How Do Canadian Hours Worked Respond to a Technology Shock?

*Lawrence J. Christiano, Martin Eichenbaum, and Robert Vigfusson**

Introduction

This paper investigates the response of hours worked to a permanent technology shock. Based on annual data from Canada and the United States, we argue that hours worked rises after a positive technology shock. While this result is consistent with the analysis in Christiano, Eichenbaum, and Vigfusson (CEV) (2003), it stands in sharp contrast to a large and growing literature, according to which hours worked falls after a positive technology shock (see, for example, Galí 1999 and Francis and Ramey 2001).[1]

The assumption that we make to identify a technology shock is the same as in the literature. Specifically, we assume that the only type of shock that affects the long-run level of average labour productivity is a permanent shock to technology. So the difference between our results cannot be attributed to the nature of our identifying assumptions. Instead, it is due to the way hours worked is incorporated into our statistical analysis. Using quarterly U.S. time-series data, CEV (2003) make the following argument. Suppose that the analyst assumes that per capita hours worked is a stationary stochastic process and works with the level of hours. Using this "level specification," the analyst would find that hours worked in the United States rises after a technology shock. On the other hand, suppose that the analyst assumed that hours worked is a difference-stationary process and works with the growth rate of hours worked. Using this "difference specification,"

1. CEV (2003) base their analysis on quarterly U.S. time-series data.

* Lawrence Christiano and Martin Eichenbaum thank the National Science Foundation for financial assistance.

the analyst would find that hours worked in the United States falls after a technology shock. In this paper, we show that exactly the same result holds for annual Canadian and U.S. data.

The question is: Which set of results is correct—those based on the level specification or those based on the difference specification? Not surprisingly, standard, univariate hypothesis tests do not yield much information about whether per capita hours worked has a unit root or not. These tests cannot reject either the null hypothesis that per capita hours worked are difference-stationary or the null hypothesis that they are stationary. As in CEV (2003), we assess the relative plausibility of the two hypotheses and their associated implications by asking the following encompassing question: Which specification has an easier time explaining the findings that emerge when the analyst proceeds using the competing specification?

As discussed below and in CEV (2003), we expect that the specification that will do best on the encompassing criterion is the one that predicts that the other model is misspecified. These considerations lead us to think that the level specification will do better than the difference specification. This is because, if the level specification is true, an analyst who adopts the difference specification is committing a specification error. But, if the difference specification is true, an analyst who adopts the level specification is not committing a specification error. While important, this consideration is not definitive because sampling considerations also enter. Specifically, when the difference specification is true, then an econometrician who adopts the level specification will encounter a weak-instrument problem that implies large sampling uncertainty and bias in the estimates of the response of hours worked to a technology shock.

To choose between the level and difference specifications, we use the kind of posterior odds ratio considered in Christiano and Ljungqvist (1988) and CEV (2003). The preferred specification is the one that can most easily explain three facts: (i) the level specification implies that hours worked rises after a technology shock; (ii) the difference specification implies that hours worked falls; and (iii) the outcome of a weak-instrument test that we implement. Focusing only on facts (i) and (ii), we find that the odds are roughly two to one in favour of the level specification over the difference specification. However, once (iii) is incorporated into the analysis, we find that the odds overwhelmingly favour the level specification. Indeed, in the case of Canada, the odds in favour of the level specification are greater than nine to one.

After establishing the case for the level specification, we analyze how money growth, the interest rate, and inflation in Canada respond to a technology shock. Focusing on point estimates, we find that in response to such a shock, money growth rises while the interest rate and inflation drop. Sampling uncertainty aside, these findings suggest that Canadian monetary policy makers have accommodated technology shocks.

This raises a key question: Exactly what role has monetary policy played in the expansion of aggregate economic activity and the fall in inflation that follow in the wake of a positive technology shock? To answer this question requires that we know how the economy would have reacted had the monetary authority acted differently. The only place where we can perform such a counterfactual experiment is in a structural economic model. Such an analysis lies beyond the scope of this paper, which relies on reduced-form time-series methods. The methods used in this paper can deliver estimates of how monetary policy actually reacted to technology shocks. These methods, however, cannot be used directly to ask how the economy would have reacted under alternative policy rules.

For an analysis of that issue, we refer the reader to Altig, Christiano, Eichenbaum, and Linde (ACEL) (2003). Using an estimated dynamic general-equilibrium model embodying wage and price frictions, ACEL assess the response of the economy to a technology shock under alternative monetary policy rules, e.g., a k per cent growth-rate rule for money. Their key conclusion is that, had policy-makers not accommodated technology shocks, hours worked would have fallen for a prolonged period of time after a positive technology shock. Somewhat paradoxically, actual hours worked responds positively as in a real-business-cycle model, because of the systematic way monetary policy makers respond to technology shocks. Since ACEL (2003) estimate their model using U.S. data, we cannot claim that this result holds for Canada. But we suspect that it does.

The remainder of this paper is organized as follows. Section 1 discusses our strategy for identifying the effects of a permanent shock to technology. In section 2, we present the results from a bivariate analysis using data on hours worked and the growth rate of labour productivity. Section 3 discusses our encompassing strategy, the results of which are reported in section 4. In section 5, we report results for Canada and how inflation, the growth rate of money, and the interest rate respond to a technology shock. The final section contains concluding remarks.

1 Identifying the Response to a Permanent Technology Shock

In this section, we discuss how we identify the effect of a permanent shock to technology. As in Galí (1999); Galí, López-Salido, and Valles (2002); Francis and Ramey (2001); and CEV (2003), we assume that the only type of shock that affects the long-run level of average labour productivity is a permanent shock to technology.[2] As discussed in CEV (2003), this assumption is satisfied by a large class of standard business-cycle models. Still, it is important to recognize that there are models where this assumption is not satisfied.[3]

We estimate the dynamic effects of a technology shock using a variant of the Shapiro and Watson (1988) procedure for long-run identifying assumptions. Our description of this procedure borrows heavily from the relevant portion of CEV (2003). Our starting point is the relationship

$$\Delta f_t = \mu + \beta(L)\Delta f_{t-1} + \tilde{\alpha}(L)X_t + \varepsilon_t^z. \tag{1}$$

Here, f_t denotes the log of average labour productivity and $\tilde{\alpha}(L)$, $\beta(L)$ are polynomials of order q and $q-1$ in the lag operator, L, respectively. Also, Δ is the first-difference operator and we assume that Δf_t is covariance stationary. The white-noise random variable, ε_t^z, is the innovation to technology. Suppose that the response of X_t to an innovation in some non-technology shock, ε_t, is characterized by $X_t = \gamma(L)\varepsilon_t$, where $\gamma(L)$ is a polynomial in non-negative powers of L. We assume that each element of $\gamma(1)$ is non-zero. The assumption that non-technology shocks have no impact on f_t in the long run implies the following restriction on $\tilde{\alpha}(L)$:

$$\tilde{\alpha}(L) = \alpha(L)(1-L), \tag{2}$$

where $\alpha(L)$ is a polynomial of order $q-1$ in the lag operator. To see this, note first that the only way non-technology shocks can affect f_t is by their effect on X_t, while the long-run impact of a shock to ε_t on f_t is given by:

$$\frac{\tilde{\alpha}(1)\gamma(1)}{1-\beta(1)}.$$

2. There is now a large literature in which the long-run identifying assumption is adopted. See, for example, Vigfusson (2002); Altig, Christiano, Eichenbaum, and Linde (2003); and Fisher (2002).

3. For example, the assumption is not true in an endogenous growth model where *all* shocks affect productivity in the long run. Nor is it true in an otherwise standard model when there are permanent shocks to the tax rate on capital income.

The assumption that Δf_t is covariance stationary guarantees $|1 - \beta(1)| < \infty$. This, together with our assumption on $\gamma(L)$, implies that for the long-run impact of ε_t on f_t to be zero it must be that $\tilde{\alpha}(1)$ equals zero. This in turn is equivalent to equation (2).

Substituting equation (2) into equation (1) yields the relationship:

$$\Delta f_t = \mu + \beta(L)\Delta f_{t-1} + \alpha(L)\Delta X_t + \varepsilon_t^z. \tag{3}$$

We obtain an estimate of ε_t^z by using equation (3) in conjunction with estimates of μ, $\beta(L)$ and $\alpha(L)$. If one of the shocks driving X_t is ε_t^z, then X_t and ε_t^z will be correlated. So, we cannot estimate the parameters in $\beta(L)$ and $\alpha(L)$ by ordinary least squares (OLS). Instead, we apply the standard instrumental-variables strategy used in the literature. In particular, we use as instruments a constant, Δf_{t-s} and X_{t-s}, $s = 1, 2, \ldots, q$.

Given an estimate of the shocks in equation (3), we obtain an estimate of the dynamic response of f_t and X_t to ε_t^z, as follows. We begin by estimating the following q^{th} order vector autoregression (VAR):

$$Y_t = \alpha + B(L)Y_{t-1} + u_t, Eu_tu_t' = V, \tag{4}$$

where

$$Y_t = \begin{pmatrix} \Delta f_t \\ X_t \end{pmatrix},$$

and u_t is the one-step-ahead forecast error in Y_t. Also, V is a positive-definite matrix. The parameters in this VAR, including V, can be estimated by OLS applied to each equation. In practice, we set q equal to 4. The fundamental economic shocks, e_t, are related to u_t by the following relation:

$$u_t = Ce_t, Ee_te_t' = I.$$

Without loss of generality, we suppose that ε_t^z is the first element of e_t. To compute the dynamic response of the variables in Y_t to ε_t^z, we require the first column of C. We obtain this by regressing u_t on ε_t^z by OLS. Finally, we simulate the dynamic response of Y_t to ε_t^z. For each lag in this response function, we computed the centred 95 per cent Bayesian confidence interval using the approach for just-identified systems discussed in Doan (1992).[4]

4. This approach requires drawing $B(L)$ and V repeatedly from their posterior distributions. Our results are based on 2,500 draws.

2 Empirical Results

A key issue when working with Canadian data is the limited span of the relevant quarterly data. Many studies of the effects of technology on the U.S. economy use quarterly data on Business Labor Productivity, a time series that is available starting in 1947. The analogous Canadian series starts only in 1987.[5] Similarly, Canadian quarterly data on hours worked in the business sector are available starting in 1987.[6] Since we are interested in identifying shocks that have a long-run effect on productivity, we work with annual data, which are available from 1961.[7] Our measure of the Canadian population is for people between the ages of 15 and 64.[8] We measured output as the average level of Canadian real GDP over the year.[9] GDP is a broader measure of output than the measure considered in Francis and Ramey (2002) or CEV (2003), namely private sector output. Using a broader measure, however, seems reasonable given that we also use a broader measure of hours worked.

We are interested in comparing the effects of a technology shock in Canada and the United States. The U.S. data that we use are the annual version of the data used in CEV (2003). The relevant series are business labour productivity and hours.[10] Our data on labour productivity growth and per capita hours worked in the United States and Canada are displayed in the first row of Figure 1 and Figure 2, respectively. For the United States, the average growth rate of these variables are –0.04 and 2.19 per cent, respectively. For Canada, the average growth rate of per capita hours and labour productivity is –0.03 and 1.80 per cent, respectively. The other variables in Figure 2 will be discussed in the multivariate section.

2.1 Impulse responses

We first consider results for a bivariate VAR for Y_t. The first element in Y_t is the growth rate of the log of labour productivity, f_t. The second element in Y_t is the log of per capita hours worked, h_t. Figure 3 reports the response of labour productivity and average hours to a one standard deviation long-

5. The CANSIM mnemonic for Business Sector Labor Productivity is V1409153.

6. The CANSIM series V1409155 measures quarterly hours worked in the business sector but starts only in 1987. The CANSIM series V159660 measures total hours worked in all sectors and is monthly but starts only in 1976.

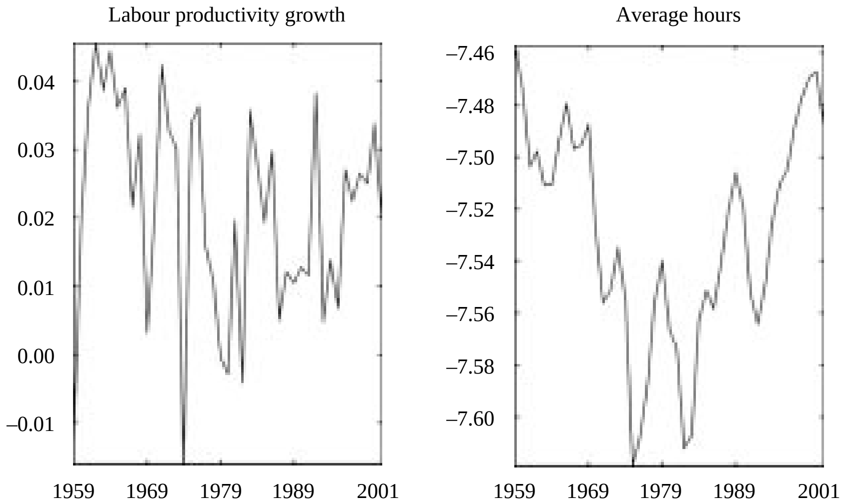
7. Annual total hours worked is measured using the CANSIM series V719842.

8. Data on this population measure are available from 1971 as CANSIM series V466971. We construct population data for the 1960s using the growth rate of the population aged 14 and over from the Canadian Historical Statistics.

9. Real GDP is measured using CANSIM series V1992067.

10. These series have DRI Economics mnemonic LBOUT and LBMN, respectively.

Figure 1
Data used in VAR, United States



run technology shock. In the United States, both labour productivity and average hours rise by roughly 1 per cent in the first year of the shock. Labour productivity continues to rise for the next eight years. Hours worked rises in a hump-shaped pattern, reaching a peak response of almost 2 per cent two years after a positive technology shock. Hours, then, slowly returns to the pre-shock level. Using Canadian data, we find similar results. In the year of the shock, labour productivity and hours worked both rise. Labour productivity rises by roughly 1 per cent while average hours rises by 0.4 per cent. The maximal rise in hours worked, 0.8 per cent, occurs a year after the shock. In Canada, it takes roughly six years for hours worked to return to its pre-shock level. For both countries, the rise in average productivity is statistically significant for a prolonged period of time. In contrast, confidence intervals about the estimated impulse-response functions for hours worked are wide. Still, the rise in U.S. hours worked is statistically different from zero for the first three years after the shock. For Canada, the rise in hours worked is statistically different from zero only in the first year after the shock.

As with the benchmark results in CEV (2003), our findings stand in sharp contrast to the literature, according to which hours worked in the United States falls after a positive technology shock (see, for example, Galí 1999 and Francis and Ramey 2001). But what accounts for this difference? It cannot be attributed to our identifying assumptions, since these are the same

Figure 2
Data used in VAR, Canada

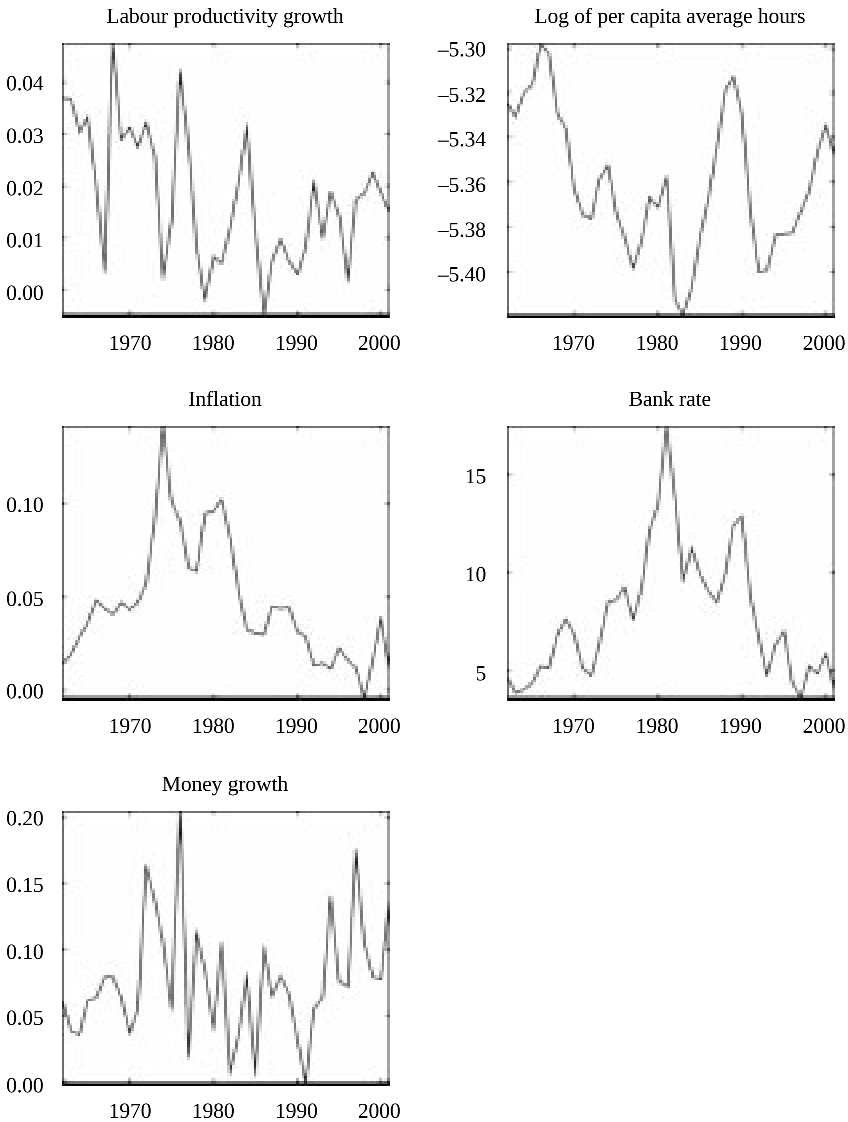
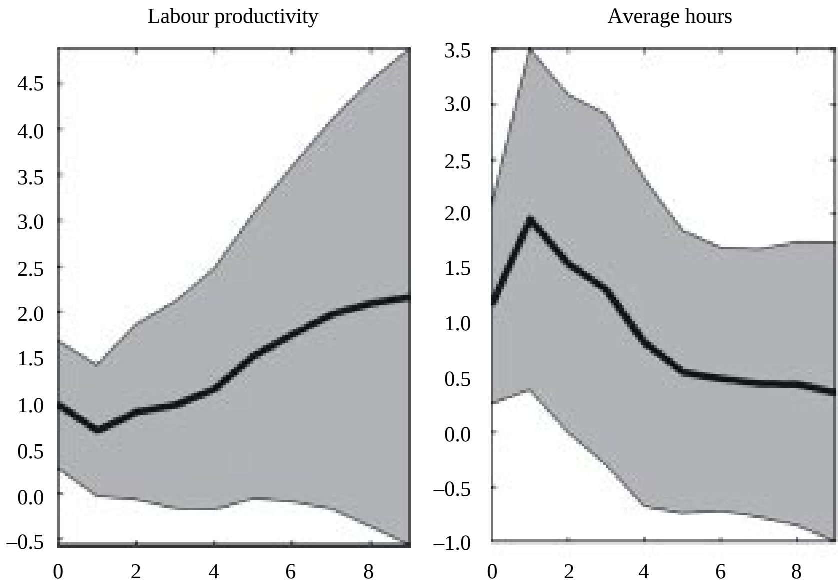
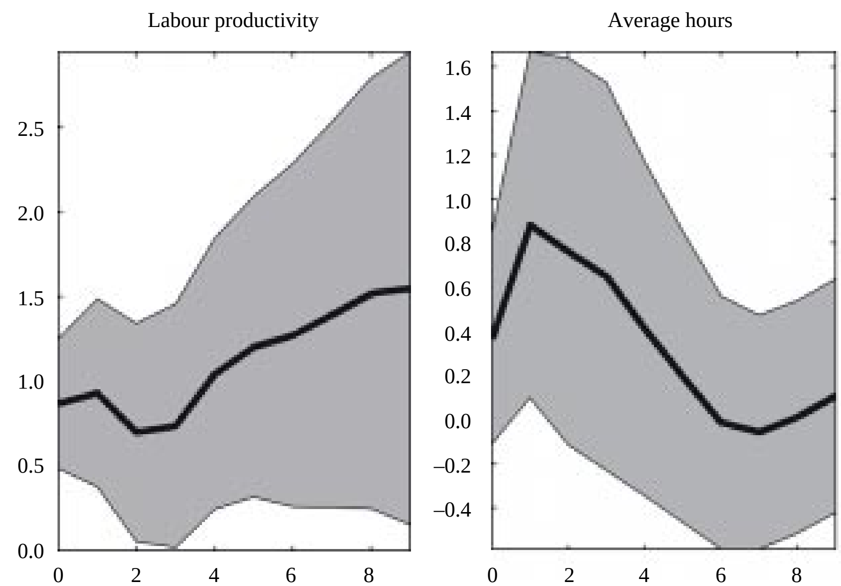


Figure 3
Impulse responses using the level specification

Panel A: United States



Panel B: Canada



Thick line: Impulse responses from level specification.
Grey area: 95 per cent confidence intervals.

as in the literature. Using quarterly U.S. data, CEV (2003) argue that the key difference has to do with how hours worked is incorporated into the analysis. There, we show that when we include Δh_t in Y_t, then hours worked falls after a technology shock. In contrast, when we include h_t in Y_t, then hours worked rises after a technology shock. For future reference, we refer to the specification when h_t is included in Y_t as the *level specification*. We refer to the specification when Δh_t is included in Y_t as the *difference specification*.

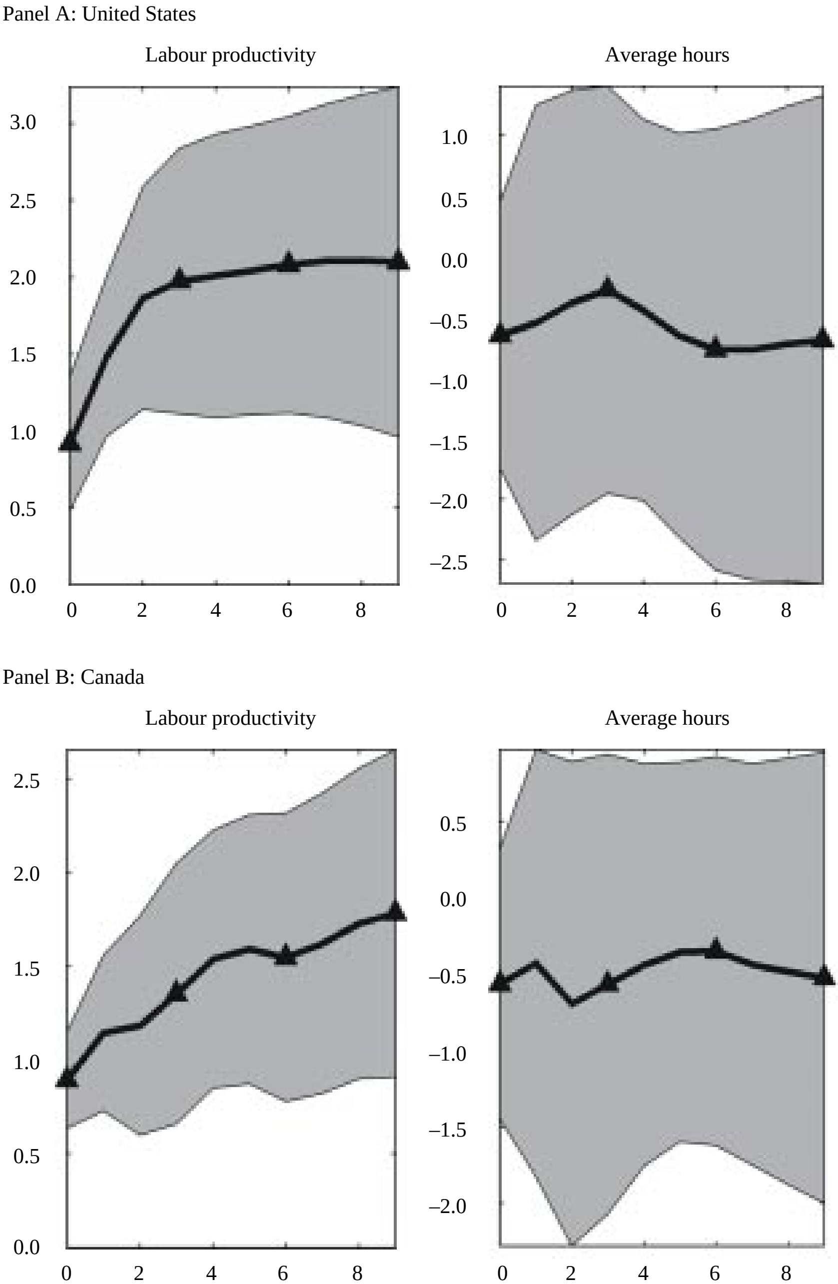
We find similar results for the annual Canadian and U.S. data. Specifically, suppose that X_t in equations (1) and (3) corresponds to the growth rate of hours worked rather than to the level of hours worked. Figure 4 reports our results for the United States and Canada. In both countries, a positive technology shock leads to a sharp, prolonged rise in labour productivity. In contrast to the results above, we now find that hours worked falls after a positive technology shock. Indeed, in both countries, according to our point estimates, hours worked never returns to the pre-shock level. Granted, confidence intervals are very large. But at least for the United States, CEV (2003) show that the initial fall in hours worked is statistically significant when the effect of a technology shock is estimated using quarterly data.

In sum, when we work with the level specification, a positive technology shock induces a large temporary increase in hours worked, both in Canada and the United States. But when we work with difference specification, a positive technology shock leads to a persistent decline in hours worked. In the next section, we address the question: Which of the competing results are more plausible, those based on the level specification or those based on the difference specification?

3 Choosing Between the Two Specifications

The level and difference specifications are based on different statistical models, corresponding to whether we assume that hours worked is difference-stationary or stationary in levels. As we saw, these specifications generated different answers to the question of what happens to hours worked after a positive technology shock. To assess which answer is more plausible, we must select between the statistical models underlying the two specifications. We first address this issue using standard classical diagnostic tests. Since these do not convincingly discriminate between the competing specifications, we then turn to the type of encompassing methods employed in CEV (2003).

Figure 4
Impulse responses using the level specification



Line with triangles: Impulse responses from difference specification.
Grey area: 95 per cent confidence intervals for simulated impulse responses.

3.1 Tests for unit roots and stationarity

In this subsection, we report the results for two well-known statistical tests of whether a univariate time series has a unit root. The first test is the augmented Dicky-Fuller (ADF) test, which tests the null hypothesis that hours worked has a unit root. The second test is the KPSS test (from Kwiatkowski et al. 1992), which tests the null hypothesis that hours worked is stationary. For the United States hour series, the ADF test fails to reject, at the 10 per cent significance level, the null hypothesis that per capita hours worked has a unit root.[11] At the same time, the KPSS test fails to reject the null hypothesis, at the 10 per cent significance level, that per capita hours worked is stationary.[12] For Canada, the results are somewhat more supportive of the level specification. The ADF test rejects the unit-root hypothesis at the 2.5 per cent significance level but fails to reject it at the 1 per cent level.[13] The KPSS test fails to reject the null hypothesis of stationarity at the 5 per cent significance level.[14] Based on these results, we conclude that conventional standard classical diagnostic tests cannot be used to convincingly discriminate between our two competing statistical models of per capita hours worked, either in Canada or the United States.

3.2 Encompassing tests

In the preceding section, we showed that conventional classical methods are not useful for selecting between the level and difference specifications of our VAR. An alternative way to select between the competing specifications is to use an encompassing criterion. Under this criterion, a model must not just be defensible on standard classical diagnostic grounds. It must also be able to predict the results based on the opposing model. If one of the two views fails this encompassing test, the one that passes is to be preferred.

In what follows, we review the impact of specification error and sampling uncertainty on the ability of each specification to encompass the other. Our

---

11. The ADF test statistic (with two lags) equals –1.6014. The critical value corresponding to a 10 per cent significance level is –2.57.

12. The value of the KPSS test statistic is 0.3221. The asymptotic critical value at the 10 per cent significance level is 0.347. In implementing this test, we set the number of lags in our Newey-West estimator of the relevant covariance matrix to two.

13. The ADF test statistic has a value of –2.5768 with three lags, while the small sample critical values are –2.26 and –2.66 at the 10 per cent and 5 per cent significance level, respectively. Asymptotic critical values for the ADF test statistic are –2.23 and –2.58 at the 10 per cent and 5 per cent significance level, respectively.

14. With two lags, the KPSS test statistic has a value of 0.3577. Therefore, one would not reject the null at the 5 per cent significance level using the asymptotic critical value of 0.463.

discussion here closely parallels the analysis in CEV (2003), who argue that, other things being equal, the specification that will do best on the encompassing test is the one that predicts that the other specification is misspecified.

We show in the next section that the level specification predicts that the difference specification is misspecified. We therefore expect that the level specification will do better than the difference specification. But as noted in CEV (2003), this consideration is not definitive because sampling considerations also enter. After discussing these issues, we present our bivariate encompassing results.

3.2.1 A priori considerations when the level specification is true

If the level specification is true and the econometrician adopts the difference specification, he is committing a specification error. To see why, recall the two steps involved in estimating the dynamic response of a variable to a technology shock. The first involves the instrumental-variables equation used to estimate the technology shock itself. The second involves the VAR used to obtain the actual impulse responses.

Suppose the econometrician estimates the instrumental-variables equation under the mistaken assumption that hours worked is a difference-stationary variable. In addition, assume that the only variable in X_t is log hours worked. The econometrician would difference X_t twice and estimate μ along with the coefficients in the finite-ordered polynomials, $\beta(L)$ and $\alpha(L)$, in the system:

$$\Delta f_t = \mu + \beta(L)\Delta f_{t-1} + \alpha(L)(1-L)\Delta X_t + \varepsilon_t^z .$$

Suppose that X_t has not been overdifferenced, so that its spectral density is different from zero at frequency zero. Then, in the true relationship, the term involving X_t is actually $\bar{\alpha}(L)\Delta X_t$, where $\bar{\alpha}(L)$ is a finite ordered polynomial. In this case, the econometrician commits a specification error because the parameter space does not include the true parameter values. The only way $\alpha(L)(1-L)$ could ever be equal to $\bar{\alpha}(L)$ is if $\alpha(L)$ has a unit pole, i.e., if $\alpha(L) = \bar{\alpha}(L)/(1-L)$. This is impossible, however, since no finite lag polynomial, $\alpha(L)$, has this property. So, incorrectly assuming that X_t has a unit root entails specification error.

We now turn to the VAR used to estimate the response to a shock. A stationary series that is first differenced has a unit moving average root. It is well known that there does not exist a finite-lag vector autoregressive representation of such a process. So here too, proceeding as though the data are difference-stationary entails a specification error.

3.2.2 A priori considerations when the difference specification is true

Suppose the difference specification is true but the econometrician works with the level specification. Here, the econometrician is not committing a specification error. To see this, first consider the instrumental-variables regression:

$$\Delta f_t = \mu + \beta(L)\Delta f_{t-1} + \alpha(L)\Delta X_t + \varepsilon_t^z, \tag{5}$$

where the polynomials, $\beta(L)$ and $\alpha(L)$, are of order q and $q-1$, respectively. The econometrician does not impose the restriction $\alpha(1)$ equals zero when it is, in fact, true. This is not a specification error, because the parameter space does not rule out $\alpha(1)$ equal to zero. In estimating the VAR, the econometrician also does not impose the restriction that hours worked is difference-stationary. This also does not constitute a specification error, because the level VAR allows for a unit root (see Sims, Stock, and Watson 1990).

The fact that the econometrician is not committing a specification error does not necessarily imply that the level specification can encompass the difference results. This is because sampling considerations must be taken into account. CEV (2003) stress that the difference specification implies that the level specification suffers from a weak-instrument problem. Weak instruments can lead to large sampling uncertainty, as well as bias. These considerations may help the difference specification explain the results of the level specification.

To see why a weak instrument arises, recall that the econometrician who adopts the level specification uses lagged values of X_t as instruments for ΔX_t. If X_t, however, actually has a unit root, this results in a weak-instrument problem. Lagged X_t's are poor instruments for ΔX_t, because ΔX_t is driven by relatively recent shocks, while X_t is heavily influenced by shocks that occurred long ago. At least in large samples there is little information in lagged X_t's for ΔX_t.

A different way to see why a weak-instrument problem arises when the econometrician mistakenly adopts the level specification is as follows. Consider the regression

$$\Delta h_t = a + \Pi h_{t-1} + p(L)\Delta h_{t-1} + q(L)\Delta f_{t-1} + \varepsilon_t^z. \tag{6}$$

A test of the hypothesis that Π is equal to zero has two interpretations. First, it is the covariates ADF test of Hansen (1995) for whether h_t has a unit root $(\Pi = 0)$. Second, it is the standard *F*-test for weak instruments discussed in Staiger and Stock (1997) for whether the lagged level of hours is a weak instrument for $\Delta h_t(\Pi = 0)$. So here, testing for the presence of a unit root in h_t is the same as testing for whether lagged hours is a weak instrument for the Δh_t. If the difference specification is true, then at least asymptotically one could not reject either hypothesis.

To summarize, when the level specification is true, the difference specification is misspecified. When the difference specification is true, the level specification is not misspecified, but the econometrician will encounter a weak-instrument problem and there will likely be large sampling uncertainty, as well as bias, associated with parameter estimates.

4 Encompassing Results

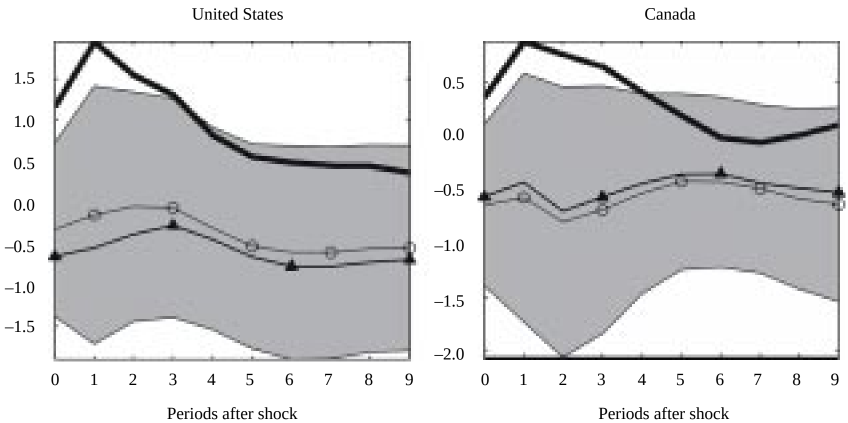
We base our encompassing tests on the ability of the level and difference specifications to match three observations. The first two observations come from the empirical-hours response that arises from the two different specifications. For the level specification, the average-hours response following a technology shock is positive. For the difference specification, the average-hours response is negative. The encompassing test compares the ability of each specification to account for both of these findings. The third observation is the empirical value of the weak instrument *F*-test. Specifically, we assess the ability of the level and difference specification to account for the observed *F*-test (equation 6). If the level specification is true then the lagged level of hours will be a good instrument. If the difference specification is true, then the lagged level of hours will be a poor instrument. Therefore, the two specifications ought to have different implications for the weak-instrument *F*-test.

4.1 Does the level specification encompass the difference specification results?

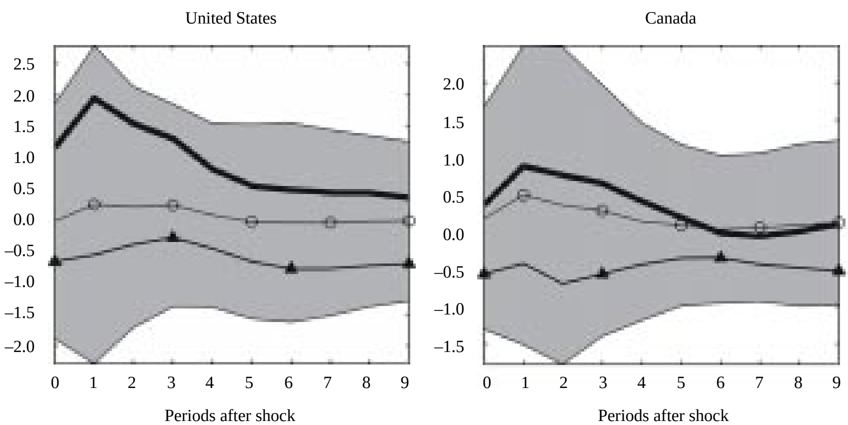
To determine whether the level specification can encompass the difference specification, we proceed as in CEV (2003). For each country, we use the estimated level specification VAR as the data-generating process (DGP). With this DGP, we simulate by bootstrap 1,000 artificial data sets, each of length equal to our actual sample size. For each simulated data set, we then (incorrectly) assume that the difference specification is true, and estimate a bivariate VAR in which hours worked appears in growth rates, and compute the impulse responses to a technology shock. Panel A of Figure 5 reports the

Figure 5
Encompassing results

Panel A: Level specification as the DGP



Panel B: Growth rate specification as the DGP



Thick line: Impulse responses from level specification.
Line with triangles: Impulse responses from difference specification.
Line with circles: Average impulse response for simulations from designated DGP.
Grey area: 95 per cent confidence intervals for simulated impulse responses.

distribution of impulse responses that arise from the simulation. The mean impulse responses appear as the thin line with circles. They correspond to the prediction of the level specification for the impulse responses that one would obtain with the (misspecified) difference specification. The lines with triangles are reproduced from Figure 4 and correspond to our point estimate of the relevant impulse-response function generated from the difference specification. The thick lines are reproduced from Figure 3 and correspond to the point estimate of the relevant impulse-response function generated from the level specification. The gray area represents the 95 per cent confidence interval of the simulated impulse-response functions.[15]

From Figure 5, we see that, for both countries, the average of the impulse-response functions emerging from the "misspecified" growth rate VAR is very close to the actual estimated impulse response generated using the difference specification. In particular, hours worked are predicted to *fall* after a positive technology shock even though they *rise* in the actual DGP. Evidently, the specification error associated with incorrectly adopting the difference specification can explain the estimated fall in hours found using the difference specification. In other words, the level specification attributes the decline in hours in the estimated VAR with differenced hours to over-differencing. We conclude that the level specification convincingly encompasses the difference specification.

4.2 Does the difference specification encompass the level results?

To assess the ability of the difference specification to encompass the level specification, we proceed as above except we now take as the DGP the estimated VARs in which hours worked appears in growth rates. Panel B in Figure 5 reports results analogous to those in Panel A. The thick solid lines, reproduced from Figure 3, are the impulse responses associated with the estimated level specification. The thin lines with the triangles are reproduced from Figure 4 and are the impulse responses associated with the difference specification.

The thin lines with circles in Panel B are the mean impulse-response functions that result from estimating the level specification of the VAR using the artificial data. They represent the difference specification's prediction for the impulse responses that one would obtain with the level specification. The grey area represents the 95 per cent confidence interval of the simulated impulse-response functions.

---

15. Confidence intervals were computed point-wise as the average simulated response plus or minus 1.96 times the standard deviation of the simulated responses.

Two results are worth noting. First, for the United States, the hours response is nearly zero. This result is closer to the difference specification than the level specification and therefore suggests that the distortion associated with not imposing a unit root in hours worked is not very large. The Canadian data have results that are somewhat different. Here the mean impulse response for hours is actually close to the level specification's result.[16] In part, this reflects the small sample-bias issues discussed in CEV (2003). Apparently, given the small sample of Canadian data, the level specification has some difficulty recovering the true impulse responses. Consistent with this, there is a great deal of sampling uncertainty associated with the estimated impulse-response function. Indeed, the confidence intervals in Panel B of Figure 5 are substantially much wider than those reported in the other figures.

Recall from our discussion above that if the difference specification is true, then an econometrician who works with the level specification ought to encounter large sampling uncertainty. This prediction faces a basic problem: It rests fundamentally on the difference specification's implication that there is a weak-instrument problem. But as we show below, when we apply a standard test for weak instruments to the data, we find little evidence of this problem. Moreover, the actual estimated confidence intervals associated with impulse responses obtained using the level specification are relatively narrow (see Figure 3).

4.3 Testing for weak instruments

To assess whether there is evidence of weak instruments in the data, we use a standard F-test for weak instruments. Specifically, we regressed ΔH_t on a constant, H_{t-1}, and the predetermined variables in the instrumental-variables regression, equation (5). These are ΔH_{t-s} and Δf_{t-s}, $s = 1, 2, 3$. Our weak-instrument F-statistic is the square of the t-statistic associated with the coefficient on H_{t-1}. In effect, this F-statistic measures the incremental information in H_{t-1} about ΔH_t.[17] If the difference specification is correct, the additional information is zero.

For the United States, the weak-instrument F-test is 5.95. This is below Staiger's and Stock recommended threshold value of 10, which suggests that, for the U.S. data, there may be a weak-instrument problem. In contrast,

16. These results are somewhat different than those reported for the quarterly U.S. data in CEV (2003).

17. As noted above, our F-test is equivalent to a standard ADF test with additional regressors. In the unit-root literature, this test is referred to as the covariate ADF test (Hansen 1995).

CEV (2003) report that there is little evidence of this problem with the U.S. quarterly data. For Canada, the weak-instrument *F*-test is 11.60, which suggests that there is no weak-instrument problem with the Canadian data.[18] As discussed above, this calls into question a basic implication of the view that hours worked has a unit root.

4.4 The relative odds of the two specifications

The results of the previous subsections indicate that the level specification can easily account for the estimated impulse-response functions obtained with the difference specification. The difference specification has a somewhat harder time accounting for the level specification results. As in CEV (2003), we quantify the relative plausibility of the two specifications by using the type of posterior odds ratio considered in Christiano and Ljungqvist (1988). Christiano and Ljungqvist developed their statistic for a similar situation where differences and levels of data lead to very different inferences.[19] In our context, we claim that the more plausible of the two VARs is the one that has the easiest time explaining the facts: (i) the level specification implies that hours worked rises after a technology shock; (ii) the difference specification implies that hours worked falls; and (iii) the value of the weak-instrument *F*-statistic.

The odds ratio that we use is calculated as follows. We simulated 1,000 artificial data sets using each of our two estimated VARs as the DGP. For an event Q, we asked what was the probability of observing that event for the level specification, $P(Q|A)$, where A denotes the level specification being true, and what was the probability of observing that event for the difference specification, $P(Q|B)$, where B denotes the difference specification being true. The relative plausibility of the two different specifications can then be assessed as the odds ratio of the level specification being true versus the difference specification being true given the observed event

---

18. The evidence against the difference specification reported here is stronger than we obtained using the ADF test in section 3.1. This is consistent with the analysis of Hansen (1995) and Elliott and Jansson (2003), who show that incorporating additional variables into unit-root tests can dramatically raise their power.

19. Eichenbaum and Singleton (1986) found in a VAR analysis that when they worked with first differences of variables, there was little evidence that monetary policy plays an important role in business cycles. However, when they worked with a trend stationary specification, monetary policy seems to play an important role in business cycles. Christiano and Ljungqvist argued that the preponderance of the evidence supported the trend stationary specification.

$$\frac{P(A|Q)}{P(B|Q)} = \frac{P(Q|A)P(A)}{P(Q|B)P(B)}.$$

If one had a prior distribution that put equal weight on A and B, then the odds ratio is just the ratio of the conditional probabilities

$$\frac{P(Q|A)}{P(Q|B)}.$$

We consider five definitions of the event Q:

(i) the difference specification is true, and the impact effect of a technology shock on hours worked is negative;

(ii) the level specification is true, and the impact effect of a technology shock on hours worked is positive;

(iii) both (i) and (ii) are true;

(iv) the weak instrument F-statistic test is greater than or equal to the F-statistic obtained with the actual data;

(v) events (i), (ii), and (iv) occur.

Table 1 reports the frequency with which the different events were observed in the two simulated data sets. For the U.S. case, the difference specification does slightly better at predicting event (i). But the level specification does much better at predicting events (ii) and (iv). The overall plausibility of the two specifications can be most easily assessed in terms of event (v), in which case, the odds ratio favours the level specification by over five to one. Similar results hold for the Canadian case. Surprisingly, the fall in hours associated with the difference specification is actually observed more frequently when the DGP corresponds to the level specification. Focusing on event (v), the odds ratio favours the level specification by over nine to one. So the odds in favour of the level specification are even higher for the Canadian case than for the U.S. case. This is consistent with our F-test results indicating less of a weak-instrument problem for Canada than for the United States.

5 Multivariate Results for Canada

In this section, we discuss how other Canadian variables (money growth, inflation, and the interest rate) respond to a technology shock.[20] To estimate these response functions, one could proceed as in CEV (2003) and estimate

20. See CEV (2003) and ACEL (2003) for multivariate results based on quarterly U.S. data.

Table 1
Simulations results and odds ratios

Specification	United States: Percentage true Level	United States: Percentage true Difference	United States: Odds	Canada: Percentage true Level	Canada: Percentage true Difference	Canada: Odds
Event						
(i) Impact negative, difference specification	67.7	71.4	0.949	91.6	81.1	1.130
(ii) Impact positive, level specification	97.1	54.3	1.788	96.9	66.8	1.451
(iii) Both impact responses	65.6	37.3	1.757	88.5	53.3	1.660
(iv) *F*-test	70.1	22.9	3.066	65.0	10.7	6.070
(v) Events (i), (ii), and (iv) true	47.6	8.9	5.343	57.6	6.2	9.290

the large simultaneous system. However, given our small sample size and the large number of parameters that would have to be estimated, we are reluctant to do so. For example, a five-variable VAR with four lags requires estimating 110 coefficients with only 40 years worth of annual data. Instead, we adopt the following sequential approach. Suppose we are interested in estimating the dynamic response functions of a set of variables, $X_1, X_2, X_3 \ldots$. For a given variable, X_i, we estimate the technology shocks $\varepsilon_t^{z,i}$ using the following version of our basic IV equation:

$$\Delta f_t = \mu + \beta(L)\Delta f_{t-1} + \alpha_H(L)\Delta H_t + \alpha_{X_i}(L)\Delta X_{it} + \varepsilon_t^{z,i}. \qquad (7)$$

We then estimate the VAR,

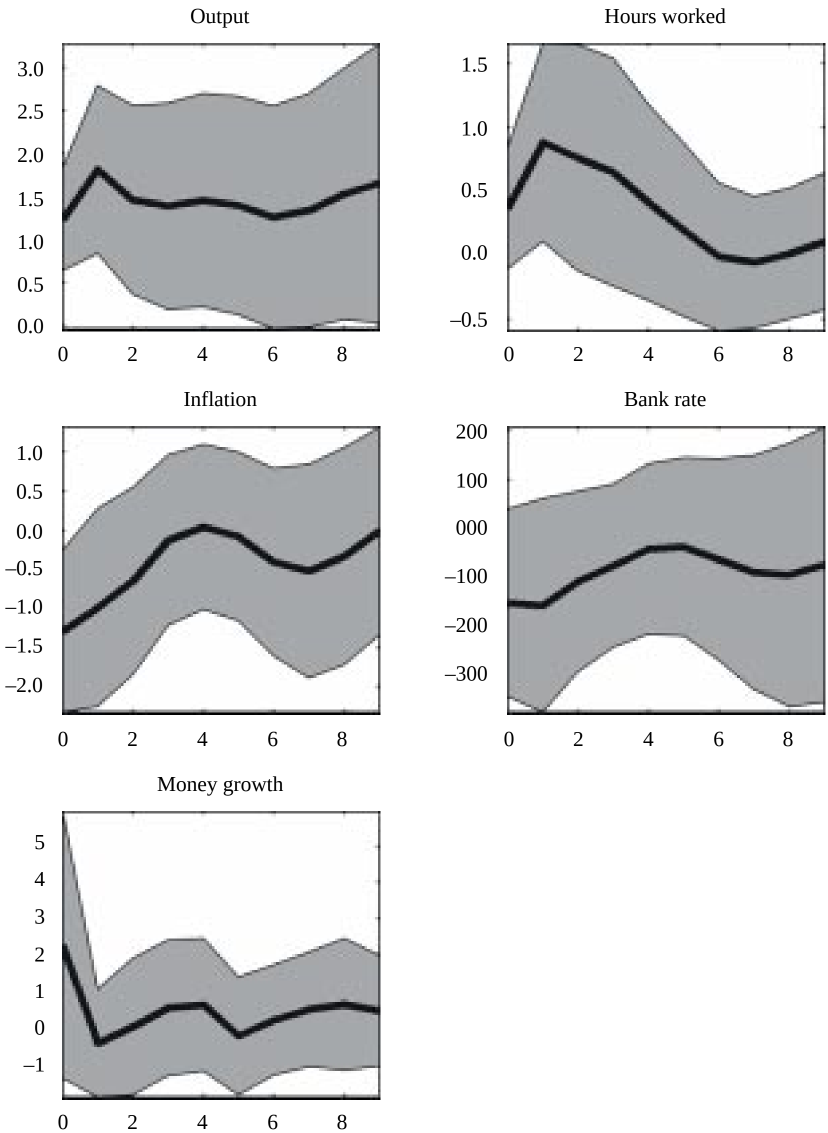
$$\begin{pmatrix} \Delta f_t \\ H_t \\ X_{it} \end{pmatrix} = B(L)\begin{pmatrix} \Delta f_{t-1} \\ H_{t-1} \\ X_{it-1} \end{pmatrix} + \gamma\varepsilon_t^{z,i} + \nu_t .$$

Finally, we derive the impulse response of X_{it} to a technology shock from the estimated VAR. Given the results of the previous section, we confine our attention to the level specification.

Figure 2 displays the time series on Canadian money growth, inflation, and the interest rate.[21] Figure 6 reports the estimated response of these variables to a technology shock. For convenience, the hours responses are repeated

21. The Bank Rate is measured using CANSIM series B14006. The money supply is M2, with CANSIM mnemonic B1630. Inflation is measured as the growth rate of the GDP deflator D15612.

Figure 6
Sequential analysis



Thick line: Impulse responses from level specification.
Grey area: 95 per cent confidence intervals for simulated impulse responses.

from Figure 3 as well as the output response implied by the bivariate VAR reported in Figure 3. Two key results emerge here. First, our point estimates indicate that inflation and the interest rate fall after an expansionary technology shock, while money growth rises. This suggests that the Bank of Canada accommodated technology shocks over our sample period. Second, the confidence intervals around the estimated impulse-response functions are quite wide. Still, the initial fall in inflation is statistically significant.[22]

Conclusions

Using annual Canadian and U.S. data, this paper argues that a positive technology shock leads to a rise in Canadian output and hours worked, as well as a fall in inflation. In CEV (2003), using similar methods and quarterly U.S. data, we argued that a technology shock also leads to a rise in aggregate consumption and investment. On the face of it, these findings are consistent with the predictions of a real-business-cycle model. But, in our view, the rise in hours worked and overall expansion in aggregate activity that follows in the wake of a technology shock reflect how monetary policy makers reacted to the technology shock. While our empirical results are suggestive on this point, they are not definitive. To make the case convincingly requires that we know how the economy would have reacted had the monetary authority acted differently. The only place where we can perform such a counterfactual experiment is in a structural economic model.

ACEL (2003) conduct this type of experiment in a dynamic general-equilibrium model embodying wage and price frictions. They argue that an estimated version of the model can account for how the U.S. economy reacted to monetary policy and technology shocks in the post-war era. They use this model as a laboratory to investigate how the U.S. economy would have reacted to a technology shock with a different monetary policy. For example, they consider what would have happened under the assumption that the Federal Reserve had not accommodated technology shocks but rather had followed a k per cent money-growth rule. The key conclusion in ACEL is that, with this counterfactual policy, hours worked would have *fallen* for a prolonged period of time after a positive technology shock. In addition, compared to the actual outcomes, output would have risen by far less and inflation would have fallen by far more. We suspect that the same result would be true for Canada. Based on these results and similar findings in Galí, López-Salido, and Valles (2003), it is clear that policy-makers ought to be vitally interested in the supply-side developments of the economy.

22. Note that the confidence intervals become extremely wide in the long run. This may reflect a near unit root in our measure of the Canadian interest rate.

Simple formulations of policy, like Taylor rules, often push discussions of the output gap into the background. Knowing why the gap moves is a critical input into policy decisions. Taken together, the results of this paper, CEV (2003) and ACEL (2003), suggest that policy-makers have, in fact, been successful at identifying technology shocks and have reacted in a way that has improved aggregate economic performance.

References

Altig, D., L.J. Christiano, M. Eichenbaum, and J. Linde. 2003. “An Estimated Dynamic, General Equilibrium Model for Monetary Policy Analysis.” Manuscript.

Christiano, L.J. and L. Ljungqvist. 1988. “Money Does Granger-Cause Output in the Bivariate Money-Output Relation.” *Journal of Monetary Economics* 22 (2): 217–35.

Christiano, L.J., M. Eichenbaum, and R. Vigfusson. 2003. “What Happens After a Technology Shock?” Board of Governors of the Federal Reserve System International Finance Discussion Paper No. 768.

Doan, T.A. 1992. *Rats User’s Manual*. Evanston, IL: Estima.

Eichenbaum, M. and K.J. Singleton. 1986. “Do Equilibrium Real Business Cycle Theories Explain Postwar U.S. Business Cycles?” *NBER Macroeconomics Annual 1986* 91–135.

Elliott, G. and M. Jansson. 2003. “Testing for Unit Roots with Stationary Covariates.” *Journal of Econometrics* 115 (1): 75–89.

Fisher, J.D.M. 2002. “Technology Shocks Matter.” Federal Reserve Bank of Chicago Working Paper No. WP–02–14.

Francis, N. and V.A. Ramey. 2001. “Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited.” Manuscript, University of Sourthern California, San Diego.

Galí, J. 1999. “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?” *American Economic Review* 89 (1): 249–71.

Galí, J., J.D. López-Salido, and J. Valles. 2002. “Technology Shocks and Monetary Policy: Assessing the Fed’s Performance.” NBER Working Paper No. 8768.

Hansen, B.E. 1995. “Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power.” *Econometric Theory* 11 (5): 1148–71.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin. 1992. “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root.” *Journal of Econometrics* 54 (1/2/3): 159–78.

Shapiro, M. and M. Watson. 1988. "Sources of Business Cycle Fluctuations." *NBER Macroeconomics Annual* 111–48.

Sims, C.A., J.H. Stock, and M.W. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica* 58 (1): 113–44.

Staiger, D. and J.H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–86.

Vigfusson, R.J. 2002. "Why Does Employment Fall After a Positive Technology Shock?" Manuscript.