

The Reliability of Output Gap Estimates in Canada

Jean-Philippe Cayen¹ and Simon van Norden²

¹ Bank of Canada
jcayen@bankofcanada.ca

² HEC (Montréal) and CIRANO
simon.van-norden@hec.ca

In this paper, we measure, with Canadian data, the scope of the revisions to real-time estimates of the output gap generated with several univariate and multivariate technics. We also make an empirical evaluation of the usefulness of the output gap estimates for predicting inflation. Our findings suggest that, for all technics, the standard deviation of the revisions is of the same order of magnitude as the output gap itself. We also find that, with exception of the Beveridge-Nelson technic, all revisions are very persistent, which means that there is a long lag before the scope of the output gap revisions is fully known. Finally, we found that the output gap estimates do not significantly improve the inflation forecasts. We infer from these results that estimates of the output gap are not very reliable.

* The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

Contents

1. Introduction	1
2. Output Gap and Estimation Errors	2
2.1 <i>Definition</i>	3
2.2 <i>Estimation techniques</i>	3
2.2.1 Deterministic Trends	4
2.2.2 The Beveridge-Nelson decomposition	5
2.2.3 Hodrick-Prescott filter	6
2.2.4 Structural VAR with long-run restrictions	7
2.2.5 Unobserved component models	8
2.3 <i>Factors explaining real-time estimations errors</i>	10
2.4 <i>Studies that measured errors in output gap estimates</i>	11
3. Evaluation Methods	12
3.1 <i>Data</i>	12
3.1.1 Data revisions	13
3.1.2 Other variables	14
3.2 <i>First evaluation method: revisions of output gap estimates</i>	14
3.2.1 Final estimates	15
3.2.2 Real-time estimates	15
3.2.3 Quasi-real estimates	16
3.2.4 Quasi-final estimates	16
3.2.5 Reliability indicators used to evaluate the revisions	17
3.3 <i>Second evaluation method: inflation forecasts</i>	17
3.3.1 Procedure used to forecast inflation	18
3.3.2 Criteria for selecting the number of lags	19
3.3.3 Evaluation of forecasts	20
4. Results	21
4.1 <i>Output Gap estimates</i>	21
4.1.1 First evaluation method: revisions of output gap estimates	29
4.1.2 Reliability indicators	31
4.1.3 Decomposition of revisions	33
4.2 <i>Second evaluation method: inflation forecasts</i>	39
5. Conclusion	45
Appendix	47
References	49

1. Introduction

The output gap, defined as the discrepancy between actual output and potential output, is an important variable of models used by monetary authorities to project and analyze the main macroeconomic variables. For countries like Canada, committed to maintaining a low and stable rate of inflation, the output gap is considered to be a useful tool for gauging the extent of inflationary pressures present in goods and services markets in the economy.

Unfortunately, the output gap is not a directly observable measure. Many techniques have been developed to estimate it, but these estimations are subject to considerable uncertainty, particularly at the end of the sample. Moreover, different techniques often produce different estimates of the output gap.

Many authors evaluate how monetary policy decisions can be affected by the uncertainty that surrounds output gap estimates.¹ Research on this topic usually involves simulating of a macroeconomic model in which the authors evaluate the effect that the uncertainty around the output gap estimates has on monetary policy rules such as the Taylor (1993) rule or inflation-forecast-based rules.²

To conduct these simulations, the authors must beforehand make some important assumptions on the output gap. They must first decide how to model the output gap and specify how it is linked with the other variables of the model. They must also formulate assumptions on the nature and the magnitude of the errors that affect the output gap estimates. But the validity of these assumptions is not always tested.³ Indeed, many authors recognize that their results depend on the assumptions they made on the specification of the model and on the type and the level of uncertainty they introduce in their model.

For the monetary authorities, it is important that the conclusions that emanate from these studies are reliable. Therefore, the assumptions introduced in these simulations must be tested. In particular, it is important to know whether there are some techniques that produce better real-time estimates of the output gap than others; that is, techniques that produce reliable end-of-sample estimates. It is also important to understand the nature and the magnitude of the errors that affect

¹ See, for example, Orphanides (1998, 2000), Orphanides et al. (2000), Smets (1998), Ehrmann and Smets (2001), Rudebusch (1999), Swanson (2000), Svensson and Woodford (2000), Gaiduch and Hunt (2000), Yetman (2000), and McCallum (2001).

² See Haldane and Batini (1998) and Amano et al. (1999) for an overview of inflation-forecast-based rules.

³ Gaiduch and Hunt (2000) is one of the rare studies that tests these assumptions.

the output gap estimates. Finally, policymakers need to know whether there is a stable predictive relationship between the real-time estimates of the output gap and inflation.

The goal of this study is to answer these questions. We measure, with Canadian data, the reliability of different univariate and multivariate techniques of estimating the output gap. In particular, we investigate the revisions to real-time estimates of the output gap over time. We also test whether the addition of output gap estimates in a simple equation helps improve inflation forecasts.

Orphanides and van Norden (1999, 2001a, b) applied the same methodology on U.S. data and found that, in general, the reliability of output gap estimates is quite low. For all the techniques they tested, the magnitude of the revisions is similar to the size of the gap estimates themselves. They also found that output gap estimates do not improve out-of-sample inflation forecasts.

Our results for Canada are similar to those of Orphanides and van Norden. We also find that the reliability of the real-time estimates of the output gap tends to be quite low. In fact, for all the techniques that we tested, the ex post revisions are of the same order of magnitude as the ex post estimates of the output gap, the real-time estimates frequently misclassify the sign of the gap, and the estimation errors appear to contain a highly persistent component that is substantial in size. Also, the output gap estimates generated from the different techniques, except for the Beveridge-Nelson decomposition, do not improve out-of-sample inflation forecasts.

This paper is organized as follow. Section 2 briefly describes the different techniques of estimating the output gap, as well as the different types of errors that can affect these estimates. Section 3 describes the methodology used to evaluate the reliability of the different estimation techniques. Section 4 presents the output gap estimates generated by the different estimation techniques and the reliability of these estimates based on two evaluation methods. Section 5 offers some conclusions.

2. Output Gap and Estimation Errors

This section describes the different approaches of estimating the output gap, as well as the errors that can affect the estimates. In section 2.1, we define an important concept for the analysis of output gap: potential output. In section 2.2, we describe the different estimation techniques studied in this paper. In section 2.3, we discuss the different sources of errors that may affect the output gap estimates. Section 2.4 presents studies that evaluate errors that surround output gap estimates.

2.1 Definition

The output gap represents the difference between the observed level of output and potential output. Thus, the key concept that must be defined is potential output. Laxton and Tetlow (1992) note that the definition of potential output has changed over time. In the 1960s and the early 1970s, potential output was considered as the maximum level of output that the economy can generate. Therefore, the analysis of business cycles consisted of identifying the cyclical peaks and explaining the factors that caused the economy to approach or move away from the cyclical peaks. Under such a definition, the output gap is always negative.

At the end of the 1960s and early 1970s, potential output started to be defined as the maximum level of output that the economy can sustain without creating inflationary pressures. This definition brings out the relationship between excess demand and inflation. As the observed level of output increases relative to its potential level, excess demand increases, which encourages economic agents to increase the prices of the goods and services that they supply. This definition of potential output is still the one generally used today. It is also the one that we use for this paper.

2.2 Estimation techniques

Claus, Conway, and Scott (2000) identify three approaches for estimating the output gap. The first approach consists of surveying businesses on their production capacity. The survey makes it possible to construct measures of capacity utilization, giving an overview of potential output when paired off with production data. However, this approach is plagued by some important problems. First, it is not clear that all the surveyed businesses have the same interpretation of the survey's questions. In particular, it is typically hard to incorporate the notion of work intensity in questions that refer to production capacity. Second, surveys often cover only a small portion of the economy, since they usually target the manufacturing sector. Therefore, we do not evaluate this approach in this paper.

A second approach consists of measuring a production function, which reflects the relationships between the observed level of output and the amount of resources used in the production process. This relationship can be used to calculate the level of output that would be produced if resources were fully employed and used at normal intensity; that is, at a level of intensity that would not create inflationary pressures. This approach also has important problems. First, it is very difficult to determine a normal level of intensity. Second, the exact form of the production function is not clear. Finally, an important component of the production function is technical progress. Just like

potential output, technical progress is not observable and is therefore hard to measure. For these reasons, we do not evaluate this approach in this paper.

The third approach, which is used in this paper, incorporates all the times-series techniques of decomposing output into two elements: potential output and the output gap. The number of existing techniques is really large, and so we limit ourselves to a number of them. Because we want to evaluate both univariate and multivariate techniques, we choose nine techniques that can be regrouped into the following five categories:

- (i) deterministic trends
- (ii) Beveridge-Nelson decomposition
- (iii) Hodrick-Prescott filter
- (iv) structural VAR with long-run restrictions
- (v) unobserved component models

Within each category, we can retain many specifications. We decide to keep specifications that have been used in other studies. Thus, for the deterministic trend models, we examine the linear trend and the quadratic trend, while for unobserved component models, we study the Watson (1986), Clark (1987), Harvey-Jaeger (1993) and Kichian (1999) models. The Beveridge-Nelson model could also be considered as an unobserved component model, but since it is estimated differently than the other unobserved component models, we separate it from them. For the structural VAR model, we use the specification proposed by Lalonde, Page, and St-Amant (1998).

Other estimation techniques, such as the multivariate filter used at the Bank of Canada (see Butler 1996), will probably be evaluated in future work. The rest of section 2.2 describes the different techniques that we have chosen.⁴

2.2.1 Deterministic Trends

The first two univariate detrending methods we consider are the linear and the quadratic trend. Both methods assume that we can decompose output into two components: a deterministic trend and a cycle component, which corresponds to the output gap. The general form of deterministic trends is:

$$y_t = \alpha + \sum_{i=1}^I \beta_i t^i + c_t, \quad (1)$$

⁴ Laxton and Tetlow (1992), Butler (1996), St-Amant and van Norden (1998), and Cerra and Saxena (2000) give a more complete picture of the existing techniques and of their advantages and disadvantages.

where y_t is our chosen measure of output (in logarithms), α is a constant, t is a time trend and c_t is the output gap. When I is equal to 1, equation (1) corresponds to the linear trend. When it is equal to 2, it represents the quadratic trend. To obtain an estimate of the output gap, one only has to do an ordinary least-squares estimation of equation (1). The estimated residuals correspond to the output gap series.

2.2.2 The Beveridge-Nelson decomposition

Beveridge and Nelson (1981) consider the case of an $ARIMA(p,1,q)$ series y_t , which is to be decomposed into a trend and a cyclical component. For simplicity, we can assume that all deterministic components belong to the trend component and have already been removed from the series. Since the first-difference of the series is stationary, it has an infinite-order moving average representation of the form

$$\Delta y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots = e_t, \quad (2)$$

where ε_t is assumed to be an innovations sequence. The change in the series over the next s periods is simply:

$$y_{t+s} - y_t = \sum_{j=1}^s y_{t+j} = \sum_{j=1}^s e_{t+j}. \quad (3)$$

The trend is defined to be:

$$\lim_{s \rightarrow \infty} E_t(y_{t+s}) = y_t + \lim_{s \rightarrow \infty} E_t \left(\sum_{j=1}^s e_{t+j} \right). \quad (4)$$

From equation (2), we can see that

$$E_t(e_{t+j}) = E_t(\varepsilon_{t+j} + \beta_1 \varepsilon_{t+j-1} + \beta_2 \varepsilon_{t+j-2} + \dots) = \sum_{i=0}^{\infty} \beta_{j+i} \varepsilon_{t-i}. \quad (5)$$

Since changes in the trend are therefore unforecastable, this has the effect of decomposing the series into a random walk and a cyclical component, so that

$$y_t = \tau_t + c_t, \quad (6)$$

where the trend is

$$\tau_t = \tau_{t-1} + e_t, \quad (7)$$

and e_t is white noise.

To use the Beveridge-Nelson decomposition, we must therefore:

- (i) identify p and q in our ARIMA($p,1,q$) model
- (ii) estimate the parameters in equation (2)
- (iii) choose some large-enough but finite value of s to approximate the limit in equation (4)⁵
- (iv) for all t and for $j = 1, 2, \dots, s$, calculate $E_t(e_{t+j})$, from equation (5)
- (v) calculate the trend at time t as in the right-end side of equation (4), and the cycle as the difference between observed output and the trend.

For the specification of the ARIMA($p,1,q$) model of output, we have not found any papers that study, for Canada, the optimal values for p and q . We therefore follow Box and Jenkins' (1976) recommendation of using a parsimonious model. We apply the Schwarz (1978) criterion⁶ on the full sample and find that Canadian output is following an ARIMA(1,1,0) process.

2.2.3 Hodrick-Prescott filter

Another mechanical detrending technique is the HP filter, developed by Hodrick and Prescott (1997). The HP filter decomposes a time series (y_t) into an additive cyclical component (c_t) and a growth component (τ_t), such as $y_t = c_t - \tau_t$, and then chooses the series τ_t to minimize the variance of the cyclical component subject to a penalty for the variation in the second difference of the growth τ_t . Formally, the HP-filtered trend is given by:

$$\{\tau_t\}_{t=0}^{T+1} = \arg \min \sum_{t=1}^T \left\{ (y_t - \tau_t)^2 + \lambda [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right\}. \quad (8)$$

and c_t is the resulting measure of the output gap. λ is called the “smoothness parameter” and penalizes the variability in the growth component. The larger the value of λ , the smoother the growth component and the greater the variability of the output gap. As λ approaches infinity, the growth component corresponds to a linear time trend. For quarterly data, Hodrick and Prescott propose setting λ equal to 1600.

⁵ Beveridge and Nelson (1981) use $s = 100$. We also use that value even if the forecasts converge before s equal to 100.

⁶ This criterion is described in section 3.3.2.

2.2.4 Structural VAR with long-run restrictions

Another way of estimating the output gap is to impose long-run restrictions on the results obtained for a multivariate autoregressive model (VAR), to recover the structural representation from which we can deduct a measure of the output gap.

Blanchard and Quah (1989) were the first to recover the structural representation of a model from a reduced-form VAR, based on a priori assumptions on the long-run relationship between supply and demand. They made the simple assumption that demand shocks have only temporary effects on real output, while supply shocks may have permanent effects. Dupasquier, Guay, and St-Amant (1997), Lalonde (1998), and Lalonde, Page, and St-Amant (1998) use the Blanchard and Quah assumptions to recover estimates of the output gap.⁷ Since Lalonde, Page, and St-Amant (1998) (hereafter LPS) apply this methodology to Canadian data, we use their specification to estimate the output gap.⁸

As with LPS, we estimate the reduced form of an eight-lags VAR, which incorporates the following variables: the first differences of (the logarithms of) real output (Δy_t), of the inflation rate ($\Delta \pi_t$), and of the real interest rate (Δr_t):

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} \mu_t \\ \mu_t \\ \mu_t \end{bmatrix} + \begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) & \phi_{13}(L) \\ \phi_{21}(L) & \phi_{22}(L) & \phi_{23}(L) \\ \phi_{31}(L) & \phi_{32}(L) & \phi_{33}(L) \end{bmatrix} \cdot \begin{bmatrix} \Delta y_{t-1} \\ \Delta \pi_{t-1} \\ \Delta r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{r,t} \end{bmatrix}. \quad (9)$$

The three series of residuals ($\varepsilon_{y,t}$, $\varepsilon_{\pi,t}$, and $\varepsilon_{r,t}$) estimated with this reduced-form model include at the same time, both permanent and transitory shocks that affect real output. To isolate the structural shocks, LPS impose restrictions similar to the one proposed by Blanchard and Quah (1989): that a first group of shocks has permanent effects on all variables (η^p); a second group has transitory effects on output, but permanent effects on inflation (η^{cp}); and a third group has transitory effects on both inflation and output (η^c).⁹ Since these restrictions are made only on the long-run dynamic of the model, the short-run movements are not constrained.

Once the structural shocks are identified, it is possible to decompose output into the three shocks, as follows:

⁷ These authors replace the terms supply and demand shocks by permanent and transitory shocks.

⁸ Their study presents a more detail discussion of the chosen specification. Different results could be obtained with a different specification, such as the one proposed by Lalonde (1998) for the U.S. for example.

⁹ Other technical restrictions must be imposed to identify the different structural shocks. These are described in the appendix.

$$\Delta y_t = \mu_y + \Gamma_y^p(1)\eta_t^p + \Gamma_y^{p*}(L)\eta_t^p + \Gamma_y^{cp}(L)\eta_t^{cp} + \Gamma_y^c(L)\eta_t^c, \quad (10)$$

where μ_y is the deterministic component on the trend, and Γ_y^i is the moving average representation of the different shocks. In their study, LPS explain that the term $\Gamma_y^p(1)$ is the long-run multiplier of permanent shocks, while $\Gamma_y^{p*}(L) = \Gamma_y^p(L) - \Gamma_y^p(1)$ represents the transitory components of the permanent shocks. Thus, potential output, which is represented by $\mu_y + \Gamma_y^p(1) \cdot \eta_t^p + \Gamma_y^{p*}(L) \cdot \eta_t^p$, can fluctuate over time.

LPS also explain that two measures of the output gap can be extracted from this specification. The first measure, RLTP, correspond to the term $\Gamma_y^{cp}(L)\eta_t^{cp} + \Gamma_y^c(L)\eta_t^c$, the sum of the transitory components that affect real output. The second measure, RLTP1, corresponds to the term $\Gamma_y^{cp}(L)\eta_t^{cp}$. Thus, RLTP1 exclude from the output gap transitory movements that are not due to changes in the trend of inflation. LPS find that the output gap measured by the RLTP1 method is a good (in-sample) predictor of inflation.

2.2.5 Unobserved component models

The last type of models considered are the unobserved component models, which are also known as dynamic factor models and state-space models. In this category, we evaluate three univariate models, the Watson (1986), the Clark (1987), and the Harvey-Jaeger (1993) models, and one multivariate model, the Kichian (1999) model.

Unobserved component models can all be represented by the same general framework, which includes the measurement equation (11) and the transition equation (12):

$$y_t = Z\alpha_t + \beta X_t + \varepsilon_t \quad (11)$$

$$\alpha_t = T\alpha_{t-1} + \delta W_t + U_t. \quad (12)$$

where α_t is an M-dimensional vector of unobserved “state variables”; y_t is an N-dimensional vector of observed variables; X_t and W_t are K- and S-dimensional vectors of observable exogenous variables; Z and T are matrices of coefficients; and ε_t and U_t are G- and M-dimensional vectors of i.i.d. gaussian errors, which are mutually independent with variance-covariance matrices H_t and Q_t respectively.

What differentiate the different models described in this section are the restrictions made to the measurement and transition equations, as well as the dynamics chosen to describe the movements of the unobserved variables.

The Watson (1986) model assumes that output can be decomposed into two elements: a trend component (τ_t) and the output gap (c_t). In this model, the measurement equation (equation 13) is therefore an identity equation. For the transition equations, the Watson model assumes that the trend component follows a random walk with drift (equation 14), while the output gap follows an AR(2) process (equation 15) to allow some persistence in the business cycle. The model can be written as follows:

$$y_t = \tau_t + c_t \quad (13)$$

$$\tau_t = \mu + \tau_{t-1} + \eta_t \quad (14)$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t. \quad (15)$$

The Clark (1987) model is similar to the Watson model, except that the drift term μ is allowed to fluctuate over time (equation 17). This variable also follows a random walk (equation 18). This represents a major difference, since the Watson model assumes that potential output increases on average at a constant rate, which is not the case for the Clark model. The Clark model can be formalized as follows:

$$y_t = \tau_t + c_t \quad (16)$$

$$\tau_t = \mu_t + \tau_{t-1} + \eta_t \quad (17)$$

$$\mu_t = \mu_{t-1} + v_t \quad (18)$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t. \quad (19)$$

The Harvey and Jaeger (1993) model is different than the Watson and Clark models in that it includes a residual term in the output equation. The measurement equation (equation 20) is therefore not an identity equation anymore. Another major difference in the Harvey and Jaeger model is that the AR(2) cycle is replaced by a sinusoidal stochastic process (equations 23 and 24). The trend component is, however, the same as in the Clark model. The model can be written as follows:

$$y_t = \tau_t + \psi_t + \varepsilon_t \quad (20)$$

$$\tau_t = \mu_t + \tau_{t-1} + \eta_t \quad (21)$$

$$\mu_t = \mu_{t-1} + v_t \quad (22)$$

$$\psi_t = \rho \cdot \cos(\lambda \psi_{t-1}) + \rho \cdot \sin(\lambda \psi_{t-1}^*) + \chi_t \quad (23)$$

$$\psi_{t-1}^* = -\rho \cdot \sin(\lambda \psi_{t-1}) + \rho \cdot \cos(\lambda \psi_{t-1}^*) + \chi_{t-1}^*. \quad (24)$$

where ρ and λ are respectively the factors that determine the amplitude and the frequency of the cycle (ψ_t).

The last unobserved component model that we evaluate is the Kichian (1999) model. This model is a modified version of the Gerlach and Smets (1997), adapted for the Canadian economy. This model has the same specification than the Clark model, except that it adds a Phillips curve that links the output gap with inflation (equation 29). This model can be written as follow:¹⁰

$$y_t = \tau_t + c_t \quad (25)$$

$$\tau_t = \mu_t + \tau_{t-1} + \eta_t \quad (26)$$

$$\mu_t = \mu_{t-1} + v_t \quad (27)$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t \quad (28)$$

$$\Delta\pi_t = \mu_\pi + \beta_0 c_t + \beta_1 c_{t-1} + \gamma(L)\omega_t + \delta(L)\varepsilon_t^\pi \quad (29)$$

where $\Delta\pi_t$ is the first difference of the inflation rate (calculate with CPI excluding food and energy), μ_p is a constant, ε_t^p are i.i.d. gaussian errors, which are uncorrelated with the other errors terms, and ω_t is a set of exogenous variables.¹¹ Kichian also introduces a moving average process MA(3) in the Phillips curve.

Since these models contain variables that are not observed, the estimation of the parameters is not trivial. In fact, the estimation of such model is done in two steps. The first one consists in constructing the likelihood function of each parameter of the model, that is, all the coefficients and the variance of the residuals; and to estimate these parameters with the maximum likelihood method. In the second step, we estimate the unobserved variables with the Kalman filter.¹²

2.3 Factors explaining real-time estimations errors

One of the reasons that explain the presence of many different techniques of estimating the output gap is that the uncertainty around the estimates is quite large. It is therefore hard to prove that one technique gives better estimates than the others. Gaiduch and Hunt (2000) distinguish three sources of errors that can affect output gap estimates and which are responsible for the large uncertainty around the output gap estimates.

¹⁰ In her study, Kichian (1999) tests different variants of the Gerlach and Smets model. She notes that the one that has the best fit is assuming a stochastic trend with a constant drift, but with a break in the drift term in 1976. However, since it is hard to detect such breaks in real time, we decide to use another variant proposed by Kichian in which the trend follows a similar process as in the Clark model.

¹¹ For the exogenous variables, Kichian (1999) includes the log changes in the real exchange rate and in the real oil prices (WTI), as well as changes in indirect taxes.

¹² The Kalman filter is a technique that recursively estimate unobserved variables ($a_{i,t}$) by optimising their mean $a_{i,t} = E_t(a_{i,t})$ and their variance-covariance matrix $P_{i,t} = E_t((a_{i,t} - a_{i,t}) (a_{i,t} - a_{i,t})')$. Harvey (1989) gives more detail on this estimation method. It should be noted that for the Kichian (1999) model, we use the GAUSS programs that she developed at the Bank of Canada (see Kichian 2000).

The first source of errors concerns the statistical uncertainty around the estimation of the parameters of the model used to estimate the output gap. This form of uncertainty is generally illustrated by the confidence intervals around the estimates of the output gap. Large confidence intervals mean that this type of error is potentially big. However, some estimation techniques, such as the Hodrick-Prescott filter, do not necessitate the estimation of any parameter. It is therefore not possible to measure this type of errors for such techniques. For this reason, and because it is generally well documented in the literature, we do not consider this form of error in our evaluation of the reliability of output gap estimates.

The second source includes errors due to the utilization of an inadequate model. For example, the omission of important variables or the utilization of incorrect assumptions in a model can cause significant errors. The main difficulty with this type of errors is that we do not perfectly know the structure of the model that explains potential output. It is therefore impossible to directly measure this type of errors.

The third source of errors identified by Gaiduch and Hunt concerns the revisions made to the real-time estimates of the output gap over time. As new data become available, analysts change their output gap estimates for the recent periods (and sometimes for longer periods). These ex post revisions are considered as errors made in the past. This form of error is closely linked with the two other types of errors described before, since these revisions can be caused by the utilization of an inadequate model or can simply reflect the normal statistical uncertainty around parameters' estimates. Data revisions can also be the cause of the revisions to the real-time estimates of the output gap.¹³

Only the third type of errors, that is, revisions to the real-time estimates of the output gap over time, is directly evaluated in this paper. However, we will also show that errors caused by the utilization of an inadequate model can potentially be large.

2.4 Studies that measured errors in output gap estimates

Even if we do not evaluate this type of errors in this paper, it is interesting to know that the statistical uncertainty around output gap estimates can be quite large. Laxton and Tetlow (1992), Kuttner (1994), and St-Amant and van Norden (1997), among others, show that the confidence intervals around the output gap estimates are quite big for multivariate techniques of estimation.

¹³ Section 3.1.2 describes the different revisions made to the data in Canada.

For the errors caused by the choice of an inadequate model, we have not find any study able to precisely evaluate them. Besides, it would almost be impossible to conduct such evaluation, since nobody knows the exact structure of the economy. Gaiduch and Hunt (2000), however, find that this type of errors can potentially be large. They observe that the differences in output gap estimates across techniques are relatively large. If we assume that one of these models is the right one, it automatically means that the other techniques are relatively bad at reproducing the “real” output gap.

We find six studies that evaluate errors link to revisions to the output gap estimates. Using Canadian data, St-Amant and van Norden (1997) show that the differences between the end-of-sample estimates and the middle-of-sample output gap estimates generated by the Hodrick-Prescott filter are rather large. Using U.S. data, Kuttner (1994) and Orphanides and van Norden (1999, 2001b) find similar results for a broader set of techniques. In fact, Orphanides and van Norden (1999, 2001b) note that the revisions are persistent and of the same order of magnitude than the estimates themselves. Gaiduch and Hunt (2000) find similar results for New Zealand. Finally, Orphanides (2000) shows that the official estimates of the output gap used by the Federal Reserve Board from the mid 1960s until the mid 1990s have also been subject to substantial revisions.

3. Evaluation Methods

We use two different methods to evaluate the reliability of real-time output gap estimates. The first one investigates the revisions of real-time estimates of the output gap over time. The second one tests whether the addition of real-time output gap estimates in a simple equation help improve inflation forecasts.

To conduct these tests, we need real-time data on Canadian output. Section 3.1 describes the data set that we use. Sections 3.2 and 3.3 describe, respectively, the two evaluation methods proposed in this paper.

3.1 Data

Real-time output data are necessary to generate real-time estimates of the output gap. Since Statistics Canada does not have such a database, we build our own one. To do this, we use the data published each quarter in the National Income and Expenditure Accounts of Statistics Canada. We also consult special catalogues when major methodological changes are made to the series.

Certain points should be noted. First, our database covers each vintage between 1972Q2 and 2001Q1.¹⁴ Statistics Canada started to publish quarterly output data in 1961. Unfortunately, the series was completely revised at the end of the 1960s and we have not found any document showing the new series. The first vintage for which we have reliable data on the entire historical period is 1972Q2.

Second, before 1997, the first observation for each vintage was 1947Q1. However, Statistics Canada made a major revision of the national accounts in 1997 and thereafter stopped publishing the data prior to 1961.¹⁵ Therefore, for the 1997Q3 vintage and the following ones, the first observation of the series is 1961Q1. For this study, we use only data after 1961Q1 for three reasons. First, quarterly output data prior to 1961 are not reliable. Second, to evaluate the impact of adding new observations in the estimations, it is preferable that all vintages start at the same date. Third, Lalonde, Page, and St-Amant (1998) and Kichian (1999) estimate their models with a data set that starts after 1961. To obtain results that coincide with theirs, we use the same sample period they use.

Third, we have to ignore a certain number of available vintages. Multivariate techniques necessitate a large number of observations, because of the number of coefficients that have to be estimated. For some of the oldest vintages, the number of observations is not large enough to make the estimation converge.¹⁶ To be able to compare the different techniques, we exclude the oldest vintage. Therefore, the first vintage for which we show results is 1982Q3.

3.1.1 Data revisions

The output data that we use are seasonally adjusted and expressed in constant dollars. These two characteristics necessarily imply that the output series will be revised periodically. In fact, Statistics Canada revises seasonal factors each year. This form of revision usually affects the past four years of data. Revisions that result from the change of the year of reference happen less often. In the past 30 years, the reference year has changed only six times.¹⁷ In the latest change of the

¹⁴ The release of output data is always done with a one-quarter lag. For example, data for the fourth quarter of 2000 were released in the first quarter of 2001. To avoid confusion in the text, we will always refer to the vintage of publication. Thus, the last observation of the 1972Q2 vintage is the first quarter of 1972.

¹⁵ The observations prior to 1961 were interpolated from annual data. Given the poor reliability of these data, Statistics Canada decided not to publish them anymore.

¹⁶ This is the case for the estimation of the Lalonde, Page, and St-Amant (1998) method using data of the 1982Q2 vintage.

¹⁷ Before the 1975Q2 vintage, real output was expressed in 1961 dollars. Between 1975Q3 and 1986Q1, it was expressed in 1971 dollars; between 1986Q2 and 1990Q1, in 1981 dollars; between 1990Q2 and 1997Q3, in 1986 dollars; between 1997Q4 and 2001Q1, in 1992 dollars; and since 2001Q2, the reference year has been 1997.

reference year, Statistics Canada also modified its methodology to calculate output at a constant rate. Since 2001Q2, Statistics Canada has used the Fisher chained index, whereas before it was using a Laspeyres index.¹⁸

The concept of real output has also evolved over time. In Canada, until 1986Q1, the benchmark series was the gross national product (GNP), and since then it has been the gross domestic product (GDP). This modification is another form of revision of the data. Other major changes have been made to the output series over time, the most recent case being in 1997, when Statistics Canada modified the definition of certain components of output to comply with international standards. This modification caused a major revision of the entire output series.

Each quarter, Statistics Canada revises the data for the current year as it receives new information from its various sources.

3.1.2 Other variables

In addition to the output series, other variables are required to estimate the multivariate methods. The structural VAR model of Lalonde, Page, and St-Amant (1998) necessitates the utilization of the total consumer price index and the modified overnight series proposed by Armour et al. (1996). Kichian's (1999) model also incorporates the total consumer price index; the consumer price index excluding food and energy; the price of West Texas Intermediate; the U.S. consumer price index; the bilateral exchange rate between Canada and the U.S.; and indirect taxes.

Some of these series can be revised, as can the different price indexes, which are seasonally adjusted. However, we do not have any real-time data for these series. We therefore use the most recent vintage available for each of these series. This means that we will underestimate the impact of data revisions for the multivariate techniques. The magnitude of revisions made to price indexes, however, is typically small relative to the revisions made to the output series.

3.2 First evaluation method: revisions of output gap estimates

The first evaluation method consists of measuring the degree to which estimates of the output gap at any point in time vary as data are revised and as data about subsequent evolution of output becomes available. Presuming that revisions “improve” our estimates, the total amount of revision gives us a lower bound on the measurement errors thought to be associated with the real-time

¹⁸ Because our last vintage is 2001Q1, our results do not capture this recent change. Normally, we would expect that this modification would reduce the magnitude of revisions to the data, since changes in the reference year should no longer affect the growth rate of real variables.

output gap.¹⁹ This is informative when and if we find that revision errors are relatively large, since we can conclude that the total error of these estimators must be larger still.

We apply each detrending method in a different number of ways, to estimate and decompose the extent of the revisions in the estimated gap series. To understand how the extent of the revisions is measured, we define several conceptually different ways in which any existing detrending method may be applied. In the remainder of this section, we describe how these methods are applied and their corresponding interpretations. We also present reliability indicators that are used to evaluate the revisions.

3.2.1 Final estimates

The first of these methods gives rise to a “final” estimate of the output gap. This simply takes the last available vintage of data we have available (in our case, the series as published in 2001Q1) and detrends it. The resulting series of deviations from trend constitutes the final estimate of the output gap. This is the typical way in which such detrending methods are employed.

3.2.2 Real-time estimates

The “real-time” estimate of the output gap is constructed in two stages. First, we detrend each and every vintage of data available to construct an ensemble of output gap series.²⁰ Of course, earlier vintage output gap series are shorter than later vintages, since the output series on which they are based end earlier. Next, we use these different vintages to construct a new series that consists entirely of the first available estimate of the output gap for each point in time.

This new series is the “real-time” estimate of the output gap. It represents the most timely estimate of the output gap that policy-makers could have constructed at any point in time. The difference between the real-time and the final estimate gives us the total revision in the estimated output gap at each point in time. We use the statistical properties of these revisions as our guide to reliability and accuracy of estimated output gaps, recalling, of course, that this is an overestimate of the true reliability of the real-time estimates, since it ignores the estimation error in the final series.

¹⁹ This lower bound comes from the fact that other sources of errors can affect the estimates of the output gap. In particular, it is reasonable to assume that some uncertainty remains with long-past historical estimates of the output gap.

²⁰ For the multivariate techniques, only output data are available in real-time. For the other variables, we use the most recent vintage available.

3.2.3 Quasi-real estimates

The differences between the real-time and the final estimates have several sources, one of which is the ongoing revision of published data. To isolate the importance of this factor, we define a third output gap measure, the “quasi-real” estimate. Like the real-time estimate, the quasi-real estimate is constructed in two steps.

The first step is to construct an ensemble of “rolling” estimates of the output gap. That is, we begin by taking the final vintage of the output series, but use only the observations up to and including 1982Q2 to compute the quasi-real estimate for 1982Q2. Next, we extend the sample period by one observation and repeat the detrending. We continue in this way until we use the full sample period for the final output series and have a full set of corresponding output gap series.

The second step is the same as that used to construct the real-time series: We collect the first available estimate of the output gap at each point in time from the various series we constructed in step one. This sequence of output gaps is the quasi-real series. The difference between the real-time and the quasi-real series is entirely due to the effects of data revision, since estimates in the two series at any particular point in time are based on data samples that cover exactly the same time period.

3.2.4 Quasi-final estimates

For unobserved component models, we are able to further decompose the revision in the estimated gap by defining another estimate of the output gap. This “quasi-final” estimate uses more information than the quasi-real estimate (which uses subsamples of final data) but less than the final estimate (which uses the full sample of final data.) This is relevant because unobserved component models use the data in two distinct phases. First, they use the available data sample to estimate the parameters of a time-series model of output. Next, they use these estimated parameters in the Kalman filter to arrive at estimates of the output gap. However, they distinguish between “filtered” and “smoothed” estimates of the output gap. The smoothed estimate uses the full sample parameter estimates and data from 1 to T to form an optimal estimate of the gap in quarter T . The filtered estimate uses only data from 1 to t with the full sample parameter estimates to make an optimal estimate of the output gap at t ($1 \leq t \leq T$). For this class of models, smoothed estimates of the output gap are used to construct the final series, while filtered estimates are used for the quasi-final series.

The difference between the quasi-final and the quasi-real series then reflect solely the effects of using different parameter estimates for the model to filter the data (i.e. the full-sample ones versus

the partial sample ones). The extent of the difference will reflect the importance of parameter instability in the underlying unobserved component model. The difference between the quasi-final and the final series reflects the importance of ex post information in estimating the output gap given the parameter values of the process generating output.

3.2.5 Reliability indicators used to evaluate the revisions

The mean of the absolute value of total revisions and the root-mean-square of the total revision series are two good measures of the magnitude of output gap revisions over time. But because the various techniques may have substantial variation in the size of the cyclical component they produce, it is easier to compare their reliability in real-time by looking at comparably scaled measures of the revisions.

A good scaled measure is the noise-to-signal ratio. There are two different ways to measure this indicator. The first one (NS1) divides the standard deviation of the total revision series over the standard deviation of the final output gap series. The second one (NS2) divides the root-mean-square of the total revision series over the standard deviation of the final output gap series. High values for these ratios mean that the size of the revisions is large relative to the magnitude of the output gap estimates. Thus, a reliable estimation technique should generate low values for both measures of the noise-to-signal ratio.

Another indicator of the reliability of estimation techniques is the correlation between the real-time and the final estimates of the output gap (COR). This indicator shows whether the real-time business cycle has the same short-term fluctuations as the final one. The correlation coefficient would be 1 in the ideal case where no revisions to the real-time estimates are ever required. A last reliability indicator is the frequency with which the real-time and final gaps are of opposite signs (OPSIGN). Reliable estimation techniques would have an OPSIGN statistic close to 0.

3.3 Second evaluation method: inflation forecasts

The second evaluation method is used to determine whether output gap estimates generated by the different techniques help improve out-of-sample forecasts of inflation. To do this, we incorporate the gap estimates in a simple equation and compare the out-of-sample forecasts produced by this equation with those produced by a similar equation that excludes the output gap. Our approach also allows us to determine whether certain techniques generate output gap estimates that do a better job at forecasting inflation than the other estimation techniques.

To perform such an evaluation, we must make certain choices and assumptions. First, to forecast inflation, we use a quasi-reduced form equation that includes lags of inflation and of output gap as independent variables. Of course, the omission of certain key variables, such as inflation expectations, will probably reduce the quality of the forecasts. It is therefore possible that output gaps might improve on simple univariate forecasts of inflation, but not on forecasts using a broader range of inputs. For this reason, we feel that the experiment that we perform may actually overstate the utility of empirical output gap models.

Second, we perform the test for a forecasting horizon of four quarters. This test is justified by the fact that monetary policy actions require time to take effect. We could, however, repeat in future work the same procedure for different forecasting horizons, with a more complete set of variables in the forecasting equation.

3.3.1 Procedure used to forecast inflation

Based on the assumptions explained in the previous section, we construct two forecasting equations. One includes the output gap and the other does not:

$$\pi_{t+4} = \tilde{\alpha} + \sum_{l=1}^L \tilde{\beta}_l \pi_{t-l} + \sum_{l=1}^L \tilde{\lambda}_l c_{t-l}^i + \tilde{\varepsilon}_{t+4} \quad (30)$$

$$\pi_{t+4} = \hat{\alpha} + \sum_{l=1}^L \hat{\beta}_l \pi_{t-l} + \hat{\varepsilon}_{t+4}, \quad (31)$$

where π_t is the level of inflation at quarter t (year-over-year inflation measured with total CPI) and c_t^i is the output gap estimate generated by technique i at quarter t .²¹ Thus, the dependent variable that we want to forecast is the level of inflation in four quarters. Because of reporting lags, information at quarter t is not available before quarter $t+1$. This is why contemporaneous values of inflation and output gap are not used as right-end side variables.

In our out-of-sample forecasting exercise, we use quarterly data from 1961Q1 to 1982Q2²² to estimate the parameters of each equation, which are then used to produce a forecast of inflation for 1983Q3. The sample is then extended to 1982Q3, the equations are re-estimated and we produce a forecast for 1983Q4. This procedure is repeated until we have a forecast for 2000Q4.

²¹ For equation (30), we estimate a different equation for each method of estimating the output gap. Since each unobserved component model provides two estimates, one filtered and the other smoothed, and the structural VAR method also generates two estimations (RLTP and RLTP1), this makes 14 different forecasting equations.

²² This represents the data set that was available in 1982Q3.

We perform the same exercise with final output gap estimates. Normally, we would expect forecasts using final estimates of the output gap to be more precise, given that these estimates incorporate a broader set of information.

3.3.2 Criteria for selecting the number of lags

For each equation, we must choose the number of lags for β and λ . We used three different criteria: the Akaike (1973) information criterion, the Schwarz (1978) information criterion, and a criterion based on the out-of-sample forecast errors (OSFE).

The Akaike (AIC) and Schwarz (SBC) information criteria both use the entire sample available. These criteria are illustrated by equations (32) and (33):

$$AIC = T \cdot \log|\Sigma| + 2N \quad (32)$$

$$SBC = T \cdot \log|\Sigma| + N \cdot \log(T), \quad (33)$$

where T is the number of observations, N is the number of coefficients to estimate, and $|\Sigma|$ is the determinant of the covariance matrix of the residuals. For both techniques, the optimal specification is the one that minimizes the value of the test.

The OSFE criterion formalized by Murchison (1999)²³ is totally different than the ones proposed by Akaike and Schwarz. First, we have to divide the sample into two parts. With observations of the first part of the sample, we estimate the coefficients of the model and use them to forecast inflation in the second part of the sample. The optimal specification is the one that minimizes forecasts errors in the second part of the sample. This criterion can be formalized as follows:

$$\min_{k \in [1, p]} \left\{ \phi^{-1} \sum_{t=s+v}^T \Gamma_t' \Gamma_t \right\} \left| \frac{\partial \sum_{t=1}^s (y_t - \sum_{h=1}^p \sum_{l=1}^k \hat{\beta}_l y_{t-l})^2}{\partial \hat{\beta}_{l,h}} \right| = 0; \Gamma_t = (y_t - \sum_{i=1}^k \hat{\beta}_i y_{t-1}), \quad (34)$$

where k is the number of lags, p is the maximum number of lags that we want to test, T is the total number of observations in the full sample, s is the number of observations in the first subsample, v is the forecasting horizon, Γ_t is the forecasting error, y_t is the variable that we want to forecast, and $\phi = T - (s + v)$.

Again, we have to make some choices. First, we set the parameter p at 20, which is quite large for a forecasting model that uses quarterly data. Of course, we fix v at 5 [since $(t+4) - (t-1) = 5$]. For

²³ Murchison mentions that this is a methodology commonly applied by forecasters. The goal of his paper is to compare this procedure with other selection criteria.

the parameter s , Murchison finds that the formula $s = 0.7T$ gives good results. However, we adapt Murchison's approach to better meet our needs. We arbitrarily adapt Murchison's rule to $s = T - \nu - 15$, since his rule does not allow enough observations for the estimation of the first vintages.

3.3.3 Evaluation of forecasts

The evaluation procedure in this case consists of comparing the out-of-sample forecasts with actual inflation. To determine whether the output gap estimates significantly improve inflation forecasts, we use a test similar to the one proposed by Diebold and Mariano (1995). They build a test that follows a normal distribution, $N(0,1)$, and that shows whether one series of forecast errors is significantly different from another one.²⁴ This test can be written as follows:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}}, \quad (35)$$

where
$$\bar{d} = \frac{1}{T} \sum_{t=1}^T [g(e_{it}) - g(e_{jt})] \quad (36)$$

$$2\pi\hat{f}_d(0) = \sum_{\tau=-(T-1)}^{T-1} 1 \cdot \left(\frac{\tau}{S(T)} \right) \hat{\gamma}_d(\tau) \quad (37)$$

$$\hat{\gamma}_d(\tau) = \frac{1}{T} \sum_{\tau=|T|+1}^T (d_t - \bar{d})(d_{t-|\tau|} - \bar{d}). \quad (38)$$

The two series of forecast errors are designated by the terms e_{it} and e_{jt} . The expression $g()$ designates the transformation made to the series of errors. For this paper, we use squared errors [$g(e_{it}) = e_{it}^2$]. The expression $2\pi\hat{f}_d(0)$ represents the weight sum of sample autocovariances and is used to correct serial correlation problems. Since the forecasting horizon is longer than one period (it is five quarters), we follow Diebold and Mariano's recommendation and use a rectangular kernel.

Harvey, Leybourne, and Newbold (1997) propose two small modifications to the Diebold Mariano test, to improve the properties with a small sample. In this paper, we use this modified version of the Diebold and Mariano test. The first modification that they propose is as follows:

$$MDM = DM \cdot \frac{\sqrt{n+1-2h+h(h-1)/n}}{n}, \quad (39)$$

²⁴ Of course, both series must forecast the same variable.

where n is the number of observations in each series of forecast errors and h is equal to the forecasting horizon minus 1. The second modification that Harvey, Leybourne, and Newbold propose is to use a t_{n-1} distribution instead of a normal distribution to determine the critical value of the test.

By applying the modified version of the Diebold and Mariano test, we can verify whether equations that incorporate an output gap measure significantly improve inflation forecasts relative to an equation that excludes such a measure. This test also allows us to determine whether certain methods generate output gap estimates that do a better job of forecasting inflation than the other estimation techniques.

4. Results

This section presents the results obtained for the two evaluation methods. First, we show the output gap estimates generated by the different methods.

4.1 Output Gap estimates

Figures 1 and 2 compare the final and real-time output gap estimates for the nine estimation methods described in section 2.2 (the structural VAR methods generate two output gap estimates: RLTP and RLTP1). Table 1 provides descriptive statistics on the various real-time, quasi-real, quasi-final, and final estimates.

First, we can clearly see that the linear trend and the Watson model do not generate realistic business cycles. In particular, their real-time estimates are negative for the entire period presented in Figures 1 and 2.

Second, the AR1 statistics of Table 1 show that all techniques generate persistent business cycles, with the exception of the Beveridge-Nelson decomposition, which has a first-order autocorrelation coefficient in the range of 0.55 to 0.58. In fact, the Beveridge-Nelson business cycle shows many more turning points than any other estimation techniques (see the TP statistics in Table 1).

Figures 1 and 2 also show a lot of variability in terms of the magnitude of the gap estimates. This is true for both final and real-time estimates. These figures also show that different output gap estimates do not necessarily display the same short-term fluctuations. In fact, some techniques frequently display positive output gaps while others display negative ones.

Figure 1

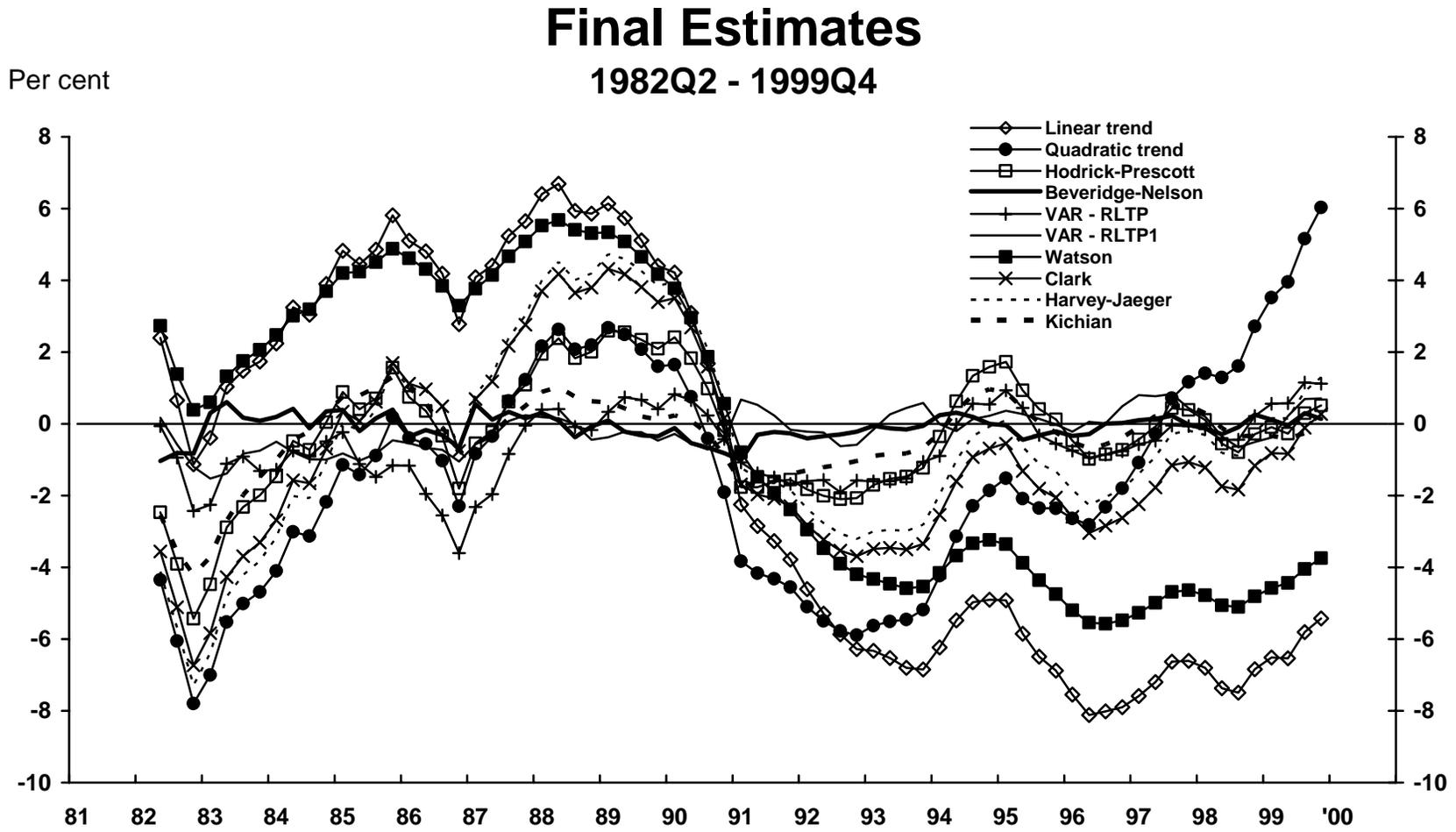
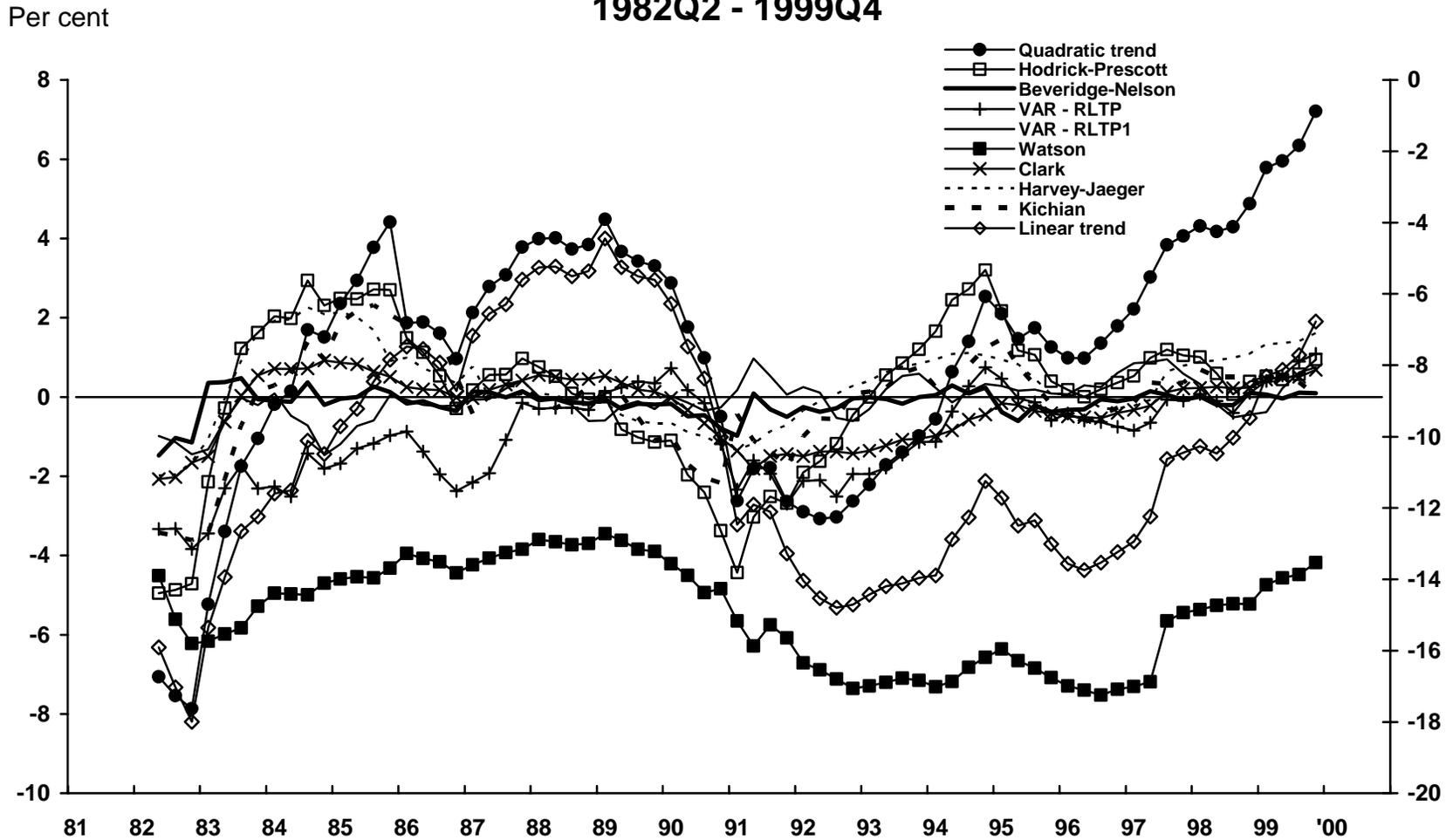


Figure 2

Real-Time Estimates 1982Q2 - 1999Q4



*Notes: The linear trend refers to the right scale, while all other series refer to the left scale. Therefore, the line representing an output gap of zero does not apply for the linear trend since this series is always negative in real-time.

Table 1 : Statistics on output gap estimates							
1982Q2 – 1999Q4²⁵							
Methods	MEAN	SD	MIN	MAX	AR1	COR	TP
Linear trend							
Finale	-1.26	5.23	-8.12	6.69	0.99	1.00	17
Temps réel	-10.59	3.35	-17.99	-4.44	0.97	0.62	17
Quasi-réelle	-9.14	2.93	-14.08	-3.95	0.97	0.68	15
Quadratic trend							
Finale	-1.49	3.14	-7.80	6.03	0.97	1.00	17
Temps réel	1.15	3.22	-7.87	7.21	0.97	0.91	19
Quasi-réelle	1.72	3.35	-7.18	7.91	0.98	0.92	15
Hodrick-Prescott							
Finale	-0.25	1.68	-5.43	2.59	0.92	1.00	21
Temps réel	0.12	1.90	-4.95	3.20	0.94	0.45	22
Quasi-réelle	0.15	1.95	-4.49	3.32	0.94	0.48	19
Beveridge-Nelson							
Finale	-0.09	0.36	-1.04	0.61	0.55	1.00	35
Temps réel	-0.14	0.34	-1.47	0.48	0.55	0.82	45
Quasi-réelle	-0.09	0.25	-0.90	0.35	0.58	0.97	34
VAR – RLTP							
Finale	-0.65	1.00	-3.61	1.16	0.87	1.00	26
Temps réel	-1.01	1.16	-3.84	1.09	0.92	0.81	32
Quasi-réelle	-1.00	1.12	-3.48	1.19	0.90	0.91	27
VAR – RLTP1							
Finale	-0.21	0.56	-1.53	0.84	0.83	1.00	25
Temps réel	-0.08	0.54	-1.46	0.97	0.81	0.83	27
Quasi-réelle	-0.15	0.49	-1.46	0.88	0.82	0.95	26
Watson							
Finale	-0.33	4.08	-5.57	5.68	0.99	1.00	11
Temps réel	-5.46	1.29	-7.52	-3.46	0.96	0.81	21
Quasi-réelle	-4.75	0.89	-6.33	-3.34	0.96	0.84	15
Quasi-finale	-2.26	1.22	-2.84	2.67	0.90	0.70	9
Clark							
Finale	-0.85	2.58	-6.73	4.32	0.96	1.00	21
Temps réel	-0.26	0.81	-2.07	0.93	0.96	0.62	16
Quasi-réelle	-0.20	0.73	-1.79	0.78	0.97	0.63	16
Quasi-finale	-0.55	0.77	-2.61	0.67	0.97	0.70	15
Harvey-Jaeger							
Finale	-0.51	2.70	-7.29	4.73	0.96	1.00	21
Temps réel	0.33	1.00	-1.98	2.25	0.95	0.04	18
Quasi-réelle	-0.02	1.23	-5.73	1.82	0.61	0.12	18
Quasi-finale	-0.25	1.57	-3.33	1.78	0.99	0.84	9
Kichian							
Finale	-0.39	1.13	-4.22	1.31	0.94	1.00	13
Temps réel	-0.07	1.29	-3.61	2.31	0.92	0.78	28
Quasi-réelle	0.02	1.22	-2.84	2.67	0.90	0.70	30
Quasi-finale	-0.19	1.00	-3.22	1.37	0.91	0.88	26

²⁵ Notes: The statistics shown for each variable are: MEAN, the mean; SD, the standard deviation; MIN and MAX, the minimum and maximum values; AR1, the first-order serial correlation; COR, the correlation with the final estimate for that method; and TP, the number of turning points.

These differences observed across techniques show that the errors that result from choosing an inadequate model (second type of error described in section 2.3) can potentially be large. To see this, one only has to assume that one technique produces the "real" output gap estimates. The difference between these estimates and those produced by other techniques would therefore indicate the magnitude of the errors that result from the choice of an inadequate model.

Tables 2 to 7 show three different statistics that help to evaluate potential errors that result from choosing a wrong model for both final and real estimates. In these tables, we assume that each line has the "real" model while each column has the wrong one. This allows us to consider a variety of potential errors.

Tables 2 and 3 show the noise-to-signal ratio statistics; that is, the ratio of the standard deviation of errors²⁶ over the standard deviation of the estimates of the good model. A statistic close to 0 means that the wrong model is good at reproducing the estimates generated by the good model. We can see that these ratios are relatively high for all techniques, which means that the standard deviation of the errors is of the same order of magnitude as the estimates of the good model.

Tables 4 and 5 show the correlation between the good and the wrong models. A statistic close to 1 means that the wrong model reproduces the short-run fluctuations of the good model fairly accurately. If it is true that the estimates of some techniques are closely correlated together, we can see that, in general, correlation coefficients are quite low.

Finally, Tables 6 and 7 show the frequency with which the good and the wrong models have opposite signs. A statistic close to 0 means that the wrong model estimates are a good indicator of output gap. These values are generally quite large, particularly for real-time estimates.

The different statistics reported in Tables 2 to 7 clearly show that errors due to the choice of an inadequate model can potentially be large. Even by using final estimates, the differences across models are large. Based on this analysis, however, we cannot draw any conclusion on the reliability of the different methods. The next two sections present evaluation criteria for the reliability of output gap estimates.

²⁶ In this case, errors are defined as the difference between estimates of the good model and estimate of the wrong model.

Table 2: Noise-to-signal ratio											
Final estimates											
(1982Q2 – 1999Q4)											
		Wrong model									
		LT	QT	HP	BN	RLTP	RLTP1	Watson	Clark	HJ	Kichian
Good model	LT	-	1.04	0.94	0.99	1.02	1.05	0.22	0.78	0.87	0.98
	QT	1.74	-	0.68	0.97	0.82	0.96	1.47	0.63	0.56	0.80
	HP	2.92	1.27	-	0.97	0.74	0.96	2.30	0.79	0.79	0.48
	BN	14.50	8.52	4.53	-	2.82	1.85	11.34	7.11	7.43	2.96
	RLTP	5.35	2.57	1.24	1.01	-	0.91	4.24	2.27	2.30	1.09
	RLTP1	9.85	5.42	2.89	1.19	1.64	-	7.82	4.64	4.72	1.99
	Watson	0.28	1.13	0.95	0.99	1.04	1.07	-	0.78	0.99	0.98
	Clark	1.58	0.77	0.51	0.99	0.87	1.00	1.23	-	0.18	0.74
	HJ	1.68	0.66	0.49	0.99	0.85	0.97	1.35	0.18	-	0.72
	Kichian	4.50	2.21	0.70	0.94	0.96	0.98	3.55	1.68	1.72	-

Table 3: Noise-to-signal ratio											
Real-time estimates											
(1982Q2 – 1999Q4)											
		Wrong model									
		LT	QT	HP	BN	RLTP	RLTP1	Watson	Clark	HJ	Kichian
Good model	LT	-	0.64	1.00	0.97	0.84	1.00	0.83	0.83	0.98	0.92
	QT	0.66	-	0.80	0.95	0.72	0.95	1.69	0.80	0.88	0.80
	HP	1.76	1.36	-	0.87	0.93	0.98	2.21	0.77	0.58	0.55
	BN	9.45	8.88	4.84	-	3.19	1.64	11.49	2.07	2.43	3.35
	RLTP	2.43	2.00	1.52	0.94	-	0.89	3.43	0.87	1.13	1.16
	RLTP1	6.12	5.61	3.42	1.03	1.89	-	7.69	1.78	2.04	2.34
	Watson	2.17	4.24	3.26	3.07	3.09	3.26	-	2.88	3.19	3.20
	Clark	3.45	3.20	1.81	0.88	1.26	1.20	4.61	-	0.93	1.21
	HJ	3.30	2.84	1.11	0.84	1.32	1.11	4.11	0.75	-	0.76
	Kichian	2.39	2.01	0.80	0.89	1.04	0.99	3.19	0.76	0.59	-

Table 4: Correlation between estimates of the good and the wrong models											
Final estimates											
(1982Q2 – 1999Q4)											
		Wrong model									
		LT	QT	HP	BN	RLTP	RLTP1	Watson	Clark	HJ	Kichian
Good model	LT	-									
	QT	0.21	-								
	HP	0.34	0.76	-							
	BN	0.05	0.24	0.23	-						
	RLTP	-0.04	0.67	0.68	0.13	-					
	RLTP1	-0.48	0.26	0.26	-0.02	0.41	-				
	Watson	1.00	0.19	0.32	0.03	-0.06	-0.49	-			
	Clark	0.63	0.77	0.89	0.13	0.49	0.07	0.62	-		
	HJ	0.49	0.82	0.92	0.12	0.55	0.19	0.47	0.98	-	
	Kichian	0.19	0.67	0.91	0.33	0.48	0.28	0.17	0.73	0.77	-

Table 5: Correlation between estimates of the good and the wrong models											
Real-time estimates											
(1982Q2 – 1999Q4)											
		Wrong model									
		LT	QT	HP	BN	RLTP	RLTP1	Watson	Clark	HJ	Kichian
Good model	LT	-									
	QT	0.79	-								
	HP	0.27	0.59	-							
	BN	0.29	0.47	0.71	-						
	RLTP	0.58	0.83	0.41	0.31	-					
	RLTP1	0.07	0.34	0.19	0.25	0.45	-				
	Watson	0.85	0.47	0.01	0.01	0.20	-0.28	-			
	Clark	0.74	0.82	0.68	0.46	0.51	-0.01	0.62	-		
	HJ	0.18	0.50	0.89	0.59	0.26	0.03	0.06	0.67	-	
	Kichian	0.38	0.63	0.85	0.50	0.39	0.23	0.13	0.65	0.81	-

Table 6: Frequency with which the estimates of the good and the wrong models have opposite signs											
Final estimates (1982Q2 – 1999Q4)											
		Wrong model									
		LT	QT	HP	BN	RLTP	RLTP1	Watson	Clark	HJ	Kichian
Good model	LT	-									
	QT	0.41	-								
	HP	0.34	0.27	-							
	BN	0.42	0.44	0.45	-						
	RLTP	0.45	0.21	0.28	0.42	-					
	RLTP1	0.69	0.42	0.44	0.44	0.38	-				
	Watson	0.04	0.45	0.38	0.44	0.49	0.73	-			
	Clark	0.17	0.27	0.23	0.45	0.31	0.58	0.18	-		
	HJ	0.24	0.23	0.24	0.44	0.24	0.54	0.25	0.07	-	
	Kichian	0.32	0.28	0.07	0.44	0.32	0.42	0.37	0.24	0.28	-

Table 7: Frequency with which the estimates of the good and the wrong models have opposite signs											
Real-time estimates (1982Q2 – 1999Q4)											
		Wrong model									
		LT	QT	HP	BN	RLTP	RLTP1	Watson	Clark	HJ	Kichian
Good model	LT	-									
	QT	0.69	-								
	HP	0.68	0.21	-							
	BN	0.30	0.54	0.46	-						
	RLTP	0.21	0.48	0.61	0.31	-					
	RLTP1	0.48	0.52	0.39	0.46	0.46	-				
	Watson	0.00	0.69	0.68	0.30	0.21	0.48	-			
	Clark	0.48	0.30	0.31	0.44	0.41	0.65	0.48	-		
	HJ	0.63	0.34	0.15	0.45	0.59	0.49	0.63	0.35	-	
	Kichian	0.58	0.34	0.24	0.42	0.51	0.49	0.58	0.30	0.17	-

4.2 First evaluation method: revisions of output gap estimates

Figure 3 shows the total revisions to the output gap estimates over time for each method. Table 8 provides descriptive statistics on these revisions. Some interesting facts should be noted.

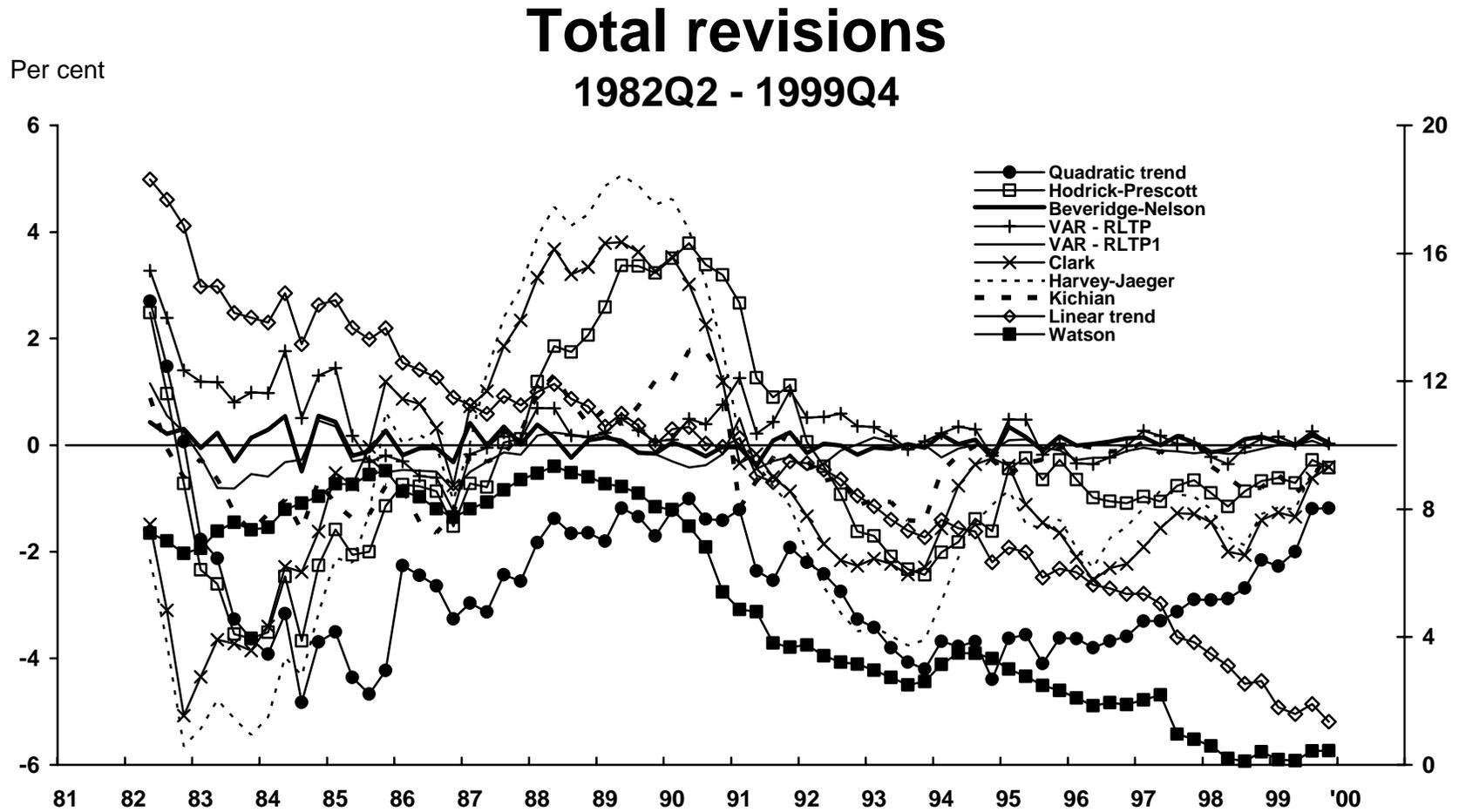
Methods	ABSM	SD	MIN	MAX	AR1
Linear trend	9.33	4.11	1.35	18.31	0.99
Quadratic trend	2.76	1.33	-4.93	2.71	0.88
Hodrick-Prescott	1.59	1.88	-3.67	3.79	0.94
Beveridge-Nelson	0.17	0.21	-0.49	0.55	-0.16
VAR – RLTP	0.51	0.68	-1.24	3.28	0.77
VAR – RLTP1	0.26	0.32	-0.84	1.16	0.65
Watson	5.13	3.14	0.11	9.34	0.99
Clark	1.91	2.18	-5.07	3.81	0.96
Harvey-Jaeger	2.46	2.84	-5.66	5.06	0.97
Kichian	0.72	0.82	-1.65	1.78	0.81

Notes: The statistics shown for each variable are: ABMS, the absolute mean; SD, the standard deviation; MIN and MAX, the minimum and maximum values; and AR1, the first-order serial correlation coefficient.

First, total revisions to the estimates generated by the linear trend and the Watson model clearly show that potential output has not grown at a constant rate in Canada in the past 20 years. Revisions for these two methods are large (in absolute values) and are highly persistent given the large first-order autocorrelation coefficients. Also, their revisions are always positive, which means that these techniques systematically underestimate the output gap. At the opposite, the quadratic trend systematically overestimates the output gap over time. This proves that deterministic trends are not appropriate to estimate potential output in Canada.²⁷

²⁷ We have not tested a breaking linear trend model, given the difficulty in determining the period of the break in real-time. Orphanides and van Norden (1999), however, test this method with U.S. data and do not find significant improvement in the results.

Figure 3



*Notes: The linear trend and the Watson model refer to the right scale, while all other series refer to the left scale. The zero line does not apply for these two series since their revisions are always positive.

Second, the AR1 statistics reported in Table 8 show that revisions for all the univariate techniques, with the exception of the Beveridge-Nelson, are very persistent. This strong persistence means that the time lag before people can realize the full extent of the revisions is quite long. Although lower than for the univariate techniques, revisions to the output gap generated by the multivariate models are also persistent, with AR1 coefficients ranging from 0.65 for the RLTP1 model to 0.81 for Kichian's model. In fact, the only method that generates output gap estimates that do not show too much persistence when revised is the Beveridge-Nelson, having an AR1 coefficient of -0.16. The low persistence of revisions to the Beveridge-Nelson estimates seems to be directly linked to the low persistence of its estimates (see the AR1 statistic in Table 1).

For the magnitude of the revisions, we can see in Figure 3 that they are generally smaller for multivariate methods. This is also confirmed by the first two columns of Table 8. Indeed, only revisions to the Beveridge-Nelson method have lower absolute mean errors and standard errors than the multivariate techniques. But, in some case, the magnitude of revisions does not really matter, especially if the magnitude of the estimates is small. Other indicators can give a better picture of the reliability of estimates. The next section presents such indicators.

4.2.1 Reliability indicators

Table 9 compares the different techniques based on the reliability criteria described in section 3.2.5. As stated earlier, these indicators do not allow us to conclude that a technique is reliable, since total revisions underestimate total errors that affect output gap estimates. Bad results for some of these statistics, however, would clearly show that a method is not reliable.

For all techniques, total revisions are of the same order of magnitude as the final estimates of the output gap. This translates into large values for both measures of the noise-to-signal ratio (NS and NSR). Even the best methods (the Beveridge-Nelson and RLTP1 models) have rather large ratios, judging by these criteria. Some methods even show noise-to-signal ratios higher than 1, which means that the standard deviation and/or root-mean-square of the revisions is larger than the estimates themselves.

Methods	NS	NSR	COR	OPSIGN
Linear trend	0.79	2.00	0.62	0.45
Quadratic trend	0.43	0.96	0.91	0.37
Hodrick-Prescott	1.12	1.17	0.45	0.42
Beveridge-Nelson	0.59	0.61	0.82	0.21
VAR – RLTP	0.68	0.79	0.81	0.07
VAR – RLTP1	0.58	0.64	0.83	0.15
Watson	0.77	1.51	0.81	0.49
Clark	0.84	0.89	0.62	0.28
Harvey-Jaeger	1.05	1.12	0.04	0.56
Kichian	0.73	0.78	0.78	0.39

Most techniques show relatively good results for the correlation criterion (COR). There are some exceptions, however, such as the Harvey-Jaeger method (0.04) and the Hodrick-Prescott filter (0.45). In the case of the quadratic trend (0.91), this criterion seems misleading, since we showed in section 4.2 that this technique systematically overestimates the output gap over time.

In the case of the OPSIGN statistic, most methods perform very poorly. In fact, more than half of the techniques predict the wrong sign in real-time once every three quarters. The only methods that do not show too much difference between the real-time and final estimates are the structural VAR methods.

To summarize, these results show that the errors associated with real-time estimates of the output gap are substantial. None of the techniques perform well for all the criteria. In fact, ex post revisions are of the same order of magnitude as the ex post estimates of the output gap, the real-time estimates frequently misclassify the sign of the gap, and the estimation errors appear to contain a highly persistent component that is substantial in size.

4.2.2 Decomposition of revisions

Figures 4 to 13 show the real-time estimates of the output gap and the decomposition of the revisions over time for each estimation method. As stated in the previous section, we can clearly see that total revisions are of the same order of magnitude as the real-time estimates of the output gap. For some techniques, such as the Clark and Harvey-Jaeger models (Figures 8 and 9), the total revisions are even larger than the real-time estimates.

As explained in section 3.2.3, one source of revision is data revision. We can see that, for all the techniques except the Beveridge-Nelson decomposition and the structural VAR models, data revisions represent a small portion of total revisions. In fact, for all techniques, data revisions are typically less than 2 per cent (in absolute value). This, in turn, means that most of the revisions are due to the addition of new points to our data sample. For the Beveridge-Nelson method and the structural VAR models (Figures 11, 12, and 13), however, we can see that data revision is responsible for most of the revisions in real-time estimates of the output gap, which means that if data were never revised, total revisions of gap estimates for these techniques would be close to 0.

For the unobserved component models (Figures 5, 8, 9, and 10), we can also measure the part of the revisions that is due to changes in the coefficients over time. This also gives an idea of coefficient stability over time. The Clark model (Figure 8) seems to be least affected by instability in the coefficients. In fact, all the instability observed for this model happened in the 1980s.

The Watson models (Figure 5), on the other hand, clearly show an important problem of coefficient instability. A break happened in the early 1990s. We can also see that the coefficients have gradually been revised over time. Again, this would tend to support the idea that potential output has not grown at a constant rate in Canada. Coefficients of the Harvey-Jaeger model (Figure 9) are also not stable. In fact, we can see many dramatic revisions in the coefficients over time.

The Kichian model also displays a lot of instability in its coefficients. This instability was most pronounced before 1993. In fact, coefficient instability explains a large part of the total revisions for this model.

Figure 4

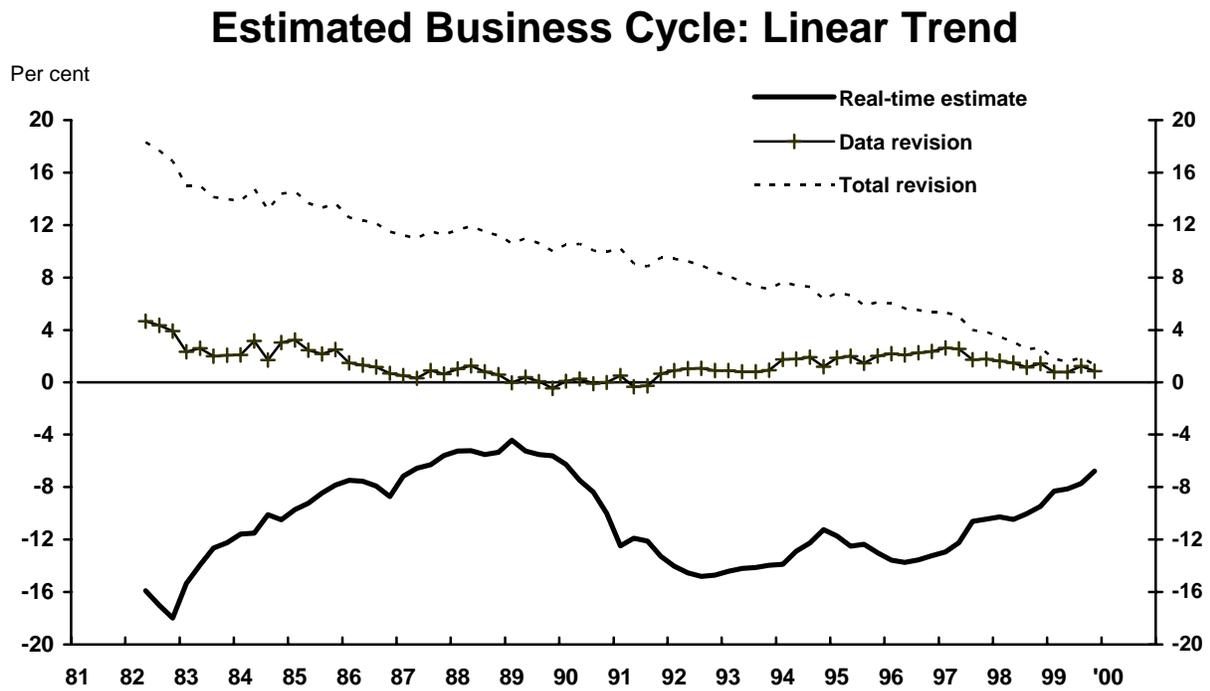


Figure 5

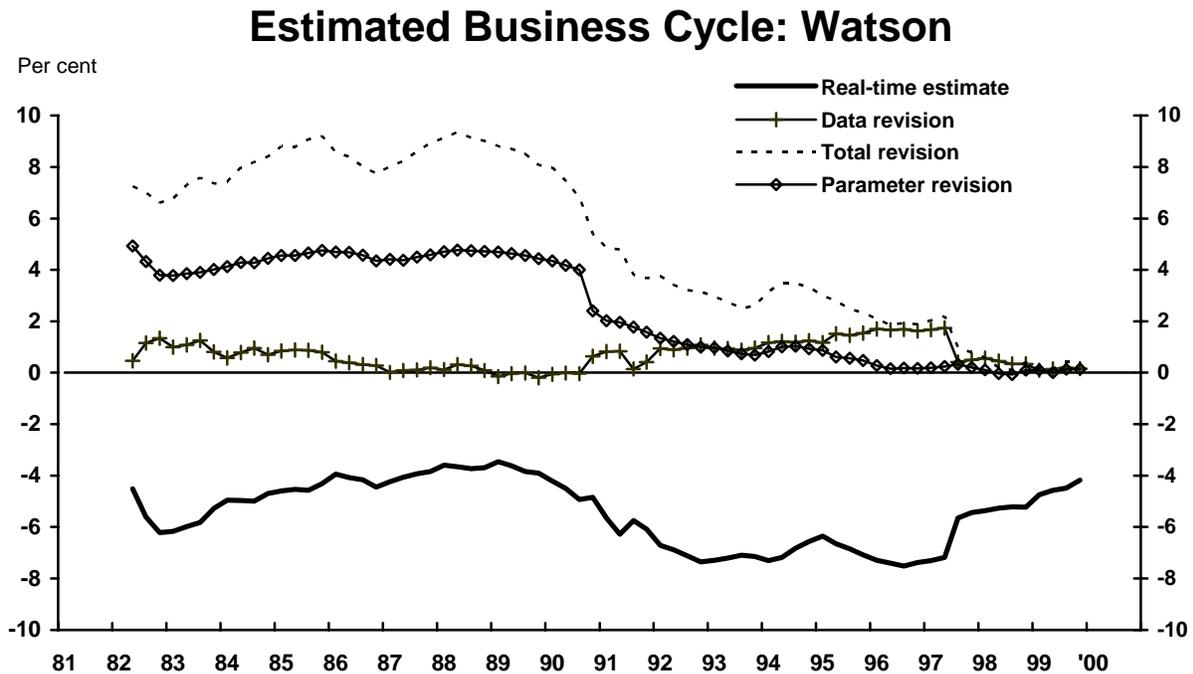


Figure 6

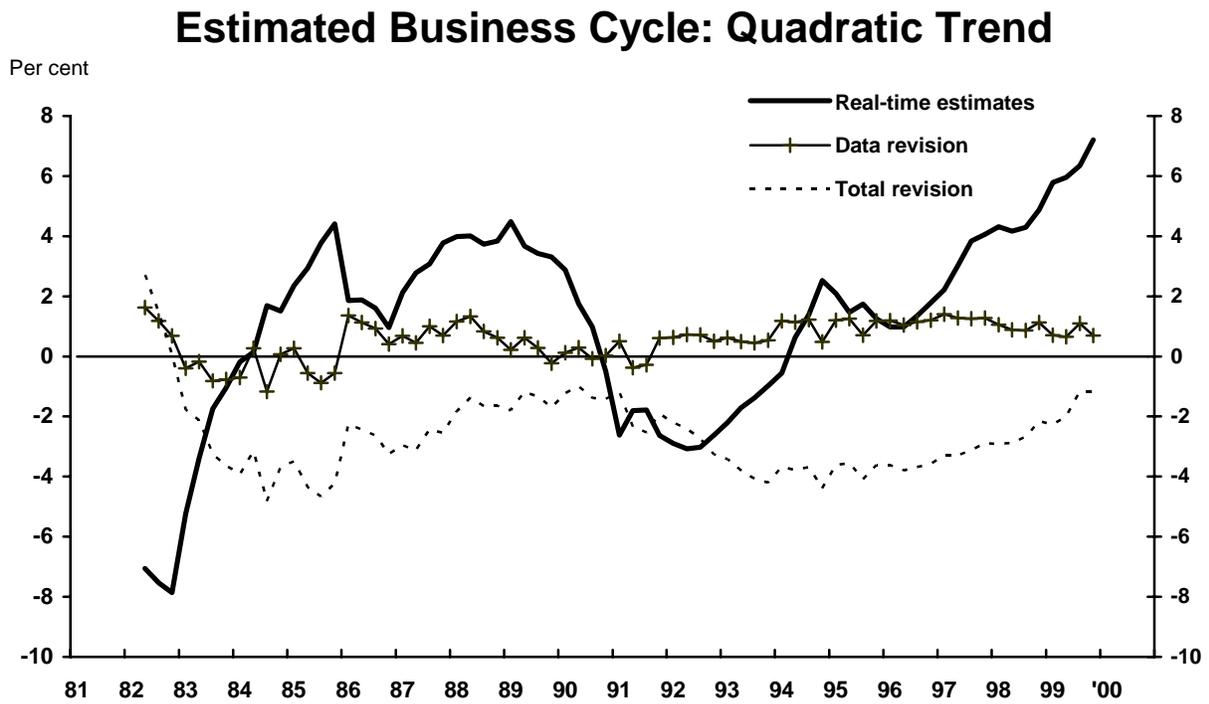


Figure 7

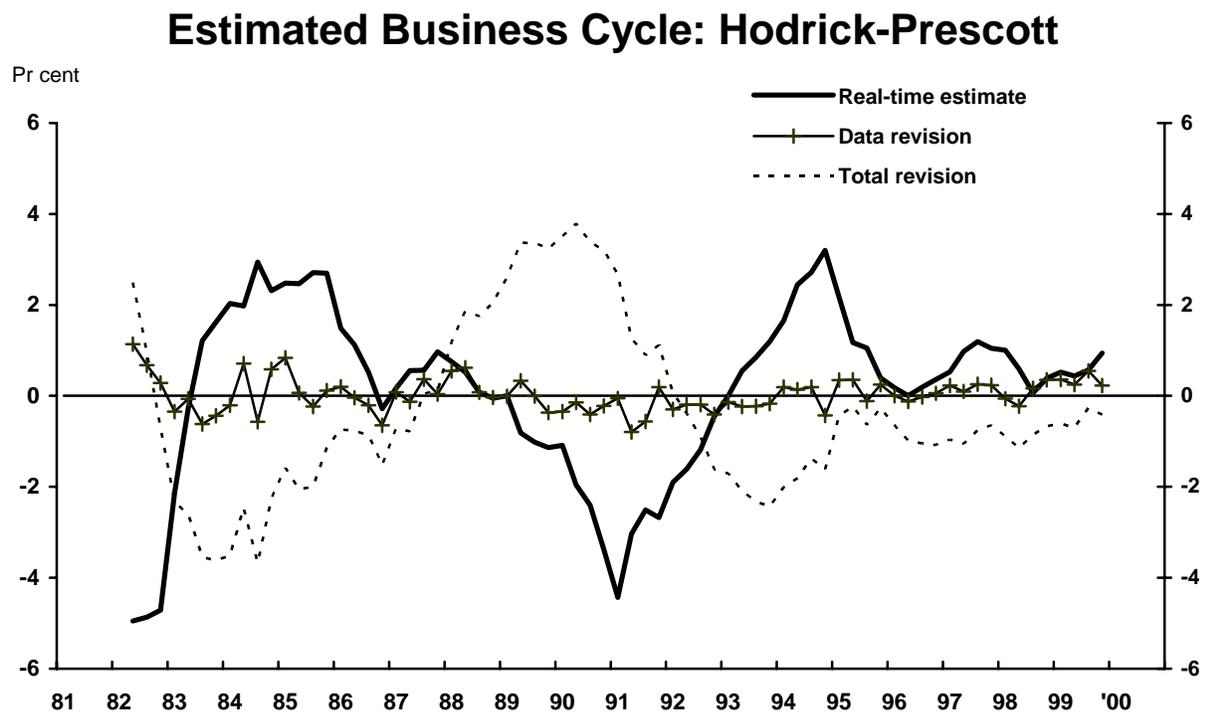


Figure 8

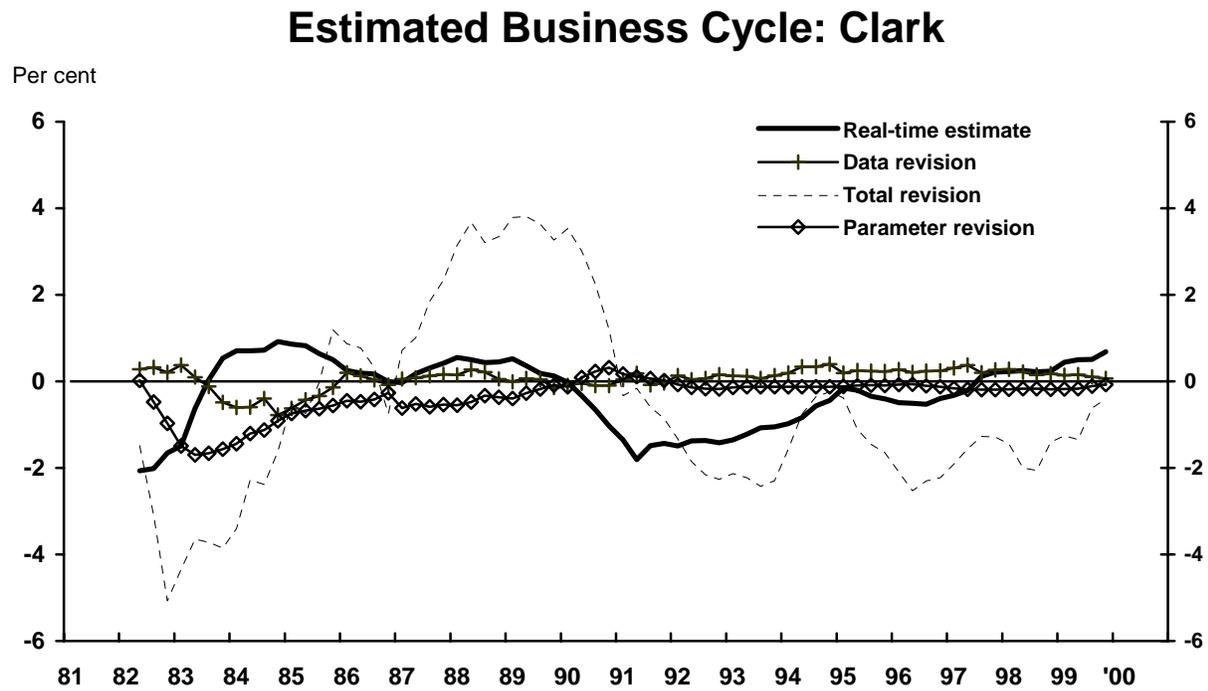


Figure 9

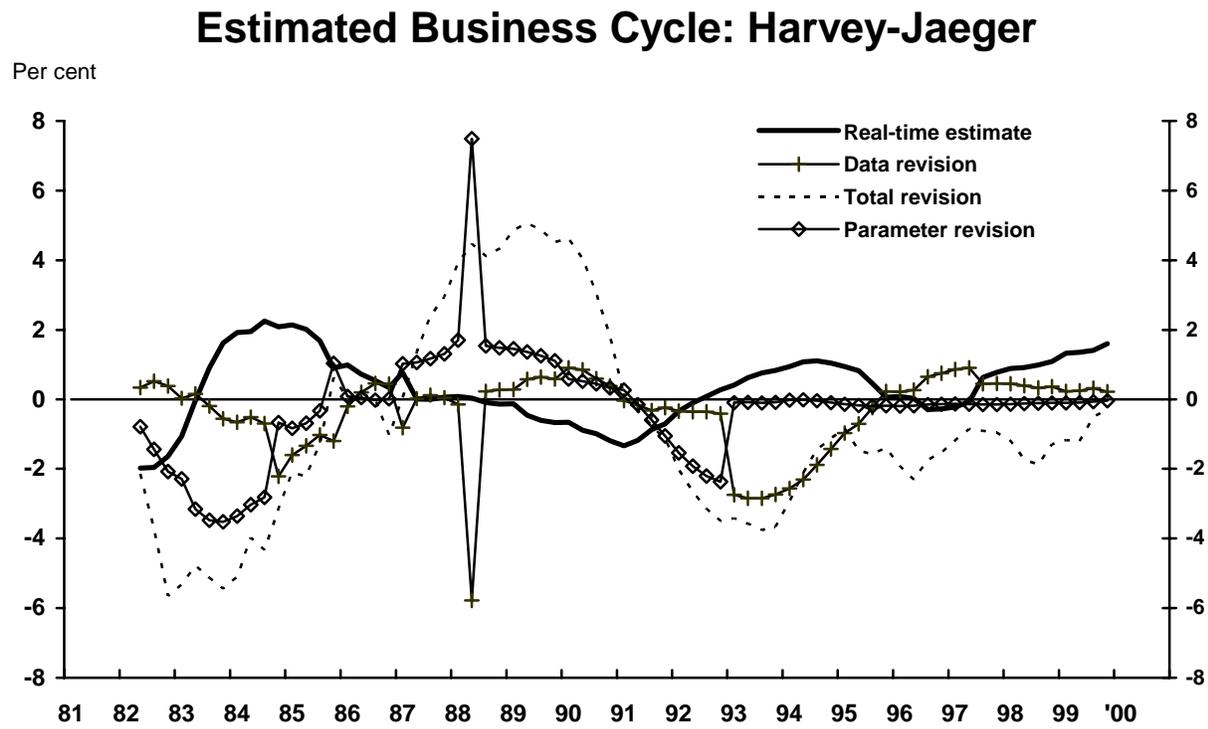


Figure 10

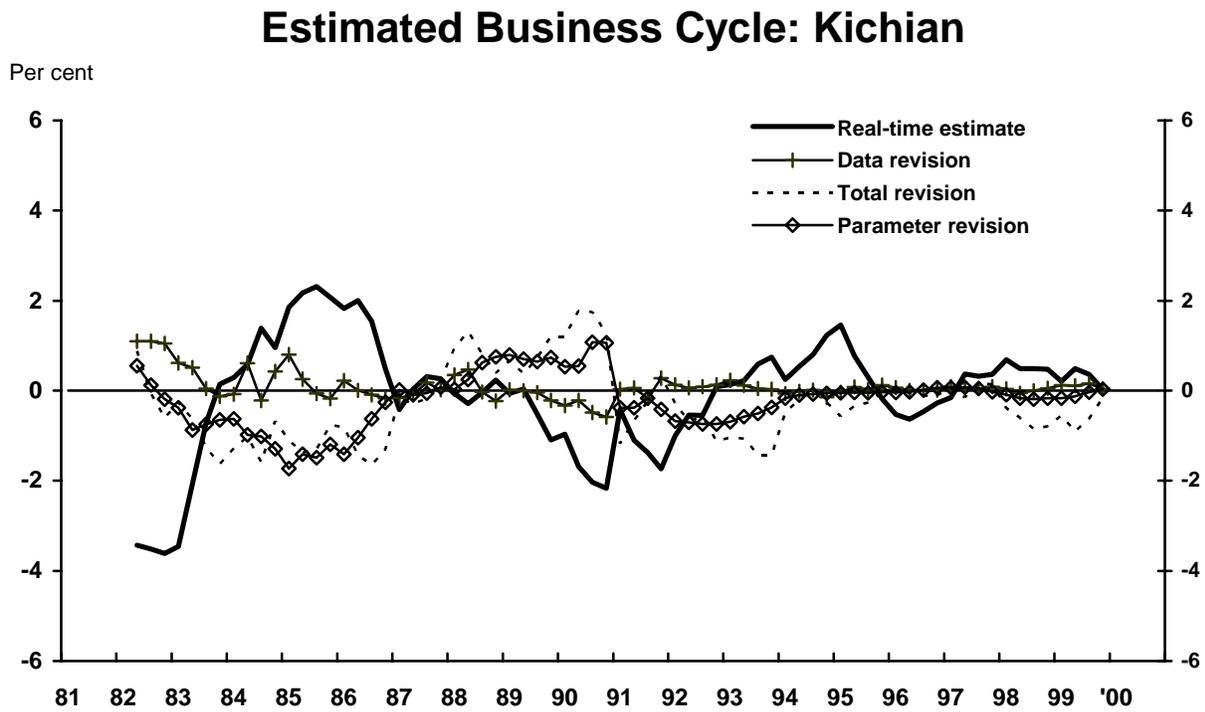


Figure 11

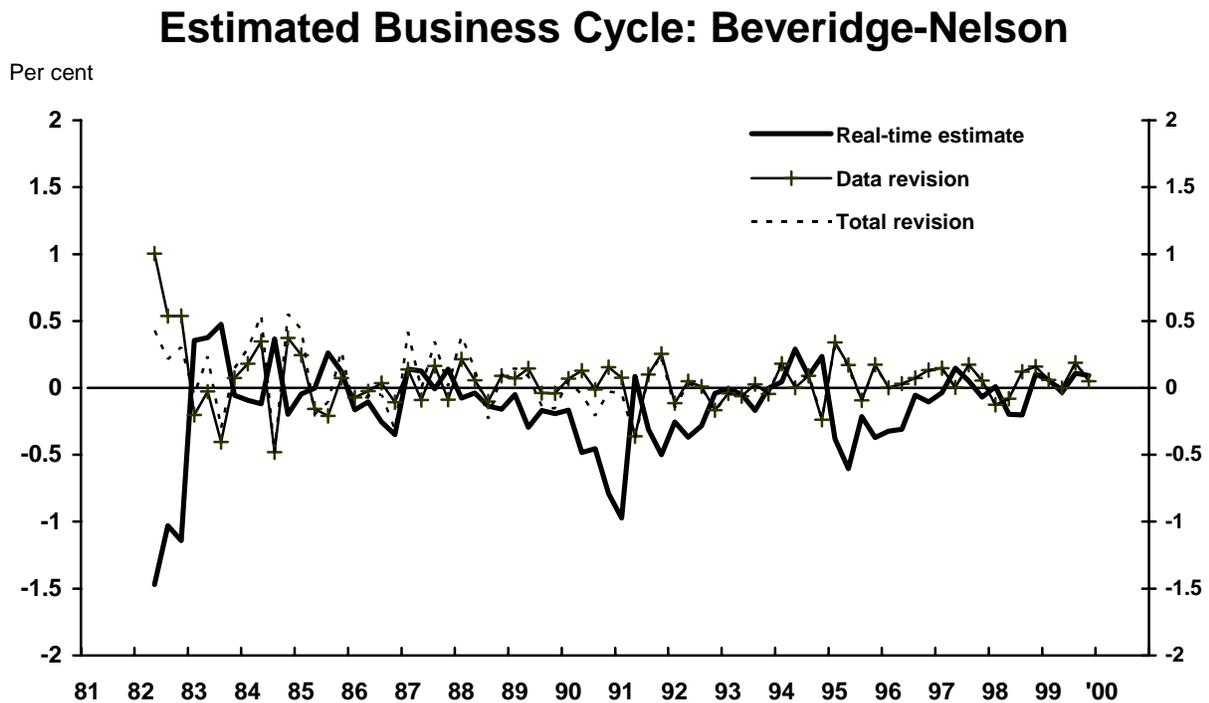


Figure 12

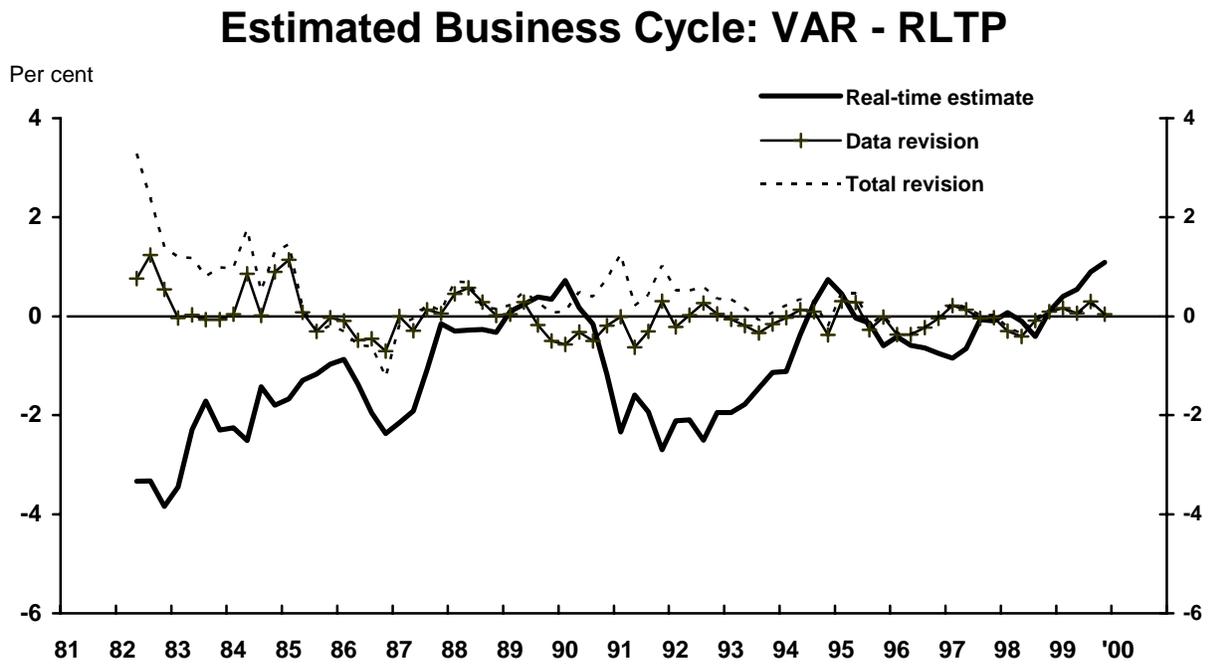
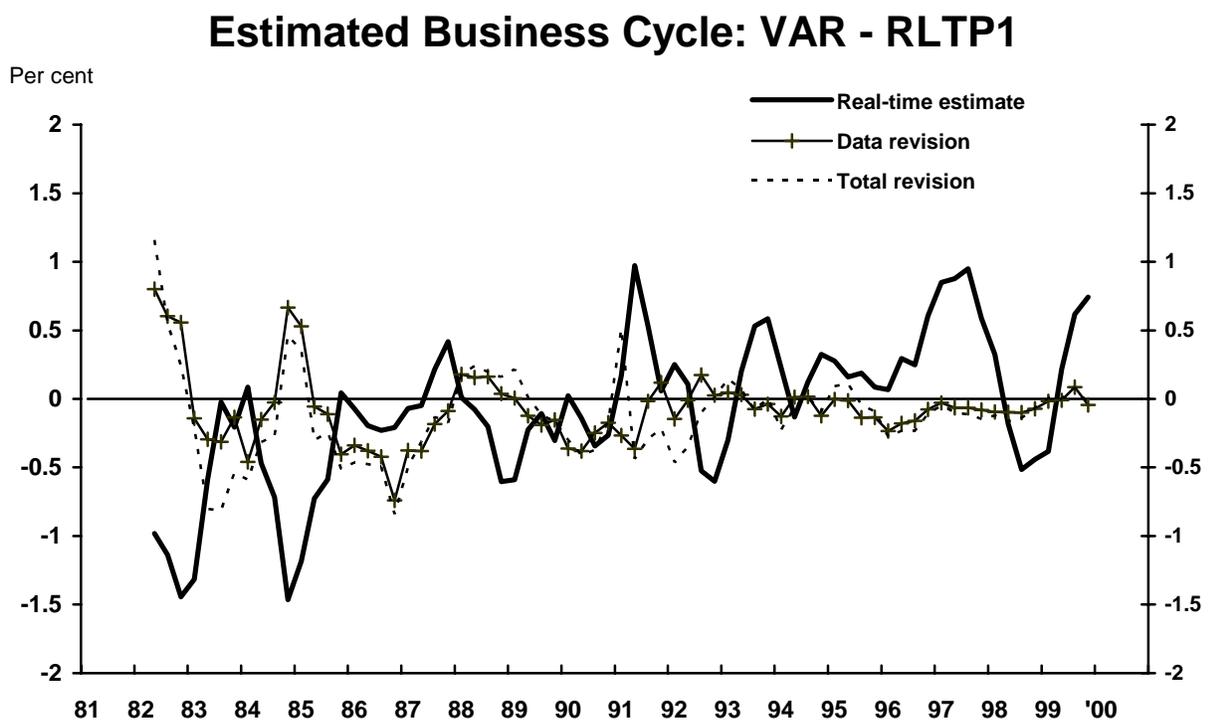


Figure 13



4.3 Second evaluation method: inflation forecasts

Table 10 shows the mean-square errors of the inflation forecast obtained for the 1983Q3 to 2000Q4 period using the procedure described in section 3.3. Each line in the table reflects the various estimated equations. For example, the first line shows the results obtained when the equation used to forecast inflation includes output gap estimates generated by the linear trend, while the last line shows the results obtained when only lags of inflation are included in the equation. The columns in the table compare the performance of different information criteria used to determine the number of lags of each variable used in the forecasting equation. They also compare the out-of-sample forecast performance of real-time and final output gaps.

MÉTHODES	REAL-TIME ESTIMATES			FINAL ESTIMATES		
	AKAIKE	SCHWARZ	OSFE	AKAIKE	SCHWARZ	OSFE
LT	15.87	10.41	6.51	14.35	6.59	4.31
QT	19.33	9.87	6.85	50.04	3.89	3.87
HP	21.90	7.51	5.96	56.36	36.84	3.79
BN	11.47	4.97	2.22	42.76	3.77	2.95
RLTP	12.36	3.91	5.36	25.44	4.27	6.01
RLTP1	14.65	6.83	4.22	8.61	7.26	4.75
W-F	25.23	7.23	4.42	99.47	46.59	5.96
W-S	19.41	9.87	6.12	25.43	7.49	5.18
C-F	24.27	10.28	8.42	25.79	6.50	7.04
C-S	17.00	7.16	5.02	37.09	4.48	4.41
HJ-F	22.90	7.81	11.41	26.87	12.71	11.54
HJ-S	16.13	8.36	8.02	32.74	4.87	4.23
K-F	24.95	11.31	14.16	33.50	6.91	10.34
K-S	26.30	8.51	6.79	23.77	11.28	13.67
NO OUTPUT GAP	4.59	4.45	4.11	4.59	4.45	4.11

* Notes: W-F and W-S represent the filtered and smoothed estimates obtained with the Watson method; C-F and C-S, the filtered and smoothed estimates of the Clark model; HJ-F and HJ-S, the filtered and smoothed estimates of the Harvey-Jaeger model; and K-F and K-S, the filtered and smoothed estimates of the Kichian method.

Table 10 shows that the inclusion of output gap estimates in an equation does not necessarily improve the quality of inflation forecasts. Most of the mean-square-error forecasts of equations that include output gap estimates are larger than the ones that exclude such estimates. In fact,

when real-time estimates are used, only the Beveridge-Nelson and the RLTP methods generate output gap estimates that improve inflation forecasts (this is true only for certain selection criteria). When final output gap estimates are used, only estimations generated by the quadratic trend, the HP filter, and the Beveridge-Nelson decomposition improve inflation forecasts, compared with an equation that excludes such estimates.

Certain information criteria used to determine the number of lags are better than others. For each estimation method, we shade the result of the information criteria that minimizes the mean-square error. We can see that the Akaike technique does not perform as well as the other two methods. The OSFE criterion provides the best forecasts for most techniques.

Also, using final estimates of the output gap does not necessarily improve inflation forecasts. On some occasions, real-time gap estimates provide lower mean-square errors. The fact that both kinds of estimates produce poor inflation forecasts, however, would tend to support the idea that it is more the instability in the inflation-output gap relationship that is responsible for the poor inflation forecast, than the relative imprecision of real-time estimates.²⁸

To assess whether certain forecasts are significantly better than others, we use the modified Diebold and Mariano test. Table 11 compares forecast errors of equations that include output gaps relative to those that exclude such estimates.²⁹ Positive tests mean that output gaps help improve inflation forecasts, while negative numbers mean the opposite. Also, large absolute values for the test mean that the two forecasts are different. Since the test follows a t distribution, the critical value at which we consider the two tests significantly different from each other at 95 per cent is 2.0.

As Table 11 shows, only real-time output gap estimates generated with the Beveridge-Nelson method significantly improve inflation forecasts. Some other techniques also generate output gap estimates that improve the inflation forecast, but these improvements are not significant (at 95 per cent). In fact, most techniques cause a deterioration of inflation forecasts (although they are not significant). These results clearly show that output gap estimates are not a reliable predictor of inflation.³⁰ Indeed, even forecasts obtained with the Beveridge-Nelson output gap are

²⁸ The omission of key variables in the forecast equation could also be partly responsible for the poor forecasting performance of the output gap.

²⁹ We use for each method the forecast obtained with the lag selection criteria, which minimizes the mean-square error. Thus, for most forecasts, the OSFE criteria have been used. A small number of forecasts, however, use the Schwarz criteria (Table 10).

³⁰ Using U.S. data, Orphanides and van Norden (2001a) find that output gap estimates do not significantly improve out-of-sample forecasts of inflation. They also find, however, that output gap estimates do improve in-sample forecasts.

questionable, since there is no consistency between forecasts that use real-time and final output gap estimates.

Table 11: Modified Diebold-Mariano Test (DMD)		
(1983Q3 – 2000Q4)		
Forecast errors of equation with output gap vs forecast errors of equation without output gap		
	REAL-TIME ESTIMATE	FINAL ESTIMATE
Linear trend	-1.03	-0.09
Quadratic trend	-1.67	0.17
Hodrick-Prescott	-0.95	0.27
Beveridge-Nelson	2.22	1.15
VAR – RLTP	0.32	-0.31
VAR – RLTP1	-0.19	-1.14
Watson – filtered est.	-0.15	-0.44
Watson – smoothed est.	-0.58	-0.37
Clark – filtered est.	-1.01	-1.26
Clark – smoothed est.	-0.55	-0.22
Harvey-Jaeger – filtered est.	-1.21	-1.59
Harvey-Jaeger – smoothed est.	-1.12	-0.09
Kichian – filtered est.	-1.10	-0.96
Kichian –smoothed est.	-0.85	-0.81

A natural extension of this work is to compare the inflation forecast produced by the output gap of the different techniques. Tables 12 and 13 show such a comparison for results using real-time and final data, respectively. Tests that have a positive sign mean that the output gap techniques presented in the left column are better predictors of inflation than the techniques presented on the top line. Tests that have a negative sign mean the opposite. For example, the first test in the top left cell of Table 12 shows that the forecasts obtained with the output gap generated by the linear trend are better than those obtained with a quadrati-trend output gap. Again, the critical value for the test is 2.0.

In general, there are no significant differences across forecasts. In fact, only equations that use the Beveridge-Nelson output gap estimates produce significantly better forecasts than other techniques. Also, the multivariate techniques of generating output gap estimates do not help produce significant better forecasts of inflation than univariate methods.

Results shown in Tables 11, 12 and 13 are somewhat surprising. The univariate Beveridge-Nelson method was not very popular during the 1990s, compared with other techniques described in this paper. In fact, we initially thought that multivariate techniques would outperform the other techniques, since they incorporate more information with which to construct output gap estimates. More work should be done to explain why this technique outperforms the others. It should be noted, however, that this results could be unique to Canada. Indeed, Orphanides and van Norden (2001a) show that the Beveridge-Nelson decomposition does not perform as well with U.S. data.

Table 12: Modified Diebold-Mariano Test (DMD)

Real-time estimates

(1983Q3 – 2000Q4)

	LT	QT	HP	BN	RLTP	RLTP1	W-F	W-S	C-F	C-S	HJ-F	HJ-S	K-F
LT	-												
QT	-0.14	-											
HP	0.23	0.53	-										
BN	2.00	4.27	2.52	-									
RLTP	1.21	1.55	1.13	-2.08	-								
RLTP1	0.97	1.30	0.72	-1.47	-0.32	-							
W-F	0.68	0.93	0.64	-1.16	-0.27	-0.08	-						
W-S	0.23	0.19	-0.05	-1.17	-0.71	-0.56	-0.43	-					
C-F	-0.37	-0.40	-0.88	-1.63	-1.09	-0.88	-1.02	-0.38	-				
C-S	0.55	1.54	1.16	-2.54	-0.65	-0.36	-0.27	0.28	1.21	-			
HJ-F	-0.35	-0.38	-1.23	-2.07	-1.22	-1.02	-0.99	-0.34	0.27	-1.88	-		
HJ-S	-0.38	-0.42	-1.01	-1.85	-1.14	-0.96	-0.92	-0.38	0.17	-1.47	-0.17	-	
K-F	-0.66	-0.63	-0.83	-1.40	-1.15	-1.03	-1.53	-0.66	-0.44	-0.98	-0.51	-0.46	-
K-S	-0.07	0.02	-0.49	-1.66	-0.94	-0.70	-0.80	-0.13	1.41	-1.04	0.68	0.65	0.73

* Notes: W-F and W-S represent the filtered and smoothed estimates obtained with the Watson method; C-F and C-S, the filtered and smoothed estimates of the Clark method; HJ-F and HJ-S, the filtered and smoothed estimates of the Harvey-Jaeger method; and K-F and K-S, the filtered and smoothed estimates of the Kichian method.

Table 13: Modified Diebold-Mariano Test (DMD)

Final estimates													
(1983Q3 – 2000Q4)													
	LT	QT	HP	BN	RLTP	RLTP1	W-F	W-S	C-F	C-S	HJ-F	HJ-S	K-F
LT	-												
QT	0.22	-											
HP	0.31	0.07	-										
BN	0.70	1.47	1.13	-									
RLTP	0.02	-0.29	-0.37	-1.22	-								
RLTP1	-0.19	-0.60	-0.67	-1.40	-0.62	-							
W-F	-0.33	-0.50	-0.48	-0.74	-0.43	-0.29	-						
W-S	-0.26	-0.45	-0.52	-0.80	-0.34	-0.14	0.16	-					
C-F	-0.72	-1.27	-1.20	-2.17	-1.24	-0.79	-0.16	-0.38	-				
C-S	-0.08	-0.48	-0.62	-1.98	-0.10	0.22	0.39	0.25	1.06	-			
HJ-F	-1.38	-1.68	-1.68	-2.06	-1.52	-1.36	-1.01	-1.08	-1.52	-1.63	-		
HJ-S	0.07	-0.33	-0.56	-1.52	0.02	0.33	0.39	0.31	1.00	0.37	1.58	-	
K-F	-0.66	-0.96	-0.93	-1.41	-0.99	-0.70	-0.63	-0.43	-0.23	-0.86	1.13	-0.81	-
K-S	-0.76	-0.85	-0.87	-0.99	-0.78	-0.71	-0.52	-0.63	-0.61	-0.79	0.06	-0.80	-0.50

* Notes: W-F and W-S represent the filtered and smoothed estimates obtained with the Watson method; C-F and C-S, the filtered and smoothed estimates of the Clark method; HJ-F and HJ-S, the filtered and smoothed estimates of the Harvey-Jaeger method; and K-F and K-S, the filtered and smoothed estimates of the Kichian method.

5. Conclusion

In this study we have evaluated, using Canadian data, the reliability of real-time estimates of the output gap obtained with different estimation techniques. To do this, we used two evaluation methods of reliability. The first investigated the revisions of real-time estimates of the output gap over time, and the second tested whether the addition of output gap estimates in a simple equation helped improve inflation forecasts.

The results obtained with these two evaluation methods are very conclusive. First, the output gap estimates vary considerably across techniques. This variability indicates that output gap estimates are largely dependent on the assumptions used to explain potential output. It also shows that the errors that result from choosing the wrong model can potentially be large.

Second, the results of the two evaluation methods show that the reliability of real-time estimates of the output gap generated by the different estimation techniques tends to be quite low. In fact, ex post revisions are of the same order of magnitude as the ex post estimates of the output gap and the real-time estimates frequently misclassify the sign of the gap. Also, the real-time output gap estimates generated from the different techniques, except for the Beveridge-Nelson decomposition, do not improve inflation forecasts. In fact, some techniques generate output gap estimates that worsen inflation forecasts.

Despite the inability of the different techniques to produce reliable estimates, some of their characteristics are worth noting. First, for all estimation techniques except the Beveridge-Nelson decomposition, the revisions to the real-time estimates of the output gap are very persistent. This means there is quite a long time lag before the full magnitude of the estimate errors can be realized.

Although important, the revision of published data does not appear to be the primary source of revisions for the univariate methods (except for the Beveridge-Nelson decomposition). Rather, the bulk of the problem is due to the pervasive unreliability of end-of-sample estimates of the output trend. Thus, even if the reliability of the underlying real-time data were to improve, real-time estimates of the output gap would remain unreliable for these methods.

Conversely, estimates generated by a structural VAR and by the Beveridge-Nelson decomposition are more affected by data revisions. In fact, this form of revision is responsible for a large part of the total revision of these techniques. If data were never revised, estimates generated by these techniques would face small revisions compared with the magnitude of the output gap estimates.

For the Kichian model, total revisions seem to be caused mostly by revisions to the coefficients, which means that the coefficients of the model are not stable over time.

The results of this study will be useful for researchers who incorporate the output gap in their work and for policy-makers who use output gap estimates as an indicator of inflationary pressures in the economy. First, economists should not ignore the strong uncertainty that surrounds output gap estimates as a result of the large and persistent revisions that affect end-of-sample estimates of the output gap when they incorporate this variable in their models. This is particularly true for models that evaluate the performance of fiscal or monetary policies over time, or that test different policy rules. Decisions that seem optimal in real-time may actually be harmful for the economy when more information is available.

In future research, it would be interesting to study how different assumptions that surrounds the persistence of output gap estimates and the persistence of revisions affect the dynamic of complex macroeconomic models. It would also be interesting to apply our methodology to other methods of estimating the output gap, such as the multivariate filter used at the Bank of Canada, or other types of gaps such as the gap between the unemployment rate and the non-accelerating inflation rate of unemployment (NAIRU), and the gap between the capacity utilization rate and the non-accelerating inflation capacity utilization rate (NAICU).

Appendix

Restrictions Imposed to Identify Structural Shocks in the VAR

To recover the structural shocks from the reduced-form residues, Lalonde, Page, and St-Amant (1998) impose a certain number of identification restrictions. First, consider the following reduced form of the VAR:

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_\pi \\ \mu_r \end{bmatrix} + \begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) & \phi_{13}(L) \\ \phi_{21}(L) & \phi_{22}(L) & \phi_{23}(L) \\ \phi_{31}(L) & \phi_{32}(L) & \phi_{33}(L) \end{bmatrix} \cdot \begin{bmatrix} \Delta y_{t-1} \\ \Delta \pi_{t-1} \\ \Delta r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{r,t} \end{bmatrix}, \quad (\text{A1})$$

which can be rewritten as: $Z_t = \mu + \phi(L) \cdot Z_{t-1} + \varepsilon_t$. Since vector Z_t is composed only of stationary variables, using the Wold theorem we can decompose it as follows:

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_\pi \\ \mu_r \end{bmatrix} + \begin{bmatrix} C_{11}(L) & C_{12}(L) & C_{13}(L) \\ C_{21}(L) & C_{22}(L) & C_{23}(L) \\ C_{31}(L) & C_{32}(L) & C_{33}(L) \end{bmatrix} \cdot \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{r,t} \end{bmatrix}, \quad (\text{A2})$$

which corresponds to $Z_t = \mu + C(L) \cdot \varepsilon_t$. $C(L) = \sum_{i=0}^{\infty} C_i L^i$ is the moving average representation of the reduced form of the VAR and ε_t is the corresponding vector of residuals estimated with the reduced form of the VAR. In this equation, $C(0) = I_n$, where I_n is the identity matrix. Also, $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Omega$, where Ω is the positive definite covariance matrix of the residuals.

The goal of the identification process is to recover the following structural moving average representation: $Z_t = \mu + \Gamma(L) \cdot \eta_t$, or

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_\pi \\ \mu_r \end{bmatrix} + \begin{bmatrix} \Gamma_{11}(L) & \Gamma_{12}(L) & \Gamma_{13}(L) \\ \Gamma_{21}(L) & \Gamma_{22}(L) & \Gamma_{23}(L) \\ \Gamma_{31}(L) & \Gamma_{32}(L) & \Gamma_{33}(L) \end{bmatrix} \cdot \begin{bmatrix} \eta_{y,t} \\ \eta_{\pi,t} \\ \eta_{r,t} \end{bmatrix}, \quad (\text{A3})$$

where $\Gamma(L)$ is the structural moving average representation and η_t is the vector of structural shocks. It is also assumed that $E(\eta_t) = 0$ and that $E(\eta_t \eta_t') = I_n$, thus the shocks are normalized and orthogonal.

From equations (A2) and (A3), we can see that $C(L) \cdot \varepsilon_t = \Gamma(L) \cdot \eta_t$. From this equality, we can deduct the following relationship: $\Gamma(0) \cdot \Gamma(0)' = \Omega$; $\eta_t = \Gamma(0)^{-1} \cdot \varepsilon_t$; and $\Gamma(L) = C(L) \cdot \Gamma(0)$. Thus, to

isolate the structural shocks (η_t) and the structural moving average representation ($\Gamma(L)$), we must first identify the matrix $\Gamma(0)$. Such identification can be done by making similar assumptions as those proposed by Blanchard and Quah (1989).

Let $\Gamma(I)$ and $C(I)$ be the long-run multiplier matrices of structural and reduced-form shocks respectively. The Blanchard and Quah approach imposes that certain shocks do not have long-run effects on output. In our case, $htcp$ and htc would be such shocks. If we also assume that some shocks do not have long-run effects on inflation (η_t^c), the matrix $\Gamma(I)$ becomes lower triangular. From equations (A2) and (A3), we can find the following equality: $C(I) \cdot \Omega \cdot C(I)' = \Gamma(I) \cdot \Gamma(I)'$. Thus, by applying the Choleski decomposition to the left side of this identity, we obtain a lower triangular matrix that corresponds to $\Gamma(I)$. It is also possible to rewrite this identity in the following form: $\Gamma(0) = C(I)^{-1} \cdot \Gamma(I)$, since $\Gamma(0) \cdot \Gamma(0)' = \Omega$. Knowing $\Gamma(0)$, it is possible to isolate the structural shocks (η_t) and the structural moving average representation ($\Gamma(L)$). With this information, we can also recover the following equation (see section 2.2.4):

$$\Delta y_t = \mu_y + \Gamma_y^p(1)\eta_t^p + \Gamma_y^{p*}(L)\eta_t^p + \Gamma_y^{cp}(L)\eta_t^{cp} + \Gamma_y^c(L)\eta_t^c. \quad (A4)$$

References

- AKAIKE, H. (1973). « Information Theory and an Extension of the Maximum Likelihood Principal ». In : 2nd International Symposium on Information Theory, publié sous la direction de B. N. Petrov et F. Csaki, Budapest, p. 267-281.
- AMANO, R., D. COLETTI et T. MACKLEM (1999). « Monetary Rules When Economic Behaviour Changes », document de travail 99-8, Banque du Canada, Ottawa.
- ARMOUR, J., W. ENGERT et B. S.C. FUNG (1996). « Overnight Rate Innovations as a Measure of Monetary Policy Shocks in Vector Autoregressions », document de travail 96-4, Banque du Canada, Ottawa.
- BEVERIDGE, S., et C. R. NELSON (1981). « A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle », *Journal of Monetary Economics*, vol. 7, no 2, p. 151-174.
- BLANCHARD, O., et D. QUAH (1989). « The Dynamic Effects of Demand and Supply Disturbances », *The American Economic Review*, vol. 79, no 4, p. 655-673.
- BOX, G., et G. JENKINS (1976). *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco.
- BUTLER, L. (1996). « The Bank of Canada's New Quarterly Projection Model (QPM), Part 4. A Semi-Structural Method to Estimate Potential Output : Combining Economic Theory with a Time-Series Filter », Rapport technique no 77, Banque du Canada, Ottawa.
- CERRA, V., et S. C. SAXENA (2000). « Alternative Methods of Estimating Potential Output and the Output Gap : An Application to Sweden » IMF Working Paper 00/59, Fonds monétaire international, Washington.
- CLARK, P. K. (1987). « The Cyclical Component of U.S. Economic Activity », *Quarterly Journal of Economics*, vol. 102, no 4, p. 797-814.
- CLAUS, I., P. CONWAY, A. SCOTT (2000). « The Output Gap : Measurement, Comparisons and Assessment », Research Paper no 44, Reserve Bank of New Zealand, Wellington.
- CROUSHORE, D., et T. STARK (2000). « A Funny Thing Happened on the Way to the Data Bank : A Real-Time Data Set for Macroeconomist », *Federal Reserve Bank of Philadelphia Business Review*, septembre/octobre, Philadelphia.
- DIEBOLD, F. X., et R. S. MARIANO (1995). « Comparing Predictive Accuracy », *Journal of Business & Economic Statistics*, vol. 13, no 3, p. 253-263.
- DUPASQUIER, C., A. GUAY et P. ST-AMANT (1997). « A Comparison of Alternative Methodologies for Estimating Potential Output and the Output Gap », document de travail 97-5, Banque du Canada, Ottawa.

- EHRMANN, M. et F. SMETS (2001). « Uncertain Potential Output : Implications for monetary Policy », ECB Working Paper Series, no 59, Banque centrale européenne, Francfort.
- ENDERS, W. (1995). Applied Econometric Time Series, John Wiley & Sons, inc, États-Unis.
- GAIDUCH, V., et B. HUNT (2000). « Inflation Targeting Under Potential Output Uncertainty », IMF Working Paper 00/158, Fonds monétaire international, Washington.
- GERLACH, S., et F. SMETS (1997), « Output Gaps and Inflation : Unobservable-Components Estimates for the G-7 Countries », document non publié, Banque des Règlements Internationaux, Bâle.
- HALDANE, A. G., et N. BATINI (1998). « Forward-Looking Rules for Monetary Policy », NBER Working Paper Series, no 6543, National Bureau of Economic Research, Cambridge.
- HARVEY, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge.
- HARVEY, A. C., et A. JAEGER (1993). « Detrending, Stylized Facts and the Business Cycle », Journal of Applied Econometrics, vol 8, no 3, p. 231-247.
- HARVEY, D. I., S. J. LEYBOURNE et P. NEWBOLD (1997). « Testing the Equality of Prediction Mean Squared Errors », International Journal of Forecasting, vol. 13, no 2, p. 281-291.
- HODRICK, R. J., et E. C. PRESCOTT (1997). « Post-war Business Cycle : An Empirical Investigation », Journal of Money, Credit, and Banking, vol. 29, no 1, p. 1-16.
- KICHIAN, M. (1999). « Measuring Potential Output within State-Space Framework », document de travail 99-9, Banque du Canada, Ottawa.
- KICHIAN, M. (2000), « GaussTM Programs for the Estimation of State-Space Models with ARCH Errors : A User's Guide », document de travail 2000-2, Banque du Canada, Ottawa.
- KUTTNER, K. N. (1994). « Estimating Potential Output as a Latent Variable », Journal of Business and Economic Statistics, vol. 12, no 3, p. 361-368.
- LALONDE, R. (1998). « Le PIB potentiel des États-Unis et ses déterminants : la productivité de la main d'œuvre et le taux d'activité », document de travail 98-13, Banque du Canada, Ottawa.
- LALONDE, R., J. PAGE et P. ST-AMANT (1998), « Une nouvelle méthode d'estimation de l'écart de production et son application aux États-Unis, au Canada et à l'Allemagne », document de travail 98-21, Banque du Canada, Ottawa.
- LAXTON, D., et R. TETLOW (1992), « A Simple Multivariate Filter for Measurement of Potential Output », Rapport technique no 59, Banque du Canada, Ottawa.
- McCALLUM, B. T. (2001), « Should Monetary Policy Respond Strongly to Output Gap », NBER Working Paper Series, no 8226, National Bureau of Economic Research, Cambridge.

MURCHISON, S. (1999), « NAOMI - A New Quarterly Forecasting Model - Part 1 : Proposed Model Selection Strategy », Working paper 2001-19, Ministère des Finances du Canada, Ottawa.

NELSON, C., et C. PLOSSER (1982). « Trends and Random Walks in Macroeconomic Time Series : Some Evidence and Implications », *Journal of Monetary Economics*, vol 10, p. 130-162.

ORPHANIDES, A. (1998). « Monetary Policy Evaluation With Noisy Information », *Finance and Economics Discussion Series*, 1998-50, Federal Reserve Board, Washington.

ORPHANIDES, A. (2000). « The Quest for Prosperity Without Inflation », *ECB Working Paper Series*, no 15, Banque centrale européenne, Francfort.

ORPHANIDES, A., R. D. Porter, D. Reifschneider, R. Tetlow et F. Finan (2000). « Errors in the Measurement of the Output Gap and the Design of Monetary Policy », *Journal of Economics and Business*, vol. 52, no 1-2, p. 117-141.

ORPHANIDES, A., et S. VAN NORDEN (1999). « The reliability of Output Gap Estimates in Real Time », *Finance and Economics Discussion Series*, 1999-38, Federal Reserve Board, Washington.

ORPHANIDES, A. et S. VAN NORDEN (2001a). « The reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time », mimeo.

ORPHANIDES, A. et S. VAN NORDEN (2001b). « The Unreliability of Output Gap Estimates in Real Time », *Working Paper 2001s-57*, CIRANO, Montréal.

PERRON, P. (1989). « The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis », *Econometrica*, vol. 57, p. 1361-1401.

ROBERTSON, J. C. et E. TALLMAN (1998). « Data Vintages and Measuring Forecast Model Performance » *Federal Reserve Bank of Atlanta Economic Review*, vol. 83, no 4, Atlanta.

RUDEBUSCH, G. D. (1999). « Is the Fed too Timid? Monetary Policy in an Uncertain World », *Federal Reserve Bank of San Francisco Working Paper 99-05*, San Francisco.

SCHWARZ, G. (1978). « Estimating the Dimension of a Model », *The Annals of Statistics*, vol. 6, no 2, p. 461-464.

SMETS, F. (1998). « Output Gap Uncertainty : Does it Matter for the Taylor Rule? », *BIS Working Paper no 60*, Banque des règlements internationaux, Bâle.

ST-AMANT, P., et S. VAN NORDEN (1997). « Measurement of the Output Gap : A discussion of recent research at the Bank of Canada », *Rapport technique no 79*, Banque du Canada, Ottawa.

STATISTIQUE CANADA. « Comptes nationaux des revenus et dépenses », no 13-001-XPB, Document publié à chaque trimestre depuis 1961, Ottawa.

STATISTIQUE CANADA (1972). Supplément annuel de la section 1 - Revue statistique du Canada 1971, Ottawa.

STATISTIQUE CANADA (1976). « Les comptes nationaux des revenus et des dépenses Volume 2 - Les estimations trimestrielles 1947-1974 », Catalogue 13-533 Hors série, Ottawa.

STATISTIQUE CANADA (1989). « Comptes nationaux des revenus et dépenses - Estimations trimestrielles 1947-1986 », Catalogue 13-533 Hors série, Ottawa.

SVENSSON, L. E. O., et M. WOODFORD (2000). « Indicator Variables for Optimal Policy », ECB Working Paper Series, no 12, Banque centrale européenne, Francfort.

SWANSON, E. T. (2000). « On Signal Extraction and Non-Certainty-Equivalence in Optimal Monetary policy Rules », Finance and Economics Discussion Series, 2000-32, Federal Reserve Board, Washington.

TAYLOR, J. (1993). « Discretion versus policy rules in Practice », Carnegie-Rochester Conference Series on Public Policy, vol. 39, p. 195-214.

VAN NORDEN, S. (1995). « Why Is It So Hard to Measure the Current Output Gap », Banque du Canada, mimeo.

YETMAN, J. (2000). « Probing Potential Output : monetary Policy, Credibility, and Optimal Learning under Uncertainty », document de travail 2000-10, Banque du Canada, Ottawa.