Testing and Estimating Persistence in Canadian Unemployment.

Ossama Mikhail* Curtis J. Eberwein† Jagdish Handa‡

October 20, 2003

Abstract

A vital implication of unemployment persistence applies to the Bank of Canada's disinflation policies since it adversely influences unemployment and considerably lengthens recessions. This paper tests for persistence in Canadian sectoral unemployment, using the modified rescaled-range test. Our results show evidence of persistence in sectoral unemployment that translates to persistence in aggregate unemployment. To quantify this aggregate-level persistence, we estimate it within the framework of Bayesian ARFIMA class of models. The results conclude that Canadian unemployment exhibits persistence in the short and intermediate run.

JEL CLASSIFICATION: C11, C22, E24.

*Corresponding Author. Department of Economics, College of Business Administration, University of Central Florida. E-mail: omikhail@bus.ucf.edu. I am grateful to the McGill Major Fellowship for financial support, to John W. Galbraith and seminar participants at the University of Central Florida for helpful comments.

†Center for Human Resource Research, Ohio State University. E-mail: ceberw@hotmail.com

‡Department of Economics, McGill University. E-mail: Jagdish.Handa@mcgill.ca

1 Introduction

Persistence in unemployment has long been documented, explored and investigated at the theoretical and applied levels. Unemployment returns each year about one third of the way to its normal level after a shock displaces it (Hall 1998, p. 34). This is the case for both the U.S.A. and Canada. Whenever evidence of persistence is found, there exists room to decrease the unemployment rate without changing any structure in the organization of the labor market. How fast can the unemployment rate be decreased depends on the persistence mechanism. Also, disinflation policies based on the unemployment rate will prove very costly in terms of lost output. If unemployment exhibits persistence, it will never go back to its original starting point. This is a vital implication of unemployment persistence, and applies to the Bank of Canada's disinflation policies. Whenever unemployment persists, the short-run adjustment of the economy can take place over a very long period.

This paper addresses the testing for persistence in sectoral Canadian unemployment data and the estimation for persistence in Canadian total unemployment data. Using the modified rescaled range test, we test for persistence in Canadian unemployment. To quantify this persistence, we estimate it within the framework of Bayesian ARFIMA class of models.

Following a brief review of the literature, Section 3 reports descriptive statistics for Canadian unemployment by industry. Section 4 tests for persistence in unemployment by industry data. Section 5 estimates persistence in aggregate Canadian unemployment using the class of Bayesian ARFIMA models. Section 6 presents our conclusions.

2 Literature Review

In this section, we document the stylized fact that unemployment exhibits persistence. A number of studies debated this empirical observation. McCallum (1989), Fortin (1989) Cozier and Wilkinson (1991) and Poloz and Wilkinson (1992) argued and reported evidence suggesting the absence of persistence in Canadian unemployment. Fortin (1989, 1991) tested for the presence of hysteresis in Canadian data covering the period from 1957 to 1990. By adding and modeling expected inflation, Fortin was able to undertake a more accurate test for hysteresis. The Phillips curve tested was $\pi_t = \alpha_1\pi_{t-1} + \alpha_2\pi_t^e + \beta[(1-\eta)U_t + \eta\Delta U_t] - \beta\gamma Z_t$, where π and π_t^e denotes inflation and expected inflation, respectively. U_t refers to unemployment and Z_t denotes other explanatory variables. Fortin defined positive hysteresis as $\eta < 0$ and negative hysteresis as $\eta > 0$. The cases of $\eta = 0$ and $\eta = 1$ are no hysteresis and full hysteresis, respectively. Fortin (1991) reported the presence of negative hysteresis for the data from 1957 to 1972. Positive hysteresis was detected for the data covering the period from 1973 to 1990. Full hysteresis was not rejected for the latter period. Fortin pointed to the Canadian unemployment insurance benefits, productivity slowdown, and union density as possible sources for hysteresis.

Coe (1990) attributed the cross-country differences in unemployment to differences in the industry wage-determination process and found evidence that the institutional structure for the determination of industry wages contributes more to the persistence of unemployment in Europe than North America and Japan. Benassi, Chirco and Colombo (1994, p. 100) used insider-outsider models to explain persistent unemployment rates. Winter-Ebmer (1991) summarized different tests of persistence, used in the literature. Heckman and Borjas

(1980) asked whether current unemployment causes future unemployment and found evidence of unemployment persistence. They drafted four ways of modeling state dependence: Markovian, occurrence, duration and lagged duration dependence.

Jones (1995) investigated the hysteresis hypothesis in Canadian data at the microeconomic and macroeconomic level. He concluded that the overall picture is not one of hysteresis, but did not rule out the presence of persistence (dependence) in unemployment rates. Nott (1996) did not find evidence of hysteresis in Canadian data. Yet, a non-linear Phillips curve was not rejected. The method followed Fortin (1991) in testing for the presence of hysteresis by estimating a linear Phillips curve equation. Using data from 1954 to 1995, Nott's results contradicted Fortin's findings of hysteresis and showed how sensitive the latter's results were to the sample period used. Wilkinson (1997) investigated the hysteresis hypothesis in Canadian data using the labor Market Activity Survey (LMAS). Defining hysteresis as irreversibility in the change of the unemployment rate and by testing evidence of negative duration dependence in unemployment spells, the study concluded that there is evidence of hysteresis at the micro data level. Wilkinson attributed the evidence of hysteresis to the loss of skills hypothesis of human capital. The intuition is that prolonged periods of unemployment erode the skill level of the unemployed which decreases the probability of exiting the unemployment spell and finding a job. Therefore, unemployment spells will exhibit negative duration dependence. Using the LMAS data, single-risk hazard rates were estimated, then aggregated to estimate hysteresis at the macro level. The study concluded that hysteresis accounts for three percent to eight percents of the Canadian unemployment rate. This small upperbound points to the difficulty of estimating hysteresis in the aggregate data.

Among others, Carey (1997) argued that cyclical unemployment rise if expected inflation is persistently higher than its current level, especially for the post-1993 data period. Therefore, persistent unemployment is caused by persistent excess inflationary expectations. Other studies also have investigated the relationship between unemployment, policy variables and labor market rigidities (see Nickell (1997) and Riddell (1999) for excellent expositions). Gil-Alana (2001) examined the persistence of unemployment in the US and four European countries by means of fractionally integrated ARMA (ARFIMA) models. The results provided a persistence ranking across the studied countries.

Koustas and Veloce (1996) used time series long-memory modeling to test for the presence of hysteresis in Canada and in the U.S.A. Using disaggregated data for unemployment rates by age and gender, they concluded that unemployment is more persistent for adult males relative to adult females. They reported that unemployment persistence in Canada is higher than in the U.S.A. Here, we extend their research by testing for unemployment persistence at sectoral level.

The fact that 'unemployment exhibits persistence' is well documented. Explaining persistence has been and still is a challenging task for macroeconomists. Many directions have been pursued, each of which contains some truth, but none is a completely satisfactory explanation. A puzzling lack of a standard definition for persistence exists. Currently, there does not exist a consensus in the empirical literature on the definition of 'persistence' in unemployment. Different authors use different definitions for this term. This paper adopts the following definition (see Mikhail, Eberwein and Handa (2003)). 'Unemployment persistence' is defined as the 'effect of a shock on unemployment felt for a minimum period of two years'.

3 Data

The variables of interest are total unemployment, and unemployment in the goods, manufacturing and services sectors for Canada. All the series are in log-level form and were detrended using the Hodrick-Prescott (HP) filter. Table 1 reports the source of the data. Table 2 reports basic descriptive statistics for Canadian unemployment by industry. Manufacturing unemployment is the most variable while service unemployment is the least variable. Table 3 presents the correlation matrix between Canadian sectoral unemployment. Services unemployment is highly correlated (0.836) with total unemployment. Also, manufacturing unemployment is highly correlated with goods sector unemployment (0.877). All unemployment series are positively correlated with each other.

Table 4 reports the sample autocorrelations for all unemployment series and for a maximum lag of 6. All monthly unemployment series exhibit slow decay. Faced with a shock, all monthly unemployment series qualify for persistence. If one analyses only the annual level data, one is bound to miss this evidence of persistence. The $Q_{LB}(6)$ statistics for all cyclical series are significant at the 5 percent level[1] (see Endnotes, no. 1). Table 5 reports the cyclical properties of the monthly unemployment series. All series are coincident with total unemployment. The $Q_{LB}(1 \text{ to } 3)$, $Q_{LB}(-3 \text{ to } -1)$ and the $Q_{LB}(-3 \text{ to } 3)$ statistics for all cyclical series are significant at the 5 percent level

[1] In small sample, the Q_{LB} suffers from lack of power. This is the principal reason for not pursuing annual data persistence analysis.

4 Testing for Persistence

Using time series analysis, the slow decline in the sample autocorrelation function is generally viewed as an indication of an integrated process. Cochrane (1988) proposed using the variance ratio to test for non-stationarity. We computed the Cochrane variance ratio for Canadian unemployment data, and concluded that after 3 years, the persistence ranking of the Canadian unemployment series (from highest to lowest) is: total unemployment, services, manufacturing and goods unemployment.[2] For the subsequent sections, we focus on documenting this persistence in aggregate unemployment.

In this paper, we present an alternative approach to testing and estimating long-range dependence. The rescaled range test statistic[3] presented here has a distinct advantage over the Cochrane variance ratio test, for determining long-range dependence.[4] There is a growing literature on long-memory processes.[5] Most of this literature treat long-memory processes as fractionally integrated processes (McLeod and Hipel (1978), see Endnotes, no. 2).

Often in economics, time series processes exhibit a hyperbolic rate of decay that is neither consistent with an I(1) process nor an I(0) process. A fractionally differenced (i.e., long memory) process can be regarded as a midpoint [labelled as a "halfway house" by Baillie (1996, p. 6)] between I(0) and I(1) processes. The attractive feature of long-memory processes is their long run predictions and effects of shocks. These predictions are very different from

[2]The results are not reported for space and are available upon request. Also, we save space by not pursuing persistence analysis for the least persistent series, namely goods unemployment.

[3]Doukhan (2003) contains an excellent survey on semiparametric estimation.

[4]See Mandelbrot (1972) for the superiority of the rescaled range test statistic over analyzing variance ratios. An application of 8 test statistics to 6 exchange rate daily series, showed that the R/S statistic is the most robust test against short-run effects but has the disadvantage of relatively smaller power (Hauser 1997, p. 269).

[5]It focuses on the hyperbolically decaying autocorrelations and impulse response weights properties of the time series under investigation. Note that a hyperbolic decay rate is lower than the exponential rate observed with the ARMA class of models.

those conventional ARMA class modeling which considers only the exponential or geometric rate of decay on the Wold decomposition coefficients.

Granger (1980) showed that the contemporaneous aggregation of panel data resulted in fractionally integrated processes (see Endnotes, no. 3). Therefore, given that total unemployment is a contemporaneous sum of N sectors unemployment, it is appropriate to treat the Canadian total unemployment[6] level data as following a long-memory process.

Our empirical analysis starts by testing for unit roots using the Augmented Dickey-Fuller test (see Table 6 for the results and Endnotes, no. 4 for the test). Including a trend and 12 lags,[7] the null hypothesis of a unit root was not rejected at the 1 percent level for all logs of the monthly level of unemployment series covering the period 1976:1 to 1998:12.

However, any evidence of a unit root is weak, since I(1) is the null and the result could be attributable to structural breaks in the series. These breaks are well documented and apparent in the graph of the series - for example the recessions of 1981-1982 and 1991-1992. As noted by Rappoport and Reichlin (1989), among others, most unit roots tests have difficulty discriminating between an I(1) process and an I(0) process with a shift in its mean. They found that most unit root tests tend to favor the difference-stationary (DS) model whenever the true process is a segmented trend. As a result, the DS model provides the better 'fit' in most applications. In our view, another important finding of their paper is that "Many quantity series appear to be adequately parametrized by segmented trends, which undergo intermittent shocks, between which they behave as trend stationary processes" (Rappoport and Reichlin 1989, p. 176). For our analysis, this quote suggests that the unemployment series might behave as a segmented-trend type rather than a difference

[6] Using Statistics Canada Input-Output Tables, the L-level of unemployment includes 112 industries.
[7] Using the AIC criterion, we experimented with other lag lengths and similar results were obtained.

stationary type. In summary, structural breaks induce a bias towards non-rejection of the null hypothesis of a unit root.[8]

Next, we shift the focus towards the midpoint between I(1) and I(0) processes, i.e., long-memory processes. To start, we investigate the shape of the sample autocorrelations to assess if further long-memory analysis is to be carried out. Tables 7, 8 and 9 illustrate the correlogram of monthly[9] *level* data for total, manufacturing and services unemployment. The apparent pictorial evidence of a hyperbolic decay rate of the sample autocorrelations will be formally tested in subsection 4.1. We then proceed to estimate and test this long-memory behavior. Note that, Newbold and Agiakloglou (1993) argued that the detection of long-memory properties through the examination of the sample autocorrelations is almost impossible.

In short, we test for the presence of long-range dependence using the modified rescaled range test and we quantify persistence by estimating the fractional integration parameter using a Bayesian ARFIMA model.[10]

---

[8] Enders (1995, pp. 243-248) has a good exposition of unit root tests in the context of structural breaks.

[9] The reason for reporting the monthly data correlogram is that - for the monthly data - the hyperbolic decay is more accentuated and less apparent at higher data frequency. For other frequencies, the decay is present. We analyzed all data frequencies and for space limitation, we choose to report quarterly data results henceforth.

[10] The most widely used notion of short-range dependence is the concept of 'strong mixing' due to Rosenblatt (1956). It measures the decline of statistical dependence between events separated by successively longer spans of time. As the time span increases and the maximal dependence between events becomes trivially small, then the time series is a strong-mixing one, such as the class of ARMA models wherein the autocorrelations decay exponentially. Dependence between events over a long span defines long-range dependence, such as long-memory processes (or fractionally integrated processes given the definition in equation (30)).

4.1 The Rescaled Range Statistic (R/S)

The rescaled range statistic (Hurst 1951) is set to detect long-range dependence. It is defined as R_T/s_T,

(1)
$$R_T \equiv \max_{1 \le k \le T} \left\{ \sum_{j=1}^{k} (y_j - \overline{y}) \right\} - \min_{1 \le k \le T} \left\{ \sum_{j=1}^{k} (y_j - \overline{y}) \right\}$$

(2)
$$s_T = \left\{ \frac{1}{T} \sum_{t=1}^{T} (y_t - \overline{y})^2 \right\}^{0.5}$$

where R_T is the range, s_T is the sample standard deviation, and $\overline{y}$ denotes the sample mean. Under the null hypothesis of a simple i.i.d. process, Lo (1991, pp. 1287-1288) showed that $T^{-1/2}R_T/s_T$ is asymptotically distributed as the range of a standard Brownian Bridge on the unit interval and has expectation of $(\pi/2)^{1/2} = 1.253$ and a standard deviation of $[(\pi/2)(\pi - 3)/3]^{1/2} = 0.272$.

Sensitivity to short-range dependence is the most important shortcoming of the rescaled range. For example, if the process is an AR(1), i.e., purely short-range dependent, then the mean of the rescaled range limiting distribution will be biased. To counter and to correct for the impact of short-range dependency on the test statistic, Lo (1991, p. 1289) devised a modified rescaled range statistic.[11] The proposed modified statistic is robust to many forms of heterogeneity and weak dependence. Also, it is able to discriminate between short- and

[11] By correcting for short-range dependency, the limiting distribution of the modified statistic is invariant to many forms of short-range dependency but sensitive to the presence of long-range dependency.

long-range dependency. The modified rescaled range statistic is defined as,

(3) $$Q_T \equiv \frac{R_T}{\sigma_T(m)}$$

where,

(4) $$\sigma_T^2(m) = c_0 + 2\sum_{j=1}^{m} w_j(m)c_j$$

c_j denotes the jth order sample autocovariance of y_t and $w_j(m)$ are the Newey and West (1987) weights using a Bartlett window defined as,

(5) $$w_j(m) = 1 - \left[\frac{j}{m+1}\right] \qquad m < T$$

In the presence of long memory, the normalized statistic $T^{-1/2}Q_T$ weakly converges to the range of a Brownian Bridge. The distribution is given by,

(6) $$F(x) = \sum_{j=-\infty}^{\infty} (1 - 4x^2j^2)\exp[-2x^2j^2]$$

This distribution is positively skewed and its fractiles are tabulated in Lo (1991, p. 1288). The modified rescaled range statistic is robust to short-range dependence and consistent with a general class of long-range dependent stationary Gaussian alternatives (see Baillie (1996, p. 28)).

The choice of m is a subject open to debate. $m \equiv [k_T]$, where $[k_T]$ denotes the greatest integer less than or equal to k_T. As defined and proposed by Lo (1991, p. 1302), $[k_T]$ is a data-dependent approach for the choice of m,

$$k_T \equiv \left(\frac{3T}{2}\right)^{1/3}\left(\frac{2\widehat{\rho}}{1-\widehat{\rho}^2}\right)^{2/3} \tag{7}$$

where $\widehat{\rho}$ is the estimated first-order autocorrelation coefficient of the data. For our analysis, and using quarterly data, we computed the modified statistic at $m = 1, 2, 3, 4, 5, 6, 7, 8$.

Table 10 reports the results of the modified rescaled range (Q_T) statistic for the first difference of the log of total unemployment, and manufacturing and services unemployment. All series are in log form and y_t denotes the log of the time series. For each value of m, the first column reports the $\sum_{j=1}^{m} w_j(m)c_j$, i.e., the sum of the weighted autocovariances. Subsequent columns report the logarithm of Q_T and the normalized test statistic value $\frac{Q_T}{\sqrt{T}}$.

Given the reported critical values in Lo (1991, p. 1288), we test the null hypothesis of short-range dependence. Table 10 computes the normalized test statistic values for Δy_t and Table 11 computes the same statistic for the Hodrick-Prescott filtered y_t. Note that the normalizing factor $\sqrt{T}$ is different in both tables. For Tables 10 and 11, the sample size is 91 and 92 observations, respectively. The reason for computing both tables is to investigate the sensitivity of the modified test statistic results to the method of detrending. Also, to check the sensitivity of the statistic to the lag length, the normalized test statistic is computed for several different values of m. Given that the normalized test statistic follows a Brownian Bridge process, the null hypothesis is examined at the 95 percent confidence level by not rejecting or rejecting according to whether the normalized test value is or is not contained in the interval [0.809 , 1.862].

Table 10 significantly rejects the simple null hypothesis of i.i.d. process at most values of

m. Long-range dependence is evident in Canadian manufacturing, services and total unemployment. Table 11 gives similar results. However, the persistence of total unemployment is less evident at the data-dependent value of m. Shorter values of m are picking up the short-range dependence. Using the Hodrick-Prescott filter increases the m lag where the first evidence of persistence is reported. For example, the evidence of persistence for total unemployment is first reported at $m = 4$ when using Δy_t and at $m = 5$ when using the HP filtered y_t. This one lag delay holds for total and manufacturing unemployment. For services unemployment, the lag delay is longer.

Given the strong evidence of long-range dependence in Δy_t series, we decided to continue our analysis of long-range dependence. The next section proposes a Bayesian approach to estimate several ARFIMA models in order to quantify the fractional integration parameter.

5 Bayesian ARFIMA

The main reasons for undertaking the long-memory analysis of the quarterly aggregate Canadian unemployment are: 1) the evidence of persistence reported by the Cochrane variance ratio test and the modified rescaled range test statistic; 2) the non-rejection of the null hypothesis of a unit root (that might be due to structural breaks); and 3) long memory may still appear at the macro level due to contemporaneous aggregation. Finally, in support of our argument, we quote Koop, Ley, Osiewalski and Steel (1997, p. 150) "when analyzing aggregated data, we should keep the possibility of long memory open."

Adopting a Bayesian approach to estimate ARFIMA has some advantages over the classical techniques.[12] First, it provides exact finite sample distributions for the impulse response

[12]There are quite a few non-Bayesian statistical techniques to estimate ARFIMA class of models. The most

and the fractional differencing parameter. Second, for predictive purposes, the Bayesian approach allows one to average across models instead of just picking one model. Third, one can perform small-sample tests of memory properties to discriminate between ARIMA and ARFIMA models simply by attaching a positive prior to the point where the fractional integration parameter equals one. The notation and derivations in this section closely follow Koop, Ley, Osiewalski and Steel (1997).

Since the modified rescaled range test statistic pointed to strong evidence of persistence in Δy_t, and to avoid any artificial distortion of the statistical properties of the data induced by the Hodrick-Prescott filter,[13] we focus our analysis on the first difference of the quarterly log of total Canadian unemployment-level. Rewrite the ARFIMA process as,

$$z_t \equiv \Delta y_t - \mu \tag{8}$$

The ARFIMA(p, δ, q) representation of this process is,

$$\phi(L)(1-L)^{\delta} z_t = v(L)\varepsilon_t \tag{9}$$

commonly used techniques can be classified as follows, 1) Maximum Likelihood methods (Sowell (1992)); 2) Approximate Maximum Likelihood methods (Baillie and Chung (1993), Li and Mcleod (1986), or using the Whittle approximation as outlined by Fox and Taqqu (1986)), where the estimation of the parameter 'd' is done at the same time as the estimation of the other parameters (coefficients of the AR and the MA parts); 3) Two-Step procedures (Geweke and Porter-Hudak (1983) and Janacek (1982)), and finally, 4) The non-iterative approximation based estimators as in Durbin (1959, pp. 307-308) and Galbraith and Zinde-Walsh (1994, pp. 144-147). This method relies on approximating the moving average process by an autoregressive model and uses the pattern of autoregressive coefficients to deduce estimates of the parameters of the underlying process. The Galbraith and Zinde-Walsh estimator have a lower bias than Durbin's for a given approximation order. This class of estimators are asymptotically efficient and more robust - regarding misspecification - to maximum likelihood based methods.

[13] The HP filter "removes important time series components that have traditionally been regarded as representing business cycle phenomena" King and Rebelo (1993, p. 208). For a complete discussion of the negative effects of the Hodrick-Prescott, see Stadler (1994, pp. 1768-1769) and Cogley and Nason (1995, p. 276). For spurious cyclical behaviour induced by the filter, see Harvey and Jaeger (1993, pp. 233-235). The results without the HP filter were computed and not reported for space.

where $v(L) = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)$ and $\phi(L) = (1 + \phi_1 L + \phi_2 L^2 + ... + \phi_p L^p)$ are polynomials in the lag operator and the roots lie outside the unit circle. Let $\theta \in \mathbb{C}^q$ and $\phi \in \mathbb{C}^p$. The errors[14] ε_t are i.i.d. $N(0, \sigma^2)$, $\delta \equiv d - 1$ and $\delta \in (-1, 0.5)$.Let $IMP(n)$ denote the impulse response function of this y_t process. The impulse response function measures the impact of a shock of size equal to 1 at time t on y_{t+n}. For a stationary process z_t, $I(n)$ equals the coefficients of the infinite-MA representation of the process. See Endnotes, no. 5, page 34, for the derivations. In the limit,

$$\lim_{n\to\infty} IMP(n) = 0 \qquad \text{if } \delta < 0, \text{ i.e., } d < 1 \tag{10}$$

$$= A(1) = \frac{v(1)}{\phi(1)} \qquad \text{if } \delta = 0, \text{ i.e., } d = 1 \tag{11}$$

$$= \infty \qquad \text{if } \delta > 0, \text{ i.e., } d > 1 \tag{12}$$

The problem at hand is the following. Whenever δ deviates from 0, $IMP(\infty)$ equals 0 or ∞. Since finding an estimate for δ that is different from zero is highly likely, an impulse response that is infinite or zero will also be highly likely. This theoretical weakness of ARFIMA is documented in Hauser, Pötscher and Reschenhofer (1999, pp. 250-252). They argued that ARFIMA modelling is inappropriate for the purpose of estimating persistence (defined as $IMP(\infty)$).[15]

Here we consider only the class of ARFIMA models. Given the above notation, the

[14] In other words, we are restricting the space of the fractional differencing parameter to $d \in (0.0, 1.5)$. In the case where $\delta = 0$, d equals 1 and the process y_t is modeled as an ARIMA$(p, 1, q)$, i.e., z_t is an ARMA(p, q). The restriction on the space of δ merits some explanation. The lower bound of δ (-1) ensures that Δy_t is invertible (see Odaki (1993)). Also, from Table 7, page 43, the autocorrelations are positive and decay hyperbolically. Therefore, restricting the lower bound of d to zero is coherent. A reasonable implication of the unit root test (Table 6, page 42) is to restrict the upper bound to 0.5 which ensures that Δy_t is stationary. Whenever $d \in (0.0, 0.5)$, y_t is said to be trend-stationary with long memory. Whenever $d \in (0.5, 1.5)$, Δy_t is stationary with intermediate memory for $d < 1$ and with long memory for $d > 1$.

[15] We adopt the answer to this criticism given by Koop, Ley, Osiewalski and Steel (1997, p. 154). Since $IMP(\infty)$ is of little relevance to the economic forecaster, they defined the following. If the frequency of

problem we are facing is a standard Bayesian one. The parameter space is partitioned into μ, σ^2 and $\omega^T \equiv (\delta, \Theta^T, \Phi^T)$, where $\Theta \equiv (\theta_1, ..., \theta_q)^T \in \mathbb{C}^q$ and $\Phi \equiv (\phi_1, ..., \phi_p)^T \in \mathbb{C}^p$. Let w denote the observed vector of data, with $w^T \equiv (\Delta y_1, ..., \Delta y_N)^T$, and let w^{*T} denote the predictions of the observed data, with $w^{*T} \equiv (\Delta y_{N+1}, ..., \Delta y_{N+n})^T$. The model is defined as,

$$(13) \qquad \begin{pmatrix} w \\ w^* \end{pmatrix} = \mu \begin{pmatrix} \iota_N \\ \iota_n \end{pmatrix} + \begin{pmatrix} \xi \\ \xi^* \end{pmatrix}$$

where

$$(14) \qquad \begin{pmatrix} \xi \\ \xi^* \end{pmatrix} \sim N(0_{N+n}, \sigma^2 \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix})$$

$$(15) \qquad \begin{pmatrix} \xi \\ \xi^* \end{pmatrix} \sim N(0, \sigma^2 V)$$

ι_N refers to an $N \times 1$ vector of ones. The elements of V are given by $v_{ij} = \sigma^{-2}\gamma(i-j)$ for $i, j = 1, ..., N+n$. $\gamma(s)$ denotes the autocovariance function given in Sowell (1992, p. 173, equation (8)). The sampling density distribution of w is given by,

$$(16) \qquad p(w|\omega, \mu, \sigma^2) = f^N_{Normal}(w|\mu\iota_T, \sigma^2 V_{11})$$

the data at hand is quarterly, and we refer to $IMP(4)$, $IMP(12)$ and $IMP(40)$ as the short-run, medium run and long-run impact of a shock respectively, then an economist is only interested in these quantities: $IMP(4)$, $IMP(12)$ and $IMP(40)$. Note that our definition (see Mikhail, Eberwein and Handa (2003)) of 'unemployment persistence' is consistent with their definition. If one accepts our definition of unemployment persistence, computing the suggested impulse responses will quantify persistence. Also, our definition makes no distinction between the intermediate and the long run impact, as classified by Koop, Ley, Osiewalski and Steel (1997, p. 154).

where f^N_{Normal} is the N-variate Normal density function[16]. The prior over the parameters is assumed to be as follows,

$$p(\omega, \mu, \sigma^2) = p(\omega)p(\mu)p(\sigma^{-2}) \propto \sigma^2 p(\omega) \tag{17}$$

where $\omega \in \Omega \equiv (-1, 0.5) \times \mathbb{C}^q \times \mathbb{C}^p$. $\mu \in R$ and $\sigma^{-2} \in R_+$. The improper prior on σ^{-2} leads to perfect robustness with respect to all $(N+n)$-variate elliptical densities with the same location and scale.[17] Given the sampling distribution and the prior, we integrate out μ and σ^{-2}, which yield the posterior density for ω.

$$p(\omega|Data) = K^{-1}|V_{11}|^{-\frac{1}{2}}(\iota_N^T V_{11}^{-1}\iota_N)^{-\frac{1}{2}} SSE^{-\frac{N-1}{2}} p(\omega) \tag{18}$$

where ι_N^T refers to the transpose of $N \times 1$ vector of ones,

$$K \equiv \int_\Omega |V_{11}|^{-\frac{1}{2}}(\iota_N^T V_{11}^{-1}\iota_T)^{-\frac{1}{2}} SSE^{-\frac{T-1}{2}} p(\omega) d\omega \tag{19}$$

$$SSE = (w - \widehat{\mu}\iota_N)^T V_{11}^{-1}(w - \widehat{\mu}\iota_N) \tag{20}$$

$$\widehat{\mu} = (\iota_N^T V_{11}^{-1}\iota_N)^{-1}\iota_N^T V_{11}^{-1} w \tag{21}$$

The posterior density is computed using Monte-Carlo simulations. The procedure is as follows. We draw a value for δ from a uniform distribution over the interval [-1.0 , 0.5], then we compute its antithetic replication by projecting the value through the mean.[18] We

[16] See Zellner (1987, p. 379, equation (B.1))
$p(w|\omega, \mu, \sigma^2) = \frac{|\sigma^2 V_{11}|^{-\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2}(w - \mu\iota_N)^T(\sigma^2 V_{11})^{-1}(w - \mu\iota_N)\right\}$.

[17] See Osiewalski and Steel (1993) for the proof.

[18] The antithetic replication is computed by projecting the draw through the mean of the uniform distribution [-1.0,0.5], i.e., -0.25. Therefore, the antithetic value equals $-0.25 - [draw - (-0.25)]$. Formally, the antithetic value $\delta^{-i} = E(\delta) - [\delta^i - E(\delta)] = 2E(\delta) - \delta^i$. See Dorfman (1997, p. 21) for more details.

also draw the values for p and q from a uniform distribution that is bounded to ensure the stationarity and invertibility of the process. This procedure efficiently enforces the ARMA stationarity part as outlined in Monahan (1984, p. 403, equation (1)). Then, we compute the likelihood and the variance-covariance matrix. Next we evaluate the log of the posterior for the parameters using the Sowell code.[19]

The algorithm used is importance sampling combined with antithetic replications. Whenever the posterior density is nonstandard - from which it is difficult or impossible to generate random draws - importance sampling allows the random draws ω to be generated from a substitute density $f(\omega)$. The empirical density is then adjusted to account for the differences between the substitute density and the actual posterior distribution $p(\omega|y,X)$ of ω. To increase the efficiency of the numerical approximation, the algorithm relies on antithetic replications. Generating random draws from the substitute density results in an empirical density that is not a random sample from the posterior distribution. Therefore, the simple averages cannot be used to estimate the posterior mean. Instead, one corrects the simple averages by computing weighted averages as follows,

$$\widehat{g}^{IS}(\omega) = \frac{\sum_{i=1}^{50000} g(\omega^i)p(\omega^i|y,X)/f(\omega^i)}{\sum_{i=1}^{50000} p(\omega^i|y,X)/f(\omega^i)} = \frac{\sum_{i=1}^{50000} g(\omega^i)s(\omega^i)}{\sum_{i=1}^{50000} s(\omega^i)} \tag{22}$$

where $s(\omega^i)$ refers to the importance weight for the ith observation in the empirical distribution, $g(\omega^i)$ denotes any quantity function of interest (e.g., the sample average) and the superscript IS denotes that the estimator is based on the importance sampling density.[20]

[19] We repeated this exercise 25,000 times. The simulation was carried out on i686 machine running LINUX 2.2.14-5.0. The FORTRAN 77 code is from Koop, Ley, Osiewalski and Steel (1997) with modifications to fit our problem. Conditional on the number of parameters in each model, the average time for simulating one model was 23 minutes.

[20] For regularity conditions ensuring the convergence of $g(\omega)$, see Dorfman (1997, p. 24). The general criteria for choosing an importance function are discussed in Bauwens, Lubrano and Richard (1999, pp.

Note that the combination of importance sampling[21] with antithetic replications increases the numerical efficiency for any symmetrical or near-symmetrical posterior distribution (Dorfman 1997, p. 25).

Our objective is to assess the relative importance of persistence in aggregate Canadian unemployment. Therefore, we focus on reporting and analyzing the results for the parameter δ and the impulse responses. As presented here, a caveat of Bayesian inference regarding the fractional differencing parameter is that the order of the ARFIMA is assumed to be fixed, i.e., known. Therefore, we consider a range of values for the orders p and q to cover model uncertainty.

5.1 Model Comparison, Sensitivity and Robustness

The posterior distribution provides a basis for "estimation of parameters conditional on the adequacy of the entertained model" and the predictive distribution enables "criticism of the entertained model in light of current data" (Box 1980, p. 383). The scope of Bayesian model comparison and model assessment is quite broad. In the literature on Bayesian model comparison, there are 1) the marginal likelihood approach, 2) the 'super-model' or 'sub-model' approach and 3) the criterion-based methods such as the L measure and the calibration distribution.[22] The second approach is efficient whenever the posterior means or modes are not far from zero. The last approach does not require proper prior distributions over the models. Here, we adopt the marginal likelihood approach. This approach (outlined later) is

77-82).

[21] We did not focus on other methods of simulation, such as the Gibbs sampling algorithm. The Gibbs sampling algorithm has been used for the analysis of univariate time series by Barnett, Kohn and Sheather (1996), Chib and Greenberg (1995), McCulloch and Tsay (1994), and for ARFIMA processes by Pai and Ravishanker (1996, 1998). For the predictive distributions see Appendix A, no. 6.

[22] See Chen, Shao and Ibrahim (2000) for an excellent exposition of all methods and the references therein.

essentially the same as the Bayes factor approach.

Let the joint density for potential data y and parameters ω be

$$p(y, \omega|M) = p(y|\omega, M)p(\omega|M) \tag{23}$$

where M indicates conditionality on the model specification. This model can also be factored as

$$p(y, \omega|M) = p(\omega|y, M)p(y|M) \tag{24}$$

where $p(y|M) = \int p(y|\omega, M)p(\omega|M)d\omega$ denotes the predictive distribution. With an actual data vector y_d,

$$p(y_d, \omega|M) = p(\omega|y_d, M)p(y_d|M) \tag{25}$$

where the first term on the right hand side of equation (25) refers to the posterior distribution of ω, given y_d, as

$$p(\omega|y_d, M) \propto p(y_d|\omega, M)p(\omega|M) \tag{26}$$

The second term on the right hand side of equation (25) refers to the predictive density associated with the particular data type y_d actually obtained.

The posterior distribution $p(\omega|y_d, M)$ allows all estimation inferences of interest to be made regarding ω. However, if y_d was not generated by the model M, it could be assessed by reference to the density $p(y_d|M)$ to the predictive reference distribution $p(y|M)$.[23]

Other methods for examining robustness are sensitivity analysis approaches via: 1) asymptotic approximation; 2) scale mixtures of normals; and 3) prior partitioning. The last method relies on working the problem 'backward'. Rather then choosing (fixing) the priors, one chooses a set of posteriors that produce a given conclusion, and determines which prior inputs are consistent with the desired results, given the observed data.

Provided that the set of models under consideration is exhaustive, mixing over the models is optimal for forecasting purposes.[24] Here, we investigate the set of ARFIMA models up to and including the orders of ARFIMA$(3, \delta, 3)$. This set is not exhaustive, but since most of economic time series - in an analysis that neglects long-range dependence - can be well approximated by low order ARMA models,[25] we stop at the orders $p = q = 3$. The ARMA part of the ARFIMA should be able to capture short-range dependency and then we can investigate the long-memory properties of the process based on the estimate of δ.

There are 16 models under consideration $M_1, M_2 ..., M_{16}$. For model M_i $(i = 1, ..., 16)$, the posterior distribution takes the form,

$$p(\omega_i|Data, M_i) \propto L(\omega_i|Data, M_i)p(\omega_i|M_i) \tag{27}$$

where $L(\omega_i|Data, M_i)$ is the likelihood function and $p(\omega_i|M_i)$ denotes the prior distribution.

[23]The success of the Bayesian predictive distribution as a model checking device is discussed at length by Geisser (1993) and Geisser and Eddy (1979). This paper does not address the prediction aspect of the proposed model but focuses only on the persistence issue.

[24]Using a squared error loss, mixing over the models is optimal for forecasting (see Min and Zellner (1993) for the proof).

[25]Since we do not consider all conceivable models for the problem at hand, model comparisons based on the posterior odds do not change if a new unspecified third model is introduced. Given the intuitive economic argument, we suggest that the Independence of Irrelevant Alternatives (IIA) property holds. See Poirier (1997, p. 150) for the definition and for an excellent exposition.

The marginal likelihood is given by,

$$p(Data|M_i) = \int L(\omega_i|Data, M_i)p(\omega_i|M_i)d\omega_i \tag{28}$$

To compare the models, one computes the marginal likelihoods and chooses the model that yields the largest marginal likelihood. Basically, the marginal likelihood approach is the same as the Bayes factor approach. Note that $p(Data|M_i)$ is the normalizing constant of the posterior distribution $p(\omega_i|Data, M_i)$. The posterior probability of model i, M_i, is given by,[26]

$$p(M_i|data) = \frac{p(M_i)K_i}{\sum_{j=1}^{m} p(M_j)K_j} \tag{29}$$

where $p(M_j)$ is the prior model probability of M_j and K_i is as defined in equation (19).

We consider the same models investigated by Koop, Ley, Osiewalski and Steel (1997), corresponding to all possible ARFIMA(p, δ, q) for $p, q \leq 3$.

"Given the appropriate tools, the most straightforward way of demonstrating a lack of dependence on the prior is to compute the summary measures of interest for a range of plausible prior densities" (Skene, Shaw and Lee 1986, p. 282). Applying this methodology, the prior probabilities for each model M_i are all taken as equal (i.e., $p(M_i) = 1/16$ for $i = 1, ..., 16$) to reflect ignorance, i.e., 'non-informative' prior.[27] We also adopt a second 'informative' prior. The reason for assuming an 'informative' prior is the following. For the models where the AR term is zero - i.e., ARFIMA$(0, \delta, q)$ - one should expect the parameter

[26] See Box (1980, p. 408, equation (*)), Carlin and Louis (1996, p. 47, equation (2.17)) and Chen, Shao and Ibrahim (2000, p. 237, equation (8.1.3)).

[27] On the quantification of ignorance, see the excellent exposition in Bauwens, Lubrano and Richard (1999, pp. 107-109). Briefly, the approach adopted here maximizes the entropy of the model density over the parameter space.

δ to capture any short-range dependency present in the data since there is no AR term to adequately reflect it. Therefore, inference based on δ would be misleading. This 'informative' prior downweights the prior weight of ARFIMA$(0, \delta, q)$ by assigning three times less prior mass to ARFIMA$(0, \delta, q)$. For example, the prior for ARFIMA$(0, \delta, 1)$ equals 0.5/16 and the prior for ARFIMA$(1, \delta, 0)$ equals 1.5/16.

Figure 1 illustrates the posterior of a simple mix of δ over the 16 ARFIMA models. The posterior distribution is highly non-linear and reflects important mass on the positive real line for δ. However, this bias towards mass over the positive real line is due to the presence of pure moving average models. In these models, the parameter δ reflects and captures both the short- and the long-range dependence of the series.

As expected and illustrated in Figures 2 and 3, ARFIMA models without autoregressive terms - such as ARFIMA$(0, \delta, 1)$ and ARFIMA$(0, \delta, 2)$ - pull the posterior distribution towards the positive side of the real line. In these cases, δ captures both the short-range and the long-range dependency of Canadian unemployment. Autoregressive components models tend to put higher mass on the negative real line. In other words, when the autoregressive component is present, $\widehat{d}$ is smaller than 1. Canadian aggregate unemployment is a long-memory process that exhibits the mean reversion property: Figure 3 shows that as the number of autoregressive parameters increases (i.e., from ARFIMA$(1, \delta, 0)$ to ARFIMA$(2, \delta, 0)$), more probability is given to the negative real line and specifically to the range $\delta \in (-1.0, -0.5)$.

Conditional on the entertained class of models and the assumptions made regarding the priors and the sample data, is aggregate Canadian unemployment trend stationary with long memory or is it first difference stationary with intermediate memory? The answer lies in Tables 12 and 13. Table 12 reports the posterior model probabilities under the assumptions of both priors: flat and informative. Conditional on each model, Table 13 reports the posterior mean, standard deviation and the mode of δ. The reason for reporting all these descriptive statistics is that the posterior is highly non-linear and non-symmetrical. Therefore, conditional on the loss function[28] used, one is faced with a different optimal Bayesian point estimate. Choosing the zero-one loss function produces the 'most likely' estimate point but a small estimation error is treated the same as a large one. Choosing the quadratic loss function protects against outliers and skewed tails. The advantage of using the quadratic loss function is that it uses all the information present in the posterior distribution to derive the mean. We report the descriptive statistics, and to conform with the ethos of Bayesian point estimation, we adopt the quadratic loss function as our approach.

Regardless of the prior used (ignorance or informative), Table 13 points to ARFIMA$(1, \delta, 0)$ as the model with the highest posterior probability. Conditional on the prior and the sample data, this model is the most likely to adequately fit the data. Based on the posterior model probabilities, the overall ranking is as follow: ARFIMA$(1, \delta, 0)$, ARFIMA$(3, \delta, 3)$ and finally, ARFIMA$(1, \delta, 1)$. Note that the posterior model probability is scattered across all models, which cautions against choosing just one model. More specifically, the standard deviation increases with the number of parameters leading to higher uncertainty in choosing only one model (with the exception of the boundary model ARFIMA$(3, \delta, 3)$).

The posterior odds[29], in favor of $(-0.5 < \delta < 0.0)$ against $(0.0 < \delta < 0.5)$, are 0.5192 to

[28]Common choices of loss functions are: 1) the quadratic loss $L(\widehat{\delta}, \delta) = (\widehat{\delta} - \delta)^2$, 2) the absolute loss $L(\widehat{\delta}, \delta) = |\widehat{\delta} - \delta|$ and 3) the zero-one loss $L(\widehat{\delta}, \delta) = c$ if $\widehat{\delta} \neq \delta$ and $L(\widehat{\delta}, \delta) = 0$ if $\widehat{\delta} = \delta$. See Dorfman (1997, p. 10) for the derivations. Choosing the quadratic (absolute, zero-one) loss results in the mean (median, mode) as the Bayesian optimal point estimate.

[29]Here, we adopt the symmetric '0-K_i' loss function, as defined in Bauwens, Lubrano and Richard (1999, p. 29). The probability of errors of type I and II are equal. See also Zellner (1987, p. 292) where "under a symmetric loss structure, a comparison of the posterior probabilities will provide a basis for choosing between

0.4808. This evidence supports the belief that Δy_t is stationary with intermediate memory. Quantitatively, the ARFIMA$(1, \delta, 0)$ model estimates a small negative value for δ, whereas the overall model estimates a small positive value for the same parameter. Therefore, we conclude that, conditional on the entertained class of models, the prior assumptions and the sample data, the first difference of the log of Canadian unemployment is stationary with intermediate memory. Among the class of low order ARFIMA models, an ARFIMA$(1, \delta, 0)$ model is the most likely one to be observed, with $\widehat{\delta} = -0.034$, i.e., $\widehat{d} = 0.966$.

5.2 The impact of shocks and its duration

For the chosen model, Figure 4 illustrates the impulse responses for $n = 4, 12$ and 40. As expected from previous results, the posterior standard deviation increases for longer horizons. It is highly skewed and exhibits fat tails.

The Bayesian-based impulse response function $IMP(n)$ measures the impact of a shock of size equal to 1 at time t on y_{t+n}. Table 14 and Figure 4 show that unemployment persistence is present. The effect of the shock persists for at least 12 quarters. For the ARFIMA$(1, \delta, 0)$ model, the variance - and the uncertainty of drawing conclusions - grows to $n = 40$. This finding reinforces the evidence of short- and intermediate-run persistence in total unemployment. The influence of model averaging is apparent in the impulse responses of the overall model. Higher variance and lower persistence occur relative to the ARFIMA$(1, \delta, 0)$. For the longer horizon $n = 40$, the shock is responsible for a large variance. Table 14 and Figure 4 quantify and illustrate the increase in the variance of the effect of the shock at longer

H_0 and H_1."

horizons.

Hence, unemployment persistence holds in the short- and intermediate-run. However, unemployment persistence over longer horizons is uncertain due to the large variance associated with $n = 40$.

6 Conclusions

This paper tested for unemployment persistence in Canadian sectoral unemployment using the modified rescaled range statistic tests. The modified rescaled range test statistic provided evidence of persistence. We conclude that the fluctuations in the sectoral Canadian unemployment series are characterized by persistence.

Conditional on our class of models [low order ARFIMA], the prior assumptions and the sample data, the first difference of the log of Canadian unemployment is stationary with intermediate memory. Unemployment persistence holds in the short and intermediate run. However, this is uncertain over longer horizons. In summary, shocks to sectoral and aggregate unemployment have fairly resolute effects.

References

[1] Baillie, Richard T., "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics* 73 (1996), 5-59.

[2] Baillie, Richard T., and Ching-Fan Chung, "Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models," *Empirical Economics* 18 (1993), 791-806.

[3] Barnett, G., Kohn, R., and S. Sheather, "Bayesian Estimation of an Autoregressive Model using Markov Chain Monte Carlo," *Journal of Econometrics* 74 (1996), 237-254.

[4] Box, George E. P., "Sampling and Bayes: Inference in Scientific Modelling and Robustness (with Discussion)," *Journal of the Royal Statistical Society*, Series A, 143 (1980), 383-430.

[5] Bauwens, Luc, Michel Lubrano, and Jean-François Richard, *Bayesian Inference in Dynamic Econometric Models,* Advanced Texts in Econometrics, (Ed.) Granger, C.W.J. and Mizon, G.E. (New York: Oxford University Press, 1999).

[6] Benassi, Corrado, Alessandra Chirco, and Caterina Colombo, *The New Keynesian Economics* (Cambridge, U.S.A.: Basil Blackwell Inc., 1994).

[7] Carey, Kevin "High Unemployment in the OECD: Does the Diagnosis of the 1980s Apply to the 1990s?" *Unpublished Manuscript,* Department of Economics, American University, Washington D.C., 1997.

[8] Carlin, P. Bradley, and Thomas A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis* (New York: Chapman and Hall USA, 1996).

[9] Chen, Ming-Hui, and Qi-Man Shao, and Joseph G. and Ibrahim, *Monte Carlo Methods in Bayesian Computation* (New York: Springer Series in Statistics, Springer, 2000).

[10] Chib, S., and E. Greenberg "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying parameter Models," *Journal of Econometrics* 68 (1995), 339-360.

[11] Cochrane, John H., "How Big is the Random Walk in GNP?" *Journal of Political Economy* 96 (1988), 893-920.

[12] Coe, David T. "Insider-Outsider Influences on Industry Wages; Evidence From Fourteen Industrialized Countries," in Wolfgang, Franz (Eds.) *Hysteresis Effects In Economic Models* (Physica-Verlag Heidelberg, 1990).

[13] Cogley, Timothy and Nason, James M., "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research," *Journal of Economic Dynamics and Control* 19 (1995), 253-278.

[14] Cozier, B. and Wilkinson, G., "Some Evidence on Hysteresis and the Costs of Disinflation in Canada?" *Technical Report No. 55*, (Ottawa: Bank of Canada, 1991).

[15] Dorfman, Jeffrey H., *Bayesian Economics Through Numerical Methods*, (New York: Springer-Verlag, 1997).

[16] Doukhan, Paul. George Oppenheim and Murad S. Taqqu, *Theory and Applications of Long-Range Dependence* (Boston: Birkhäuser, 2003).

[17] Durbin, James, "Efficient Estimation of Parameters in Moving-Average Models," *Biometrika* 46 (1959), 306-316.

[18] Enders, Walter, *Applied Econometric Time Series.* (New York: John Wiley & Sons Inc., 1995).

[19] Fortin, Pierre, "How 'Naural' is Canada's High Unemployment Rate?" *European Economic Review* 33 (1989), 89-110.

[20] Fortin, Pierre, "The Phillips Curve, Macroeconomic Policy and the Welfare of Canadians," *Canadian Journal of Economics* 24 (1991), 774-803.

[21] Fox, Robert, and Murad S. Taqqu, "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *Annals of Statistics* 14 (1986), 517-532.

[22] Galbraith, John William and Victoria Zinde-Walsh, "A Simple Non-iterative Estimator for Moving Average Models," *Biometrika* 81 (1994), 143-155.

[23] Geisser, Seymour, *Predictive Inference: An Introduction* (London: Chapman & Hall, 1993).

[24] Geisser, Seymour, and William F. Eddy, "A Predictive Approach to Model Selection," *Journal of the American Statistical Association* 74 (1979), 153-160.

[25] Geweke, John, and Susan Porter-Hudak, "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis* 4 (1983), 221-238.

[26] Gil-Alana, Luis A., "The Persistence of Unemployment in the USA and Europe in Terms of Fractionally ARIMA Models," *Applied Economics* 33 (2001), 1263-1269.

[27] Granger, Clive W. J., "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics* 14 (1980), 227-238.

[28] Hall, Robert E., "Labor-Market Frictions and Unemployment Fluctuations," *National Bureau of Economic Research,* Working Paper Series no. 6501, 1998.

[29] Harvey, A. C., and A. Jaeger, "Detrending, Stylized Facts, and the Business Cycle," *Journal of Applied Econometrics* 8 (1993), 231-247.

[30] Hauser, Michael A., "Semiparametric and Nonparametric Testing for Long Memory: A Monte Carlo Study," *Empirical Economics* 22 (1997), 247-271.

[31] Hauser, Michael A., and Benedikt M. Pötscher, and Erhard Reschenhofer, "Measuring Persistence in Aggregate Output: ARMA Models, Fractionally Integrated ARMA Models and Non-Parametric Procedures," *Empirical Economics* 24 (1999), 243-269.

[32] Heckman, James J., and Georges J. Borjas, "Does Unemployment Cause Future Unemployment ?" *Economica* 47 (1980), 247-283.

[33] Hurst, Harold Edwin, "Long-Term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers* 116 (1951), 770-799.

[34] Janacek, Gareth J., "Determining the Degree of Differencing for Time Series via the Long Spectrum," *Journal of Time Series Analysis* 3 (1982), 177-183.

[35] Jones, Stephen R.G., *The Persistence of Unemployment.* (Montreal: McGill Queen's University Press, 1995).

[36] King, Robert G., and Sergio T. Rebelo, "Low Frequency Filtering and Real Business Cycles," *Journal of Economic Dynamics and Control* 17 (1993), 207-231.

[37] Koop, Gary, and Eduardo Ley, and Jacek Osiewalski, and Mark F.J. Steel, "Bayesian Analysis of Long Memory and Persistence using ARFIMA Models," *Journal of Econometrics* 76 (1997), 149-169.

[38] Koustas, Zisimos and William Veloce, "Unemployment Hysteresis in Canada: an Approach Based on Long-Memory Time Series Models," *Applied Economics* 28:7 (1996), 823-831.

[39] Li, W. K. and Ian A. McLeod, "Fractional Time Series Modelling," *Biometrika* 73 (1986), 217-221.

[40] Lo, Andrew W., "Long-Term Memory in Stock Market Prices," *Econometrica* 59 (1991), 1279-1313.

[41] Mandelbrot, Benoit B., "Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis," *Annals of Economic and Social Measurement* 1 (1972), 259-290.

[42] McCallum, John, "Les Taux de Chômage Canadien et Américan dans les années 1980s: un test de trois hypothéses," *Revue d'Analyse Économique* 64 (1988), 494-508.

[43] McCulloch, R. E. and R. S. Tsay, "Bayesian Analysis of Autoregressive Time Series via the Gibbs Sampler," *Journal of Time Series Analysis* 15 (1994), 235-250.

[44] McLeod, Angus Ian and Keith William Hipel, "Preservation of the Rescaled Adjusted Range: A Reassessment of the Hurst Phenomena," *Water Resources Research* 14 (1978), 491-508.

[45] Mikhail, O., Curtis J. Eberwein and Jagdish Handa, "Hysteresis and Persistence in Unemployment: a Definition," *Working Paper*, Department of Economics, University of Central Florida, 2003.

[46] Min, Chung-Ki and Arnold Zellner, "Bayesian and non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Econometrics* 56 (1993), 89-118.

[47] Monahan, John F., "A Note on Enforcing Stationarity in Autoregressive-Moving Average Models," *Biometrika* 71 (1984), 403-404.

[48] Newbold, Paul. and Christos Agiakloglou, "Bias in the Sample Autocorrelations of Fractional Noise," *Biometrika* 80 (1993), 698-702.

[49] Newey, Whitney K. and Kenneth D. West, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 (1987), 703-708.

[50] Nickell, Stephen, "Unemployment and Labour Market Rigidities: Europe versus North America," *Journal of Economic Perspectives* 11 (1997), 55-74.

[51] Nott, Loretta, "Hysteresis in the Canadian labour market: Evidence from the 1990s," *Research Report* no. 9605, Department of Economics, University of Western Ontario, 1996.

[52] Odaki, Mitsuhiro, "On the Invertibility of Fractionally Differenced ARIMA Processes," *Biometrika* 80 (1993), 703-709.

[53] Osiewalski, Jacek and Mark F.J. Steel, "Robust Bayesian Inference in Elliptical Regression Models," *Journal of Econometrics* 57 (1993), 345-363.

[54] Pai, J. S. and Nalini Ravishanker, "Bayesian Modeling of ARFIMA processes by Markov Chain Monte Carlo Methods," *Journal of Forecasting* 15 (1996), 63-82.

[55] Pai, J. S. and Nalini Ravishanker, "Bayesian Analysis of Autoregressive Fractionally Integrated Moving-Average Processes," *Journal of Time Series Analysis* 19 (1998), 99-112.

[56] Poirier, Dale J., "Comparing and Choosing Between Two Models with a Third Model in the Background," *Journal of Econometrics* 78 (1997), 139-151.

[57] Poloz, S. S. and Wilkinson, G., "Is Hysteresis a Characteristic of the Canadian Labour Market? A Tale of Two Studies," *Working Paper 92-3*, Ottawa: Bank of Canada,1992.

[58] Rappoport, Peter and Lucrezia Reichlin, "Segmented trends and Non-Stationary Time Series," *Economic Journal* 99 (Supplement 1989), 168-177.

[59] Riddell, Graig W., "Canadian labour Market Performance in International Perspective," *Presidential Address to The Canadian Economics Association*, Toronto, May 29, 1999.

[60] Rosenblatt, Murray, "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences* 42 (1956), 43-47.

[61] Skene, A. M., and J. E. H. Shaw, and T. D. Lee, "Bayesian Modelling and Sensitivity Analysis," *The Statistician* 35 (1986), 281-288.

[62] Sowell, Fallaw "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Models," *Journal of Econometrics* 53 (1992), 165-188.

[63] Stadler, George W., "Real Business Cycles," *Journal of Economic Literature* XXXII (1994), 1750-1783.

[64] Wilkinson, Gordon, "Micro-Approach to Hysteresis in Unemployment," *Working Papers* no. 97-12, Ottawa: Bank of Canada, 1997.

[65] Winter-Ebmer, Rudolf, "Some Micro Evidence on Unemployment Persistence," *Oxford Bulletin of Economic Statistics* 53 (1991), 27-43.

[66] Zellner, Arnold, *An Introduction to Bayesian Inference in Econometrics,* Original Edition 1971 (New York: John Wiley & Sons Inc., 1987).

7 Endnotes (not for publication)

1. The Ljung-Box statistic for the sample autocorrelation is computed as follows, $Q_{LB} \equiv T(T+2)\sum_{k=1}^{M}\left[\frac{r_k^2}{T-k}\right] \sim^{asy} \chi^2_{(M)}$. This test statistic is used to test H_0 : independently distributed error terms. In small sample, the Q_{LB} suffers from lack of power. This is the principal reason for not undertaking annual data persistence analysis.

2. Following the definition[30] of McLeod and Hipel (1978), a process is considered to be a long memory if the quantity

$$\lim_{n\to\infty}\sum_{j=-n}^{n} \mid \rho_j \mid \tag{30}$$

is nonfinite, where ρ_j denotes the autocorrelation at lag j. Note that this is equivalent to an unbounded spectral density [frequency domain analysis] of the process at low frequencies. Considering only linear univariate models, the process y_t is said to be integrated of order d, or I(d), if

$$(1-L)^d(y_t-\mu)=\varepsilon_t \tag{31}$$

where L is the lag operator, $E(\varepsilon_t)=0$, $E(\varepsilon_t^2)=\sigma^2$, and $E(\varepsilon_t\varepsilon_s)=0$ for $s\neq t$ and where the fractional parameter d is possibly a noninteger.[31] The process is weakly stationary for $d<0.5$ and invertible for $d>-0.5$. The infinite-order autoregressive representation of the process is given by,

$$y_t=\sum_{k=0}^{\infty}\pi_k y_{t-k}+\varepsilon_t \tag{32}$$

where the weights π_k are obtained from the binomial expansion,

$$(1-L)^d=\sum_{k=0}^{\infty}(-1)^k\binom{d}{k}L^k \tag{33}$$

$$\binom{d}{k}\equiv\frac{d(d-1)(d-2)(d-3)...(d-k+1)}{k!} \tag{34}$$

$$(1-L)^d=\{1-dL+\frac{d(d-1)L^2}{2!}+\frac{d(d-1)(d-2)L^3}{3!}+...\} \tag{35}$$

[30] For other definitions of long memory, see Resnick (1987).

[31] An autoregressive integrated moving average process ARIMA (p,d,q) process is defined as a process that requires d^{th} differences to produce a stationary ARMA (p,q) process. 'd' denotes an integer. 'p' denotes the number of autoregressive lags and 'q' refers to the number of moving average lags. Formally, an ARIMA (p,d,q) process is written as $(1-\phi_1 L-...-\phi_p L^p)(1-L)^d Y_t=c+(1+\theta_1 L+...+\theta_q L^q)\varepsilon_t$.

Therefore,[32]

$$\pi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} = \pi_{k-1}\frac{k-1-d}{k} = \prod_{0\le j\le k} \frac{j-1-d}{j} \approx^{asy} \frac{k^{-d-1}}{\Gamma(-d)} \tag{36}$$

where $\Gamma(.)$ is the gamma function.[33] $\approx^{asy}$denotes the (asymptotic) limit. Similarly, the infinite moving average representation of the process can be expressed as,

$$y_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \tag{37}$$

where,[34]

$$\psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} = \psi_{k-1}\frac{k-1+d}{k} = \prod_{0\le j\le k} \frac{j-1+d}{j} \approx^{asy} \frac{k^{d-1}}{\Gamma(d)} \tag{38}$$

Note that the cumulative impulse response to a unit innovation is given by $\psi(1) = \sum_{i=0}^{\infty} \psi_j$. Equation (38) shows that the impulse response coefficient ψ_k decays at a slower rate than the geometric decay of ARMA class.[35] For this reason, Granger and Joyeux (1980) proposed the fractionally integrated process as an approach to modeling long memories in time series. The autocorrelations of a fractional white noise process follow,

$$\rho_k = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} = \prod_{0\le j\le k} \frac{j-1+d}{j-d} \approx^{asy} \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} = \mathrm{C}\ k^{2d-1} \tag{39}$$

where C denotes a constant term. Given the definition of long memory by McLeod and Hipel (1978) (equation (30)), fractionally integrated processes are long-memory processes. The autocorrelation coefficients have the same sign as d. When $d < 0$, the process is called 'anti-persistence' or 'short memory'. When $d > 0$, the process possesses a long-memory. Note that both cases imply long-range dependence. If $-0.5 < d < 0.5$, then ε_t is a stationary and ergodic process with bounded and positively valued spectrum at all frequencies. For $0 < d < 0.5$, the process is a long-memory process satisfying equation (30), i.e., the autocorrelations are not summable. The autocorrelations are all positive and decay at a hyperbolic rate. For $-0.5 < d < 0$, the process autocorrelations sum to a constant. The process is said to be a 'short memory' process and all its autocorrelations (excluding lag zero) are negative and decay hyperbolically to zero.

3. Granger (1980) examined the time series behavior of a contemporaneously aggregated

---

[32] For the derivation, see Baillie (1996, p. 18).

[33] The gamma function is defined as, $\Gamma(q) = \int_0^{\infty} u^{q-1}e^{-u}du$ for $0 < q < \infty$ [Zellner (1987, p. 364, equation A.6)]. Here, one uses the following property of the gamma function: $\Gamma(q+1) = q\Gamma(q)$ for $q > 0$.

[34] For derivation see Baillie (1996, p. 18) and Hamilton (1994, pp. 448-452).

[35] For example, compare ψ_k for the cases where $d = 0$ versus $d \neq 0$.

panel data. Formally, let

$$z_t = \sum_{i=1}^{N} y_{it} \tag{40}$$

which is the aggregate of N component and independent processes y_{it}, such that for $i = 1, ..., N$,

$$y_{it} = \alpha_i y_{it-1} + u_{it} \tag{41}$$

Hence, each individual process is an AR(1) with the autoregressive coefficients α_i to be drawn from a Beta(0,1) distribution,[36]

$$dF(\alpha) = \frac{2}{B(p,q)} \alpha^{2p-1}(1-\alpha^2)^{q-1} \qquad 0 \leq \alpha \leq 1, \quad p > 0, \quad q > 0 \tag{42}$$

Interestingly, Granger (1980) showed that in the limit for large N, z_t is an integrated process with $I\left(1 - \frac{q}{2}\right)$. In other words, z_t is a fractional process.

4. The Augmented Dickey-Fuller test the null hypothesis of unit root as follows, $\Delta y_t = \alpha_1 y_{t-1} + \alpha_2 Trend + \sum_j \beta_j \Delta y_{t-j} + e_t$ for $j = 1, 2, ...p$. where e_t is an independent, stationary process, and p is the lag length chosen for the dependent variable. The null hypothesis of a unit root is equivalent to testing $\alpha_1 = 0$. The test statistic is then compared to MacKinnon (1990) critical values. Other unit root tests can be found in Hamilton (1994, Chapter 17, 475-543).

5. Formally, the coefficients of the infinite-MA are $A(L) = (1-L)^{-\delta}\phi^{-1}(L)v(L)$. The n^{th} order partial sum of these coefficients is the cumulative response for z_t. They also represent the impulse responses $I(n)$ for the level of y_t. Formally, $I(n)$ could be represented by the n^{th} coefficient of $A^*(L) \equiv (1-L)^{-1}A(L) = (1-L)^{-d}\phi^{-1}(L)v(L)$. The coefficients of $\phi^{-1}(L)v(L)$ are defined as,

$$J(i) = 0 \qquad i + 1 - j \leq 0 \tag{43}$$

$$J(i) = \sum_{j=0}^{q} \theta_j f_{i+1-j} \tag{44}$$

$$f_1 \equiv 1 \text{ and } f_h \equiv -(\phi_1 f_{h-1} + ... + \phi_p f_{h-p}) \text{ for } h \geq 2. \tag{45}$$

Therefore, the coefficients of $A^*(L) \equiv (1-L)^{-d}\phi^{-1}(L)v(L)$ are computed as follows,

$$I(n) = \sum_{i=0}^{n} c_i(-d) J(n-i) \tag{46}$$

[36] The standardized Beta probabilty distribution function [Zellner (1987, p. 373)] is given by, $p(z|a,b) = \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1}$, for $0 \leq z \leq 1$ and $B(a,b) \equiv \int_0^1 z^{a-1}(1-z)^{b-1}dz$.

where $c_0(.) \equiv 1$ and $c_j(a) \equiv \prod_{k=1}^{j}(\frac{k-1-a}{k})$. When $\delta = 0$, $d = 1$ and $c_i = 1$ for $i \geq 0$. In the limiting case, where $\delta = -1$, $d = 0$, so that $c_o(0) = 1$ and $c_i(0) = \Pi_{k=1}^{i}(1-\frac{1}{k}) = 0.0$, since $c_1(0) = 0.0$. In the latter case, the impulse responses coefficients $I(n)$ equal $J(n)$, i.e., they collapse to the same coefficients as an ARMA(p, q) process. We now examine the behavior of the impulse responses under different fractional parameter specifications.

6. The predictive distributions are based on $p(w^*|Data)$. Note that $y_{N+n} = y_N + \iota_n^T w^* = y_N + n\mu + \iota_n^T \xi^*$. The posterior predictive density is given by,

$$p(y_{N+n}|\omega, Data) = \int_\Omega f(y_{N+n}|\omega)p(\omega|Data)d\omega \tag{47}$$

$$= f_s^1(y_{N+n}|N-1, y_N + n\mu + \iota_n^T V_{21}V_{11}^{-1}(w - \widehat{\mu}\iota_N), \tag{48}$$

$$\frac{N-1}{SSE}\left[\iota_n^T V_{22.1}\iota_n + \frac{(n - \iota_n^T V_{21}V_{11}^{-1}\iota_T)^2}{\iota_N^T V_{11}^{-1}\iota_N}\right]^{-1}) \tag{49}$$

where $V_{22.1} = V_{22} - V_{21}V_{11}^{-1}V_{12}$, and $f_s^k(.|r, b, A)$ is the k-variate Student t density with r degrees of freedom, location vector b and precision matrix A. Formally,[37]

$$p(y_{N+n}|\omega, Data) = \frac{(N-1)^{\frac{T-1}{2}}\Gamma(\frac{N+n-1}{2})}{\pi^{\frac{n}{2}}\Gamma(\frac{N-1}{2})} \left|\frac{N-1}{SSE}\left[\iota_n^T V_{22.1}\iota_n + \frac{(n - \iota_n^T V_{21}V_{11}^{-1}\iota_N)^2}{\iota_N^T V_{11}^{-1}\iota_N}\right]^{-1}\right|^{\frac{1}{2}} \left[N - 1 + \begin{array}{c}\left(y_{N+n} - y_N - n\mu - \iota_n^T V_{21}V_{11}^{-1}(w - \widehat{\mu}\iota_N)\right)^T \\ \left[\frac{N-1}{SSE}\left[\iota_n^T V_{22.1}\iota_n + \frac{(n-\iota_n^T V_{21}V_{11}^{-1}\iota_N)^2}{\iota_N^T V_{11}^{-1}\iota_N}\right]^{-1}\right] \\ \left(y_{N+n} - y_N - n\mu - \iota_n^T V_{21}V_{11}^{-1}(w - \widehat{\mu}\iota_N)\right)\end{array}\right] \tag{50}$$

The posterior density of the parameter μ is given by,

$$p(\mu|\omega, Data) = f_s^1(\mu|N-1, \widehat{\mu}, \frac{T-1}{SSE}\iota_N^T V_{11}^{-1}\iota_N) \tag{51}$$

Note that the last two densities are conditional on ω, therefore one integrates it out through a numerical procedure.

- Granger, Clive W. J., and Roselyne Joyeux (1980) 'An Introduction to Long-Memory Time Series Models and Fractional Differencing,' *Journal of Time Series Analysis* 1, 15-29.
- Hamilton, James D. (1994) *Time Series Analysis* Princeton University Press.
- MacKinnon, James G. (1990) 'Critical Values for Cointegration,' *Working Paper*, University of California San Diego, Department of Economics, 1990-04.
- Resnick, Sidney I. (1987) *Extreme Values, Regular Variation and Point Processes* Springer-Verlag, New York.

[37] See Zellner (1987, p. 383, equation (B.20)).

APPENDIX: Tables and Figures

Table 1

CANSIM SOURCE	
MONTHLY DATA FROM 1976:1 To 1998:12	
TOTAL UNEMPLOYMENT	D980712
UNEMPLOYMENT - GOODS	D968135
UNEMPLOYMENT - MANUFACTURING	D968140
UNEMPLOYMENT - SERVICES	D968141

Label	: D980712 (UPDATED to 2000)
Title	: CDA LF CHARACTERISTICS MONTHLY SA / UNEMPLOYMENT AGE 15+ SA CDA
Subtitle	: CANADA, LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 1976, SEASONALLY ADJUSTED. INCLUDES LF CHARACTERISTICS BY AGE & SEX; LABOUR FORCE, UNEMPLOYMENT & UNEMPLOYMENT RATE BY INDUSTRY; EMPLOYMENT BY INDUSTRY, OCCUPATION & CLASS OF WORKER; HOURS OF WORK BY INDUSTRY.
Factor	: THOUSAND
Unit	: PERSONS
Source	: SDDS 3701 STC (71-001)
Update	: 11 April, 2000
Period	: January 1976 - March 2000
Frequency	: monthly

Label	: D968135 (UPDATED to 2000)
Title	: CDA LF CHARACTERISTICS MONTHLY SA / UNEMPLOYMENT GOODS-PRODUCING SECTOR SA CDA
Subtitle	: CANADA, LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 1976, SEASONALLY ADJUSTED. INCLUDES LF CHARACTERISTICS BY AGE & SEX; LABOUR FORCE, UNEMPLOYMENT & UNEMPLOYMENT RATE BY INDUSTRY; EMPLOYMENT BY INDUSTRY, OCCUPATION & CLASS OF WORKER; HOURS OF WORK BY INDUSTRY.
Factor	: THOUSAND
Unit	: PERSONS
Source	: SDDS 3701 STC (71-001)
Update	: 11 April, 2000
Period	: January 1987 - March 2000
Frequency	: monthly

Label	: D968140 (UPDATED to 2000)
Title	: CDA LF CHARACTERISTICS MONTHLY SA / UNEMPLOYMENT MANUFACTURING SA CDA
Subtitle	: CANADA, LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 1976, SEASONALLY ADJUSTED. INCLUDES LF CHARACTERISTICS BY AGE & SEX; LABOUR FORCE, UNEMPLOYMENT & UNEMPLOYMENT RATE BY INDUSTRY; EMPLOYMENT BY INDUSTRY, OCCUPATION & CLASS OF WORKER; HOURS OF WORK BY INDUSTRY.
Factor	: THOUSAND
Unit	: PERSONS
Source	: SDDS 3701 STC (71-001)
Update	: 11 April, 2000
Period	: January 1987 - March 2000
Frequency	: monthly

Label	: D968141 (UPDATED to 2000)
Title	: CDA LF CHARACTERISTICS MONTHLY SA / UNEMPLOYMENT SERVICES-PRODUCING SECTOR SA CDA
Subtitle	: CANADA, LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 1976, SEASONALLY ADJUSTED. INCLUDES LF CHARACTERISTICS BY AGE & SEX; LABOUR FORCE, UNEMPLOYMENT & UNEMPLOYMENT RATE BY INDUSTRY; EMPLOYMENT BY INDUSTRY, OCCUPATION & CLASS OF WORKER; HOURS OF WORK BY INDUSTRY.
Factor	: THOUSAND
Unit	: PERSONS
Source	: SDDS 3701 STC (71-001)
Update	: 11 April, 2000
Period	: January 1987 - March 2000
Frequency	: monthly

(M) denotes MONTHLY
(Q) denotes QUARTERLY
(A) denotes ANNUAL

Table 2

DESCRIPTIVE STATISTICS FOR CANADIAN UNEMPLOYMENT by INDUSTRY - HP-FILTERED

Series	Obs	MEAN	St-Dev	MIN	MAX
	ANNUAL				
Total UE (A)	23	0.0000	0.1487	-0.2050	0.2927
UE GOODS (A)	23	0.0000	0.1609	-0.1984	0.3719
UE MANUF. (A)	23	0.0000	0.1680	-0.2097	0.4673
UE SERVICE (A)	23	0.0000	0.1317	-0.2170	0.1985
	QUARTERLY				
Total UE (Q)	92	0.0000	0.0934	-0.2085	0.2399
UE GOODS (Q)	92	0.0000	0.1110	-0.2113	0.3813
UE MANUF. (Q)	92	0.0000	0.1235	-0.2341	0.4230
UE SERVICE (Q)	92	0.0000	0.0792	-0.1869	0.1992
	MONTHLY				
Total UE (M)	276	0.0000	0.0364	-0.1624	0.1243
UE GOODS (M)	276	0.0000	0.0585	-0.2377	0.2320
UE MANUF. (M)	276	0.0000	0.0751	-0.2306	0.2299
UE SERVICE (M)	276	0.0000	0.0353	-0.1297	0.1042

Table 3

CORRELATION MATRIX - HP FILTERED DATA

	Total UE (M)	UE GOODS (M)	UE MANUF. (M)	UE SERVICE (M)
Total UE (M)	1.000000			
UE GOODS (M)	0.836181	1.000000		
UE MANUF. (M)	0.717364	0.877487	1.000000	
UE SERVICE (M)	0.853827	0.619120	0.540794	1.000000

Table 4

AUTOCORRELATION of CANADIAN UNEMPLOYMENT
HP - FILTERED DATA

	K=1	K=2	K=3	K=4	K=5	K=6
Total UE (M)	0.7552	0.5894	0.4882	0.3193	0.1773	0.0696
UE GOODS (M)	0.7817	0.6020	0.4555	0.3154	0.1683	0.0327
UE MANUF. (M)	0.7319	0.5725	0.4449	0.3255	0.1811	0.0835
UE SERVICE (M)	0.6274	0.4275	0.3239	0.2245	0.1527	0.0813
Total UE (Q)	0.9105	0.7436	0.5469	0.3421	0.1691	0.0323
UE GOODS (Q)	0.8574	0.6195	0.3595	0.1324	0.0077	-0.0579
UE MANUF. (Q)	0.8118	0.5464	0.2490	-0.0181	-0.1383	-0.1455
UE SERVICE (Q)	0.8874	0.7170	0.5105	0.2961	0.1285	-0.0161
Total UE (A)	0.6342	0.1411	-0.2573	-0.4802	-0.5080	-0.4618
UE GOODS (A)	0.5600	0.1308	-0.2492	-0.4255	-0.4267	-0.4039
UE MANUF. (A)	0.4857	0.1471	-0.2090	-0.4413	-0.3700	-0.3801
UE SERVICE (A)	0.6225	0.1668	-0.1508	-0.3129	-0.3636	-0.3483

Table 5

CROSS CORRELATION between CANADIAN UNEMP by INDUSTRY
AT DIFFERENT LAGS - HP FILTERED MONTHLY

	T+3	T+2	T+1	T	T-1	T-2	T-3
TOTAL UNEMPLOYMET							
UE GOODS	0.3327	0.4657	0.6424	0.8362	0.7123	0.6067	0.5512
UE MANUF	0.3083	0.4166	0.5676	0.7174	0.6234	0.5564	0.5392
UE SERVICES	0.3684	0.4631	0.6042	0.8538	0.6665	0.5117	0.4446
UNEMPLOYMENT GOODS							
	T+3	T+2	T+1	T	T-1	T-2	T-3
UE MANUF	0.3989	0.5347	0.6994	0.8775	0.7085	0.5728	0.4748
UE SERVICES	0.4505	0.4972	0.5496	0.6191	0.5312	0.399	0.3287
UNEMPLOYMENT SERVICES							
	T+3	T+2	T+1	T	T-1	T-2	T-3
UE MANUF	0.4516	0.4743	0.4976	0.5408	0.5019	0.3901	0.3524

Table 6: Augmented Dickey-Fuller (ADF) Tests

	ADF Test Statistic		**MacKinnon (1990) Critical Values**
Total Unemployment	-2.040	1 percent	-3.996
Manufacturing Unemployment	-2.148	5 percent	-3.428
Services Unemployment	-1.821	10 percent	-3.137

Table 7 Correlogram of UE_TOTAL

Sample: 1976:01 1998:12
Included observations: 276

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.986	0.986	271.36	0.000
		2	0.971	-0.043	535.59	0.000
		3	0.955	-0.075	791.77	0.000
		4	0.936	-0.096	1038.7	0.000
		5	0.915	-0.066	1275.5	0.000
		6	0.892	-0.056	1501.6	0.000
		7	0.870	0.012	1717.3	0.000
		8	0.845	-0.067	1921.9	0.000
		9	0.821	-0.023	2115.4	0.000
		10	0.795	-0.034	2297.6	0.000
		11	0.768	-0.047	2468.4	0.000
		12	0.742	0.035	2628.4	0.000
		13	0.718	0.061	2778.7	0.000
		14	0.693	-0.041	2919.4	0.000
		15	0.669	0.012	3050.8	0.000
		16	0.645	-0.025	3173.5	0.000
		17	0.621	-0.019	3287.5	0.000
		18	0.596	-0.046	3393.1	0.000
		19	0.572	0.030	3490.8	0.000
		20	0.549	-0.008	3581.0	0.000
		21	0.526	-0.011	3664.1	0.000
		22	0.502	-0.034	3740.3	0.000
		23	0.478	-0.053	3809.7	0.000
		24	0.454	-0.016	3872.4	0.000
		25	0.432	0.087	3929.6	0.000
		26	0.410	-0.035	3981.2	0.000
		27	0.388	-0.034	4027.5	0.000
		28	0.365	-0.048	4068.6	0.000
		29	0.343	0.036	4105.2	0.000
		30	0.322	-0.025	4137.4	0.000

Table 8 Correlogram of UE_MANUFACTURING

Sample: 1976:01 1998:12
Included observations: 276

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.960	0.960	257.18	0.000
		2	0.922	0.009	495.47	0.000
		3	0.883	-0.038	714.79	0.000
		4	0.840	-0.071	914.03	0.000
		5	0.787	-0.160	1089.4	0.000
		6	0.733	-0.043	1242.3	0.000
		7	0.674	-0.105	1371.7	0.000
		8	0.618	0.019	1481.0	0.000
		9	0.559	-0.057	1570.7	0.000
		10	0.495	-0.096	1641.2	0.000
		11	0.437	0.055	1696.6	0.000
		12	0.392	0.125	1741.3	0.000
		13	0.357	0.129	1778.4	0.000
		14	0.320	-0.020	1808.4	0.000
		15	0.286	-0.033	1832.4	0.000
		16	0.263	0.070	1852.8	0.000
		17	0.245	0.021	1870.6	0.000
		18	0.228	-0.022	1886.1	0.000
		19	0.214	-0.025	1899.8	0.000
		20	0.201	-0.038	1911.9	0.000
		21	0.190	-0.055	1922.7	0.000
		22	0.182	0.014	1932.7	0.000
		23	0.165	-0.097	1940.9	0.000
		24	0.148	-0.002	1947.6	0.000
		25	0.149	0.218	1954.3	0.000
		26	0.143	-0.053	1960.6	0.000
		27	0.128	-0.097	1965.7	0.000
		28	0.110	-0.055	1969.4	0.000
		29	0.090	-0.064	1971.9	0.000
		30	0.062	-0.136	1973.1	0.000

Table 9 Correlogram of UE_SERVICES

Sample: 1976:01 1998:12
Included observations: 276

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.980	0.980	268.08	0.000
		2	0.963	0.057	527.79	0.000
		3	0.946	-0.010	779.13	0.000
		4	0.926	-0.056	1021.3	0.000
		5	0.906	-0.042	1253.8	0.000
		6	0.884	-0.068	1475.9	0.000
		7	0.862	0.001	1688.0	0.000
		8	0.838	-0.070	1889.2	0.000
		9	0.814	-0.029	2079.5	0.000
		10	0.788	-0.039	2258.7	0.000
		11	0.763	0.001	2427.4	0.000
		12	0.739	0.023	2586.3	0.000
		13	0.717	0.023	2736.1	0.000
		14	0.694	-0.001	2877.1	0.000
		15	0.673	0.026	3010.1	0.000
		16	0.650	-0.036	3135.0	0.000
		17	0.628	-0.030	3251.7	0.000
		18	0.605	-0.038	3360.5	0.000
		19	0.584	0.050	3462.3	0.000
		20	0.564	-0.002	3557.7	0.000
		21	0.543	-0.021	3646.6	0.000
		22	0.520	-0.087	3728.4	0.000
		23	0.500	0.030	3804.1	0.000
		24	0.478	-0.038	3873.6	0.000
		25	0.459	0.078	3937.9	0.000
		26	0.440	-0.006	3997.4	0.000
		27	0.418	-0.108	4051.1	0.000
		28	0.396	-0.010	4099.7	0.000
		29	0.378	0.072	4144.2	0.000
		30	0.361	0.015	4184.7	0.000

Table 10

Modified Range over Standard Deviation (R/S) Test Statistic (Lo (1991))
The change of the log of Quarterly Canadian Unemployment Time Series

Total Unemployment

	Sum of Weights	Log(R/S)	Normalized Test Statistic
m=1	0.00083	0.99843	1.04450
m=2	0.00132	0.94306	0.91947
m=3	0.00170	0.90835	0.84884
m=4‡	0.00197	0.88583	0.80594*
m=5	0.00215	0.87287	0.78226*
m=6	0.00227	0.86382	0.76612*
m=7	0.00236	0.85764	0.75530*
m=8	0.00244	0.85233	0.74613*

Manufacturing Unemployment

	Sum of Weights	Log(R/S)	Normalized Test Statistic
m=1	0.00095	1.00146	1.05182
m=2	0.00257	0.86828	0.77404*
m=3	0.00468	0.76656	0.61241*
m=4	0.00717	0.68730	0.51025*
m=5	0.00991	0.62423	0.44127*
m=6	0.01282	0.57275	0.39195*
m=7‡	0.01585	0.52954	0.35483*
m=8	0.01899	0.49237	0.32572*

Services Unemployment

	Sum of Weights	Log(R/S)	Normalized Test Statistic
m=1	0.00306	0.89904	0.83085
m=2	0.00462	0.84992	0.74198*
m=3‡	0.00560	0.82397	0.69896*
m=4	0.00616	0.81050	0.67761*
m=5	0.00619	0.80973	0.67641*
m=6	0.00600	0.81438	0.68369*
m=7	0.00578	0.81970	0.69211*
m=8	0.00565	0.82271	0.69692*

‡ : Denotes the value for $[k_T]$
* : Indicates significance at the 5 percent level.

Table 11

Modified Range over Standard Deviation Test Statistic (Lo (1991))
HP Filtered Log of Quarterly Canadian Unemployment Time Series

	Sum of Weights	Log(R/S)	Normalized Test Statistic
Total Unemployment			
m=1	0.00432	1.08805	1.27690
m=2	0.00837	1.00475	1.05403
m=3	0.01201	0.95025	0.92972
m=4	0.01513	0.91232	0.85197
m=5	0.01771	0.88538	0.80073*
m=6	0.01975	0.86611	0.76598*
m=7	0.02132	0.85241	0.74219*
m=8	0.02247	0.84294	0.72618*
m=25‡	0.01193	0.95127	0.93192
Manufacturing Unemployment			
	Sum of Weights	Log(R/S)	Normalized Test Statistic
m=1	0.00755	0.98624	1.01006
m=2	0.01415	0.90783	0.84321
m=3	0.01951	0.86003	0.75534*
m=4	0.02348	0.83040	0.70551*
m=5	0.02608	0.81297	0.67775*
m=6	0.02764	0.80315	0.66260*
m=7	0.02854	0.79771	0.65435*
m=8	0.02906	0.79460	0.64969*
m=14‡	0.02810	0.80035	0.65834*
Services Unemployment			
	Sum of Weights	Log(R/S)	Normalized Test Statistic
m=1	0.00310	1.06064	1.19881
m=2	0.00598	0.97848	0.99216
m=3	0.00853	0.92493	0.87708
m=4	0.01069	0.88799	0.80555*
m=5	0.01244	0.86213	0.75898*
m=6	0.01380	0.84391	0.72781*
m=7	0.01481	0.83134	0.70704*
m=8	0.01551	0.82308	0.69373*
m=21‡	0.00974	0.90338	0.83462*

‡ : Denotes the value for $[k_T]$
* : Indicates significance at the 5 percent level.

Table 12: Posterior Model Probabilities

Posterior Model Probabilities for ARFIMA$(\mathbf{p}, \boldsymbol{\delta}, \mathbf{q})$		
Model	**Flat Prior**	**Informative Prior**
$(0, \delta, 0)$	0.0733	0.0646
$(0, \delta, 1)$	0.0681	0.0300
$(0, \delta, 2)$	0.0108	0.0047
$(0, \delta, 3)$	0.0359	0.0158
$(1, \delta, 0)$	0.3139	0.4155
$(1, \delta, 1)$	0.1128	0.0995
$(1, \delta, 2)$	0.0219	0.0193
$(1, \delta, 3)$	0.0235	0.0207
$(2, \delta, 0)$	0.0584	0.0773
$(2, \delta, 1)$	0.0541	0.0478
$(2, \delta, 2)$	0.0173	0.0152
$(2, \delta, 3)$	0.0218	0.0192
$(3, \delta, 0)$	0.0090	0.0119
$(3, \delta, 1)$	0.0280	0.0247
$(3, \delta, 2)$	0.0323	0.0285
$(3, \delta, 3)$	0.1181	0.1042

Table 13: Posterior Characteristics

Posterior Characteristics of δ			
Model	**Mean**	**St-Dev**	**Mode**
$(0, \delta, 0)$	0.429	0.055	0.500
$(0, \delta, 1)$	0.313	0.113	0.325
$(0, \delta, 2)$	0.281	0.133	0.275
$(0, \delta, 3)$	0.019	0.173	-0.075
$(1, \delta, 0)$	-0.034	0.263	-0.200
$(1, \delta, 1)$	0.053	0.326	0.400
$(1, \delta, 2)$	0.052	0.313	0.250
$(1, \delta, 3)$	-0.173	0.360	0.000
$(2, \delta, 0)$	-0.189	0.352	0.025
$(2, \delta, 1)$	-0.132	0.329	-0.175
$(2, \delta, 2)$	0.019	0.335	0.150
$(2, \delta, 3)$	0.042	0.400	0.475
$(3, \delta, 0)$	-0.085	0.333	-0.100
$(3, \delta, 1)$	-0.321	0.329	-0.575
$(3, \delta, 2)$	-0.222	0.362	-0.425
$(3, \delta, 3)$	0.318	0.208	0.425
Overall Model[38]	0.036	0.359	0.425

Table 14: **Impulse Responses**

	ARFIMA($\mathbf{1}, \boldsymbol{\delta}, \mathbf{0}$)	**Overall Model**
$n = 4$	2.235	2.253
	(0.372)	(0.153)
$n = 12$	2.618	1.144
	(0.861)	(2.301)
$n = 40$	2.874	1.684
	(1.750)	(4.997)

(.) denotes the posterior standard deviations.

Posterior Density DELTA (Simple Mix)

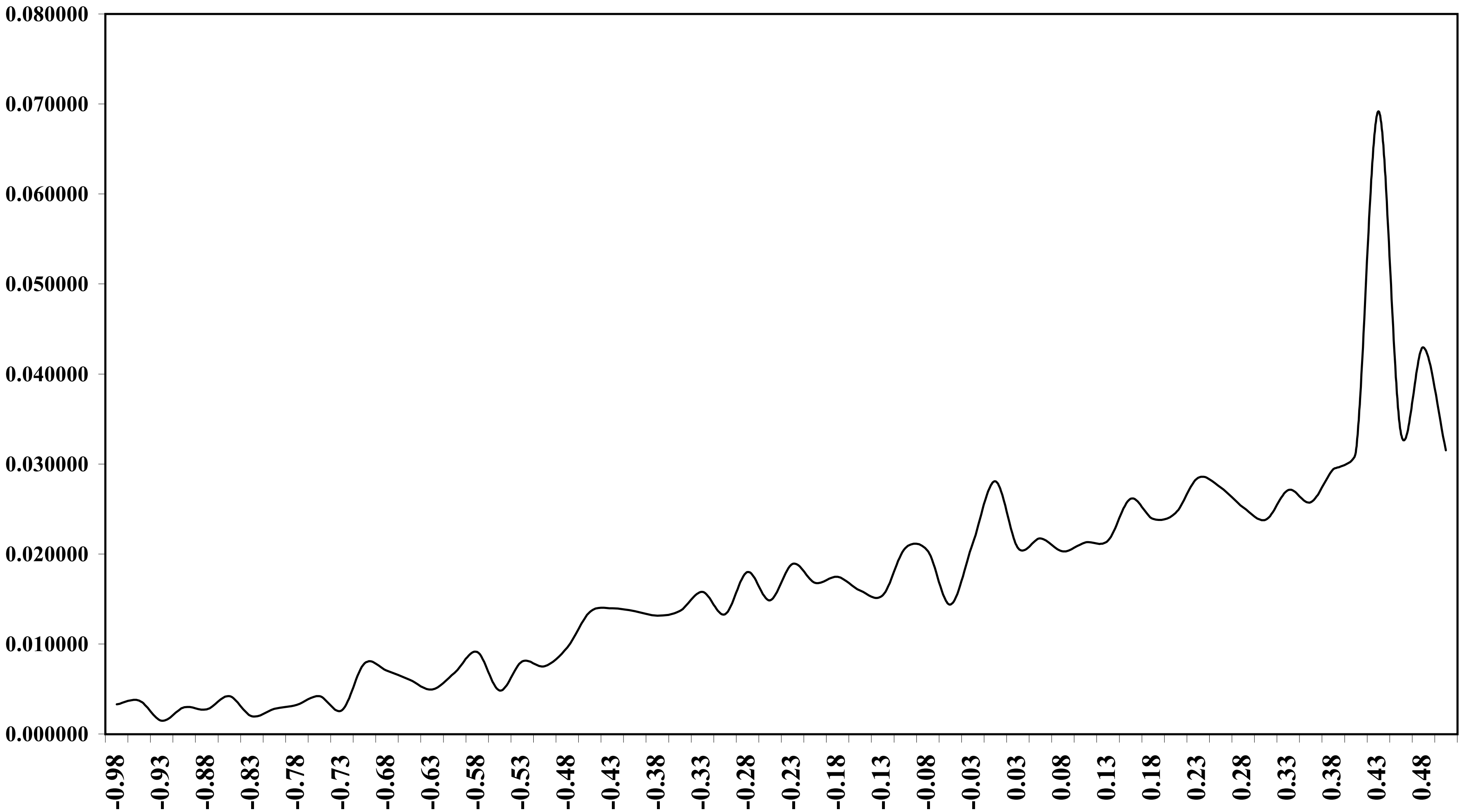

Figure 1

Posterior Densities for DELTA

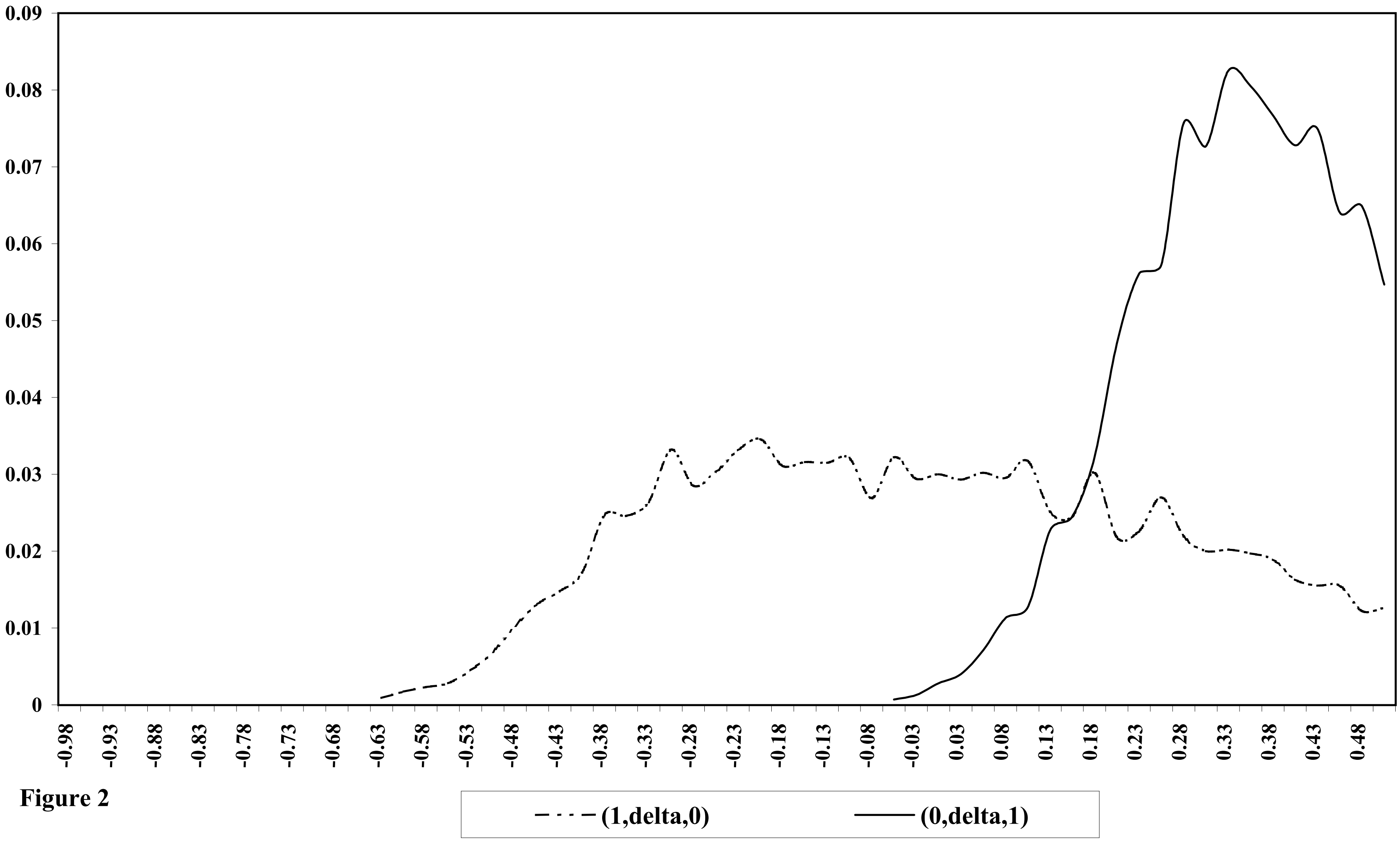

Figure 2

Posterior Densities for DELTA

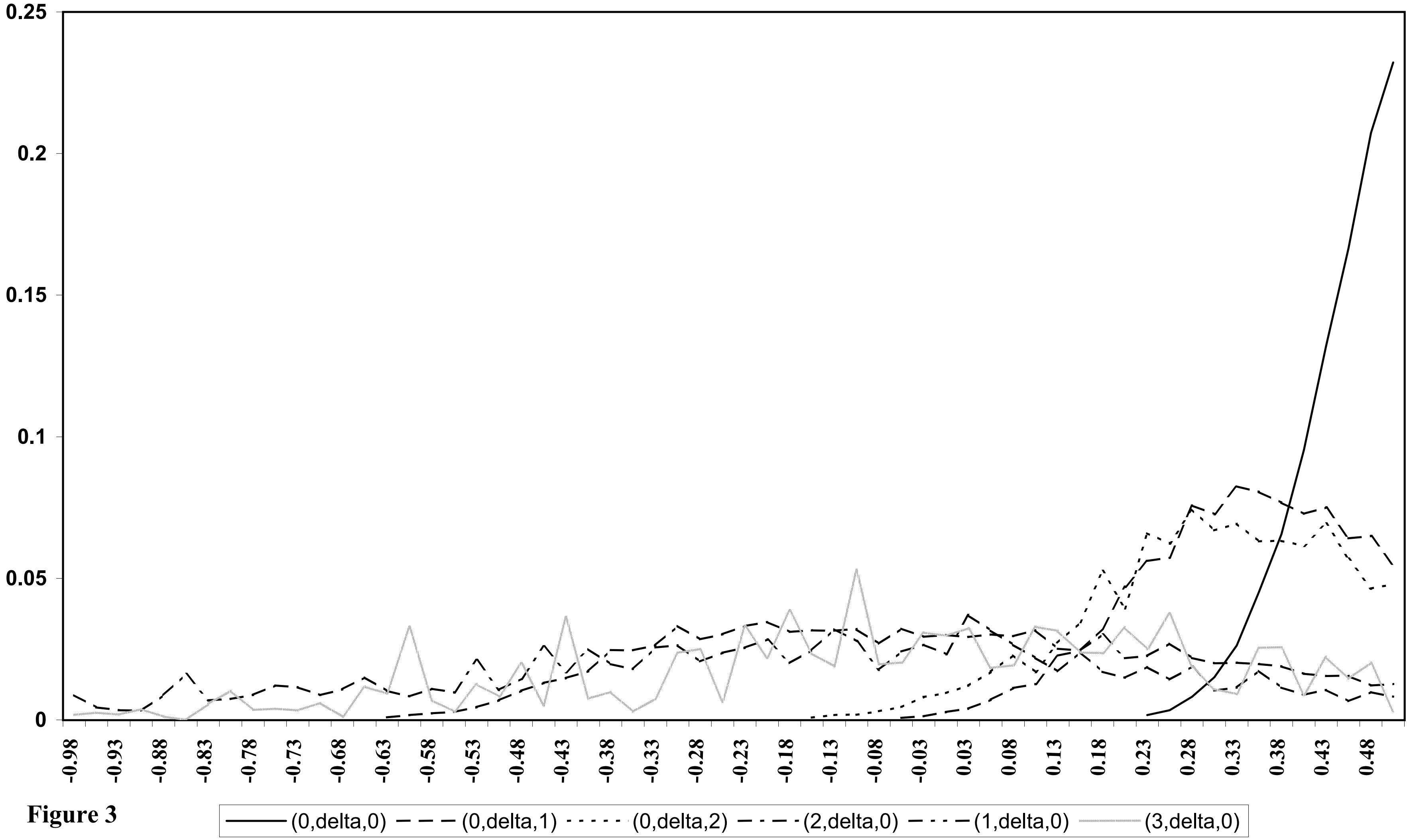

Figure 3

Posterior Densities for ARFIMA(1,d,0)

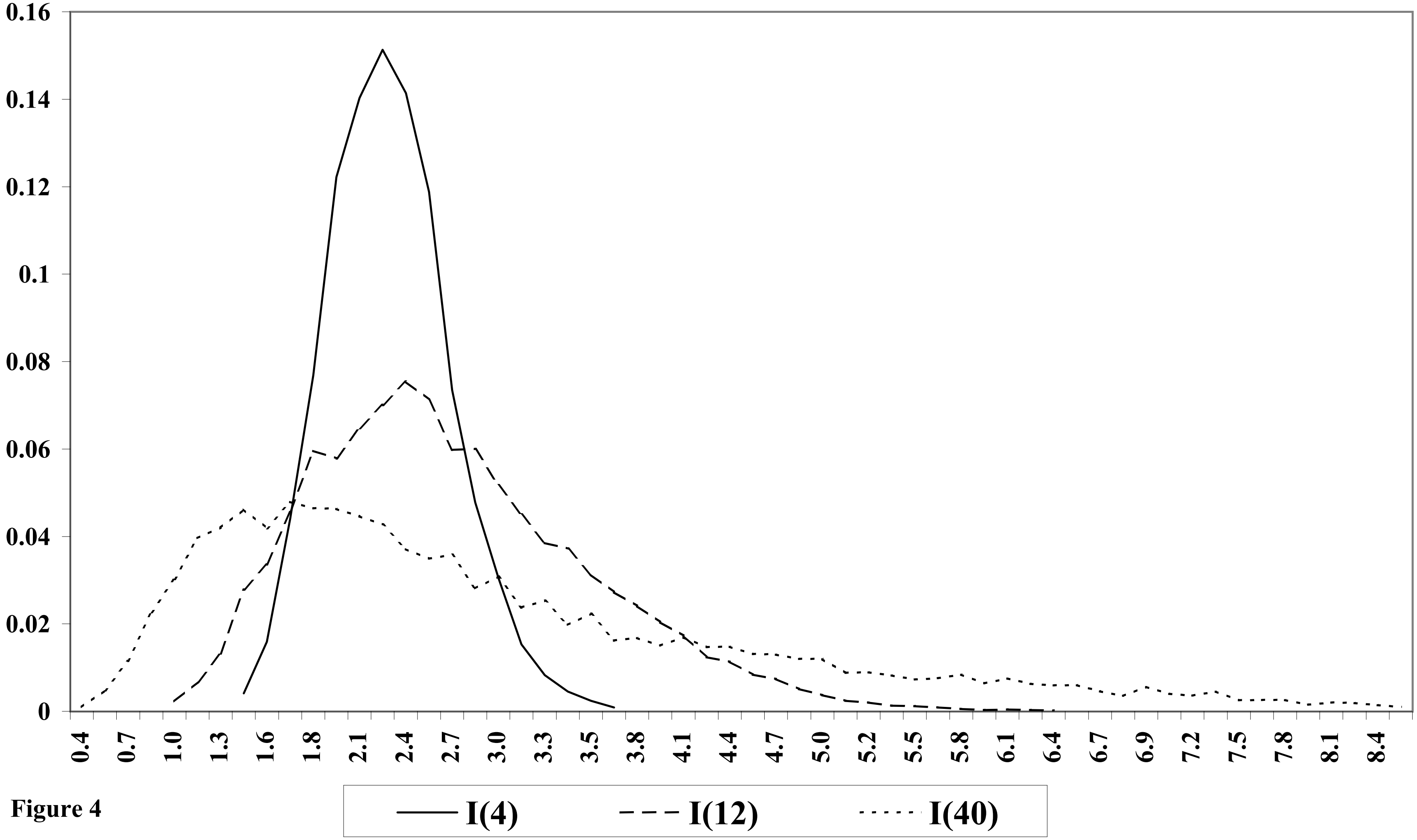

Figure 4