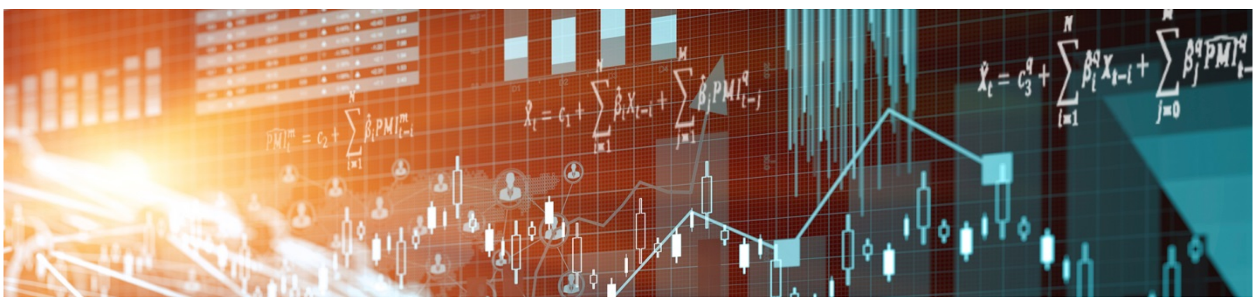
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2016-28

The Real-Time Properties of the Bank of Canada's Staff Output Gap Estimates



by Julien Champagne, Guillaume Poulin-Bellisle and Rodrigo Sekkel

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2016-28

June 2016

The Real-Time Properties of the Bank of Canada's Staff Output Gap Estimates

by

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

[1]Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
julienchampagne@bankofcanada.ca

[2]Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
gpoulin-bellisle@bankofcanada.ca

[3]Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
rsekkel@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

We thank Brigitte Desroches, Jean-Philippe Cayen, Pierre St-Amant, Sharon Kozicki, Ben Tomlin and seminar participants at the 2016 Bank of Canada Fellowship Learning Exchange conference for useful conversations and comments. We especially thank Simon van Norden for comments on an earlier version of the paper. Additionally, we thank Hope Pioro for help in retrieving some of the data used in this paper and Nicole van de Wolfshaar for editorial assistance.

Abstract

We study the revision properties of the Bank of Canada's staff output gap estimates since the mid-1980s. Our results suggest that the average staff output gap revision has decreased significantly over the past 15 years, in line with recent evidence for the U.S. Alternatively, revisions from purely statistical methods to estimate the gap have not experienced the same drop in magnitude. We then examine the usefulness of real-time gap estimates for forecasting inflation and find no deterioration in forecast performance when inflation projections are conditioned on real time rather than on final estimates of the gap.

JEL classification: C38, E17, E32
Bank classification: Potential output; Central bank research; Econometric and statistical method

Résumé

Nous étudions les propriétés de révision des estimations de l'écart de production produites par le personnel de la Banque du Canada depuis le milieu des années 1980. Nos résultats tendent à montrer que l'ampleur desdites révisions a généralement diminué de façon significative durant les 15 dernières années, ce qui est conforme à des résultats récents obtenus pour les États-Unis. Par ailleurs, nous n'observons pas la même diminution pour les révisions apportées aux estimations de l'écart de production provenant de méthodes purement statistiques. Nous examinons ensuite l'utilité des estimations en temps réel de l'écart de production pour prévoir l'inflation au Canada. Nous ne trouvons pas de détérioration dans la qualité des prévisions lorsque les projections d'inflation reposent sur des estimations en temps réel plutôt que sur des estimations finales de l'écart de production.

Classification JEL : C38, E17, E32
Classification de la Banque : Production potentielle; Recherches menées par les banques centrales; Méthodes économétriques et statistiques

Non-technical summary

The output gap—the difference between an economy's output relative to its trend or potential level—plays a central role in the conduct of monetary policy. It is the central aggregate demand link between monetary policy action and inflation pressures and plays a prominent role in central bank communications. Consequently, central banks devote large resources to its measurement.

But measurement of the output gap is challenging, especially in real time, because trend or potential output level is not directly observable. There is a vast literature arguing that measuring the output gap in real time is subject to a large degree of uncertainty, which complicates its use in policy analysis since policy-makers are more concerned with real-time estimates. For instance, the seminal paper by Orphanides and van Norden (2002) illustrated that well-known statistical methods used to measure the output gap yield estimates for the United States are subject to large subsequent revisions and thus not reliable in real time. Similar empirical analyses also finding unreliable real-time output gap estimates have been conducted for many other countries, such as Canada, the euro area and Japan, to name a few.

All in all, these studies point to substantial uncertainty in output gap estimates in real time using well-known statistical models and thus policy decisions based on such measures would be considerably biased. However, in a recent paper, Edge and Rudd (2016) argue that these issues have become less severe in more recent samples, such as the estimates produced by the Federal Reserve Board staff's (Greenbook) for the United States. Edge and Rudd (2016) show that the staff's estimates of the output gap have considerably improved real-time revision properties than those found by Orphanides and van Norden (2002) for an earlier set of estimates and thus have been more reliable recently. Furthermore, contrary to previous studies, they find that there is no deterioration in forecast performance when inflation projections are conditioned on real-time instead than on final estimates of the staff output gap.

Our paper uses a novel data set of real-time data and forecasts from the Bank of Canada's staff economic projections to study the revisions properties of the output gap in Canada over the past 30 years. We find that for the earlier part of the sample, the output gap was indeed estimated with a lot of uncertainty and revisions were large and volatile. However, over the second half of our sample, revisions properties have improved significantly, and the output gap has been estimated more reliably, consistent with the findings in Edge and Rudd (2016). We find that this improvement on the reliability of the output gap estimates is not matched by judgement-free, well-known statistical methods to estimate the output gap. We then examine the usefulness of real-time output gap estimates for forecasting inflation in Canada. We find that Phillips curve models that condition inflation forecasts on the staff's expected path of the output gap in real time do not outperform simple univariate models. Nonetheless, this result does not follow from the real-time unreliability of output gap estimates because real-time estimates of the gap perform as well or better than final estimates in Phillips curve models.

1 Introduction

The output gap—the difference between an economy's output relative to its trend or potential level—plays a central role in the conduct of monetary policy. There is a vast literature arguing that its measurement in real time is subject to a large degree of uncertainty, which complicates its use in policy analysis. For example, the prominent work by Orphanides and van Norden (2002) showed that well-known detrending methods yield output gap measures for the United States that are subject to large subsequent revisions and thus not reliable in real time. They argue that the end-point problem of statistical filters is the main cause behind these large gap revisions, more important than data revisions in official statistics. Marcellino and Musso (2011), Cayen and van Norden (2005), Kamada (2005), Bernhardsen et al. (2005) and Cusinato et al. (2013) use similar statistical methods and also find real-time output gap estimates to be unreliable for the euro area, Canada, Japan, Norway and Brazil, respectively. Moreover, Orphanides (2003) uses real-time output gap estimates from 1980 to 1992 for the United States produced by the Federal Reserve's staff (Greenbook estimates) and shows that they are subject to very large revisions and thus unreliable. However, a recent paper by Edge and Rudd (2016) argues that these problems have become less severe in more recent samples; they show that the Federal Reserve's staff estimates of the output gap have considerably better real-time revision properties than those found by Orphanides and van Norden (2002) and Orphanides (2003) for an earlier set of output gap estimates. Edge and Rudd (2016) find that revisions have become smaller since the mid-1990s, concluding that the output gap has been estimated more reliably in real time in more recent samples than what previous studies have documented.[1] Furthermore, they find that contrary to previous studies, there is no deterioration in forecast performance when inflation projections are conditioned on real time rather than on final estimates of

[1]Similar conclusions are found in two other recent papers that augment econometric models with more modern techniques and show that they can also yield lower gap revisions in more recent samples for the U.S. On the one hand, Mertens (2014) uses well-known econometric decompositions to estimate the trend and cycle of output but let the volatility of shocks vary over time. He argues that augmenting these statistical models with stochastic volatility better captures the changes in macroeconomic volatility, such as those observed over the past 30 years. He further notes that the consistency of his results with Edge and Rudd (2016) staff's estimates suggests that adjustments due to time-varying volatility could have played an important role in judgemental estimates of the output gap. On the other hand, Kamber et al. (2016) impose a more persistent trend in the traditional Beveridge-Nelson decomposition and find output gap estimates that are more intuitive and reliable.

the staff output gap.[2]

Our paper uses a novel data set of real-time data and forecasts constructed from the Bank of Canada's staff economic projections to study the revision properties of the output gap in Canada over the period from 1987:Q1 to 2015:Q4. To our knowledge, our paper is the first to study the revision properties of a central bank staff's output gap estimates outside the United States. Estimates of output gaps in central banks are usually derived from more than one model, and the staff often use soft indicators and judgement to complement statistical tools. Consequently, assessing whether or not the unreliability of statistical gap estimates in real time translates to central bank staff gap estimates is of great policy relevance. We show that the average size of the revisions and a measure of noise-to-signal ratio have diminished significantly in the more recent half of our sample. We find that this improvement on the reliability of the output gap estimates is not matched by judgement-free, well-known statistical methods to estimate the output gap.[3] We then examine the usefulness of real-time output gap estimates for forecasting inflation in Canada. We find that Phillips curve models that condition inflation forecasts on the staff's expected path of the output gap in real-time do not outperform simple univariate models. Nonetheless, this result does not follow from the real-time unreliability of output gap estimates because real-time estimates of the gap perform as well or better than final estimates in Phillips curve models.

The remainder of the paper is organized as follows. Section 2 describes our real-time data and forecasts data set and lays out the methodology. Section 3 analyzes the revisions properties of the staff output gap estimates, while Section 4 turns to the revision properties of gaps estimated from purely statistical methods. Section 5 examines inflation forecasts and Section 6 concludes.

[2]Using purely econometric methods to estimate the output gap, Orphanides and van Norden (2005) show that real-time output gaps do not perform nearly as well as ex post gap measures to forecast inflation.

[3]Relative to Edge and Rudd (2016), our sample includes the Great Recession period and the subsequent recovery, rendering the improvement in the staff's estimates of the output gap in real time even more convincing.

2 Real-Time Data and Estimates from the Staff's Projections: Details

We use a novel and proprietary database of real-time output gap estimates constructed from the Bank of Canada's staff economic projections. Bank of Canada staff produce four exhaustive projections each year, following the release of the quarterly national income and expenditure accounts, which are generally carried out around the end of March, June, September, and December. These staff projections contain quarterly forecasts as well as historical (real-time) data of numerous macroeconomic aggregates. They are a material part of the analysis presented to the Governing Council every quarter in the weeks leading up to the publication of the Bank's *Monetary Policy Report*.[4] The quarterly staff projections are analogous to the Greenbook forecasts prepared by the Federal Reserve Board staff; it is judgemental in the sense that output gap estimates are based on different sources of information and economic models.[5]

Our sample of real-time output gap vintages begins in 1987:Q1 and ends in 2015:Q4, for a total of 116 quarterly data vintages. All of those vintages contain data extending (at least) back to 1973:Q1 and include a nowcast for the current quarter as well as forecasts for at least eight quarters ahead.

We follow the same timing convention and definitions as Orphanides and van Norden (2002) and Edge and Rudd (2016): the real-time output gap estimates for quarter(t) are obtained from the forecast prepared in the quarter(t+1) vintage. For example, the March 2001 projection data vintage contains the real-time estimate for the 2000:Q4 output gap. Consequently, our sample of real-time output gap estimates starts in 1986:Q4 and ends in 2015:Q3. The corresponding final estimate of the output gap, in contrast, is defined as the estimate available two years after the last real-time estimate in the sample, again mimicking

[4] See Macklem (2002) for details about the information and analysis presented by the staff to the Governing Council. We highlight the fact that these are staff's estimates and thus may not be the same estimates provided in the *Monetary Policy Report* (MPR), the Bank of Canada publication representing the view of the Governing Council available every quarter since 2001.

[5] Methods used to estimate potential output have evolved through time at the Bank of Canada. See, for example, Pichette et al. (2015) for a description of recent methods; and Laxton and Tetlow (1992), Butler (1996) and Barnett et al. (2009) for older estimation techniques. Additionally, Barnett et al. (2009) examine how real-time shocks are transmitted to output gap staff forecasts using projection data from 1994 to 2005.

Orphanides and van Norden (2002) and Edge and Rudd (2016) for comparability.[6]

3 Revision Properties of the Staff's Real-Time Output Gap

Output gap revisions are defined as the difference between the final and real-time gap estimates. We first calculate the real-time properties of the revisions over our entire sample (i.e., 1986:Q4 to 2013:Q3). We report the mean, mean absolute, standard deviation, root mean squared error (RMSE) for the gap revisions, as well as a measure of noise-to-signal ratio, defined as the RMSE of the gap revisions divided by the standard deviation of the final gap estimate.[7] Then, we follow Edge and Rudd (2016) to assess how the properties of the output gap revisions have evolved over time and break our sample in two equal parts: the first from 1986:Q4 to 2000:Q1, and the second from 2000:Q2 to 2013:Q3. The choice of the split ensures not only that the two subsamples have an equal number of observations, but also that the number of periods separating the real-time and final gap measures is the same for both periods (i.e., two years). For the first subsample (ending with the 2000:Q2 vintage), the final gap is defined as the estimate as of 2002:Q2, while for the second subsample the final gap is defined as the estimate of 2015:Q4. We then recalculate the properties for the output gap revisions for both subsamples. Table 1 reports the results.

[6] Orphanides and van Norden (2002) and Edge and Rudd (2016) point out that there is no well-defined concept of final estimate because data are subject to constant revision. Using a final estimate computed two years after the last real-time estimate in a given sample, as in Orphanides and van Norden (2002) and Edge and Rudd (2016), ensures there are at least two complete annual revisions of the national income and product accounts between any real-time gap estimates and the final gap estimates. See the next section for the specific vintage dates of the final estimates associated with every sample used in the paper.

[7] We also computed another measure of noise-to-signal ratio, the standard deviation of the gap revision divided by the standard deviation of the final estimate. As Edge and Rudd (2016) point out, the RMSE-based measure is more relevant from a policy perspective since they reflect mean errors (or biases) in the real-time gap estimates that are absent from the standard-deviation-based ratio. Consequently, we decided to report in the text only the RMSE-based noise-to-signal ratio. Note that all the results presented below that related to our measure of noise-to-signal ratio are robust to this other definition of the noise-to-signal ratio.

Table 1: Statistics on Staff Output Gap Revisions

	Full sample	1st subsample	2nd subsample
Mean revision	2.04	1.64	0.66
Mean abs. revision	2.06	1.69	0.71
RMSE	2.77	2.06	0.87
Standard deviation	1.89	1.25	0.56
Noise-to-signal ratio	1.54	1.13	0.61
Mean of *Final Gap*	0.47	-0.96	0.12
Standard deviation *Final Gap*	1.80	1.82	1.42

Note: This table reports statistics on the Bank of Canada's staff output gap revisions. For the calculation of output gap revisions, we use the 2015:Q4 vintage as the final output gap estimate for the full sample period as well as for the 2nd subsample (2000:Q2–2013:Q3), and we use 2002:Q2 as the final estimate for the first subsample (1986:Q4–2000:Q1). The noise-to-signal ratio is defined as the ratio of the RMSE of the gap revisions over the standard deviation of the final gap estimate.

Three important observations stand out. First, the mean and mean absolute revisions are roughly similar, implying the staff has been underestimating the output gap in real-time in a systematic way.[8] Second, output gap revisions are very large for the full sample. For example, the mean revision is larger than the average estimated gap, while the standard deviation of the revisions is roughly similar to the final estimate's standard deviation (1.9 vs. 1.8). The RMSE is equal to 2.8 per cent, implying a noise-to-signal ratio of 1.54. These results are somewhat more pronounced than those from econometric models for the United States (Orphanides and van Norden 2002) and similar to those for Canada (Cayen and van Norden 2005), estimated on samples of roughly equal length.[9] Consequently, the errors associated with real-time estimates of the staff output gap are substantial. Third, this hides an important change in the revisions properties within the full sample, as shown in the next two columns of Table 1. The negative bias in staff output gap mean (and mean absolute) revisions has gone down substantially for the period from 2000:Q2 to 2013:Q3 relative to the earlier period from 1986:Q4 to 2000:Q1. The RMSE, for example, falls by about 1.2 percentage points, while the standard deviation of the revisions decreases from 1.25 to 0.56 percentage points. More importantly, the noise-to-signal ratio falls from 1.13

[8]As revisions are defined as final minus real-time estimates, a positive average revision implies a *negative* bias in the real-time output gap estimation.

[9]The sample covered by Orphanides and van Norden (2002) for the U.S. is from 1965:Q1 to 1997:Q4, while Cayen and van Norden (2005) cover the period from 1972:Q1 to 2003:Q4 for Canada. The gap revisions results presented in Table 1 lie in the middle of those in Cayen and van Norden (2005).

to 0.61. Moreover, this improvement in the real-time gap estimates is robust to different cutoff dates or different subsample sizes. For example, we compare gap revisions from the 1990s against those from the 2000s (defining the final estimate as the vintage two years after the end of each subsample, as above), and find that the noise-to-signal ratio decreases by 71 per cent. We also compare more "extreme" cases, for example using smaller sub-periods, such as one including the financial crisis (e.g., 2007–13) with a sub-period of equal length in the 1990s marked by mild economic fluctuations (e.g., 1994–2000), and again we find that the noise-to-signal ratio decreases noticeably, by over 60 per cent. Table 4 in the appendix presents the revisions statistics for these alternative subsamples.

Following Edge and Rudd (2016), we test the statistical significance of the reduction in the signal-to-noise ratio using a regression-based approach that tests for a change in the noise-to-signal ratios between subsamples. For each subsample, this approach consists of defining a variable equal to the ratio of the squared gap revisions and the variance of the final output gap. For example, we compute $\tilde{x}_t^{1986:Q4-2000:Q1} \equiv (y_t)^2/(\sigma^2_{F,1986:Q4-2000:Q1})$, where y_t is the quarter-t gap revision and $\sigma^2_{F,1986:Q4-2000:Q1}$ is the variance of the final gap estimate over the first subsample. We then compute $\tilde{x}_t$ similarly over the second subsample (2000:Q2–2013:Q3).[10]

To assess whether the improvement in the noise-to-signal ratio is significant over the two subsamples, we test if the mean of $\tilde{x}_t$ is lower in the second subsample relative to the first. A simple way to perform this test is to stack the $\tilde{x}_t$ terms from both subsamples into a single vector and regress it on a dummy variable equal to one over the second subsample. A test of the null hypothesis that the coefficient on the dummy variable is zero against the one-sided alternative that the coefficient is negative will tell whether the observed improvement in $\tilde{x}_t$ (and the noise-to-signal ratio) is statistically significant.[11] We find that the reduction in the RMSE-based noise-to-signal ratio is significant at 1 per cent (p-value of 0.01), which points to a substantial improvement in the average staff gap revision between the first and second subsamples.[12]

[10]The mean of the $\tilde{x}_t$ variable computed over each subsample will be equal to the square of the RMSE-based noise-to-signal ratio for each subsample.

[11]As in Edge and Rudd (2016), we use heteroskedasticity and autocorrelation (HAC) robust standard errors to compute the relevant t-statistics. Table 5 in the appendix presents the regression results. See Edge and Rudd (2016) for more details about this statistical test.

[12]Note that while the size of the revisions has decreased significantly over the sample we study, there is

4 Revision Properties of Statistical Gap Estimates

A natural candidate to explain the improvement in the revision properties of the Bank of Canada staff's output gap is the decrease in GDP volatility observed in our second subsample. Indeed, the Canadian economy was quite volatile during the late 1980s and early 1990s, which might have resulted in unreliable real-time estimates of the output gap.[13] Nonetheless, while the first subsample includes these volatile episodes, the latter one contains the Great Recession following the 2008 financial crisis. As we illustrate in the previous section, the improvement in the revision properties of the staff's gap estimates is robust to excluding the volatile period from 1987–93 from the earlier subsample while keeping the Great Recession in the latter. This points toward a greater ability by the staff to estimate the output gap in real-time than before. To further assess this possibility, in this section we use real-time GDP data to examine the revision properties of purely econometric estimates of output gaps that have no added subjective judgement.[14] In particular, we estimate real-time output gaps with seven well-known methods that have been used extensively in the literature: (1) a linear trend, (2) a broken-linear trend and a (3) quadratic linear trend of log real GDP, as well as the (4) Hodrick-Prescott filter, the (5) Beveridge-Nelson decomposition, and two bandpass filters, i.e., those by (6) Baxter and King (1999) and (7) Christiano and Fitzgerald (2003).[15] For each vintage of real-time GDP, we estimate those models using data from 1973:Q1 onward. We then compute gap revisions using the same timing conventions as above, and finally we calculate the noise-to-signal ratio over our two

still a lot of uncertainty around real-time estimates. These results should not be interpreted as meaning that output gap estimates in real-time are free of uncertainty.

[13]The period from 1987–90 was characterized by persistently high inflation and tight monetary policy by the Bank of Canada, even if there were clear signs that the economy was slowing down in 1989–90. This was followed in 1991 by the introduction of the federal goods and services tax (GST) in January and the announcement of the inflation-control target regime in February. See, for example, Laidler and Robson (1993) or Thiessen (1998) for a description of macroeconomic developments in Canada during the late 1980s and early 1990s.

[14]As Edge and Rudd (2016), we note that the goal of this exercise is not to find what the best statistical detrending procedure but rather to assess whether some feature of real GDP changed such that detrending has become easier to do in real-time.

[15]The Baxter and King (1999) and Christiano and Fitzgerald (2003) filters use cutoffs of 6 and 32 quarters, and the Baxter and King (1999) filter uses an AR(4) for endpoint padding. The Beveridge-Nelson decomposition assumes an ARIMA (1,1,0) process for log real GDP. We also tried to estimate the unobserved component models of Watson (1986) and Clark (1987), but the severity of the pile-up problem prevented us from getting meaningful estimates with our sample.

sample periods to assess whether the improved reliability of the staff output gap estimates extends to the purely statistical gaps. Table 2 reports the results.

Table 2: Noise-to-Signal Ratios for Different Econometric Models of the Output Gap

	1st subsample	2nd subsample
Linear trend	0.66	0.71
Broken trend	0.77	0.67
Quadratic trend	0.83	1.02
HP filter	1.20	0.96
Beveridge-Nelson	0.54	0.44
Baxter-King	1.16	1.18
Christiano-Fitzgerald	0.73	0.70

Note: This table shows the noise-to-signal ratio of various statistical models to compute output gaps. For the calculation of output gap revisions, we use the 2015:Q4 vintage as the final output gap estimate for the 2nd subsample (2000:Q2–2013:Q3), and 2002:Q2 as the final estimate for the first subsample (1986:Q4–2000:Q1). The noise-to-signal ratio is defined the ratio of the RMSE of the revisions over the standard deviation of the final estimate.

No consistent picture emerges. All seven statistical gap models show very modest increases or decreases in their noise-to-signal ratios. For those models with falling ratios, we test the significance of the decreases using the regression-based approach described in the previous section. *None* of the decreases in the RMSE-based noise-to-signal ratio is significant at the 5 per cent level. Table 5 in the appendix presents the regression results.

This analysis shows that the large and significant decrease in the noise-to-signal ratio for the staff's output gap estimates is not matched by the purely statistical gaps. We conclude that some aspect of the staff's output gap estimation process has improved over the years and has led to an improvement in the revision properties of the output gap. It is possible that adjustments to time-varying volatility played an important role in judgemental real-time estimates of the output gap for Canada, as conjectured by Mertens (2014) for the improved reliability of real-time staff output gaps for the United States.[16] The use of soft information, like the *Business Outlook Survey*, and the development of new tools to assess potential output,[17] are possible explanations for the improved judgemental real-time estimates in the recent past relative to the earlier period.[18]

[16] See the discussion in Mertens (2014).

[17] See Pichette et al. (2015) for a description of potential output estimation techniques used at the Bank. The *Business Outlook Survey* was introduced in the autumn of 1997.

[18] An example is the more accurate staff's assessment of potential output during the Great Recession

5 Inflation Forecasting

Orphanides and van Norden (2005) argue that Phillips curve (PC) models with real-time econometric estimates of the output gap perform worse than PC models based on final estimates of the gap as well as univariate models of inflation. They conclude that real-time estimates of the output gap are not reliable for forecasting inflation in the United States. Edge and Rudd (2016) revisit the link between inflation forecasts and real-time output gaps using their Greenbook estimates by generating forecasts of inflation that are conditional on the staff's forecasts of the output gap. For the period from 1996:Q2 to 2006:Q4, they find no reduction in forecasting performance when using real-time gap estimates instead of the final estimate.[19]

In this section, we examine the usefulness of the staff output gap estimates as predictors of future inflation in Canada. For this purpose, we estimate PC models linking core inflation to the staff output gaps and use the gaps forecasted by the staff for an out-of-sample dynamic simulation of future inflation. We then contrast the forecasts provided by the PC models to other commonly used univariate inflation forecasting models. While Edge and Rudd (2016) test their inflation-forecasting models on data for only 10 years (i.e., 1996–2006), we have almost 30 years of inflation and output gap forecast data, allowing us to test the accuracy of our inflation forecasts over a much longer period. We estimate the following models for inflation forecasting:

1. *The Real-Time-Gap PC model.* We estimate PC models of the form $\pi_t = \rho_0 + \sum_{i=1}^{4} \rho_i \pi_{t-i} + \gamma x_t^{RT} + \varepsilon_t$, where x_t^{RT} represents the staff's real-time estimates of the output gap. We then use staff forecasts of the real-time gap for the out-of-sample dynamic simulations that generate the forecasts for π_{T+h}, where h is the horizon of the forecast in quarters.

(2008–09) than during the 1991–92 recession. Specifically, the staff's estimate of potential were too pessimistic in 1991, possibly because of the uncertainty period of the late 1980s and early 1990s (as discussed earlier), and potential output estimates were eventually revised up in the subsequent years. The staff's estimates of potential during the Great Recession were not revised as much seven years afterward as those from the 1991–1992 recession were.

[19] Edge and Rudd (2016) also note that omitting the output gap in the PC model causes a deterioration in forecast performance over their 10-year sample.

2. *The Final-Gap PC model.* We estimate a model of the same form as (1), $\pi_t = \rho_0 + \sum_{i=1}^{4} \rho_i \pi_{t-i} + \gamma x_t^F + \varepsilon_t$, where x_t^F is the staff's final output gap estimate. This *final estimate* of the output gap is then used for the out-of-sample dynamic simulations that generate the forecasts for π_{T+h}, where h is the horizon of the forecast in quarters.

3. *The PC with Real-time GDP Growth model.* This model substitutes the staff's real-time estimates of the gap with actual real-time GDP growth.[20] The dynamic simulations are carried out with the staff's real-time forecasts of GDP growth. Orphanides and van Norden (2005) show that inflation forecasting models with real-time GDP growth tend to outperform models with real-time output gaps.

4. *A univariate AR(4) model.* We estimate the model $\pi_t = \rho_0 + \sum_{i=1}^{4} \rho_i \pi_{t-i} + \varepsilon_t$, and iterate it forward to generate forecasts for π_{T+h}, where h is the horizon of the forecast, in quarters. This specification omits the output gap but is otherwise identical to the baseline model in (1). 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.

5. *The Atkenson-Ohanian Random Walk.* This model simply takes $\frac{1}{4}\sum_{i=1}^{4} \pi_{T-i}$ as a forecast for π_{T+h}, where 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 PC models.

6. *The Unobserved Components Stochastic Volatility (UCSV) model of Stock and Watson (2007).* Initially proposed by Stock and Watson (2007), the model decomposes inflation into trend and cycle components and uses the filtered estimates of the trend as a forecast for π_{T+h}. Let $\pi_t = \tau_t + \eta_t^T$ 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* $N(0, I)$.

[20] We also add lags of GDP growth to this specification because it could be argued that a distributed lag of GDP growth is a better substitute for the output gap.

Our historical data on the real-time output gap dates back to 1973:Q1, thus dictating the starting estimation period for all the models.[21] Also note that we use the most recent vintage of core CPI inflation because we are interested to assess how well the Bank of Canada's staff output gap predicts the inflation rate currently observed.[22] Note that the results for total inflation are qualitatively similar as those presented below. See Table 6 in the appendix for details.

Table 3 shows the RMSE for all the models for the current quarter nowcast (T0), and for the following one, two, four, and six quarters-ahead horizons. We show the results for the full sample, 1987:Q1 to 2015:Q4.[23] Several interesting results can be highlighted. First, PC models with the staff's assessment and forecasts of the output gap do not outperform simple univariate inflation forecasting models, as the Atkeson and Ohanian (2001) random walk or the Stock and Watson (2007) UCSV model. Nonetheless, in line with Edge and Rudd (2016), this inability to improve on this univariate model does not follow from the unreliability of real-time estimates of the output gap. In fact, PC models conditioned on real-time estimates of the output gap perform slightly better than the ones conditioned on the final estimates, particularly at longer horizons.[24] Second, the PC model that condition the forecast on staff's real-time GDP growth forecasts, instead of gaps, actually leads to lower forecast errors, on average, than the ones from the PC with real-time gap.[25] In contrast with output gaps, real-time output growth in the PC model does not outperform the same model with the actual output growth, as real-time output gaps do. Furthermore, Table 8 in the appendix presents PC forecasts results with distributed lags of output gap and output growth. The results are qualitatively the same but the models perform worse overall.[26] Overall, the PC models do not perform better than the simple univariate AR(4),

[21]Recall that our first data vintage is 1987:Q1, and all our vintages have data dating back to 1973:Q1.

[22]Our measure of core inflation is the Bank of Canada's CPIX, which excludes the eight most volatile components of CPI. See Macklem (2001) for details.

[23]As mentioned in Section 2, the first real-time estimate of the output gap in our data set is for 1986:Q4, and is available from the March 1987 vintage. Consequently, our first set of forecasts use this first vintage of the data to estimate a given inflation model and then forecast out-of-sample inflation for 1987:Q1, 1987:Q2, ..., 1988:Q4, and so forth for every vintage.

[24]In the appendix (Table 7), we show that this result is robust to different choices of lag length for inflation.

[25]This result is consistent with the findings of Orphanides and van Norden (2005), although they do not use forecasts of the output gap in their dynamic simulations.

[26]See Table 8 in the appendix for the inflation forecasting results with lags of output gaps and growth in the PC models.

the Atkeson and Ohanian (2001) or Stock and Watson (2007) UCSV models. These last two models perform notably better over longer horizons.

Table 3: RMSE for Core Inflation Models

	$T0$	$T1$	$T2$	$T4$	$T6$
Real-Time Gap	0.94	1.02	1.07	1.14	1.21
Final Gap	0.95	1.05	1.17	1.37	1.58
Real-Time Growth	0.91	0.97	1.05	1.11	1.23
Final Growth	0.91	0.95	1.03	1.07	1.17
AR	0.91	0.96	1.03	1.08	1.20
AO RW	0.88	0.90	0.92	0.92	0.98
UCSV	0.87	0.89	0.93	0.93	0.96

Note: This table shows the RMSE for all the CPIX inflation forecasting models over the full sample (1987:Q1–2015:Q4). T0 is the forecast for the current quarter (nowcast), while T1 is the forecast one period ahead, and so forth. The first two models are based on the staff's estimate of the real-time and final output gaps, while the third and fourth are based on real-time and final output growth, respectively. The remaining three models are the univariate models described in the text.

6 Conclusion

The output gap plays a central role in the conduct of monetary policy, and thus its measurement in real-time is of great importance. In prominent papers, some authors have argued that it has not been measured accurately in real time and consequently its use is problematic for policy purposes. However, a recent paper by Edge and Rudd (2016) argues that these problems have become less severe in more recent samples; using Greenbook data from 1996 to 2006, they show that the staff at the Federal Reserve Board produced output gap estimates with revision properties that were significantly better than those previously found in the literature. This paper uses a novel data set of real-time data and forecasts from the Bank of Canada's staff economic projections to study the properties of the output gap revisions for Canada during the period from 1987:Q1 to 2015:Q4. We show that the average size of revisions to the output gap, as well as a measure of noise-to-signal ratio, have diminished significantly in the more recent part of our sample. We find that this improvement on the reliability of the output gap estimates is not matched by judgement-free, statistical methods to estimate the output gap. These results are in line with those found by Edge and Rudd (2016) for the U.S., even if our sample includes the volatile period of the Great Recession and the subsequent recovery.

We then examine the usefulness of real-time output gap estimates for forecasting inflation in Canada. We find that Phillips curve forecasting models that condition forecasts on staff's expected path of the output gap do not outperform simple univariate models. Nonetheless, this result does not follow from the real-time unreliability of output gap estimates, as real-time estimates of the gap perform as well or better than final estimates of the gap in PC models.

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.

Barnett, R., S. Kozicki, and C. Petrinec (2009): "Parsing shocks: real-time revisions to gap and growth projections for Canada," *Federal Reserve Bank of St. Louis Review*, 91(4), 247–65.

Baxter, M. and R. G. King (1999): "Measuring business cycles: approximate band-pass filters for economic time series," *Review of Economics and Statistics*, 81(4), 575–593.

Bernhardsen, T., Ø. Eitrheim, A. S. Jore, and Ø. Røisland (2005): "Real-time data for Norway: challenges for monetary policy," *The North American Journal of Economics and Finance*, 16(3), 333–349.

Butler, L. (1996): "The Bank of Canada's new Quarterly Projection Model Part 4. A semi-structural method to estimate potential output: combining economic theory with a time-series filter," Technical Report No. 77, October.

Cayen, J.-P. and S. van Norden (2005): "The reliability of Canadian output-gap estimates," *The North American Journal of Economics and Finance*, 16(3), 373–393.

Christiano, L. J. and T. J. Fitzgerald (2003): "The band pass filter," *International Economic Review*, 44(2), 435–465.

Clark, P. K. (1987): "The cyclical component of U.S. economic activity," *The Quarterly Journal of Economics*, 102(4), 797–814.

Cusinato, R. T., A. Minella, and S. d. S. P. Júnior (2013): "Output gap in Brazil: a real-time data analysis," *Empirical Economics*, 44(3), 1113–1127.

Edge, R. M. and J. B. Rudd (2016): "Real-time properties of the Federal Reserve's output gap," *Review of Economics and Statistics*, Forthcoming.

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

Kamada, K. (2005): "Real-time estimation of the output gap in Japan and its usefulness for inflation forecasting and policymaking," *The North American Journal of Economics and Finance*, 16(3), 309–332.

Kamber, G., J. Morley, and B. Wong (2016): "Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter," *Unpublished manuscript*.

Laidler, D. E. and W. B. P. Robson (1993): *The Great Canadian Disinflation: The Economics and Politics of Monetary Policy in Canada, 1988–93*, vol. 19, CD Howe Institute.

Laxton, D. and R. Tetlow (1992): *A simple multivariate filter for the measurement of potential output*, 59, Bank of Canada Ottawa, ON, Canada.

Macklem, T. (2001): "A new measure of core inflation," *Bank of Canada Review*, Autumn, 3–12.

——— (2002): "Information and analysis for monetary policy: coming to a decision," *Bank of Canada Review*, Summer, 11–18.

Marcellino, M. and A. Musso (2011): "The reliability of real-time estimates of the euro area output gap," *Economic Modelling*, 28(4), 1842–1856.

Mertens, E. (2014): "On the reliability of output gap estimates in real time," *Unpublished manuscript, Federal Reserve Board*.

Orphanides, A. (2003): "Monetary policy evaluation with noisy information," *Journal of Monetary economics*, 50(3), 605–631.

Orphanides, A. and S. van Norden (2002): "The unreliability of output-gap estimates in real time," *Review of Economics and Statistics*, 84(4), 569–583.

——— (2005): "The reliability of inflation forecasts based on output gap estimates in real time," *Journal of Money, Credit, and Banking*, 37(3), 583–601.

Pichette, L., P. St-Amant, B. Tomlin, and K. Anoma (2015): "Measuring potential output at the Bank of Canada: the Extended Multivariate Filter and the Integrated Framework," *Bank of Canada Discussion paper*, 2015-1.

Stock, J. H. and M. W. Watson (2007): "Why has U.S. inflation become harder to forecast?" *Journal of Money, Credit and banking*, 39(s1), 3–33.

Thiessen, G. G. (1998): "The Canadian experience with targets for inflation control," *Canadian Public Policy/Analyse de Politiques*, 415–428.

Watson, M. W. (1986): "Univariate detrending methods with stochastic trends," *Journal of Monetary Economics*, 18(1), 49–75.

7 Appendix

Table 4: Statistics on Bank's Staff Output Gap Revisions Using Different Subsamples

	1990s	2000s	1993:Q3–2000:Q1	2007:Q1–2013Q3
Mean revision	1.52	0.08	1.12	0.52
Mean abs. revision	1.59	0.70	1.22	0.60
RMSE	2.07	0.83	1.67	0.80
Std. dev.	1.42	0.84	1.25	0.62
Noise-to-signal ratio	1.27	0.67	1.46	0.51
Mean of final estimate	-1.58	-0.47	-1.09	-0.57
Std. dev. of final estimate	1.62	1.25	1.14	1.57

Note: This table reports statistics on the Bank of Canada's staff output gap revisions. The first two subsamples correspond to the 1990s (1991–2000) and the 2000s (2001–2010) referred in the text. We use 2003:Q1, 2013:Q1, 2002:Q2 and 2015:Q4 as final estimates for the displayed subsamples, respectively. The noise-to-signal ratio is defined as the ratio of the RMSE of the revisions over the standard deviation of the final estimate.

Table 5: Test of Statistical Significance for Changes in Noise-to-Signal Ratios

Estimation method		Coefficient	Newey-West Std. Error	t-statistic	p-value
	dummy	-0.902	0.392	-2.30	0.01
BoC Staff	constant	1.277	0.370		
	dummy	0.072	0.135	0.54	0.71
Linear trend	constant	0.429	0.120		
	dummy	-0.146	0.161	-0.91	0.18
Broken trend	constant	0.597	0.148		
	dummy	0.337	0.282	1.20	0.88
Quadratic trend	constant	0.693	0.141		
	dummy	-0.526	0.571	-0.92	0.18
HP filter	constant	1.439	0.494		
	dummy	-0.100	0.065	-1.54	0.06
Beveridge-Nelson	constant	0.296	0.055		
	dummy	0.059	0.570	0.10	0.54
Baxter-King	constant	1.356	0.329		
	dummy	-0.053	0.275	-0.19	0.42
Christiano-Fitzgerald	constant	0.538	0.192		

Note: This table shows regression results for the significance test in the decrease in the noise-to-signal ratio of output gap revisions from the staff and different econometric models. See text for details. N = 108.

Table 6: RMSE for Total Inflation Models

	$T0$	$T1$	$T2$	$T4$	$T6$
Real-Time Gap	2.05	2.26	2.28	2.22	2.41
Final Gap	2.07	2.29	2.32	2.22	2.34
Real-Time Growth	2.07	2.30	2.32	2.23	2.38
Final Growth	2.07	2.31	2.35	2.27	2.43
AR	2.07	2.29	2.33	2.27	2.44
AO RW	2.12	2.20	2.14	2.11	2.22
UCSV	2.03	2.23	2.26	2.18	2.23

Note: This table shows the RMSE for all the CPI inflation forecasting models over the full sample (1987:Q1–2015:Q4). T0 is the forecast for the current quarter (nowcast), while T1 is the forecast one period ahead, and so forth. The first two models are based on the staff's estimate of the real-time and final output gaps, while the third and fourth are based on real-time and final output growth, respectively. The remaining three models are the univariate models described in the text.

Table 7: AR and Phillips Curve Forecasts with Different Inflation Lags

Lags	Model	$T0$	$T1$	$T2$	$T4$	$T6$
	AR	0.91	0.97	1.04	1.10	1.22
3	Real-Time Gap	0.95	1.05	1.11	1.20	1.30
	Final Gap	0.96	1.07	1.21	1.44	1.69
	AR	0.91	0.96	1.03	1.08	1.20
4	Real-Time Gap	0.94	1.02	1.07	1.14	1.21
	Final Gap	0.95	1.05	1.17	1.37	1.58
	AR	0.86	0.88	0.93	0.96	1.01
5	Real-Time Gap	0.89	0.95	0.98	0.98	1.00
	Final Gap	0.91	0.99	1.09	1.22	1.40
	AR	0.93	0.97	1.04	1.11	1.20
6	Real-Time Gap	0.93	1.00	1.04	1.08	1.12
	Final Gap	0.94	1.02	1.13	1.28	1.47

Note: This table shows the RMSE of CPIX inflation forecasting over the full sample (1987:Q1–2015:Q4) with different choices of lags for inflation. T0 is the forecast for the current quarter (nowcast), while T1 is the forecast one period ahead, and so forth. The first model is the simple AR() while the last two models are the PC models based on the staff's estimate of the real-time and final output gaps, respectively.

Table 8: Phillips Curve Forecasts with Different Output Gap or Output Growth Lags

Lags	Model	T0	T1	T2	T4	T6
	RT Gap	0.94	1.02	1.07	1.14	1.21
0	Final Gap	0.95	1.05	1.17	1.37	1.58
	RT GDP growth	0.91	0.97	1.05	1.11	1.23
	Final GDP growth	0.91	0.95	1.02	1.07	1.17
	RT Gap	0.98	1.09	1.18	1.30	1.36
2	Final Gap	1.00	1.08	1.22	1.48	1.74
	RT GDP growth	0.91	0.97	1.04	1.12	1.25
	Final GDP growth	0.92	0.96	1.03	1.08	1.17
	RT Gap	1.04	1.20	1.30	1.50	1.56
4	Final Gap	1.08	1.21	1.40	1.78	2.14
	RT GDP growth	0.93	0.98	1.02	1.07	1.05
	Final GDP growth	1.00	1.05	1.12	1.19	1.31

Note: This table shows the RMSE of CPIX inflation forecasting over the full sample (1987:Q1–2015:Q4) using the same PC model as in the text, but adding lags of output gaps or output growth in the model. This exercise is done for both the real-time and final estimates of the output gap and output growth. Lags = 0 corresponds to the baseline model in the text. T0 is the forecast for the current quarter (nowcast), while T1 is the forecast one period ahead, and so forth.