Asymmetric output cost of lowering inflation: empirical evidence for Canada

Hyeon-seung Huh  School of Economics, Hallym University
Hyun-Hoon Lee  Division of Economic and International Trade, Kangwon National University

Abstract. A strand of theoretical and empirical evidence in the literature suggests non-linearity in the output-inflation relationship, viz. a non-linear Phillips curve. We develop a VAR model of output, inflation, and terms of trade augmented with logistic smooth transition autoregression specifications. Empirically, the model captures non-linear features present in the data. Output costs of reducing inflation vary, depending on the economy, size of inflation change, and whether policy makers seek to disinflate or prevent inflation from rising. Thus, inferences based on the conventional linear Phillips curve may provide misleading signals about the cost of lowering inflation and the appropriate policy stance. JEL Classification: C32, E52

Le coût asymétrique en termes de production de la réduction de l’inflation : résultats pour le Canada. Ce mémoire a son origine dans les travaux qui suggèrent une certaine non linéarité dans la relation production-inflation i.e. une courbe de Phillips qui serait non-linéaire. Les auteurs utilisent un modèle VAR de la production, de l’inflation et des termes d’échange, enrichi de spécifications autogressives définissant une transition logistique souple. Il appert que les coûts en termes de perte de production de la réduction de l’inflation varient grandement selon l’état de l’économie, la taille des changements recherchés dans le taux d’inflation, et selon que les autorités cherchent créer une déflation ou simplement à empêcher l’accélération de l’inflation. Voilà qui implique que les inférences tirées d’un modèle construit sur la courbe linéaire traditionnelle de Phillips peuvent fournir des signaux trompeurs quant aux coûts de la réduction de l’inflation et suggérer de mauvaises politiques.

1. Introduction

One of the key costs of achieving low inflation is the short-term output loss that generally accompanies a permanent decline in inflation. Particularly stark examples

The authors wish to thank two anonymous referees for very constructive comments. This research was supported by the Hallym Academy of Sciences at Hallym University. Email: hshuh@hallym.ac.kr.

Canadian Journal of Economics / Revue canadienne d’Economique, Vol. 35, No. 2
May / mai 2002. Printed in Canada / Imprimé au Canada
0008-4085 / 02 / 218–238 / © Canadian Economics Association
of this output cost were seen in the early 1980s and 1990s when disinflations in both periods were accompanied by severe recessions. Obviously, policymakers' decisions on the timing and extent of inflation reduction depend on balancing the costs and benefits of moving to a new, lower level of inflation. The issue has become more relevant since the beginning of the 1990s, when several countries, including Canada, explicitly committed themselves to low inflation targets. Under this inflation targeting regime, policy makers have the tasks of fulfilling low inflation and at the same time, minimizing the necessity of generating large recessions down the track.

Undoubtedly, accurately assessing the output cost of reducing inflation is crucial as only with accurate measures can the net benefits of reducing inflation reliably be assessed. A standard approach in the literature is to use the Phillips curve to estimate the so-called 'sacrifice ratio,' which measures the cumulative loss in output associated with a one percentage point permanent reduction in inflation. A sacrifice ratio of 10 per cent reported by Okun (1978) for the United States is a well-known example. Typically, this short-run trade-off between output and inflation is assumed to be constant under the proposition that the shape of the Phillips curve is linear. Subsequent estimates of the output cost have been conducted in a linear framework.

A strand of the theoretical literature, however, suggests the non-linear nature of the Phillips curve. In the monopolistic competitive model, for example, the reduced sensitivity of inflation as the economy strengthens implies that the shape of the Phillips curve is concave. In the capacity constraints model, on the other hand, increased sensitivity of inflation to the economy's strength is in accordance with the convex Phillips curve. There are also models suggesting that the relationship between output and inflation may vary with the level of inflation. The menu cost model, for example, implies that firms increase the frequency and size of price adjustment as inflation rises. The resultant lesser effect on output and greater effect on the price level are consistent with the convex Phillips curve. (See Dupasquier and Ricketts 1998 for a comprehensive review of the microfoundations.).

Empirical evidence supporting non-linearities in the output-inflation relationship has also been mounted in recent times. Laxton, Rose, and Tambakis (1993), Debelle (1996), and Dupasquier and Ricketts (1998) are among the studies that report such non-linear relationships for Canada. Laxton et al., and Dupasquier and Ricketts find that the output cost of reducing inflation is much smaller in periods of excess demand than in periods of excess supply. Using quarterly data, for example, Dupasquier and Ricketts report that the output cost per percentage point reduction in inflation is 1.96 per cent in periods of excess demand, compared with 12.5 per cent in periods of excess supply. They also report that the output cost is estimated to be 3.22 at high levels of inflation, but 7.69 per cent at low or moderate levels of inflation. This negative relationship between the initial level of inflation and the resulting output cost was also reported in Debelle, who examined the last three Canadian disinflations. Ball (1994) and Jordan (1997) conclude in their multi-country studies that the output costs for reducing inflation vary with the initial level of inflation in various countries, including Canada. Depending on particular epi-
sodes, the output costs for Canada range between 0.14 and 2.9 per cent in a disinflation regime, but between −3.1 and 9.3 per cent in an inflation regime. All of these studies were performed in a single-equation framework. As Cecchetti (1994) and Rowe (1998) comment, the well-known problem of endogeneity in estimating the Phillips curve is an outstanding issue for these studies. Further, their studies are limited in the sense that the output-inflation relationship is allowed to vary only with the states of the economy. In a recent study, Filardo (1998) presents a more comprehensive approach by employing Tong’s (1983) threshold autoregression models in a multivariate framework. In his model, the output costs of lowering inflation are allowed to vary with signs and sizes of the shocks in addition to the states of the economy. With application to the U.S. data, he finds that the output costs of reducing inflation are indeed dependent on these factors.

Our purpose in this paper is to extend Filardo’s study to the case of Canada. However, there is a key difference. Our VAR model is constructed to accommodate a potentially important departure from linearity through a logistic smooth transition autoregression (LSTAR) specification. The LSTAR specification allows the model to alternate between different regimes, with linear and discrete non-linear cases as extreme ends. It is important to note that the transition is carried out in a smooth manner so that there can be a continuum of states between the regimes. This contrasts with Filardo’s model, where regime switching is entirely discrete. The use of smooth transition in the present application, as opposed to discrete regime switching, is justified particularly by the fact that slow adjustments and inertia in inflation and consumers’ expectations are the main reasons for the output cost of lowering inflation. Further, consumers may exert different degrees of inertia and so will adjust with different time lags. When aggregate behaviour is considered, the time path of regime changes is likely to be better captured by a model which permits gradual rather than instantaneous adjustment.

Our baseline VAR model consists of real GDP, the inflation rate and the terms of trade in line with Gordon and King (1982), King and Watson (1994), Cecchetti (1994), and Cecchetti and Rich (1999). The additional inclusion of the terms of trade variable in the model is motivated by Schelde-Anderson (1992) and Ball (1994), who identified the changes in that variable as a potentially important factor influencing the output costs through the inflationary process. The underlying shocks in the model are identified by imposing long-run restrictions and exogeneity con-

---

1 Inflation tends to move slowly over time, generating a great deal of persistence and inertia. Consumers’ expectations may also adjust slowly over time, perhaps being based on some sort of adaptive mechanism. Because decisions about wages and prices depend on expectations of future changes, slow adaptation is self-fulfilling, creating inertia.

2 A rise in import prices, for example, will feed into consumer price inflation possibly with a lag, which would increase output costs to achieve a given disinflation. Further, if it sparks consumers’ inflation expectations through the associated depreciation in the exchange rate, lowering inflation could be more costly. Given that Canada is a small open economy, those effects of terms of trade on the output costs may be particularly relevant.
ditions of the type used for VAR models by extending the procedure by Shapiro and Watson (1988) to the case involving LSTAR specifications.\footnote{3}

The remainder of this paper is organized as follows. In section 2 we present further discussion about the issues the paper seeks to answer. In section 3 we discuss a structural model, which constitutes our empirical analysis. The empirical results of this study are given in section 4. In section 5 we draw policy implications from the empirical results, and present some conclusions.

2. Some preliminary considerations

If the Phillips curve is linear, the cost of a deflation in terms of temporarily reduced output is exactly equal to the benefit of the higher output that caused inflation in the first place. The costs of reducing inflation are also the same, regardless of when the response of monetary policy is taken. As the timing is of little consequence to the final outcome, there is no first-order cost of delaying response to inflationary pressures. Moreover, the variability of output will not affect the average level of output. This implies that monetary policy may affect the variability of output, but not its average level. Hence, policy mistakes are unlikely to induce first-order welfare losses unless output variability is directly in the policy maker’s objective function (De Long and Summers 1988).

Different results emerge if the Phillips curve is non-linear. In the case of a convex Phillips curve (i.e., capacity constraints model), the output costs of a deflation are larger than the benefits of greater output from a similarly sized inflation. This implies that early action to counteract inflationary pressures can reduce the need to take stronger disinflationary action later. Inversely, delaying the response results in higher inflation and necessitates a significantly stronger monetary tightening to bring inflation back under control. Moreover, the variability of output has a direct effect on its average level. As Clark, Laxton, and Rose (1996) and Nobay and Peel (2000) highlight, a monetary policy that lowers the variability of output will also increase the average level of output.\footnote{4} Thus, success at stabilizing the output cycle will generate first-order welfare gains. Policy errors that raise the variability of output would have the opposite result.

The pictures reverse if the Phillips curve is concave (i.e., a monopolistic competitive model). The increase in output that generates higher inflation would be greater than the subsequent decline in output that would return inflation to its previous rate. Hence, there is little incentive to move early to combat inflationary pressures. Instead, there may be preferences towards aggressive probing for the limits of capacity because the consequences of overshooting decline on the margin.

3 The use of long-run identifying restrictions is different from Filardo (1998), who adopts Cholesky-type contemporaneous identifying restrictions.

4 In related analyses, Debelle and Laxton (1997), and Laxton, Rose, and Tambakis (1999) illustrate this argument in terms of the unemployment costs of lowering inflation.
Moreover, concavity implies that increasing the variability of output (i.e., larger amplitudes of business cycles) will also raise the average level of output.

These diametrically different implications render the task of inferring the shape of a non-linear Phillips curve especially important. Yet there is no consensus in the literature regarding the precise non-linear form of the Phillips curve. Non-linear shapes implied by explicit micro underpinnings may also be intractable in practice. Further, Dupasquier and Ricketts (1998) suggest that a model nesting more than one type of non-linearity may be needed to fit the data better. For these reasons, we do not attempt to estimate directly one specific model or another. We take a rather broad approach through the VAR-LSTAR model. This specification is sufficiently flexible to allow various non-linear Phillips curve shapes. It also allows shapes that are convex in one region and concave in another region (i.e., a kinked curve).

Following previous studies, the output cost of reducing inflation is defined as a ratio of the output response relative to the inflation response with respect to an innovation to domestic demand. Hereafter, this is referred to as the COFI ratio, following Filardo (1998). In the model, the impulse response functions (hence COFI ratios) are dependent on the signs and sizes of the shocks and the initial states before these shocks hit the system. The results of the estimated LSTAR model are summarized along three dimensions. First, do positive shocks to domestic demand have different effects from negative shocks? Second, do domestic demand shocks have different effects at different points in the business cycle, for example, when output growth is initially low (high) or when inflation is initially falling (rising)? Third, do shocks of different magnitudes have disproportionate effects? A typical linear model produces a single coefficient, since it is independent of those factors. Our specification can draw a richer picture of the output cost than a standard linear Phillips curve. Further, if the COFI ratio is indeed shown to have such asymmetries, inferences from the linear Phillips curve may provide misleading signals about the cost of reducing inflation and thus the appropriate policy stance to be taken.

The first and second questions are directly related to the conventional concepts of sacrifice and benefice ratios. The sacrifice ratio typically refers to the output cost for a permanent decline in inflation during a disinflation episode, whereas the benefice ratio typically refers to the output gain for a permanent rise in inflation during an episode of accelerating inflation. Both ratios are equal in absolute value for a linear case (i.e., a linear Phillips curve). In our impulse response framework, the response of output relative to the response of inflation following a negative shock to domestic demand can be regarded as a measure for the output cost of lowering inflation. The conventional sacrifice ratio corresponds to this measure when the economy is initially in a disinflation regime. On the other hand, the output response relative to the inflation response following a positive shock to domestic demand can be regarded as a measure for the output gain by allowing for higher inflation. The conventional benefice ratio corresponds to this measure when the economy is initially in an accelerating inflation regime. We follow Filardo (1998) and relate the negative of this benefice ratio to the cost of resisting incipient inflation. It can be thought of as the forgone loss of output that would have accompanied the rise in inflation.
The third question is related to the macroeconomic policy central to the choice between ‘gradualism’ and ‘cold turkey.’ One view is that gradualism is less costly because of wages and prices inertia and the need for time to adjust to monetary tightening. This view has been formalized by Taylor (1983), who presents a model of staggered wage adjustment in which quick disinflation reduces output but slow disinflation does not. A contrary view, that disinflation is less costly if it is quick, is known as the cold turkey strategy. According to Sargent (1983) and Ball (1994), rapid disinflation produces credibility and hence a shift in expectations that makes disinflation less costly. While there is no consensus regarding the choice between gradualism and cold turkey, we shed some light on the matter, since the output cost of reducing inflation is allowed to vary with the speed of a given disinflation (or inflation).

A caveat should be mentioned before we proceed to empirical analysis. While various theoretical models suggest the presence of non-linearities in the Phillips curve, there is good reason to believe that the empirically estimated Phillips curve may have changed its slope in different monetary policy regimes. If the changes in regime happened to coincide with different macroeconomic outcomes (as seems very likely), then changes in the slope of the Phillips curve that were due to regime shifts may be attributed instead to non-linearities. This raises a possibility that empirical evidence in favour of the departure from linearity may reflect merely changing structure resulting from those regime shifts, not the endogenously generated non-linearities in the Phillips curve. Nevertheless, we note that there are two major difficulties in tackling this possibility. First, there is no official source for the dates of changes in monetary policy regime. Accordingly, identifying different monetary policy regimes itself will involve a degree of arbitrariness. Second, given the typical size of macroeconomic data, it will be very difficult to draw an empirical distinction between a linear model with regime switching and a purely non-linear model. We conjecture that most of empirical investigations drawing on the non-linear Phillips curve, including ours, are likely to be subject to this critique. In this light, we do not pursue the matter further, but acknowledge it as a possible concern.5

3. Empirical model

3.1. Baseline linear VAR model

Consider a three-variable VAR model that comprises the terms of trade ($\tau t$), real output ($y$) and the inflation rate ($\pi$), all expressed in their first differences. There are assumed to be three structural disturbances in the system: a terms of trade shock ($\epsilon_{\tau t}$), a domestic supply shock ($\epsilon_{yt}$), and a domestic demand shock ($\epsilon_{\pi t}$). They are assumed to have a mean of zero and to be contemporaneously uncorrelated; that is, $E(\epsilon_i \epsilon_j) = 1$ for $i = j$, and $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$. It is also assumed that the terms of trade are exogenous in the short run and long run, which reflects the fact that

5 We thank the referee for this comment.
Canada is a small open economy. Under this assumption, none of the domestic shocks, $\epsilon_{yt}$ and $\epsilon_{\pi_t}$, will have any impact on the terms of trade at all horizons. As a consequence, the dynamic interactions of real output and inflation in response to $\epsilon_{yt}$ and $\epsilon_{\pi_t}$, which constitute our major focus, can be equivalently evaluated from the following two-variable model:

$$\Delta y_t = \sum_{i=0}^{p} a_{yT,i} \Delta T_{t-i} + \sum_{i=1}^{p} a_{y\pi,i} \Delta y_{t-i} + \sum_{i=0}^{p} a_{y\pi,i} \Delta \pi_{t-i} + \epsilon_{yt}$$  (1a)

$$\Delta \pi_t = \sum_{i=0}^{p} a_{\pi T,i} \Delta T_{t-i} + \sum_{i=1}^{p} a_{\pi y,i} \Delta y_{t-i} + \sum_{i=1}^{p} a_{\pi \pi,i} \Delta \pi_{t-i} + \epsilon_{\pi t},$$  (1b)

where $\Delta = (1 - L)$, and $L$ is the lag operator. By construction, $\Delta T_{t-i}$ is uncorrelated with the domestic disturbances $\epsilon_{yt}$ and $\epsilon_{\pi t}$. Note that the above model specification has a smaller number of free parameters relative to the unrestricted three-variable model. This feature will prove beneficial in dealing with the non-linear version of the model below, in which the number of free parameters is doubled.

Now, consider a vector moving average (VMA) expression corresponding to the system described by (1a) and (1b) as

$$\begin{pmatrix} \Delta y_t \\ \Delta \pi_t \end{pmatrix} = \Gamma(L) \epsilon_t = \begin{pmatrix} \Gamma_{11}(L) & \Gamma_{12}(L) \\ \Gamma_{21}(L) & \Gamma_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{yt} \\ \epsilon_{\pi t} \end{pmatrix},$$  (2)

where $\Gamma(L)$ is a polynomial function in the lag operator $L$. To exactly identify the two structural shocks, one additional restriction is needed. For this, it is assumed that domestic demand shocks have no long-run effects on the levels of real output by setting element $(1,2)$ of the long-run multiplier matrix $\Gamma(1)$ in (2) to zero, that is $\Gamma_{12}(1) = 0$. Cecchetti (1994) and Cecchetti and Rich (1999) used the same identifying restriction to estimate the U.S. sacrifice ratio in their two-variable structural VAR model. When the long-run multiplier matrix $\Gamma(1)$ is lower triangular, as in this case, the Blanchard and Quah (1989) procedure can be used conveniently to accomplish the identification of the structural shocks. In this paper, however, we use an alternative procedure proposed by Shapiro and Watson (1988). In their method, the identifying restrictions are imposed directly on the VAR model itself without recourse to its VMA counterpart. This property allows us to perform the structural analysis (i.e., COFI ratios) in a straightforward manner even when the VAR is prompted with LSTAR specifications. This will be fully explained in section 3.3.

Following Shapiro and Watson, the long-run restriction that $\Gamma_{12}(1) = 0$ can be imposed by restricting the sum of the coefficients on $\Delta \pi_t$ in (1a) to equal zero (i.e., $\sum_{i=0}^{p} a_{y\pi,i} = 0$), which yields

$$\Delta y_t = \sum_{i=0}^{p} a_{yT,i} \Delta T_{t-i} + \sum_{i=1}^{p} a_{y\pi,i} \Delta y_{t-i} + \sum_{i=0}^{p-1} b_{y\pi,i} \Delta^2 \pi_{t-i} + \epsilon_{yt},$$  (1a')
where \( b_{y activating, i} \) are functions of \( a_{y activating, i} \) in (1a). The variable \( \pi_t \) only enters in second differences with the maximum number of lags set to \( p - 1 \). Equation (1a') cannot be estimated by OLS, owing to the contemporaneous value of \( \Delta^2 \pi_t \) in the right-hand variables. However, the parameters in (1a') can be estimated consistently using an Instrumental Variables (IV) procedure. An appropriate set of instruments is lags one through \( p \) of \( D_y t \) and \( D\pi_t \), and lags zero through \( p \) of \( DTT_t \). Equation (1b) can also be estimated consistently using the same set of instruments as (1a') plus the estimated residual from (1a').\(^6\) Once both (1a') and (1b) are estimated, they are inverted to have \( G \) in (2). It is then possible to use the estimated structural moving-average representation to examine the impact of changes in \( \varepsilon_y \) and \( \varepsilon_{\pi_t} \) on output and inflation. When the model is exactly identified, the results from Shapiro and Watson are identical to those from Blanchard and Quah.\(^7\)

3.2. A logistic smooth transition autoregression (LSTAR) representation

Collect equations (1a') and (1b) in

\[
\begin{pmatrix}
\Delta y_t \\
\Delta\pi_t
\end{pmatrix} = \begin{pmatrix}
\sum_{i=0}^{p} a_{y, i} L^i \\
\sum_{i=0}^{p} a_{\pi, i} L^i
\end{pmatrix} \begin{pmatrix}
\sum_{i=1}^{p} \Delta y_{-i} L^{i-1} \\
\sum_{i=1}^{p} \Delta\pi_{-i} L^{i-1}
\end{pmatrix} \begin{pmatrix}
1 - L \\
L
\end{pmatrix} \sum_{i=0}^{p-1} b_{y, i} L^i \\
\sum_{i=0}^{p-1} a_{\pi, i} L^i
\begin{pmatrix}
\varepsilon_y \\
\varepsilon_{\pi_t}
\end{pmatrix} + \begin{pmatrix}
\varepsilon_y \\
\varepsilon_{\pi_t}
\end{pmatrix}
\]  

or, in a compact notation,

\[ x_t = \Psi(L)x_t^* + \varepsilon_t. \]  

Then, the set of structural equations (3) can be augmented with an LSTAR component, such as

\[ x_t = \Psi(L)x_t^* + \Xi(L)x_t^* F(z_t) + \varepsilon_t, \]  

where \( \Xi(L) \) has the exactly same components as \( \Psi(L) \) of (3).\(^8\) Under this specification, the LSTAR component obeys the long-run identifying restriction in the same manner that the linear part does. The logistic function \( F(z_t) \) is assumed to have the following form:

\[ F(z_t) = (1 + \exp\{-\lambda(z_t - c)/\delta_z\})^{-1}, \]  

where \( F(z_t) \) lies in the range of 0 and 1, and \( \lambda > 0 \). The variable \( z_t \) is a switching indicator that represents the state of the economy, and the parameter \( c \) represents the threshold around which the dynamics of the model change. The parameter \( \lambda \) is the smoothness parameter. If \( \lambda \) approaches zero, \( F(z_t) \) converges to a constant, and

\[ Equation (1b) can be equivalently estimated by OLS after substituting (1a') into (1b).\(^6\) See Pagan and Robertson (1998) for a good exposition on the use of instrumental variables estimation in the structural VAR literature.\(^7\) Note that the intercept terms are suppressed in both \( \Psi(L) \) and \( \Xi(L) \).\(^8\)
the model becomes linear. If $\lambda$ approaches infinity, the model becomes a threshold autoregression model along the lines of Tong (1983): the model’s dynamics change abruptly, depending on whether $z_t$ is greater or less than $c$. The parameter $\delta_z$ is the standard deviation of the switching variable $z_t$. The smoothness parameter $\lambda$ is not scale free, since its value depends on the magnitude of the switching variable $z_t$. Dividing by $\delta_z$ normalizes the deviations of $z_t$ from the threshold value and facilitates interpretation of the smoothness parameter. This also makes it easier to find sensible starting values in initiating the optimization process for estimation. See Granger and Teräsvirta (1993), and Teräsvirta (1998) for a detailed exposition on LSTAR models.

3.3. Output cost of reducing inflation (COFI ratio)

For the case of our linear baseline model, the output cost can be computed based on the structural impulse response functions from (2) in the usual manner. For inflation, the sum of the coefficients in $\Gamma_{22}(L)$ measures the effects of a demand shock on its level. In the case of output, however, the COFI ratio requires us to consider the cumulative effect on its level resulting from the incidence a demand shock. Taken together, the output cost over the time horizon $r$ can be calculated as

$$\text{COFI} = \sum_{i=0}^{\rho} \left( \sum_{j=0}^{\rho} \frac{\partial y_{t+j}/\partial \epsilon_{\pi t}}{\partial \pi_{t+j}/\partial \epsilon_{\pi t}} \right) = \frac{\sum_{i=0}^{\rho} \sum_{j=0}^{i} \Gamma_{12,j}}{\sum_{j=0}^{\rho} \Gamma_{22,j}},$$

where $\Gamma_{22,j}$ are the responses of the series to the demand shock at a horizon of $j$ in (2); that is, $L = j$.

An alternative measure we use here is to adopt the concept of generalized impulse response functions by Koop, Pesaran, and Potter (1996), which can be used in both linear and non-linear cases. An impulse response function is defined as the effect of a one-time shock on the forecast of variables in a system. The response of a variable following a shock is then compared with a baseline ‘no shock’ scenario. That is, the generalized impulse response function of a variable $x$, $\text{GI}_x$, is defined as the difference between two conditional expectations as

$$\text{GI}_x(n, \epsilon_{\pi t}, \omega_{t-1}) = E[x_{t+n}|\epsilon_{\pi t}, \omega_{t-1}] - E[x_{t+n}|\omega_{t-1}], \quad n = 0, 1, 2, \cdots,$$

where $n$ is the forecast horizon, $\epsilon_{\pi t}$ is the shock generating the response, $\omega_{t-1}$ is the history or initial values of the variables in the model, and $E[\cdot]$ is the expectations operator. Then, the COFI ratio can be expressed as

$$\text{COFI} = \sum_{i=0}^{\rho} \left( \sum_{j=0}^{i} \text{GI}_y(j, \epsilon_{\pi t}, \omega_{t-1}) \right) / \sum_{j=0}^{\rho} \text{GI}_y(j, \epsilon_{\pi t}, \omega_{t-1}),$$

where $\text{GI}_y(j, \epsilon_{\pi t}, \omega_{t-1})$ is the generalized response of $y$ to a demand shock $\epsilon_{\pi t}$ at a horizon of $j$ when the initial state before the shock was $\omega_{t-1}$. The same notation
applies to the case of $G_{\pi}(j, e_{\pi t}, \omega_{t-1})$. For a linear model, the impulses are invariant to history so that $\omega_{t-1}$ can be zeroed out. That is, estimated GIs from (7) will be identical to those estimated from the conventional impulse response function, as in (2). Consequently, COFI ratios from both (6) and (8) will be identical.

For the case of non-linear models, such as our LSTAR model, however, the responses are dependent on the signs and sizes of the shocks and the initial states before these shocks hit the system. In this context, Koop, Pesaran, and Potter treat the two conditional expectations in the right-hand side of (7) as random variables, implying that GI, is also random. Accordingly, the GI functions and the COFI ratios must be computed by simulating the model. As alluded to earlier, the GI functions in adjunct with the Shapiro and Watson model may be usefully exploited in drawing structural interpretations such as the COFI ratio from our long-run identified LSTAR model. To see this, recall that in (4) with (5), the restrictions required for identification are imposed directly on the VAR components. Accordingly, the forecast values made from this estimated LSTAR will also preserve those restrictions. These forecasted values are then used to form the GI, which can be interpreted as the model’s structural responses to the shocks. Following (8), the COFI ratio can be obtained from forecasting the estimated LSTAR model directly without recourse to their VMA counterparts. For the simulation procedure, this paper follows the procedure used by Koop, Pesaran, and Potter (1996) and Weise (1999). The appendix in Weise, in particular, provides a detailed account of simulation of VAR-LSTAR models.

4. Empirical results

4.1. Estimating the baseline linear model

The baseline linear VAR model summarized in (3) is estimated for quarterly data over the sample period 1961:Q1 to 1997:Q4. The lag length chosen was $p = 4$ on the basis of the Sims likelihood ratio test. Definitions of the data are as follows. The measure of real output ($y$) is the log of real GDP seasonally adjusted, in chained 1992 prices. The data on the terms of trade (TT) are defined as the log ratio of the implicit price deflator of exports of goods and services to imports of goods and services and is seasonally adjusted at chained values ($1992 = 100$). The rate of inflation ($\pi$) is measured as the quarterly percentage change in the consumer price index with the base year $1992 = 100$ and is seasonally adjusted using $X - 11$ procedure. All data were drawn from DATASTREAM. The mnemonics for real output, export, and import price deflators and the consumer price index are CNGDP...D, CNIPDEXPE, CNIPDIMPE, and CNC...F.

Before estimating a linear VAR model, Augmented Dickey-Fuller tests were used to determine the order of integration of the series. The test indicated that real GDP, the terms of trade, and the inflation rate are characterized as I(1) processes. We then applied the Johansen procedure to test for evidence of cointegration between output and inflation with the terms of trade entering as an exogenous variable. Both the trace and the maximum eigenvalue tests indicated no cointegrating relationships in the model at conventional levels.
4.2. Testing linearity against LSTAR

The tests against a specific non-linear alternative are usually formulated as Lagrange multiplier-type tests because they only require the estimation of the model under the null hypothesis, that is, the linear model. Applying this, the null hypothesis of linearity in (4) and (5) is \( H_0; \lambda = 0 \) against \( H_1; \lambda > 0 \). There is a problem, however, since the model is identified under the alternative but not under the null hypothesis. When the null hypothesis is true, \( \Xi(L), \lambda, \) and \( c \) in the non-linear component can take any value. To circumvent this problem, Luukkonen, Saikkonen, and Teräsvirta (1988) and Granger and Teräsvirta (1993) suggest the use of the first-order Taylor series approximation for the LSTAR component.

Following their method, the LM tests can be performed, equation by equation, in three steps. In step (1), collect residuals \( \tilde{\mu}_i \) for \( i = \Delta y \) and \( \Delta \pi \) from estimating (3) and compute the residual sum of squares \( SSR_i^0 = \Sigma \tilde{\mu}_i^2 \). In step (2), run the following auxiliary regression of \( \tilde{\mu}_i \) on \( \Psi(L), x_i^* \) and \( z_i \Xi(L), x_i^* \), where \( \Psi(L) \), and \( \Xi(L) \) are the rows of \( \Psi(L) \) and \( \Xi(L) \) corresponding to equation \( i \) in (4), and \( z_i \) is a switching variable. As before, collect residuals \( \tilde{\nu}_i \), and calculate \( SSR_i^1 = \Sigma \tilde{\nu}_i^2 \). In the final step (3), compute the test statistics \( LM_i = T(SSR_i^1 - SSR_i^0)/SSR_i^0 \) for each equation \( i \), where \( T \) is the number of observations. Under the null hypothesis, \( LM_i \) is distributed \( \chi^2(m_i) \), where \( m_i \) is the number of the regressors in the term \( z_i \Xi(L), x_i^* \). In small samples, Granger and Teräsvirta recommend the use of an F test in order to improve size and power properties. The equivalent F statistic is \( F_i = [(SSR_i^0 - SSR_i^1)/m_i]/[SSR_i^0/(T - 2m_i - 1)] \). While these tests are performed within a single-equation context, Weise (1999) recently used a log-likelihood ratio test for the test of linearity in the system as a whole; that is, \( H_0; \lambda = 0 \) in all of the equations. Let \( \Omega^{0} = \Sigma \tilde{\mu}_i \tilde{\mu}_i^T / T \) and \( \Omega^{1} = \Sigma \tilde{\nu}_i \tilde{\nu}_i^T / T \) be the estimated variance-covariance matrices of residuals from the restricted and unrestricted regressions, respectively, where \( \tilde{u}_i = (\tilde{u}_{1i}, \tilde{u}_{2i}) \) and \( \tilde{\nu}_i = (\tilde{\nu}_{1i}, \tilde{\nu}_{2i}) \). Then the log-likelihood statistic is \( LR = T \{ \log |\Omega^{0}| - \log |\Omega^{1}| \} \), which is asymptotically distributed \( \chi^2 \) with the degree of freedom \( (m_1 + m_2) \).

To perform these tests, the switching variable \( z_i \) must be assumed a priori. Several candidates are considered for the switching variable. They are the growth rate of real output (\( \Delta y \)) and the change in the rate of inflation (\( \Delta \pi \)), one to five lagged values. We also experiment with one to five lagged values of changes in the terms of trade (\( \Delta \tau \)) as possible switching variables. The results of these linearity tests are reported in table 1 in terms of their marginal significance levels (\( p \) values).

10 As explained in section 3.1, estimation should be carried in a sequential manner, that is, from (1a) to (1b).
11 Granger and Teräsvirta (1993) point out that an LM-type test mentioned above would have low power against the alternative hypothesis when \( \Xi(L) \) is close to zero such that the only non-linear element in (4) is the constant. They describe an adjustment proposed by Luukkonen, Saikkonen, and Teräsvirta (1988), which involves a third-order Taylor approximation to the regression in step (2). In a multivariate set-up, however, the addition of higher-order terms is very costly, and given the typical size of macroeconomic samples, it is not practical. Further, estimates of \( \Xi(L) \) reported in the following section are far from zero.
12 If \( z_i \) is assumed to be unknown, the test statistics must be generalized to account for this. See Tsay (1986) and Luukkonen, Saikkonen, and Teräsvirta (1988) for this case.
When the change in terms of trade is used as the switching variable, the evidence against linearity is very weak. Except for two cases in the inflation rate equation, the null hypothesis of linearity is never rejected in both equations and the system as a whole. By contrast, linearity is rejected in favour of an LSTAR specification when output growth and the change in the rate of inflation are used as switching variables. The evidence is particularly strong with $Dy_2$ and $Dp_2$ as the switching variables. In each case, linearity is rejected unanimously by the LR test for the system as a whole, and by the LM test for individual equations in the system. Overall, these tests provide solid evidence against linearity in favour of LSTAR specifications.

A different LSTAR model can be estimated for a host of switching variables, including other variables not considered in this section. Obviously, one cannot estimate every possible model, and there is no strong theoretical consensus on which variable to choose. In this sense, we restrict our consideration to the results of table 1 in search of a plausible candidate for a switching variable. In particular, we follow Granger and Teräsvirta (1993) and Teräsvirta (1998), who suggest using the variable associated with the smallest $p$ value if that value is sufficiently small to reject the null hypothesis of linearity. The rationale behind this decision rule is that the linearity test has the highest power against the correctly specified alternative. If an inappropriate switching variable is selected, the resulting test may still have power against the alternative, but the power is less than if the correct switching variable is used. Thus, the strongest rejection of the null hypothesis suggests that the corresponding switching variable should be selected. Based on this criterion, we use $Dy_{-1}$ and $D\pi_{-1}$ as the switching variables for the subsequent analysis.

### 4.3. Estimating the structural LSTAR model

In principle, the parameters of the LSTAR model can be estimated by applying non-linear least squares. In practice, however, there could be convergence problems...
because some of the parameters are redundant and their estimates highly associated with those of other parameters. This problem may be more severe in a VAR-type LSTAR model like ours, which could result in the possible danger of under-identification. One way of circumventing this problem is to use the method suggested by Teräsvirta (1998) and Teräsvirta, Tjøstheim, and Granger (1994), which imposes coefficient restrictions that set certain elements of \( \Psi(L) \) and \( \Xi(L) \) equal to zero or each other. This suggestion may not be very helpful for VAR-type LSTAR models, however, since it could distort dynamic interactions among the variables in the model. Weise (1999) raises a similar concern, arguing that the results with such a method may be sensitive to which restrictions are imposed, and the choice of restrictions cannot be guided by economic theory.

Weise used an alternative procedure to estimate a VAR-type LSTAR model. First, the threshold parameter \( c \) is fixed at a constant, which is sensibly chosen from an economic standpoint. The LSTAR model is then estimated equation-by-equation OLS using this value for \( c \), allowing \( \lambda \) to vary. The value of \( \lambda \) that minimizes the log of the determinant of the variance-covariance matrix of residuals from these regressions is selected for the final regressions. This method avoids unjustifiable coefficient restrictions, but sacrifices efficiency. In addition, standard errors are not computed for \( c \) and \( \lambda \).\(^{13}\) We follow this method, but also take up several plausible values of \( c \) for consideration. For the switching variable \( \Delta y_{t-1} \), the parameter \( c \) is selected in the range of \(-3.0 \) to \(3.0 \) per cent with an incrementing order of 0.2 percentage points. While others can be considered, this prespecified range is reasonable, given that the switching variable is expressed in quarterly changes. It amounts to covering about 98 per cent of the quarterly growth rates observed in the Canadian GDP. With the switching variable \( \Delta \pi_{t-1} \), the parameter \( c \) is selected from \(-2.0 \) to \(2.0 \) percentage points in an incrementing order of 0.2 percentage points. This range covers all of the quarterly changes in the Canadian inflation rate, except one at 1991:Q2.

Table 2 reports the estimates of \( \lambda \)

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Switching & Estimates of \( \lambda \) with & Estimates of & & \\
variables & \( c \) being fixed & \( c \) with \( \lambda = \infty \) & & \\
\hline
\( \Delta y_{t-1} \) & \( c = 1.0 \) & \( \lambda = 15.64 \) & \( \lambda = \infty \) & \( c = 0.92 \) \\
\( \Delta \pi_{t-1} \) & \( c = 0.0 \) & \( \lambda = 1.96 \) & \( \lambda = \infty \) & \( c = 0.49 \) \\
\hline
\end{tabular}
\end{table}

\(^{13}\) Teräsvirta, Tjøstheim, and Granger (1994) also considered this procedure in a univariate LSTAR model.
along with the selection of $c$ produced by this procedure. When the switching variable $\Delta \gamma_{t-1}$ is used, the selected choice is $c = 1.0$ with $\lambda = 15.64$. When $\Delta \pi_{t-1}$ is used as the switching variable, $c = 0.0$ is chosen and the smoothness parameter is much smaller at 1.96. The effect of different sizes of $\lambda$ is visualized in the upper panel of figure 1, which plots the estimated transition function $F(z_t)$ against its switching variable. For the switching variable $\Delta \gamma_{t-1}$, the transition between the two extreme regimes (characterized $F = 0$ and $F = 1$) appears to be almost instantaneous. On the other hand, $F$ changes quite slowly with the switching variable $\Delta \pi_{t-1}$, indicating a fairly smooth transition from one regime to another.\textsuperscript{14} The lower panel

\textsuperscript{14} Designate $F(z_t) = 0$ to be a pure ‘falling inflation’ regime, and $F(z_t) = 1$ to be a pure ‘rising inflation’ regime. Then, a smoothness parameter $\lambda = 1.96$ implies that when $\Delta \pi_{t-1}$ moves one standard deviation above (below) zero, $F(z_t)$ equals approximately 0.87 (0.13). This means that the regime is a linear combination of the pure rising and falling inflation regimes, with a weight of 0.87 (0.13) on the former and 0.13 (0.87) on the latter. Compare this with the value of $\lambda = 15.64$, for which minute deviations of the switching variable from the threshold level essentially place the economy entirely on one regime or the other.
of figure 1 depicts these transition functions over time. For the switching variable \( \Delta y_{t-1} \), extreme values of \( \hat{F} \) (either 0 or 1) are quite common over time. This is not the case for the switching variable \( \Delta \pi_{t-1} \), owing to slower transitions between the regimes.

As a way of checking the sensitivity of the estimates, the third column of table 2 reports the results obtained from estimating a threshold autoregression \( (\lambda \to \infty) \). The parameter \( c \) was chosen as the value for which the log determinant of the variance-covariance matrix of residuals was minimized. The estimated threshold value for \( \Delta y_{t-1} \) is indeed close to the maintained value of 1.0 in the second column. This is not surprising, since the estimated smoothing parameter in the second column is large. Little would be lost by estimating this system using a threshold autoregression specification. For the case of \( \Delta \pi_{t-1} \), however, the threshold value is estimated to be 0.49 and differs from the selected value of \( c \) in the second column. Hence, threshold autoregression specification is likely to produce different results from those reported using the LSTAR model.

Table 3 reports results of several diagnostic tests to check the statistical adequacy of the estimated LSTAR models in terms of their marginal significance levels \( (p \text{ values}) \). The F test statistics in the second column reject the null hypothesis of linearity for all equations. This confirms our earlier evidence against linearity and in favour of the type of non-linearity represented by the LSTAR specification. All individual equations also pass a battery of misspecification tests, pointing to no evidence of any serious model inadequacy. Results of the LM test in the fourth column indicate no serial correlation, nor is there any evidence of ARCH effects in the fifth column. Not do the results of the Jarque-Bera test for normality in the sixth

<table>
<thead>
<tr>
<th>Equations</th>
<th>F test</th>
<th>( \hat{\sigma}^2/\hat{\sigma}^2 _\text{L} )</th>
<th>Auto (4)</th>
<th>ARCH (4)</th>
<th>Normality</th>
<th>LSTAR</th>
<th>Constancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switching variable = ( \Delta y_{t-1} )</td>
<td>( \Delta y )</td>
<td>0.00</td>
<td>0.78</td>
<td>0.39</td>
<td>0.18</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>( \Delta \pi )</td>
<td>0.01</td>
<td>0.89</td>
<td>0.27</td>
<td>0.12</td>
<td>0.06</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Switching variable = \( \Delta \pi_{t-1} \)

| \( \Delta y \) | 0.04 | 0.83 | 0.36 | 0.16 | 0.19 | 0.25 | 0.23 |
| \( \Delta \pi \) | 0.00 | 0.75 | 0.54 | 0.31 | 0.23 | 0.47 | 0.32 |

NOTES

Figures reported are the marginal significance levels \( (p \text{ values}) \). Figures reported in the second column are from the F tests for the null hypothesis that the coefficients on the \( F(z_t) \) terms are jointly equal to zero. The third column, denoted by \( \hat{\sigma}^2/\hat{\sigma}^2_\text{L} \), reports the ratio of the estimated variance of the model relative to that of the corresponding linear model. For the following two columns, the terms Auto (4) and ARCH (4) refer to the F versions of the LM test for fourth-order serial correlation and the LM test for fourth-order ARCH effects, respectively. The sixth column reports the results of the Jarque-Bera test for normality. The seventh column termed as LSTAR reports the results of testing linearity against the LSTAR specification using the estimated residuals. Figures in the last column are from the F tests for the null hypothesis of parameter constancy. See Eitrheim and Teräsvirta (1996), and Teräsvirta (1998) for details.
column pose any major concern. A mild exception is when the switching variable is $\Delta y_{t-1}$. There is some evidence against the normality of the errors for both output and inflation equations, but only at the 10 per cent level of significance. The seventh column checks for remaining non-linearity in residuals using the LSTAR test described in section 4.2. No further evidence to support non-linearity was found in any equation.

Testing parameter constancy is another important way of checking the model’s adequacy, since our LSTAR models were estimated assuming constant parameters. Most well-known parameter constancy tests, such as the CUSUM test and its variants, are tests against a non-specified alternative or a single structural break. In this paper, instead, we use the procedure proposed by Eitrheim and Teräsvirta (1996) and Teräsvirta (1998) in which the alternative hypothesis to parameter constancy is a set of smoothly changing parameters. In their studies, three types of functional forms are suggested to characterize parameter inconstancy under the alternative hypothesis. We choose the most flexible one among them, which allows monotonically as well as non-monotonically changing parameters. To save space, the technical details of this procedure are not repeated here. Under the null hypothesis of parameter constancy, the test statistic follows an asymptotic $\chi^2$ with $3m$ degrees of freedom ($m$ is the number of the regressors in each equation). The F test whose degrees of freedom are $3m$ and $(T - 4m - 2)$, respectively, is recommended for use in small samples.15

The results from applying this test to the estimated LSTAR models are reported in the last column of table 3. Clearly, none of the equations rejects the null hypothesis of parameter constancy at standard levels. It is interesting to apply the same test to the linear models. The null hypothesis of parameter constancy is rejected strongly for both output and inflation equations, since the test statistics are $F(39, 94) = 1.86$ and $F(42, 90) = 2.22$, respectively, with their marginal significance levels being less than 1 per cent. Taking both results together, the failure of parameter constancy in the linear models could be due partly to an effect of neglected non-linearity on the tests. Our evidence in favour of our LSTAR specifications is reinforced at the same time. Finally, table 3 reports the residual variances of the equations relative to those of their linear counterparts. When $\Delta y_{t-1}$ is used as the switching variable, the residual variances of the equations are only 78 and 89 per cent, respectively, of those from the linear models. When the switching variable is $\Delta \pi_{t-1}$, the residual variances of the estimated models fall by 17 and 25 per cent, respectively, compared with those from the linear models.

4.4. Asymmetric COFI ratio estimates

Table 4 reports the estimated COFI ratios using equation (8) over the horizon of twenty quarters; that is, $\rho = 20$. They should be interpreted as the cumulative output

\begin{eqnarray}
\Phi_i(L)(t^i x_t) + \Phi_i(L)(t^{i+1} x_{t+1}) + f_i(L)(t^{i+1} x_{t+1}) + f_i(L)(t^{i+2} x_{t+2}) + f_i(L)(t^{i+3} x_{t+3}) + f_i(L)(t^{i+4} x_{t+4}) + e_i(t^i x_t) \quad (i = \Delta y, \Delta \pi, \Delta y_{t-1})
\end{eqnarray}

for $i = \Delta y$ and $\Delta \pi$, and testing $\Phi_i(L) = \Phi_j(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = 0$ in each equation for the null hypothesis of parameter constancy.

---

15 In our application, the test involves estimating the model $x_t = \Psi(L, x_t^* + \Phi_i(L)(t^{i+1} x_{t+1}) + f_i(L)(t^{i+1} x_{t+1}) + f_i(L)(t^{i+2} x_{t+2}) + f_i(L)(t^{i+3} x_{t+3}) + f_i(L)(t^{i+4} x_{t+4}) + e_i(t^i x_t) \quad (i = \Delta y, \Delta \pi, \Delta y_{t-1})$ and testing $\Phi_i(L) = \Phi_j(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = \Phi_{ij}(L) = 0$ in each equation for the null hypothesis of parameter constancy.
loss associated with a permanent 1 percentage point decline in the rate of inflation, as shown in equation (6). The COFI ratios are calculated for two initial states when the switching variable is \( \Delta \pi_{t-1} \) (referred to as model 1): falling inflation (\( \Delta \pi_{t-1} < 0 \)) and rising inflation (\( \Delta \pi_{t-1} > 0 \)). In the case of \( \Delta y_{t-1} \) as the switching variable (referred to as model 2), the ratios are calculated for three regimes: low growth (\( \Delta y_{t-1} < 0 \)), moderate growth (\( 0 < \Delta y_{t-1} < 1.82 \)), and high growth (\( \Delta y_{t-1} > 1.82 \)).

The figures under ‘1 per cent’ are computed from the impulse response functions to the demand shock of 0.53. This size of the shock is set to the standard deviation of demand shocks computed from their corresponding linear models. The figures under ‘2 per cent’ are computed from the impulse response functions to the demand shock of 1.06. These impulse responses are then divided by two, respectively, so that they can be compared with the responses to one-standard-error shocks. Table 4 also reports the results from estimating the linear models as in (3) for the sake of comparison.

Simulations show that COFI ratios vary systematically depending on the signs of the shocks. For both models 1 and 2, the COFI ratios for deliberate disinflations (negative shocks) are different from those for pre-emptive interventions against incipient inflation pressures (positive shocks). They are additionally dependent on the initial state of the economy. Looking at model 1 in the case of disinflation first, we see that the output cost in a rising inflation state is much lower than that in a falling inflation state. A 1 per cent disinflation costs 4.25 per cent of output if initiated when the economy is in a rising inflation state and 6.86 per cent when the economy is in a falling inflation state. In a rising inflation state, a tighter monetary

\[16\] The moderate-growth state is classified as the period in which its GDP growth rate lies in the range of 16 and 84 per cent when the whole sample GDP growth rates are ordered from lowest to highest. The low-growth state corresponds to the period's having its growth rate in the lower 16 per cent band, while the high-growth state corresponds to the period's having its growth rate in the upper 16 per cent. This way of devising the initial states is arbitrary, but it may not be avoidable in nature.
policy could stifle economic activity to a lesser extent than in a falling inflation state, making resultant output costs less in this case. Looking at the case of disinflation with model 2, we see that the output cost decreases as the initial strength of the economy increases. A 1 per cent disinflation costs 9.86 per cent of output if initiated when the economy is in a low-growth state, 5.86 per cent when the economy is in a medium-growth state, and 2.55 per cent when the economy is in a high-growth state. Intuitively, the stronger the economy, the less the effect on output of a tighter monetary policy, which results in fewer output costs to achieve a given disinflation.

The output cost of pre-emptively resisting incipient inflation varies with the state of the economy in a similar manner. In the case of model 1, the output cost in a rising inflation state is much lower than that in a falling inflation state. Preventing a 1 percentage point increase in inflation costs 3.96 per cent of output when the economy is in a rising inflation state, but 6.32 per cent when the economy is in a falling inflation state. The authority must relax monetary policy more extensively in a falling inflation state than in a rising inflation state if both would produce a given rise in inflation. Accordingly, the forgone output as a result of pre-emptive monetary policy will be larger in a falling inflation state than in a rising inflation state. In the case of model 2, the output cost of pre-emptively resisting incipient inflation also decreases as the initial strength of the economy increases. Preventing a 1 percentage point increase in inflation costs 9.46 per cent of output if initiated when the economy is in a low-growth state, 5.31 per cent when the economy is in a medium-growth state, and 2.16 per cent when the economy is in a high-growth state. The strength of economic forces associated with a weak economy must exceed those associated with a strong economy if both were to produce a given rise in inflation. This requires a tighter preemptive monetary policy in a low-growth state in order to prevent rising inflation, thus leading to a larger forgone output.

Simulation results in the table also indicate that the COFI ratio increases as the size of the desired inflation change increases. This finding is shown to be unanimous irrespective of the model specifications. In the case of disinflation, for example, model 1 suggests that when the economy is in a rising (falling) inflation state, the output cost per percentage point reduction in inflation is 4.25 (6.86) per cent for a 1 percentage point reduction, but 4.94 (7.59) per cent for a 2 percentage point reduction. Model 2 draws a similar line, indicating that the cost for a 1 percentage point disinflation is lower than for each percentage point reduction of a 2 percentage point disinflation, irrespective of the initial strength of the economy. The same result can also be inferred in the case of a pre-emptive intervention against incipient inflation. Irrespective of the initial states of the economy, both models, 1 and 2, indicate that preventing a potential 1 percentage point rise in inflation costs less per percentage point of incipient inflation than a 2 percentage point rise.

5. Policy implications and conclusions

Accurately assessing the output cost of reducing inflation plays an important role in determining the appropriate policy to achieve and ultimately maintain price stabil-
ity. As its contribution, this paper takes explicit account of possible non-linearity in estimating the output cost of lowering inflation. While the precise non-linear form of the Phillips curve implied by explicit micro underpinnings may be intractable, in our paper we show that its distinct features may be parsimoniously represented in terms of an LSTAR model. The adoption of the LSTAR specification as a non-linear component was justified by its flexibility in allowing slow adjustments in association with price/wage rigidities and consumers’ expectations. The findings of this paper suggest that the non-linear model provides a richer picture of output cost than a standard linear Phillips curve. Specifically, the COFI ratio is related in a non-linear way, depending critically on the state of the economy, the size of the inflation change, and whether policy makers seek to disinflate or prevent inflation from rising. This evidence has valuable policy implications for traditional views on the output cost of reducing inflation.

First, the traditional linear model tends to either overestimate or underestimate the COFI ratio, depending on the initial states of the economy. For example, model 1 indicates that the COFI ratio implied by the linear model overestimates when the economy is in a rising inflation state, but underestimates when the economy is in a falling inflation state. If the economy were in the former regime, the non-linear model’s lower cost of reducing inflation would hence suggest a more aggressive stance of monetary policy than would have been implied by a linear Phillips curve model. A less aggressive stance is recommended if the economy were in the latter regime. Looking across the other model specifications reported in table 4, a similar inference can be drawn in that our non-linear COFI ratios vary case by case. This suggests that inferences based on the linear Phillips curve may provide misleading signals about the cost of lowering inflation and, hence, the appropriate policy stance.

The second implication is related to the benefit of pre-emptive policy. Our empirical evidence reveals that the COFI ratios for the case of deliberate disinflation are higher than those calculated from the case of pre-emptively preventing incipient inflation. That is, a policy of pre-emptively preventing rising inflation of a given size turns out to be less costly than a policy to deliberately disinflate. Pre-emptive policies are often justified by the familiar finding that monetary policy affects the economy with long and variable lags. This study provides a further incentive for pre-emptive monetary responses to inflationary pressures. A pre-emptive tightening helps to prevent the economy from moving too far up to the levels where inflation begins to rise more rapidly, thereby avoiding the need for a more aggressive tightening in the future to reverse this large rise in inflation.

The third implication is due to the fact that the output cost of disinflation may vary with the speed of a given disinflation, central to the choice between gradualism and cold turkey. Our study shows that the average output cost of a 1 percentage point disinflation is in all cases less costly per percentage point of disinflation than a 2 percentage point disinflation. The higher output cost of rapid disinflation indicates that gradualism is a lower output cost strategy than cold turkey. This finding corroborates that of Taylor (1983) on the cost advantage of the gradualist approach.

The final implication from our study is related to the issue of policy experimentation. In a recent volume of the Journal of Economic Perspectives, a number of
authors argue that the Federal Reserve Board could test the limits of the economy without any adverse long-run consequences. In particular, Stiglitz (1997) highlights that such policy experimentation can enhance economic welfare on the grounds that the U.S. Phillips curve is perhaps concave. The estimated shape of the Phillips curve presented in this paper, however, exhibits a form of convexity when the economy is operating at a high capacity. Accordingly, such a prescription may not be desirable for Canada. Our empirical evidence suggests, instead, that if the central bank, in testing out the limits of the economy, induces overheating, the resultant periods of depressed activity to bring down inflation would more than offset the gains from the periods of higher activity.

References


17 Clark, Laxton, and Rose (1996), Filardo (1998), and Laxton, Rose, and Tambakis (1999) argue against the attractiveness of policy experimentation by finding evidence that the U.S. Phillips curve is convex.