, defined as actual minus forecast values, for real GDP growth and CPI inflation, at the 1- and 4-quarters-ahead horizons. Forecasted dates range from 1982Q2 to 2015Q4.

Mankiw and Shapiro (1986), small sample biases tend to skew the test results to reject rationality too often. We thus estimate the simpler version

$$e_{t+h} = y_{t+h} - y^F_{t+h} = \alpha + \varepsilon_{t+h}. \tag{3}$$

A test of $\alpha = 0$ in Equation (3) is equivalent to jointly testing $\alpha = 0$ and $\beta = 0$ in Equation (2). Equation (3) is thus a simple test of the zero mean of the forecast error.

Table 2 reports the estimation results for our full sample. We estimate Equation (3) not only for the staff forecasts, but also for each of the forecasting models considered in Section 3.1. Numbers in bold denote statistical significance at the 5 per cent level. We report bias tests for the nowcasting horizon (h=0) as well as h=1,2,4,6 and 8. The tests largely confirm the evidence shown in Figure 4: on the one hand, we cannot reject that the staff GDP growth nowcasts, as well as the 1- and 2-quarter-ahead forecasts, have means equal to zero. On the other hand, there is statistical significance that the staff overpredicts GDP growth at longer horizons. For example, at the 4-quarter-ahead horizon, staff overpredicts real GDP growth by 0.90 percentage points. Linear models, such as the AR and VAR, also systematically overpredict GDP growth at all horizons.

The results for the CPI inflation forecasts are in stark contrast to those for GDP growth. We cannot reject the null hypothesis that the staff forecasts for CPI inflation have a zero mean regardless of the horizon. Just as for GDP growth, linear models such as the AR or VAR also have a tendency to overestimate inflation because they adjust only slowly their forecasts to the lower levels of inflation observed after the introduction of the IT regime. For models like the AORW or the UCSV, which adapt more easily to structural changes, their average forecast errors are not statistically different from zero.

Table 2: Bias Test Results

	Horizon	Staff	AR(4)	AO RW	MSM(2)-AR(4)	UCSV	VAR
GDP growth							
	0	0.12	**-0.56**	0.18	**-0.56**	0.17	**-1.30**
	1	-0.21	**-0.76**	0.22	-0.64	0.21	**-1.66**
	2	-0.42	**-0.73**	0.29	-0.46	0.28	**-1.59**
	4	**-0.90**	**-0.84**	0.28	-0.48	0.28	**-1.87**
	6	**-1.02**	**-0.95**	0.17	-0.57	0.16	**-1.95**
	8	**-0.92**	**-1.01**	0.12	-0.63	0.11	**-1.85**
CPI inflation							
	0	-0.07	**-0.46**	-0.21	**-0.42**	-0.16	-0.07
	1	-0.05	**-0.71**	-0.28	**-0.65**	-0.23	-0.33
	2	0.06	**-0.94**	-0.32	**-0.87**	-0.27	-0.45
	4	-0.05	**-1.36**	-0.39	**-1.28**	-0.34	**-0.89**
	6	0.12	**-1.71**	-0.45	**-1.62**	-0.40	**-1.41**
	8	0.10	**-2.00**	-0.51	**-1.91**	-0.46	**-1.91**

Note: This table reports bias coefficient from the Mincer-Zarnowitz test (Equation 3) on staff and model forecasts. **Bold entries** indicate statistical significance of the reported bias coefficients at the 5 per cent level. The Newey-West lag order corresponds to the forecast horizon + 1, where the nowcast is defined as forecast horizon 0. Forecasted dates range from 1982Q2 to 2015Q4.

In Figure 4, we showed that staff cannot predict recessions 4-quarters-ahead. Consequently, recessions can cause large and serially correlated negative forecast errors. In the Appendix, we perform the same bias test, controlling for recessions (as defined by the C.D. Howe Institute). Although the size of the bias is somewhat smaller, the qualitative results of Table 2 remain robust.

4 Subsample stability

It has been well documented that U.S. macroeconomic volatility has declined since the early 1980s (e.g., Kim and Nelson (1999), McConnell and Perez-Quiros (2000) and Blanchard and Simon (2001)). In Canada, this decline was experienced about a decade later, since the period between the late 1980s and early 1990s was particularly volatile (e.g., Longworth et al. (2002)).[11] Following the February 1991 announcement by the Bank of Canada and

[11]This period was marked by different developments that clouded the economy with uncertainty. For example, following 5 years of robust economic growth and relatively stable inflation at about 4 per cent, inflation took off in 1988; the Bank Rate was very high, coupled with the historical high of the Canadian-

the federal government of the inflation-targeting regime along with the implementation of the first inflation targets for 1992–95, the volatility of macroeconomic activity declined and, apart from the recession period (2008–09), has stayed relatively low since then.

An important and related question is whether this decline in economic volatility is due to less uncertainty or a less volatile predictable component of macroeconomic activity. For instance, Tulip (2006) and Campbell (2007), using forecast data from the FRB's Greenbooks and the Survey of Professional Forecasters (SPF), respectively, have documented a change not only in volatility but also in the predictability of the U.S. economy since the mid-1980s. They argue that the decline in the predictable component of both output and inflation is important in explaining the drop in volatility. As seen in Figure 1, while Canadian GDP growth and CPI inflation have been less volatile since the early 1990s, the Bank of Canada staff forecasts have become more accurate, especially for GDP growth. In this section, we use the staff forecasts to analyze more closely the change in economic volatility and predictability in Canada.

4.1 Forecast evaluation across subsamples

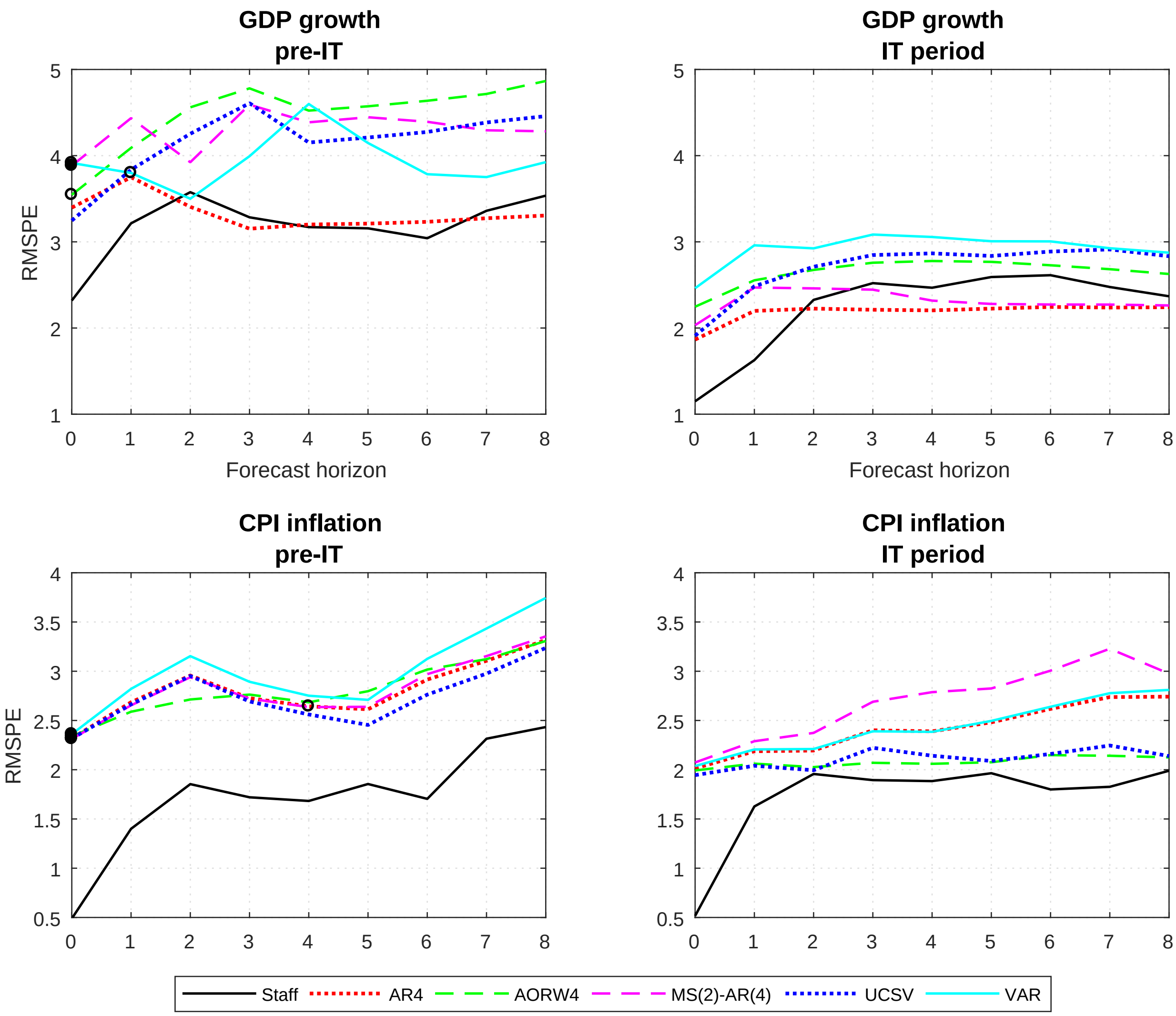
We start by computing the RMSPEs for the staff and models separately for the pre-IT (1982Q2–91Q4) and IT (1992Q1–2015) periods, analogous to Figures 2 and 3. The upper and lower panels of Figure 5 present the results.[12] For real GDP growth, the RMSPEs are substantially smaller during the IT period. For example, the RMSPE of the staff GDP growth nowcast fell from 2.3 to approximately 1.1 per cent, while the 8-quarter-ahead staff RMSPEs diminished from 3.5 to 2.4 per cent. The benchmark models' GDP growth forecasts, overall, also experienced similar declines in RMSPEs. For CPI inflation, the RMSPEs' declines following the introduction of IT are more subdued. For instance, the staff RMSPEs are relatively similar for the pre-IT and IT periods except over the longest horizons (h=7,8), where RMPSEs are smaller for the IT period. For the models, RMPSEs have generally declined following IT but the declines are less pronounced than for their

U.S. 90-day treasury bill spreads; and the USD/CAD exchange rate was very volatile. There was also political uncertainty around the unravelling of the Meech Lake Accord in 1990.

[12] We reproduce only the results of the upper left panels of Figures 2 and 3, i.e. the baseline results with no jump-off for the models.

GDP growth counterparts.

Figure 5: Staff RMSPEs for GDP *vs.* Model Benchmarks Prior to and After Inflation Targeting



Note: This figure shows the RMSPEs of real GDP growth (upper panel) and CPI inflation (lower panel) forecasts from the staff and the benchmark models up to horizon t+8. The left and right panels represent RMSPEs for the 1982Q2–1991Q4 and 1992Q1–2015Q4 forecasted periods, respectively. Circles (○ and ●) indicate that the RMSPE ratio comparing the performance between the benchmarks and the staff is statistically significant at 10 and 5 per cent, respectively.

Despite the decrease in RMSPEs, comparing the staff real GDP growth forecasts with the econometric models shows that the qualitative evidence presented earlier is essentially the same across both subsamples. For instance, the staff dominates the AR model for the nowcast and the 1-quarter-ahead forecast; after this horizon, the AR model performs as well as or better than the staff for the remaining horizons. For CPI inflation, results are markedly different since the announcement of the IT regime: staff RMSPEs have stayed relatively flat since the introduction of IT, while some models have experienced important declines in their RMSPEs (i.e. the UCSV and AORW). Unsurprisingly, the staff forecast is still more accurate than these models for the nowcast and 1-step-ahead forecasts, but essentially on par with them for longer horizons. These results signal a drop in predictability in CPI inflation after the introduction of the IT regime, but less so for GDP growth.[13] We examine this more closely next.

4.2 Predictability and uncertainty before and after IT

Building on the above findings, we follow Campbell (2007) and examine how predictability and uncertainty have changed before and after the introduction of IT. We use Equation (1) and decompose the volatility (standard deviation) of actual real GDP growth and CPI inflation, $Std(y_{t+h})$, into a predictable ($Std(y^F_{t+h})$) and an uncertainty component ($Std(e_{t+h})$) for different forecast horizons h.[14] As a result, a reduction in the volatility of GDP growth (or CPI inflation) that arises from a reduction in the average volatility of GDP growth (or CPI inflation) shocks, $Std(e_{t+h})$, is considered a reduction in uncertainty while a decline in volatility due to a fall in $Std(y^F_{t+h})$ is considered a reduction in predictability. Table 3 presents the results of the decomposition for three different forecast horizons: the current quarter (nowcast), 1- and 4-quarter-ahead forecasts. The third to fifth columns of Table 3 display $Std(y_{t+h})$, $Std(y^F_{t+h})$ and the root mean squared error (RMSE), i.e. $Std(e_{t+h})$, respectively. These measures are computed over the pre-IT (1982Q2–1991Q4) and IT (1992Q1–2015Q4) periods, along with the ratio of IT over pre-IT standard deviations.

[13] In the Appendix, we compare the staff CPI inflation forecasts for the IT period with a constant forecast of 2 per cent. We find that this constant forecast is less accurate than the staff nowcast, but slightly more accurate over all the other horizons.

[14] Of course, to get an exact decomposition of Equation (1), one should take the variance. We use standard deviations here because we are only interested in the change in the predictable and uncertainty components and because standard deviations are easier to interpret.

Table 3: Volatility Decomposition of Quarterly GDP growth and CPI Inflation

Horizon	Sample	$Std(y_{t+h})$	$Std(y^F_{t+h})$	RMSE
Real GDP growth				
Current quarter	Pre-IT	3.69	3.10	2.32
	IT	2.08	1.99	1.15
	Ratio	0.56	0.64	0.50
1-quarter-ahead	Pre-IT	3.52	2.05	3.21
	IT	2.07	1.50	1.63
	Ratio	0.59	0.73	0.51
4-quarters-ahead	Pre-IT	3.08	1.74	3.17
	IT	2.04	1.33	2.47
	Ratio	0.66	0.77	0.78
CPI Inflation				
Current quarter	Pre-IT	2.21	2.28	0.49
	IT	1.70	1.69	0.52
	Ratio	0.77	0.74	1.06
1-quarter-ahead	Pre-IT	1.97	2.00	1.40
	IT	1.69	1.38	1.63
	Ratio	0.86	0.69	1.16
4-quarters-ahead	Pre-IT	1.80	1.55	1.68
	IT	1.69	1.04	1.88
	Ratio	0.94	0.67	1.12

Note: This table reports the volatility decomposition of real GDP growth (upper panel) and CPI inflation (bottom panel) for three different horizons (current quarter, 1- and 4-quarters-ahead). The third column presents the standard deviation of the actual series while the fourth column shows the standard deviation of the forecasts. The RMSE is reported in the fifth column. For each horizon, the first row reports results for the pre-IT period (1982Q1–1991Q4), the second row reports results for the IT period (1992Q1–2015Q4) and the third reports the IT/pre-IT ratio of standard deviations.

Three observations stand out. First, the volatility of actual GDP growth and CPI inflation fell substantially after the introduction of IT. For instance, GDP growth declined by over 40 per cent while CPI inflation declined by 23 per cent. Second, we look at the next two columns to assess how this decline can be accounted for by changes in predictability and uncertainty. Over the short-horizon forecasts (h=0,1), while both the predictability and uncertainty of GDP growth have declined markedly, the decline in uncertainty is more important (36 vs. 50 per cent (for h=0), and 27 vs. 49 per cent (for h=1), respectively). For the 1-year-ahead forecasts (h=4), the decline in both predictability and uncertainty is roughly equal at about 22 per cent. This is consistent with Figure 5, where staff RMSPEs decline more than the models for horizons h=0,1 but less for the 4-step-ahead forecast (h=4). Third, the picture for CPI inflation is markedly different: while predictability of CPI inflation has fallen more than for GDP growth, uncertainty has slightly increased, especially for the 1- and 4-quarters-ahead horizons. We note that although the RMSPEs have gone up, they are still lower than those from the GDP growth forecasts.

These results differ somewhat from those found for the U.S. by Campbell (2007) and Tulip (2006). Both papers find that the decline in real GDP growth volatility is mostly due to a decline in its predictable component. The decline in uncertainty is smaller and less clear than that for volatility.[15] Using our staff forecast data, we find that both the predictable and uncertainty components of GDP growth have fallen since the introduction of IT, and that the fall in uncertainty is relatively more important than the fall in predictability. On the inflation front, we find that predictability of CPI inflation has almost disappeared: for example, the one-year-ahead forecast volatility has declined by 33 per cent since IT, hovering at about 1 per cent.[16] This echoes the findings in Stock and Watson (2007), Tulip (2006) and many others that the predictable component of U.S. inflation has also disappeared. However, while some studies find a decline in uncertainty for U.S. inflation (e.g. Tulip (2006)), we find no discernible fall in uncertainty in Canadian CPI inflation since the early 1980s.

4.3 Forecast bias across subsamples

Romer and Romer (2000) argue that the FRB's staff Greenbook inflation forecasts are rational and contain valuable information above those of the private sector. Sims (2002) extends their sample to 1995, and while he finds some upward bias in the forecasts, he corroborates Romer and Romer (2000)'s evidence that the Greenbook forecasts are of high

[15] We highlight that our sample is different from those used by Campbell (2007) and Tulip (2006), which roughly span from the late 1960s to early 2000s. Ours starts in the early 1980s until 2015Q4 and includes the Great Recession but excludes the 1970s volatile period.

[16] This is a consequence of IT: forecasts have averaged 2 per cent since 1992, hovering around the target and within the 1 and 3 per cent bands.

quality and could be valuable to the private sector. Examining a slightly longer period from 1968 to 1998, Capistrán (2008) shows that the evidence in both previously cited papers masks subsample instabilities. Until the Volcker disinflation, Federal Reserve staff forecasts tended to systematically underestimate inflation, a pattern that reversed itself in the second half of the sample. After the Volcker disinflation, Greenbook forecasts systematically overpredict inflation. Rossi and Sekhposyan (2016) propose forecast rationality tests that account for sample instabilities and confirm these time-varying biases in the Greenbook inflation forecasts.

Some evidence of time-varying biases in the staff forecasts for GDP growth are evident in Figure 4. On the one hand, before the introduction of IT (and the 1991–92 recession) staff mostly underpredicted GDP growth at the 4-quarter-ahead horizon, a pattern that reverses afterwards. Since 1990, staff has a clear tendency to overestimate GDP growth and does so more forcefully during recessions. On the other hand, it is not clear from Figure 4 that there is any time-varying bias in the staff CPI inflation forecasts.

Table 4 displays estimates of Equation 3 for GDP growth and CPI inflation for our two subperiods. Before the introduction of IT, the real GDP forecasts show no evidence of bias. During the IT period, there is clear evidence of bias in the 4-quarter-ahead forecasts. Finally, there is no evidence of biases in the CPI inflation forecasts for both subperiods.

Table 4: Bias Test Results for pre-IT and IT periods

	GDP growth		CPI inflation	
Horizon	Pre-IT	IT period	Pre-IT	IT period
0	0.41	-0.01	-0.08	-0.06
1	0.28	**-0.42**	-0.07	-0.04
2	0.83	**-0.94**	0.08	0.05
4	0.07	**-1.27**	0.23	-0.15
6	-0.12	**-1.33**	0.37	0.03
8	-0.23	**-1.14**	0.94	-0.17

Note: This table reports bias coefficients (α) for the Mincer-Zarnowitz test performed on staff forecasts for the pre-IT and IT periods, respectively. Bold entries indicate statistical significance of the reported bias coefficients at 5 per cent. The Newey-West lag order corresponds to the forecast horizon + 1, where the nowcast is defined as forecast horizon 0.

5 Comparing the staff forecasts with those of other professional forecasters

A natural question that arises following the above analysis is how the staff forecasts compare with those of other forecasters. For instance, Consensus Economics Inc. (CE) gathers and records forecasts for the Canadian economy from many professional forecasters on a timely basis.[17] Moreover, the Bank of Canada's MPR, published quarterly, contains the official forecasts from the governing council and can be different from the staff's. In this section, we compare the staff forecasts with (1) forecasts from CE and (2) the MPR forecasts. We then assess whether (1) the staff forecasts contain additional information over the professional forecasters' and (2) the MPR forecasts add value over the staff's.

5.1 Consensus Economics forecasts

We start by comparing the staff's forecast accuracy with average forecasts from CE. Because CE forecasts are year-over-year quarterly growth rates, we adjust the staff data to get analogous forecasts for comparison. CE records forecasts generally around the middle of the months of March, June, September and December, which corresponds to the timing of the staff projections. The CE sample of quarterly forecasts for Canada is shorter than our staff forecasts sample, starting in 1994Q1, and includes forecasts up to 6 quarters ahead. We thus compare the staff nowcast and 6-quarter-ahead forecasts with CE average forecasts prepared during the 1994Q1–2015Q2 period.[18] The left panel of Table 5 presents the results, shown as RMSPE ratios of CE over staff.

[17] See http://www.consensuseconomics.com for details.

[18] The sample ends in 2015Q2 because the 6-quarter-ahead forecast prepared in 2013Q4 is for 2015Q2.

Table 5: Bank of Canada Staff *vs.* Consensus Economics (mean) forecasts

Horizon (quarters)	1994Q1–2015Q2		2000Q1–2015Q2	
	GDP growth	CPI inflation	GDP growth	CPI inflation
0	1.16	1.53	1.14	1.49
1	1.13	1.12	1.10	1.08
2	1.09	1.04	1.10	1.02
3	1.04	0.98	1.08	0.99
4	1.00	0.98	1.10	1.02
5	0.95	0.92	1.06	0.93
6	0.92	0.96	1.02	1.00

Note: This table reports RMSPE ratios computed as CE RMSPE divided by Bank of Canada staff RMSPE. A ratio above 1.0 implies better staff forecast accuracy relative to CE.

Overall, staff forecasts fare relatively well compared with those of professional forecasters, especially over the short horizons. For the nowcast, and the 1-, 2-, 3-step-ahead forecasts, staff is either better than or as good as CE in predicting both GDP and CPI growth. For the 5- and 6-quarter-ahead forecasts, staff do somewhat worse. If we focus on a more recent sample, for example, 2000–15, staff forecasts become even better relative to those of CE. In the right panel of Table 5, we can see that RMSPE ratios are greater than one for all horizons for GDP growth and in 5 out of 7 horizons for CPI inflation.

We now turn to assess whether the staff forecasts contain valuable information over and above the average forecasts from CE. Imagine that market participants use the average forecast from CE to form their expectations. If the staff forecasts were available, they could base their expectations on a linear projection of both CE and the staff forecasts, i.e.,

$$y_{h,t} = c + \beta_{CE}\hat{y}_{h,t}^{CE} + \beta_{Staff}\hat{y}_{h,t}^{Staff} + \epsilon_t, \tag{4}$$

where $y_{h,t}$ is the h-quarter ahead growth rate of the variable of interest and $\hat{y}_{h,t}^{CE}$ and $\hat{y}_{h,t}^{Staff}$ are the h-quarters-ahead CE and staff forecasts prepared in quarter t, respectively. Testing whether the regression coefficients β are significantly different from zero will then tell us if each set of forecasts contains relevant information to help forecast the variable of interest. Table 6 presents the results.

For GDP growth, the results indicate overwhelmingly that staff forecasts possess rele-

Table 6: Role of Staff and CE forecasts in Predicting Actual values

Horizon (quarters)	Real GDP growth		
	c	β_{Staff}	β_{CE}
0	-0.05(0.11)	0.79(0.24)	0.23(0.27)
1	0.02(0.24)	1.04(0.26)	-0.08(0.32)
2	-0.19(0.37)	0.94(0.29)	0.03(0.30)
3	-0.59(0.73)	0.82(0.37)	0.23(0.42)
4	-0.41(1.48)	0.78(0.43)	0.11(0.48)
5	1.01(2.08)	0.56(0.37)	-0.19(0.54)
6	2.05(2.18)	0.47(0.29)	-0.46(0.60)
$\mu(0-4)$	-0.15(0.39)	0.89(0.35)	0.05(0.34)

Horizon (quarters)	CPI inflation		
	c	β_{Staff}	β_{CE}
0	-0.03(0.03)	0.91(0.13)	0.11(0.14)
1	0.06(0.16)	0.78(0.21)	0.21(0.25)
2	0.29(0.30)	0.68(0.26)	0.21(0.29)
3	0.11(0.47)	0.47(0.18)	0.51(0.30)
4	0.59(0.78)	0.43(0.21)	0.29(0.43)
5	1.85(0.97)	0.16(0.21)	-0.12(0.49)
6	2.23(1.06)	0.12(0.25)	-0.29(0.50)
$\mu(0-4)$	0.18(0.25)	0.60(0.20)	0.33(0.24)

Note: This table reports regression coefficients from Equation 4. c is the constant coefficient, while β_{Staff} and β_{CE} are the coefficients on the staff and CE forecasts, respectively. The sample period includes forecasts prepared from 1994Q1 to 2015Q2. HAC robust standard errors are shown in parentheses; the Newey-West lag order corresponds to the forecast horizon $h + 1$. The forecast horizon $\mu(0-4)$ refers to the average forecast 0 to 4 quarters ahead.

vant information not contained in the CE forecasts. The point estimates for all horizons are large, positive and statistically significant while those of CE are small, sometimes negative and in general not significant. Table 6 also reports the results using the average forecast from 0 up to 4-quarters ahead ($\mu(0-4)$); again, one would place a weight close to one (0.9) to the staff forecast and close to zero (0.05) to the CE forecast.

The results for inflation are qualitatively similar. For the short horizons (0 to 2 quarters ahead), the weight one would put on the staff forecasts is very high (between 0.7 and 0.9). Over the longer horizons, the staff's advantage becomes smaller, and for the 5- and 6-quarter-ahead forecasts, one should simply predict 2 per cent without using either the staff or CE forecasts. For instance, someone forecasting inflation 5 or 6 quarters ahead would move away from CE forecasts, as the point estimates are negative. For the average forecast from 0 to 4 quarters ahead ($\mu(0-4)$), one would put a weight of about two thirds on the staff forecast and one third on CE forecast. Our results for inflation are qualitatively similar to those of Romer and Romer (2000), who compare the Greenbooks with private forecasters; they find that the FRB's staff has considerable advantage over private forecasters when predicting future U.S. inflation and suggest that the optimal forecasting strategy of someone with access to both forecasts would be to put essentially no weight on the commercial forecast.

5.2 *Monetary Policy Report* forecasts

We then compare the staff forecasts with those from the MPR and assess whether the MPR improves upon the staff in forecasting GDP growth and CPI inflation. Three things are worth pointing out: first, MPR forecasts for GDP growth are only available since 1997 and only since 2003 for CPI inflation, on a biannual basis (in April and October). Consequently, we start our sample in 1997 for GDP growth and 2003 for CPI inflation.[19] Second, MPR forecasts are average annual growth rates for the current and following year (and not quarterly year-over-year growth as in the CE comparison). Therefore, we compute staff average annual forecasts for the current and following year from our data-set. Third, MPR forecasts are produced one month after those of the staff, a substantial advantage given that new monthly GDP and inflation data often become available during this time span.[20] Consequently, we should expect MPR forecasts to outperform the staff's. Table 7 presents the baseline results of this exercise.

[19]Note that we use the MPR forecast data gathered by Binette and Tchebotarev (2017).

[20]The staff forecasts prepared for the April and October MPRs are made in March and September, respectively.

Table 7: Bank of Canada Staff *vs.* MPR forecasts

	GDP growth	CPI inflation
Current year		
March vs. April MPR	0.98	0.77
September vs. October MPR	1.15	0.82
Next year		
March vs. April MPR	1.06	1.01
September vs. October MPR	1.03	0.90

Note: This table reports RMSPE ratios computed as MPR divided by Bank of Canada staff RMSPEs. A ratio above 1.0 implies better staff forecast accuracy relative to the MPR. Sample is 1997–2014 for GDP growth and 2003–14 for CPI inflation.

Apart from the current-year inflation forecasts, it seems that overall, the MPR forecasts are not more accurate than the staff forecasts, even with one more month of data. Although it is difficult to get any statistical significance between the forecast accuracy of the staff and these alternative forecasts due to the short samples, we note that these results are in line with those found in the U.S. for the Federal Reserve staff forecasts (Greenbook) relative to the FOMC (Romer and Romer (2008)).

To assess further whether MPR forecasts provide relevant information over the staff's, we estimate a similar regression as in Equation 4, where we substitute the CE forecasts for those from the MPR. As we just mentioned, we only have two horizons here: the current and the next year. Table 8 presents the results.

The results for GDP growth are quite striking: although the MPR forecast possesses one more month of information, it does not provide more relevant information than the staff forecast for the current-year forecast (point estimates of 0.51 vs. 0.50), and the estimate is in the wrong sign (although not statistically significant) for the next-year forecast (1.61 for the staff vs. -0.18 for the MPR). For inflation, the MPR does seem to bring relevant information: for the current year, the estimate in the MPR forecast is large (1.21) and statistically significant while the staff's is not. For next's year forecast, the MPR's estimate is still large but less significant. Overall, and keeping in mind the small sample sizes, the exercise shows that the MPR does provide relevant information over the staff to forecast CPI inflation, but does not for GDP growth.

Table 8: Role of Staff and MPR forecasts in predicting actual values

Horizon (years)	GDP growth c	β_{Stf}	β_{MPR}
0	-0.05(0.08)	0.51(0.67)	0.50(0.64)
1	-1.85(1.96)	1.61(0.49)	-0.18(0.68)

Horizon (years)	CPI inflation c	β_{Stf}	β_{MPR}
0	0.06(0.16)	-0.31(0.28)	1.21(0.30)
1	0.56(1.08)	-0.43(1.19)	1.11(0.72)

Note: This table reports regression coefficients from the regression of actual GDP growth values (CPI inflation) on staff and MPR forecasts. c is the constant coefficient while β_{Stff} and β_{MPR} are the coefficients on the staff and MPR forecasts, respectively. The forecasts are annual averages for the current and next year (horizons 0 and 1, respectively). The sample period includes forecasts prepared from 1997 to 2014 for GDP growth and 2003 to 2014 for CPI inflation. HAC robust standard errors are shown in parentheses.

6 Conclusion

Central bankers face several types of uncertainty when implementing monetary policy: there are well-known lags between changes in monetary policy instruments and their ultimate effects, as well as uncertainty about future economic conditions, to name a few. Consequently, the accuracy of central banks' staff economic forecasts plays a crucial role in the effective implementation of monetary policy.

In this paper, we present a new, rich database containing both real-time historical data and forecasts collected from archived Bank of Canada staff economic projections. Our database is comprised of quarterly vintages of many key Canadian macroeconomic aggregates, some of which go back to the early 1970s. Notably, we have a consistent set of real-time data and forecasts for GDP growth and CPI inflation since 1982. To our knowledge, no other such extensive data-set of central bank staff forecasts is available for such a long time period apart from the FRB's Greenbooks. During the sample covered by our database, Canada also experienced two severe recessions (1991–92 and 2008–09) and large swings in commodity prices and exchange rates.

We use this novel data-set to provide a thorough evaluation of the Bank of Canada staff forecasts for GDP growth and CPI inflation. Several observations stand out. First, the staff nowcasts for both real GDP growth and CPI inflation are significantly more accurate than those from standard econometric models, reflecting the fact that some indicators for the current quarter are available to the staff when the forecasts are prepared. Second, this nowcasting advantage is much less persistent for real GDP growth than for CPI inflation, as it essentially disappears after one quarter for the former but persists all through the forecasting horizons for the latter. Third, we show that when we provide the benchmark models with the staff's short-term forecasts, their longer-term forecasts for GDP growth improve relative to the staff while for CPI inflation, the staff's advantage remains. Fourth, when we consider the pre-IT and IT periods separately, we find smaller RMSPEs for both real GDP growth and CPI inflation in the latter sample. For GDP growth, the results for the full sample qualitatively remain robust for both subsamples, while for CPI inflation, the decline in RMSPEs has been much more pronounced for some econometric models. This last result suggests a substantial drop in the predictability of inflation, which we confirm in a standard variance decomposition. Finally, we perform standard statistical bias tests and find that staff GDP growth forecasts are biased at long horizons, whereas they are not for CPI inflation.

We complete our forecast evaluation by assessing how the staff compares with other professional forecasters. We show that the Bank of Canada staff forecasts are more accurate and informative than the average forecasts from Consensus Economics. Furthermore, we find that the forecasts prepared by the Bank of Canada's Governing Council and published in the MPR are usually more accurate than the staff's for the current year, but less so for the next year, although the MPR forecasts are based on one more month of data.

This paper provides new analysis of central bank forecasts and sheds light on the implications of inflation targeting for macroeconomic forecasting. Notwithstanding the difficulty of forecasting GDP, inflation and other macroeconomic aggregates in real time, the Bank of Canada staff forecasts, similar to the FBR's Greenbooks, appear to be near the frontier of predictability. We hope that the new database we constructed, which will be made publicly available and updated annually, will be useful for future research.

References

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

Binette, A. and D. Tchebotarev (2017): "Evaluating real GDP growth forecasts in the Bank of Canada Monetary Policy Report," *Bank of Canada Staff Analytical Note*, 2017-21.

Blanchard, O. and J. Simon (2001): "The long and large decline in US output volatility," *Brookings Papers on Economic Activity*, 1, 135–164.

Brayton, F., T. Laubach, and D. L. Reifschneider (2014): "The FRB/US Model: A tool for macroeconomic policy analysis," Tech. rep., Board of Governors of the Federal Reserve System (US).

Campbell, S. D. (2007): "Macroeconomic volatility, predictability, and uncertainty in the great moderation: evidence from the Survey of Professional Forecasters," *Journal of Business & Economic Statistics*, 25(2), 191–200.

Capistrán, C. (2008): "Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious?" *Journal of Monetary Economics*, 55(8), 1415–1427.

Champagne, J., G. Poulin-Bellisle, and R. Sekkel (2018): "The real-time properties of the Bank of Canada's staff output gap estimates," *Journal of Money, Credit and Banking*, 50(6), 1167–1188.

Champagne, J. and R. Sekkel (2018): "Changes in monetary regimes and the identification of monetary policy shocks: narrative evidence from Canada," *Journal of Monetary Economics*, Forthcoming.

Chauvet, M., S. Potter, et al. (2013): "Forecasting output," *Handbook of Economic Forecasting*, 2 (Part A), 141–194.

Cloyne, J., C. Ferreira, and P. Surico (2018): "Monetary policy when households have debt: new evidence on the transmission mechanism," *Review of Economic Studies*, Forthcoming.

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017): "Innocent bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 88, 70–89.

Croushore, D. (2011): "Frontiers of real-time data analysis," *Journal of Economic Literature*, 49(1), 72–100.

Croushore, D. and T. Stark (2001): "A real-time data set for macroeconomists," *Journal of Econometrics*, 105(1), 111–130.

——— (2002): "Forecasting with a real-time data set for macroeconomists," *Journal of Macroeconomics*, 24(4), 507–531.

——— (2003): "A real-time data set for macroeconomists: Does the data vintage matter?" *Review of Economics and Statistics*, 85(3), 605–617.

Croushore, D. and S. van Norden (2017): "Fiscal forecasts at the FOMC: evidence from the Greenbooks," *Review of Economics and Statistics*.

Dorich, J., M. Johnston, R. Mendes, S. Murchison, and Y. Zhang (2013): "Totem II: An updated version of the Bank of Canada's Quarterly Projection Model," *Bank of Canada Technical Report*, 100.

Faust, J. and J. H. Wright (2009): "Comparing Greenbook and reduced form forecasts using a large real-time dataset," *Journal of Business & Economic Statistics*, 27(4), 468–479.

——— (2013): "Forecasting inflation," *Handbook of Economic Forecasting*, 2 (Part A), 3–56.

Fernandez, A. Z., H. Branch, E. F. Koenig, and A. Nikolsko-Rzhevskyy (2011): "A real-time historical database for the OECD," *Globalization and Monetary Policy Institute Working Paper*, 96.

Fortin-Gagnon, O., M. Leroux, D. Stevanovic, and S. Surprenant (2018): "A large Canadian database for macroeconomic analysis," *CIRANO Scientific Series 2018S-25.*

Garratt, A. and S. P. Vahey (2006): "UK real-time macro data characteristics," *The Economic Journal*, 116 (509), F119–F135.

Gervais, O. and M.-A. Gosselin (2014): "Analyzing and forecasting the Canadian economy through the LENS model," *Bank of Canada Technical Report*, 102.

Giannone, D., J. Henry, M. Lalik, and M. Modugno (2012): "An area-wide real-time database for the euro area," *Review of Economics and Statistics*, 94(4), 1000–1013.

Groen, J. J., G. Kapetanios, and S. Price (2009): "A real time evaluation of Bank of England forecasts of inflation and growth," *International Journal of Forecasting*, 25(1), 74–80.

Helliwell, J. F. (2005): "From flapper to bluestocking: what happened to the young woman of Wellington Street?" *Bank of Canada Review*, Winter, 31–39.

Joutz, F. and H. O. Stekler (2000): "An evaluation of the predictions of the Federal Reserve," *International Journal of Forecasting*, 16(1), 17–38.

Kim, C.-J. and C. R. Nelson (1999): "Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle," *Review of Economics and Statistics*, 81(4), 608–616.

Longworth, D. et al. (2002): "Inflation and the Macroeconomy: Changes from the 1980s to the 1990s," *Bank of Canada Review*, 2002 (Spring), 3–18.

Mankiw, N. G. and M. D. Shapiro (1986): "Do we reject too often?: Small sample properties of tests of rational expectations models," *Economics Letters*, 20(2), 139–145.

McConnell, M. M. and G. Perez-Quiros (2000): "Output fluctuations in the United States: What has changed since the early 1980's?" *American Economic Review*, 90(5), 1464–1476.

McCracken, M. W. and S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34(4), 574–589.

Mincer, J. A. and V. Zarnowitz (1969): "The evaluation of economic forecasts," in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER, 3–46.

Murray, J. (2013): "Monetary policy decision making at the Bank of Canada," *Bank of Canada Review*, Autumn, 1–9.

Orphanides, A. (2003): "The quest for prosperity without inflation," *Journal of Monetary Economics*, 50(3), 633–663.

Patton, A. J. and A. Timmermann (2012): "Forecast rationality tests based on multi-horizon bounds," *Journal of Business & Economic Statistics*, 30(1), 1–17.

Poloz, S., D. Rose, and R. Tetlow (1994): "The Bank of Canada's new quarterly projection model (QPM): An introduction," *Bank of Canada Review*, 1994 (Autumn), 23–38.

Romer, C. D. and D. H. Romer (2000): "Federal Reserve information and the behavior of interest rates," *American Economic Review*, 90(3), 429–457.

——— (2004): "A new measure of monetary shocks: Derivation and implications," *American Economic Review*, 94(4), 1055–1084.

——— (2008): "The FOMC versus the staff: where can monetary policymakers add value?" *American Economic Review*, 98(2), 230–235.

Rossi, B. and T. Sekhposyan (2016): "Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts," *Journal of Applied Econometrics*, 31(3), 507–532.

Sims, C. A. (2002): "The role of models and probabilities in the monetary policy process," *Brookings Papers on Economic Activity*, 2(2), 1–40.

Stock, J. H. and M. W. Watson (1996): "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics*, 14(1), 11–30.

——— (2007): "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39(s1), 3–33.

Timmermann, A. (2007): "An evaluation of the World Economic Outlook forecasts," *IMF Staff Papers*, 54(1), 1–33.

Tulip, P. (2006): "Has the economy become more predictable? Changes in Greenbook forecast accuracy," *Journal of Money, Credit and Banking*, 41(6), 1217–1231.

Appendix A The Bank of Canada's database of historical real-time data and staff forecasts

Bank of Canada staff have long prepared an in-depth analysis of the Canadian and international economy following the quarterly release of the National Accounts at the end of March, June, September and December. In this analysis, staff use (real-time) historical data and various economic models to produce forecasts for many macroeconomic variables and provide a report to Governing Council. This report was entitled "Staff Economic Projection" between 1972 and 2002, "Canadian Economic Projection" between 2003 and 2008, and has been called "Canadian Economic Outlook" since. These reports (which are not publicly available) represent a material part of the analysis presented to the Governing Council of the Bank of Canada in the weeks leading up to monetary policy decisions (Murray (2013)).

Staff have kept electronic copies of the Staff Economic Projection reports since 1991. However, the staff reports produced between 1972 and 1991 are stored in the Bank's archives and have not been digitized. Similarly, while staff has saved electronic versions of projection data produced over the recent past, the output from older projections was also stored in hard-copy format in the Bank's archives. We imported into spreadsheets the real-time historical data and staff forecasts for some key Canadian macroeconomic variables that were produced as far back as the early 1970s. We corroborated as much as we could the older data with archived hard copies of the Staff Economic reports.

The database is available online at https://www.bankofcanada.ca?p=201766. Each variable in the database can be downloaded in a separate csv file; within each file, the data are organized with observation dates in the rows and vintage dates in the columns. The vintage date corresponds to the date the forecasts were produced by the staff. As a result, each column gives the entire time-series history available to staff at the vintage date (shown in the column header),[21] as well as the forecasts prepared by staff at this date. For example, at the 1987Q1 vintage, the observations before 1987Q1 are historical data while those from 1987Q1 onward are forecasts. At the 1987Q2 vintage, the observations before

[21] It is important to note that the historical real-time data might have been adjusted by staff at different points in time and thus might not correspond exactly to earlier data vintages from Statistics Canada.

1987Q2 are historical data while those from 1987Q2 onward are forecasts, and so forth. By going horizontally from left to right in a spreadsheet, one can track the revisions to an observation (historical data) or to a forecast. Note that because there is a five-year ban on access to staff projections data, as of this writing vintages of real-time data and forecasts are available up to 2013Q4.[22] Going forward, once per year four quarterly projections will be added to the database.

Below, we present all the macroeconomic variables included in the database. For each variable, we detail the vintages available, the start date of the historical data and other relevant information.

- **Gross Domestic Product (real)**: Vintages available from 1986Q4 to 2013Q4. Historical data start in 1967Q1 or 1973Q1, depending on vintage.

  Statistics Canada adopted gross domestic product (GDP) as its headline national accounts measure in 1986, in place of gross national product (GNP), following a major revision of its national accounts. Consequently, staff started to forecast GDP in the fourth quarter of 1986. Many historical revisions to GDP (and its components listed below) in the database correspond to those documented by Statistics Canada.[23]

- **Gross Domestic Product (nominal)**: Vintages available from 1982Q2 to 2013Q4. Historical data start in 1967Q1 or 1973Q1, depending on vintage.

  Staff produced forecasts for nominal GDP before the official adoption of GDP in the national accounts by Statistics Canada in 1986.

- **Implicit Price Deflator for GDP**: Vintages available from 1986Q4 to 2013Q4. Historical data start in 1967Q1 or 1973Q1, depending on vintage (with the exception of the 2001Q2 to 2002Q2 vintages, which have historical data starting in 1981Q1).

- **Gross National Expenditure (nominal)**: Vintages available from 1970Q4 to 1986Q4. Missing vintages in 1971Q2 and 1973Q4. Prior to 1982Q2, very few historical data available. From 1982Q2 to 1986Q4, historical data available from 1967Q1 onward. From 1986Q4 onward, staff shifted focus to gross domestic product (GDP).

[22] Since the last date to which a forecast refers to is 2015Q4 (i.e. the 8-quarters-ahead forecast prepared in 2013Q4), the last vintage we will use to get the actual data is 2016Q2.

[23] See https://www150.statcan.gc.ca/n1/pub/13-606-g/2016001/article/14617-eng.htm for a history of Canada's macroeconomic accounts.

- **Implicit Price Deflator for GNE**: Vintages available from 1970Q4 to 1986Q3. Missing vintages in 1971Q2 and 1973Q4. Prior to 1982Q2, very few historical data available. From 1982Q2 to 1986Q4, historical data available from 1967Q1 onward.

- **Personal Consumption Expenditure (PCE)**: Vintages available: 1982Q2 to 1999Q4 and 2011Q2 to 2013Q4 (between 2001Q1 and 2011Q1, PCE was bundled into another variable labelled CIRINV, detailed below). Historical data start in 1967Q1 or 1973Q1, depending on vintage.

  As for GDP and the other components, PCE was revised throughout our sample; e.g., in 1986 and 1997.

- **Deflator for Personal Consumption Expenditures**: Vintages available: 1982Q2 to 1999Q4 and 2011Q2 to 2013Q4 (between 2001Q1 and 2011Q1, a deflator for CIRINV, detailed below, was created). Historical data start in 1967Q1, 1968Q4 or 1972Q4, depending on vintage.

- **Investment (Gross fixed capital formation)**: Vintages available from 1982Q2 to 2013Q4. Historical data start in 1967Q1, 1973Q1 or 1981Q1, depending on vintage. As for GDP and the other components, investment was revised throughout our sample: e.g., in 1986 and 1997.

- **Inventory Investment**: Vintages available: 1982Q2 to 1999Q4 and 2011Q1 to 2013Q4 (between 2001Q1 and 2011Q1, personal consumption expenditure was bundled into another variable labelled CIRINV, detailed below). Historical data start in 1967Q1 or 1973Q1, depending on vintage. Note shorter forecast horizons in some vintages (nowcast and 1-quarter-ahead forecast for 1993Q3 to 1999Q4).

- **Residential Investment**: Vintages available from 1982Q2 to 2001Q1, 2001Q3 to 2004Q4 and 2011Q2 to 2013Q4 (between 2000Q1 to 2011Q1, it was also bundled in CIRINV, detailed below). Historical data start generally in 1967Q1 and 1973Q1, but are missing in some vintages, notably between 2001Q2 to 2003Q2. Note that forecast horizons are shorter in some years (e.g., 1984Q1 and 1986Q3).

- **Government Expenditures**: Vintages available from 1982Q2 to 2013Q4. Historical data start in 1967Q1 (vintages up to 2005Q3, except 2001Q2–02Q3 vintages which start in 1981Q1) or 1973Q1 (for 2005Q4–2013Q4 vintages). Note some missing historical data in the 1986Q3–86Q4 vintages.

- **Exports of Goods and Services**: Vintages available from 1982Q2 to 2013Q4. Historical data start in 1967Q1 (vintages up to 2005Q3, except 2001Q2–02Q3 vintages which start in 1981Q1) or 1973Q1 (for 2005Q4–2013Q4 vintages).

- **Imports of Goods and Services**: Vintages available from 1982Q2 to 2013Q4. Historical data start in 1967Q1 (vintages up to 2005Q3, except 2001Q2–02Q3 vintages which start in 1981Q1) or 1973Q1 (for 2005Q4–2013Q4 vintages).

- **CIRINV**: Variable created by staff as the sum of personal consumption expenditures, residential investment and investment in inventories. Vintages available from 1993Q3 to 2011Q1. Historical data start in 1967Q1 or 1973Q1, depending on vintage.

- **Deflator for CIRINV**: Variable created by staff. Deflator for the sum of personal consumption expenditures, residential investment and investment in inventories. Vintages available from 1993Q3 to 2011Q1. Historical data start in 1967Q1 or 1973Q1, depending on vintage.

- **Output gap**: The output gap is defined as the difference between the actual output of the economy and its potential output (or trend). Numbers are in log differences (actual GDP minus potential GDP). Vintages available from 1987Q1 to 2013Q4. Historical data start dates vary between 1967Q1 and 1973Q1, depending on vintage.

- **Consumer Price Index (CPI)**: Vintages available from 1973Q3 to 2013Q4, with a missing vintage in 1973Q4. Few historical data in the 1973Q3 to 1982Q1 vintages. For most of the remaining vintages, historical data begin in 1967Q1. Note some missing historical data in the 1993Q3–93Q4 vintages and few forecasts in the 1986Q3 and 1995Q3 vintages. For more information regarding revisions to CPI, see Appendix C in Statistics Canada's Consumer Price Index reference paper.[24]

[24] See link: https://www150.statcan.gc.ca/n1/pub/62-553-x/62-553-x2015001-eng.pdf.

- **Core Consumer Price Index (CPIX)**: Defined as the CPI excluding Food and Energy (CPIXFET) from 1980Q1 to 2001Q1, and as the CPI excluding the 8 most volatile components (CPIX) from 2001Q2 to 2013Q4. Both exclude effects of changes in indirect taxes. Vintages available from 1980Q1 to 2013Q4. Historical data start in 1967Q1 for most vintages, although they are missing in the 1980Q1 to 1982Q1 vintages (only forecasts are available). Note that few forecasts are available in the 1986Q3 vintage.

- **Unemployment Rate**: Vintages available from 1982Q2 to 2013Q4. Historical data start in 1967Q1 for most vintages (up to the 2009Q1 vintage), and start either in 1973Q1 or 1976Q1 in the subsequent years. Note shorter forecast horizons for the 1986Q3 and 2005Q4 to 2009Q1 vintages.

- **U.S. / CAD Nominal Exchange Rate**: Quarterly average of the U.S./Canada nominal exchange rate. Vintages available from 1974Q2 to 2013Q4. From the 1982Q2 vintage onward, historical data start in 1967Q1. Between 1974Q2 and 1982Q2, historical data are generally missing and forecasts have shorter and varying horizons.

- **Staff Projection Policy Rate**: Staff used two measures for the policy rate over the sample of vintages available (1982Q2 to 2013Q4):

  - the **Bank Rate** from 1982Q2 to 1993Q2 and from 2007Q3 to 2013Q4;
  - the **3-month Commercial Paper Rate** from 1993Q3 to 2007Q2.

  Numbers represent quarterly averages. Historical data start in 1967Q1. Note the shorter forecast horizon for the 1986Q3 vintage.

- **Five-year Federal Government Bonds Rate**: Vintages available from 2011Q1 to 2013Q4. Numbers represent quarterly averages. Historical data start in 1973Q1.

More recent changes are documented at
http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getMainChange&Id=498102.

Appendix B Data adjustments for forecast evaluation

We highlight that the staff database is published in its raw form, i.e. we did not apply any adjustments on top of the staff data. We use this database in the paper to provide a thorough evaluation of the Bank of Canada staff forecasts for real GDP, CPI, core CPI and the components of GDP from 1982Q2 onward. Our sample starts in 1982Q2 as it is the first vintage where the historical data start at least as far back as 1973Q1 for all variables (with many starting as far back as 1967Q1, as documented above); this is needed to estimate in real time the benchmark models used in the forecast evaluation exercises. Nonetheless, three issues with the database warrant some data manipulations in order to perform the forecast evaluation from 1982Q1 onward: (1) the length of the historical data across vintages; (2) the forecast horizon across vintages; and (3) changes in variable definitions. We describe them in detail below.

B.1 Historical data time series

For all the variables used in the paper, most of the vintages contain historical data dating back to (at least) 1973Q1. However, as we move forward across the vintages, we notice that for a few vintages the historical data do not extend as far back as 1973; moreover, in the recent past Statistics Canada has been publishing many macroeconomic aggregates only up to 1981Q1. Consequently, staff extended the historical data using backcasting techniques (back to 1973Q1) in order to estimate their forecasting models.[25] For those vintages without historical data dating back to 1973Q1, we backcasted simply using data from the next vintage where longer historical data exist (if the overlapping historical data were identical between the current and that next vintage). If the next vintages did not contain long enough historical data, we backcasted using growth rates from the most recent vintage (i.e. prior to the current vintage) which contained data back to 1973Q1.

[25] For example, the estimation of the projection models ToTEM (Dorich et al. (2013)) or LENS (Gervais and Gosselin (2014)) required longer data series than those published by Statistics Canada. Staff generally backcasted official series using older vintages of the data which contained longer time series for the historical data.

B.2 Forecast horizons

In most cases from 1982Q2 onward, vintages include forecasts with horizons up to 8 quarters ahead.[26] Nonetheless, a few vintages have slightly shorter horizons. Consequently, we faced an important trade-off: we could either drop the longest forecast horizons or impute values to the missing forecasts. Because very few vintages have forecast horizons shorter than 8-quarters-ahead, we decided to use all the available information and proceed with imputation for missing forecasts.[27] We imputed forecasts in levels for the missing quarters using the forecasted growth rates for these quarters from the previous vintage. Overall, these manipulations are relatively minor and allow for a clean forecast evaluation from 1982Q2 onward.

B.3 Changes in variable definitions

As discussed above, most variables that are included in the database have relatively consistent definitions through time. However, there are some definition changes which we needed to deal with in order to perform the forecast evaluation. For example, as mentioned above, real GDP was only adopted by Statistics Canada in the National Accounts as its headline national accounts measure in 1986. Staff produced forecasts for nominal GDP before 1986Q4 but not for the implicit GDP price deflator. Consequently, for the 1982Q2 to 1986Q3 period, we deflate nominal GDP by the implicit GNE price deflator to compute a real measure for GDP. Note that since we evaluate staff forecasts relative to models in real time (i.e. by vintage), this data manipulation does not have material implications for our results.

PCE also experiences definition changes throughout our sample. However, unlike for real GDP, the changes in PCE arise from changes in the staff forecasting models. For instance, PCE's definition includes housing investment at different points in time. To get

[26]As a result, most vintages include at least 9 quarters of forecasts: a nowcast and forecasts for the next 8 quarters.

[27]For example, as noted above, the forecast horizon is shorter in the 1986Q3 vintage: specifically, horizons t+6 to t+8 are missing for all variables in that quarter. As a result, we use forecasts from the 1986Q2 vintage for these missing horizons. Another example is the missing forecast horizons for CPI inflation in the 1995Q3 vintage (horizons t+2 to t+8). We thus use forecast values prepared in 1995Q2 to impute the missing values.

around this problem, we defined PCE as the residual of real GDP and the other consistently available GDP components. Specifically, PCE is defined as

$$PCE_t = Y_t - (I_t + G_t + X_t - M_t), \tag{5}$$

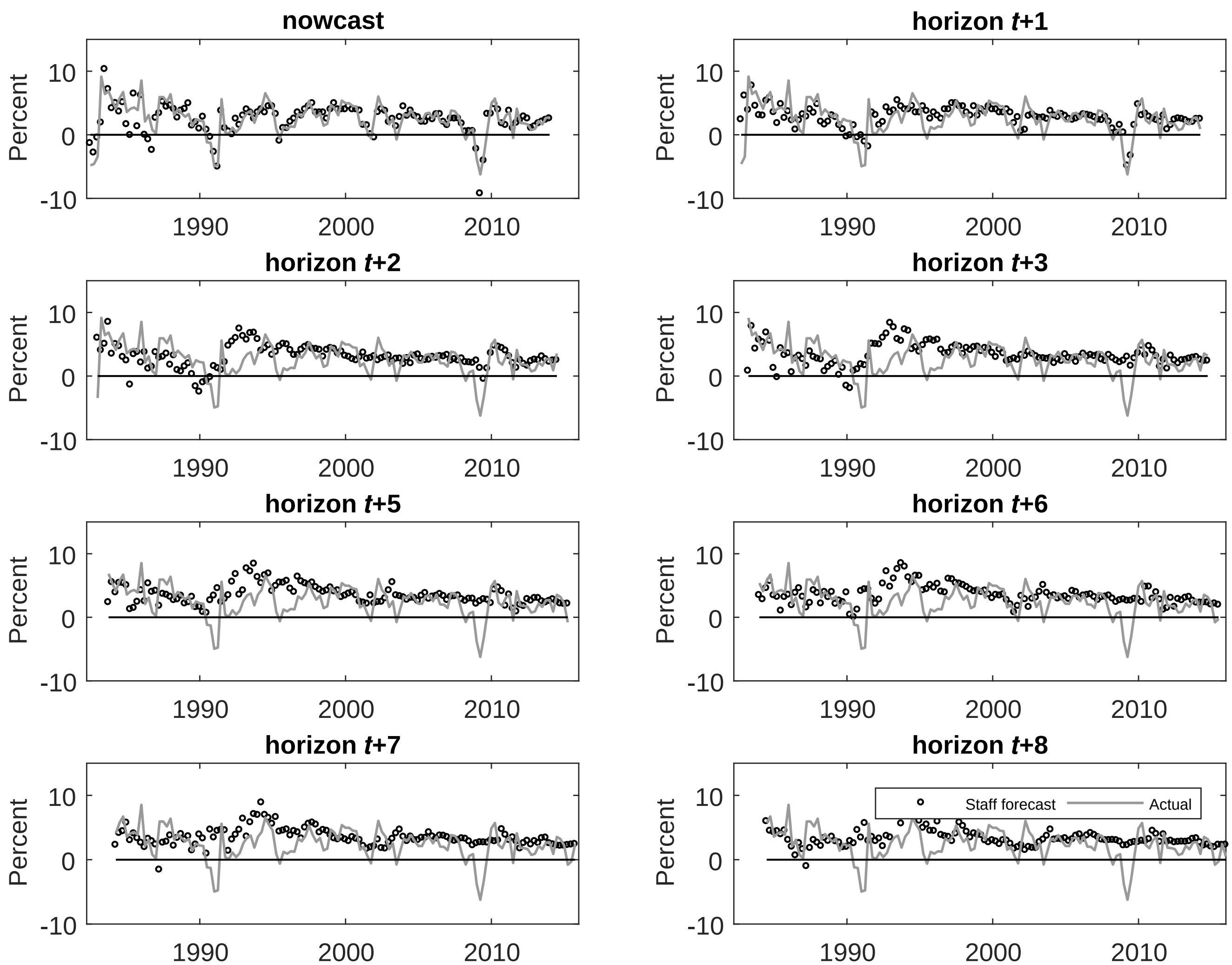
where Y_t represents the level of GDP, I_t business investment, G_t government expenditures, X_t exports and M_t imports. Therefore, in the forecasting exercise of GDP components (see Section D below), the PCE definition includes inventories as well as housing investment.[28] This way we ensure we have a consistent set of vintages for PCE that are consistent from 1982Q2 to 2013Q4.

[28] Although the above identity (Equation 5) holds in nominal terms, we compute PCE in the forecasting exercise using Equation (5) in real terms.

Appendix C The staff GDP growth and CPI inflation forecasts and actual values

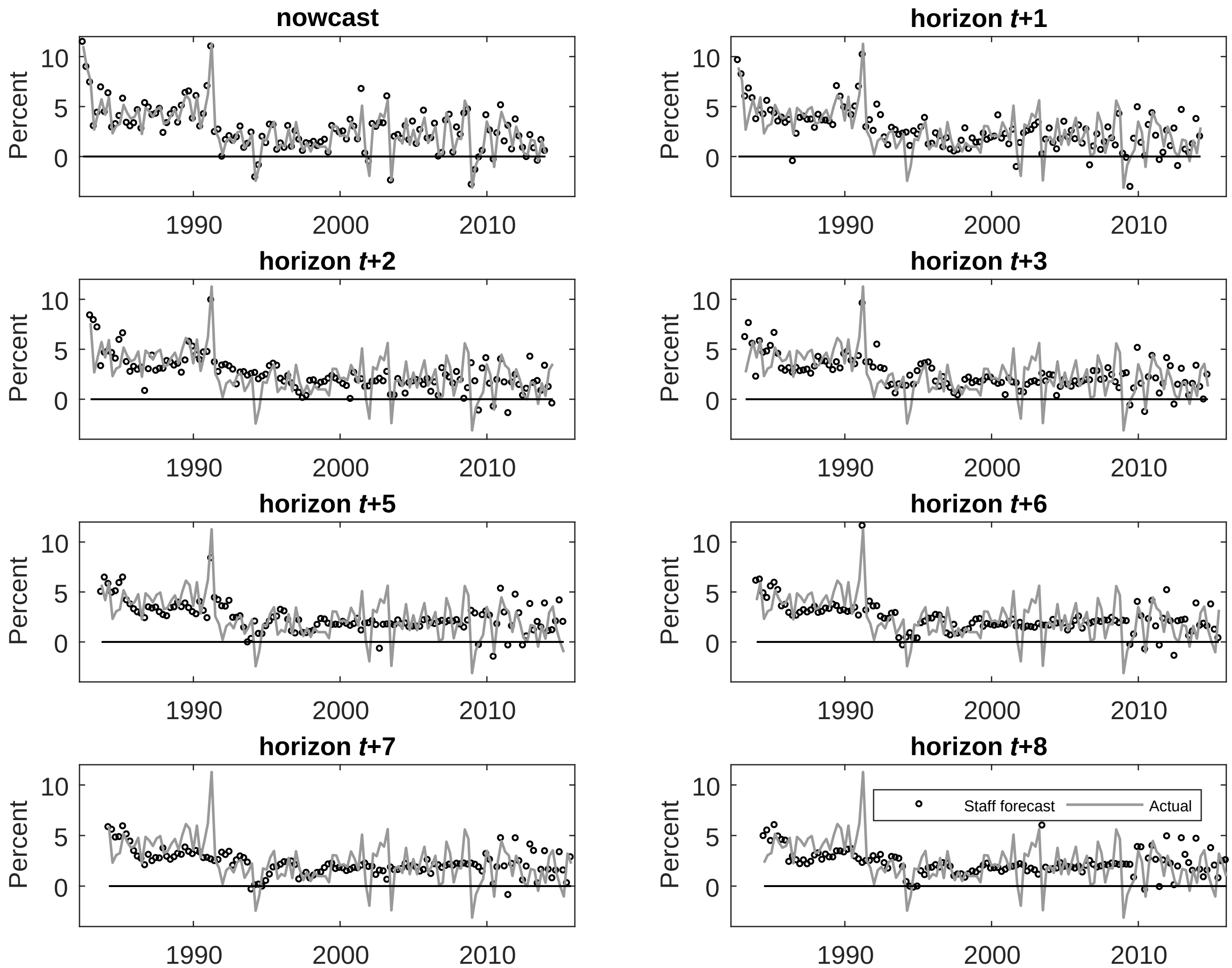
Figures C.1 and C.2 below plot the staff forecasts for GDP growth and CPI inflation, respectively, for horizons h=0 up to 8 together with the real-time actual data. We omit the 4-quarter-ahead horizon, as it is presented in Figure 1 in the main text.

Figure C.1: GDP growth: staff forecasts *vs.* actuals



Note: This figure shows the staff GDP growth forecasts and the corresponding actual values for the nowcast (h=0) up to 8 quarters ahead, omitting the 4-quarters-ahead horizon (see Figure 1 in main text).

Figure C.2: CPI inflation: staff forecasts *vs.* actuals



Note: This figure shows the staff CPI inflation forecasts and the corresponding actual values for the nowcast (*h=0*) up to 8 quarters ahead, omitting the 4-quarter-ahead horizon (see Figure 1 in main text).

Appendix D Comparing the staff forecasts for GDP components with the AR model

In Section 3 of the main text, we showed that the staff short-term forecasts were significantly more accurate than the econometric models for GDP growth. However, the simple AR(4) model produced forecasts for horizons beyond 1-quarter-ahead that were on par with the staff forecast. Here, we examine how the staff forecasts for each of the GDP components, i.e. consumption, investment, government expenditure, exports and imports, compare with the AR(4) model.

Table D.1 summarizes the results. The first row for each component depicts the staff RSMPE for each horizon. The remaining entries show the ratio of the AR(4) RMSPE over the staff's for various jumping-off points, as described in the main text. A ratio over one indicates that the staff RMSPE is lower than the AR(4) model, while numbers below one favor the AR(4) over the staff. For all of GDP components, the staff is significantly more accurate than the AR(4) for the nowcast. The advantage of the staff over the econometric model ranges from 13% (government expenditures) to 53% (imports). However, the size and statistical significance of this advantage vanishes rather quickly for all components as we consider longer horizons. For example, the staff is still more accurate than the AR(4) model when forecasting government expenditures and exports, but the differences in RMSPE are rather small and insignificant.

In short, the evidence presented in Table D.1 shows that the relative performance of the staff forecasts for GDP growth extends to the components of GDP: staff forecasts are substantially better at short horizons, but are not more accurate than a simple AR(4) for horizons over 1-quarter-ahead.

Table D.1: AR(4) Benchmarks Performance Relative to Staff for Forecasting GDP Components Growth

Model		T0	T+1	T+2	T+4	T+6	T+8
Consumption growth							
Staff	(RMSPE)	4.18	5.19	5.44	5.27	5.10	5.04
AR(4)	Baseline	1.27***	1.04	0.96	0.92	0.94	0.95
	jump-off 0	.	0.99	0.96	0.94	0.94	0.95
	jump-off +1	.	.	0.97	0.94	0.95	0.95
	jump-off +2	.	.	.	0.92	0.97	0.97
Investment growth							
Staff	(RMSPE)	7.79	9.38	9.74	9.90	10.47	9.82
AR(4)	Baseline	1.26*	1.04	0.97	0.97	0.94	1.02
	jump-off 0	.	1.01	0.98	0.97	0.94	1.02
	jump-off +1	.	.	0.99	0.99	0.94	1.02
	jump-off +2	.	.	.	0.98	0.94	1.02
Government expenditure growth							
Staff	(RMSPE)	2.51	2.55	2.35	2.52	2.60	2.24
AR(4)	Baseline	1.13	1.09	1.16	1.05	1.03	1.21
	jump-off 0	.	1.13	1.17	1.05	1.02	1.21
	jump-off +1	.	.	1.20	1.06	1.02	1.20
	jump-off +2	.	.	.	1.07	1.01	1.19
Exports growth							
Staff	(RMSPE)	8.05	9.98	11.22	10.39	10.23	9.64
AR(4)	Baseline	1.41***	1.13*	1.00	1.02	1.03	1.02
	jump-off 0	.	1.12*	1.01	1.03	1.03	1.02
	jump-off +1	.	.	0.99	1.03	1.02	1.02
	jump-off +2	.	.	.	1.03	1.03	1.02
Imports growth							
Staff	(RMSPE)	7.56	10.68	11.79	11.83	11.10	10.61
AR(4)	Baseline	1.53***	1.10	0.99	0.92*	0.98	1.01
	jump-off 0	.	1.09	1.01	0.94	0.98	1.01
	jump-off +1	.	.	1.01	0.94	0.98	1.01
	jump-off +2	.	.	.	0.94	0.98	1.01

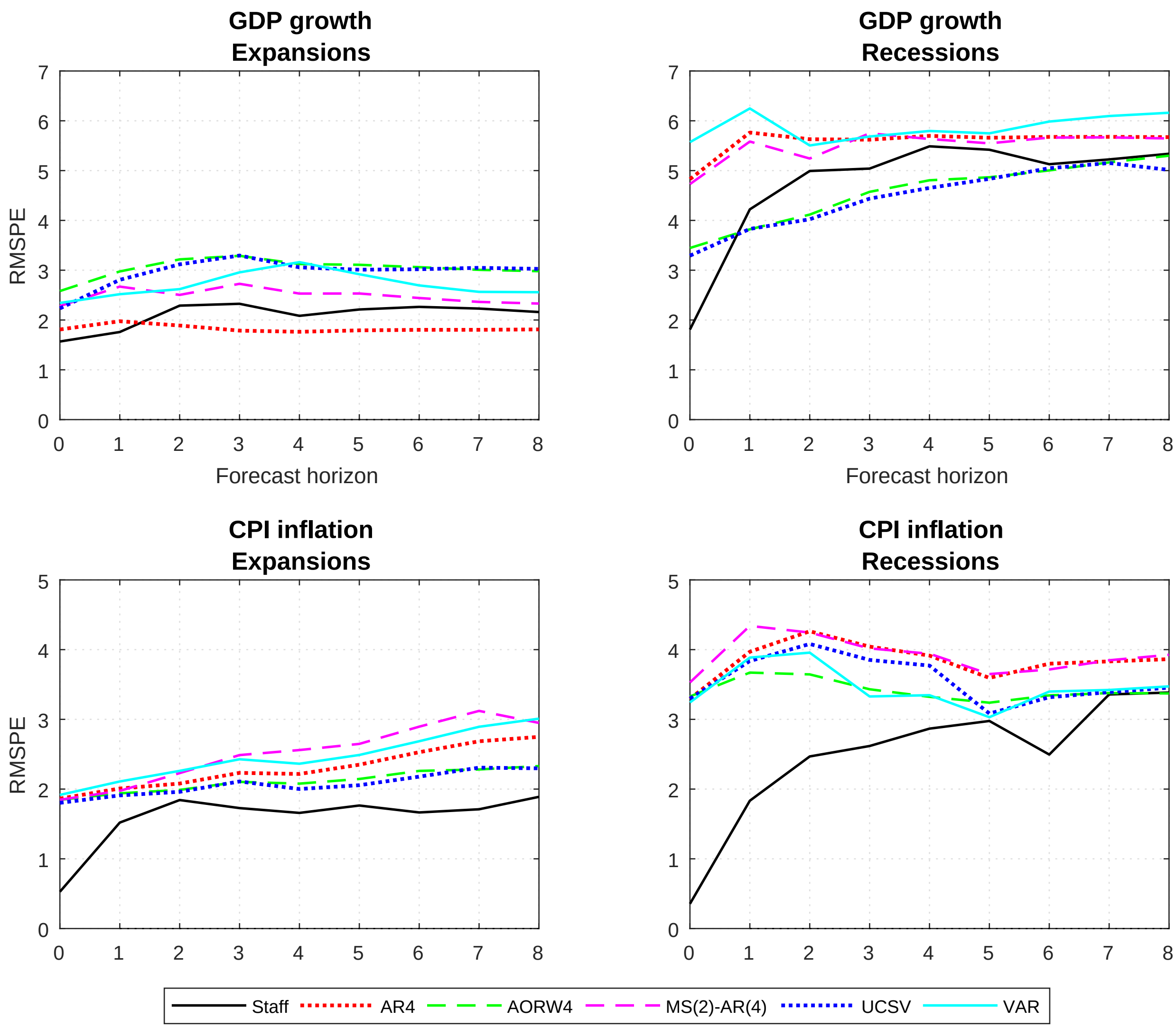
Note: This table reports RMSPEs for the staff GDP components growth forecasts, as well as RMSPE ratios defined as AR(4) over staff. Stars (***, **, *) indicate significance of Diebold-Mariano test-statistics at 1, 5 and 10 per cent, respectively. Forecast dates range from 1982Q2 to 2015Q4.

Appendix E Forecasting expansions and recessions

Here we conduct a similar RMSPE comparison as in Section 3 in the paper, but we divide the sample between periods of expansions and recessions. The recession dates are those determined by the C.D. Howe Institute. As in the paper, dates denote the forecasted period, not when the forecast was prepared. Figure E.1 shows the results.

As expected, the RMSPE for the staff and the econometric models are markedly higher for periods of recession. However, the qualitative results of Section 3 remain unchanged: the staff is more accurate than the econometric models both during expansions and recessions for the short horizons, but then slightly less so than the AR model for the longer-horizon forecasts. Interestingly, for CPI inflation the gap between the performances of the staff and the econometric models is much wider during recessions than during expansions: although predicting recessions is virtually impossible, staff is better than the models at recognizing the negative pressures on inflation at the onset of recessions.

Figure E.1: Staff RMSPEs for GDP growth and CPI inflation *vs.* econometric models during expansions and recessions



Note: This figure shows the RMSPEs of GDP growth and CPI inflation forecasts from the staff and the econometric benchmarks estimated in real time up to horizon $t+8$. The left panels show the RMPSEs for expansionary periods, while the right panels show the RMSPEs for recession periods (as defined by the C.D. Howe Institute). Forecasted dates range from 1982Q2 to 2015Q4.

Appendix F Percentage of time that econometric benchmarks are more accurate than staff's

In this section, we report an alternative evaluation metric to compare the staff and the benchmark econometric models. Specifically, we calculate the proportion of forecasts for which the econometric models have a lower forecast error than the staff. A number greater than 50 per cent implies that on average the econometric models have lower forecast errors relative to staff, while a number below 50 per cent implies that staff is usually better than the models. We report in Tables F.1 to F.3 the results for real GDP growth, total CPI inflation and core CPI inflation. In Table F.4, we report results of the AR(4) model relative to staff for the GDP components.

For GDP growth (Table F.1), the numbers are well below 0.5 for the nowcast and the 1-quarter-ahead horizons. For the remaining horizons, the staff is roughly on par with the AR(4) and MSM(2)-AR(4) models, and more accurate than the VAR model 65 to 70 per cent of the time. Table F.4 shows that a similar result is generally observed for the components of GDP: staff is relatively better than the AR(4) model over short horizons and worse over longer horizons. These results are in line with those presented in Section D above.

Tables F.2 and F.3 show the results for CPI and core CPI inflation, respectively. Given the staff's information advantage, the econometric models are never more accurate at the nowcasting and 1-quarter-ahead horizons. The AR(4), MSM(2)-AR(4) and VAR models are usually less accurate than the staff (i.e. about 40 per cent of the periods), while the UCSV and AO-RW models are sometimes on par with the staff at the 2-quarter-ahead horizon and beyond. Overall, using this alternative evaluation metric does not alter the main takeaways found in the paper: staff is relatively better over the short forecast horizons but much less so 2-quarters-ahead and beyond.

Table F.1: Percentage of the Time that Econometric Benchmarks Are More Accurate than Staff for Forecasting GDP Growth

Model		T0	T+1	T+2	T+4	T+6	T+8
AR(4)	jump-off -1	0.31	0.44	0.53	0.47	0.52	0.48
	jump-off 0	.	0.44	0.54	0.46	0.52	0.50
	jump-off +1	.	.	0.57	0.46	0.52	0.50
	jump-off +2	.	.	.	0.46	0.55	0.49
AO-RW	jump-off -1	0.30	0.36	0.40	0.43	0.46	0.46
	jump-off 0	.	0.38	0.43	0.43	0.47	0.42
	jump-off +1	.	.	0.52	0.48	0.49	0.43
	jump-off +2	.	.	.	0.56	0.54	0.53
MSM(2)-AR(4)	jump-off -1	0.35	0.38	0.50	0.41	0.50	0.48
	jump-off 0	.	0.40	0.46	0.42	0.49	0.50
	jump-off +1	.	.	0.49	0.46	0.49	0.49
	jump-off +2	.	.	.	0.45	0.50	0.47
UCSV	jump-off -1	0.34	0.38	0.44	0.42	0.46	0.43
	jump-off 0	.	0.39	0.50	0.45	0.50	0.43
	jump-off +1	.	.	0.54	0.54	0.55	0.51
	jump-off +2	.	.	.	0.53	0.68	0.54
VAR	jump-off -1	0.31	0.31	0.37	0.35	0.30	0.29
	jump-off 0	.	0.32	0.35	0.35	0.31	0.31
	jump-off +1	.	.	0.43	0.38	0.31	0.31
	jump-off +2	.	.	.	0.38	0.28	0.30

Note: This table reports the share of forecast periods in which the econometric benchmarks have a lower absolute forecast error than the staff. A value above 0.50 means that the econometric model performs better than the staff in more instances. Forecast dates range from 1982Q2 to 2015Q4.

Table F.2: Percentage of the Time that Econometric Benchmarks Are More Accurate than Staff for Forecasting Inflation

Model		T0	T+1	T+2	T+4	T+6	T+8
AR(4)	jump-off -1	0.13	0.35	0.39	0.40	0.33	0.36
	jump-off 0	.	0.43	0.42	0.37	0.36	0.38
	jump-off +1	.	.	0.48	0.43	0.36	0.38
	jump-off +2	.	.	.	0.45	0.35	0.39
AO RW	jump-off -1	0.15	0.41	0.49	0.46	0.39	0.50
	jump-off 0	.	0.46	0.49	0.50	0.43	0.47
	jump-off +1	.	.	0.49	0.50	0.45	0.46
	jump-off +2	.	.	.	0.49	0.53	0.50
MSM(2)-AR(4)	jump-off -1	0.13	0.33	0.39	0.35	0.35	0.37
	jump-off 0	.	0.42	0.38	0.39	0.36	0.38
	jump-off +1	.	.	0.51	0.40	0.34	0.37
	jump-off +2	.	.	.	0.42	0.35	0.38
UCSV	jump-off -1	0.15	0.43	0.48	0.46	0.41	0.45
	jump-off 0	.	0.43	0.47	0.45	0.50	0.44
	jump-off +1	.	.	0.54	0.51	0.43	0.52
	jump-off +2	.	.	.	0.50	0.50	0.54
VAR	jump-off -1	0.18	0.35	0.40	0.39	0.33	0.33
	jump-off 0	.	0.41	0.43	0.41	0.38	0.38
	jump-off +1	.	.	0.50	0.41	0.36	0.37
	jump-off +2	.	.	.	0.46	0.39	0.41

Note: This table reports the share of forecast periods in which the econometric benchmarks have a lower absolute forecast error than the staff. A value above 0.50 means that the econometric model performs better than the staff in more instances. Forecast dates range from 1982Q2 to 2015Q4.

Table F.3: Percentage of the Time that Econometric Benchmarks Are More Accurate than Staff for Forecasting Core Inflation

Model		T0	T+1	T+2	T+4	T+6	T+8
AR(4)	jump-off -1	0.25	0.35	0.43	0.39	0.31	0.31
	jump-off 0	.	0.41	0.37	0.40	0.35	0.31
	jump-off +1	.	.	0.44	0.46	0.38	0.34
	jump-off +2	.	.	.	0.49	0.42	0.36
AO RW	jump-off -1	0.23	0.39	0.42	0.46	0.40	0.40
	jump-off 0	.	0.39	0.42	0.44	0.39	0.41
	jump-off +1	.	.	0.43	0.47	0.45	0.43
	jump-off +2	.	.	.	0.57	0.47	0.44
MSM(2)-AR(4)	jump-off -1	0.24	0.36	0.40	0.37	0.31	0.27
	jump-off 0	.	0.43	0.37	0.38	0.35	0.31
	jump-off +1	.	.	0.46	0.43	0.35	0.31
	jump-off +2	.	.	.	0.46	0.43	0.38
UCSV	jump-off -1	0.23	0.37	0.46	0.47	0.47	0.43
	jump-off 0	.	0.39	0.37	0.49	0.44	0.46
	jump-off +1	.	.	0.44	0.52	0.45	0.43
	jump-off +2	.	.	.	0.47	0.48	0.49
VAR	jump-off -1	0.18	0.37	0.39	0.42	0.33	0.24
	jump-off 0	.	0.35	0.37	0.45	0.39	0.29
	jump-off +1	.	.	0.41	0.46	0.43	0.35
	jump-off +2	.	.	.	0.45	0.43	0.39

Note: This table reports the share of forecast periods in which the econometric benchmarks have a lower absolute forecast error than the staff. A value above 0.50 means that the specified econometric model performs better than the staff in more instances. Forecast dates range from 1982Q2 to 2015Q4.

Table F.4: Percentage of the Time that Econometric Benchmarks Are More Accurate than Staff for Forecasting GDP Components Growth

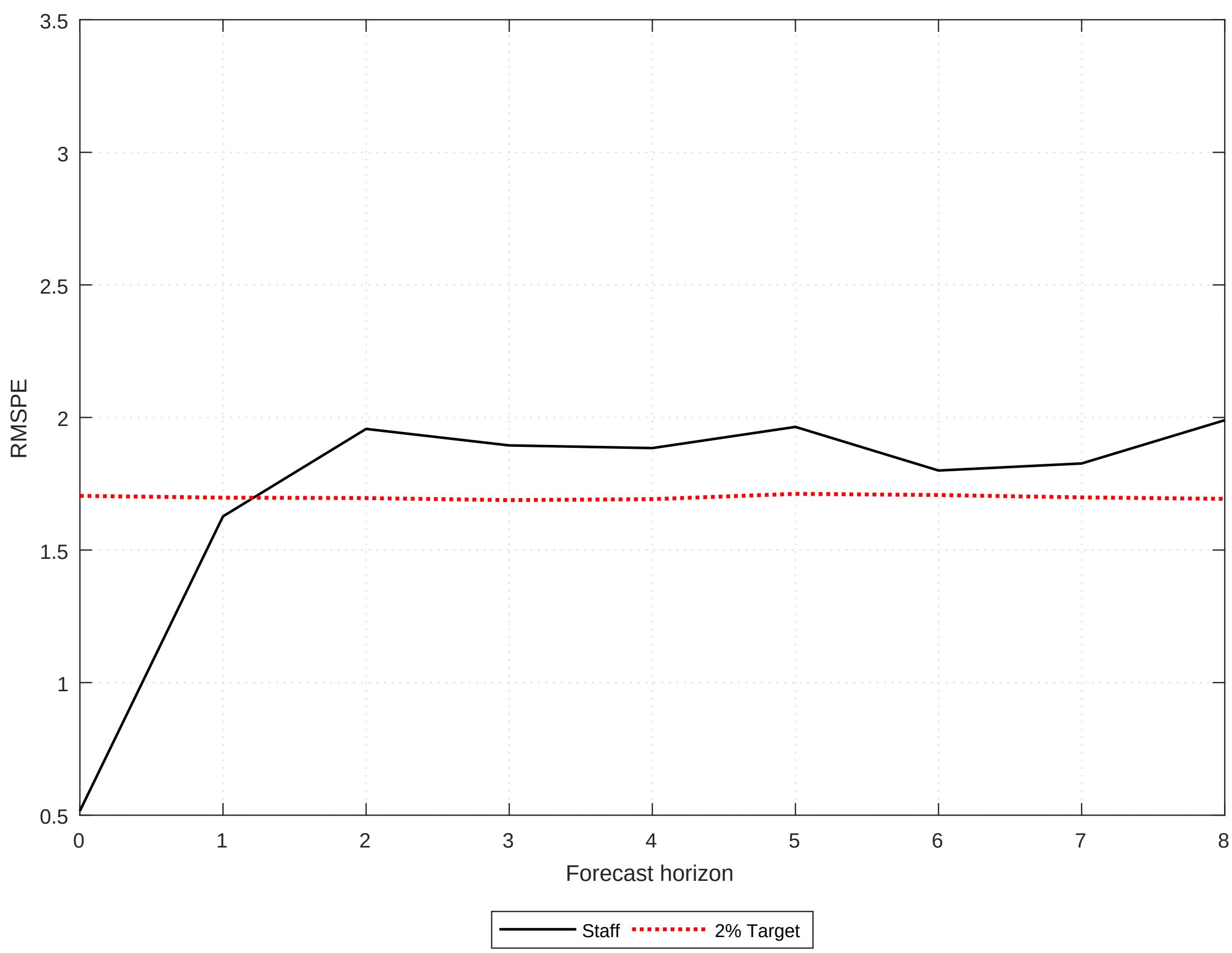
Model		T0	T+1	T+2	T+4	T+6	T+8
Consumption growth							
AR(4)	jump-off -1	0.37	0.49	0.57	0.56	0.59	0.58
	jump-off 0	.	0.50	0.53	0.60	0.59	0.58
	jump-off +1	.	.	0.49	0.60	0.57	0.58
	jump-off +2	.	.	.	0.60	0.58	0.57
Investment growth							
AR(4)	jump-off -1	0.45	0.43	0.50	0.57	0.50	0.43
	jump-off 0	.	0.38	0.48	0.56	0.49	0.43
	jump-off +1	.	.	0.47	0.52	0.50	0.43
	jump-off +2	.	.	.	0.47	0.49	0.44
Government expenditure growth							
AR(4)	jump-off -1	0.44	0.45	0.49	0.48	0.47	0.40
	jump-off 0	.	0.43	0.49	0.49	0.50	0.40
	jump-off +1	.	.	0.46	0.49	0.49	0.40
	jump-off +2	.	.	.	0.46	0.49	0.41
Exports growth							
AR(4)	jump-off -1	0.40	0.39	0.46	0.47	0.46	0.49
	jump-off 0	.	0.35	0.46	0.45	0.46	0.50
	jump-off +1	.	.	0.48	0.46	0.46	0.50
	jump-off +2	.	.	.	0.47	0.45	0.49
Imports growth							
AR(4)	jump-off -1	0.35	0.49	0.51	0.60	0.47	0.52
	jump-off 0	.	0.49	0.50	0.55	0.49	0.52
	jump-off +1	.	.	0.50	0.55	0.49	0.53
	jump-off +2	.	.	.	0.57	0.47	0.53

Note: This table reports the share of forecast periods in which the econometric benchmarks have a lower absolute forecast error than the staff. A value above 0.50 means that the econometric model performs better than the staff in more instances. Forecast dates range from 1982Q2 to 2015Q4.

Appendix G Comparing the staff CPI inflation forecasts with a constant forecast of 2 Per Cent

Figure G.1 shows the RMSPEs from the staff CPI inflation forecasts during the inflation targeting period relative (1992Q1 to 2015Q4) to a constant forecast of 2 per cent throughout.

Figure G.1: RMSPEs of CPI inflation forecasts, Staff *vs.* 2 Per Cent Target



Note: This figure shows the RMSPEs of the staff CPI inflation forecasts and a constant forecast of 2 per cent. Forecasted dates range from 1992Q1 to 2015Q4.

Appendix H Additional bias tests

As we saw in Figure 1 in the main text, staff is not able to predict recessions 4-quarters ahead. Consequently, recessions cause large and serially correlated negative forecast errors at longer horizon forecasts. Here we expand on the bias test reported in Section 3.4 of the paper and control for recession periods. We run a regression similar to Equation (3) in the main text, but add a dummy if the quarter being forecasted falls in a recession, as defined by the C.D. Howe Institute:

$$e_{t+h} = \alpha + \gamma D^{Recession} + \varepsilon_{t+h}, \tag{6}$$

where e_{t+h} is the h-quarter-ahead forecast error for either GDP growth or CPI inflation. As in Section 3.4, we perform the tests for the staff's and econometric models' forecasts. Table H.1 reports the estimates of α and bold entries denote statistical significance at the 5 per cent level. We can see that controlling for recessions attenuates the staff bias in forecasting GDP growth at the longer horizons, but does not eliminate it. For instance, even when controlling for recessions, the average forecast error for the 6-quarters-ahead GDP growth forecast is -0.68, implying that staff forecasts overestimate actual GDP growth on average by 68 percentage points. In contrast, the recession dummy in equation (6) does not have any material effect on the staff CPI inflation average forecast error. We also highlight that both the AR(4) and the VAR models generate biased forecasts for GDP growth and CPI inflation, while the MSM(2)-AR(4) model produces biased CPI inflation forecasts for all horizons.

Table H.1: Bias Test Results

Horizon	Staff	AR(4)	AO RW	MSM(2)-AR(4)	UCSV	VAR
GDP						
0	0.23	-0.07	0.34	-0.29	0.30	**-0.79**
1	0.05	-0.18	0.43	-0.20	0.40	**-1.12**
2	-0.04	-0.20	0.56	-0.02	0.51	**-1.18**
4	-0.49	-0.34	0.67	0.02	0.59	**-1.51**
6	**-0.68**	**-0.47**	0.65	0.05	0.61	**-1.56**
8	-0.52	**-0.53**	0.67	-0.11	0.63	**-1.43**
Inflation						
0	-0.07	**-0.36**	-0.05	**-0.35**	-0.03	-0.01
1	0.05	**-0.56**	-0.08	**-0.56**	-0.04	-0.19
2	0.16	**-0.77**	-0.13	**-0.69**	-0.07	-0.33
4	0.04	**-1.26**	-0.29	**-1.15**	-0.19	**-0.86**
6	0.19	**-1.66**	-0.39	**-1.55**	-0.32	**-1.44**
8	0.09	**-1.97**	-0.46	**-1.85**	-0.40	**-1.93**

Note: This table reports the constant coefficient from the Mincer-Zarnowitz test (Equations 6) on staff and model forecasts. **Bold entries** indicate statistical significance of the reported bias coefficients at 5 per cent. The Newey-West lag order corresponds to the forecast horizon + 1, where the nowcast is defined as forecast horizon 0. Forecast dates range from 1982Q2 to 2015Q4.