University of Toronto
Department of Economics

Working Paper 448

A New Structural Break Model with Application to Canadian Inflation Forecasting

By John M Maheu and Yong Song

March 13, 2012

A New Structural Break Model with Application to Canadian Inflation Forecasting

John M. Maheu* Yong Song†

January 2012

Abstract

This paper develops an efficient approach to model and forecast time-series data with an unknown number of change-points. Using a conjugate prior and conditional on time-invariant parameters, the predictive density and the posterior distribution of the change-points have closed forms. The conjugate prior is further modeled as hierarchical to exploit the information across regimes. This framework allows breaks in the variance, the regression coefficients or both. Regime duration can be modeled as a Poisson distribution. A new efficient Markov Chain Monte Carlo sampler draws the parameters as one block from the posterior distribution. An application to Canada inflation time series shows the gains in forecasting precision that our model provides.

Key words: multiple change-points, regime duration, inflation targeting, predictive density, MCMC

*Corresponding author: Department of Economics, 150 St. George St. University of Toronto, Toronto, Canada, M5S 3G7, 1-416-978-1495, and RCEA, jmaheu@chass.utoronto.ca

†CenSoC, University of Technology, Sydney, ysong1@gmail.com

1 Introduction

This paper develops an efficient Bayesian approach to model and forecast time series data with an unknown number of change-points. A conjugate prior is modeled as hierarchical to exploit information across regimes. Regime duration can be fixed or modelled as stochastic such as a Poisson distribution. The application to Canada inflation time series shows the gains in forecasting precision that our model provides.

Accounting for structural instability in macroeconomic and financial time series modeling and forecasting is important. Empirical applications by Clark and McCracken (2010), Geweke and Jiang (2011), Giordani et al. (2007), Liu and Maheu (2008), Wang and Zivot (2000), Stock and Watson (1996) among others demonstrate significant instability.

The problem of forecasting in the presence of structural breaks has been recently addressed by Koop and Potter (2007), Maheu and Gordon (2008), Maheu and McCurdy (2009) and Pesaran et al. (2006) using Bayesian methods. These approaches provide feasible solutions but are all computationally intensive.

The purpose of this paper is to forecast Canadian inflation with a change-point model suitable for out-of-sample forecasting with the attractive features of the previous approaches but which is computationally less demanding. Our approach allows for multiple structural breaks in-sample and out-of-sample, a duration-dependent structural break probability and estimation in several minutes. A new posterior simulator samples the parameters jointly as one block which results in very good mixing of the Markov chain.

We extend Maheu and Gordon (2008) and Maheu and McCurdy (2009) in five directions. First, a conjugate prior for the parameters which characterize each regime is adopted. Conditional on this prior and the time-invariant parameters, the predictive density has a closed form. The computational burden is reduced compared to Maheu and Gordon (2008), in which a non-conjugate prior is assumed.[1] Second, a hierarchical structure for the conjugate prior is introduced to allow learning and sharing of the information across regimes as in Pesaran et al. (2006). In the presence of a structural break, the new parameters are drawn independently from the hierarchical prior. Third, we show how to model the regime duration as a Poisson distribution, which implies duration dependent break probabilities. Unlike Maheu and Gordon (2008) who focus on the filtering and forecasting problem we show how to produce the smoothed distribution of the change-points. Lastly, different types of the break dynamics including having breaks in the variance, the regression coefficients or both are nested in this framework.

The differences between this paper and Koop and Potter's (2007) who also model regime durations are as follows. First, they assume a heterogeneous distribution for the duration in each regime. Their approach augments the state space by regime durations, so there are $O(T^2)$ states, which implies a large transition matrix. In contrast, we assume that the regime durations are drawn from the same distribution. This simplification results in number of states being $O(T)$ in our model. Second, Koop and Potter (2007) assume that after a structural change, the parameters in the new regime are related to those in the previous regime through a random walk. This path dependence in parameters further

[1] Maheu and Gordon (2008) assume a conditional conjugate prior and use Gibbs sampling to compute the predictive density.

increases computation time.

In this paper parameters in each regime are drawn independently from a hierarchical prior. This reflects an abrupt change of the parameters and is convenient for computation. We introduce a new MCMC sampler to draw all the parameters including the hierarchical prior, the parameters of the durations, the change-points and the parameters characterizing each regime from their posterior distribution jointly. As a result the mixing of the chain is very good.

Different versions of the model are applied to a Canada inflation series to investigate its dynamic stability. Canadian inflation is challenging to forecast as inflation targeting was introduced in 1991. This raises the question of how useful is the data prior to this date in forecasting after 1991. We also show how to incorporate exogenous subjective information from policy changes into our model.

The log-predictive likelihood is used as the criteria for model comparison. The best model is the hierarchical model which allows breaks in the regression coefficients and the variance simultaneously. This model provides large improvements compared to linear no break models and to autoregressive benchmarks with a GARCH parameterization.

We identify 4 major change-points in the Canada inflation dynamics. The model comparison shows that the duration dependent break probability is not a significant feature of the data. After controlling for the structural breaks, adding extra lags as the explanatory variables does not improve the out-of-sample forecasting. Including exchange rates and a commodity price index as predictors significantly improves linear models but is still dominated by the structural change model without these regressors.

The paper is organized as follows. Section 2 introduces the model and a Markov Chain Monte Carlo method is proposed to sample from the posterior distribution efficiently. Section 3 extends the non-hierarchical prior to a hierarchical one in order to exploit the information across regimes. Different extensions of the hierarchical model are introduced in Section 4, including a model with breaks only in the variance or in the regression coefficients. Duration dependent break probability is also discussed by assuming a Poisson distribution for the regime durations. Section 5 applies the model to a Canada inflation time series. Section 6 concludes.

2 Structural Break Model with Conjugate Prior

In the following we assume that two consecutive structural breaks define a regime. A regime consists of a set of contiguous data drawn from a data density with a fixed model parameter θ. Different regimes will have different θ which is assumed to be drawn from a specified distribution. The number of observations in a regime denotes the duration of a regime.

If time i is the starting point of the most recent regime, it is assumed that the data before time i is not informative for the posterior of the parameters θ governing the current regime.

If the most recent break is at time i ($i \leq t$) then the duration of the current regime at time t is defined as $d_t = t - i + 1$. The duration is used as a state variable in the following for two reasons. First, we wish to study not only the forecasting problem but also the ex-post analysis of multiple change-points in-sample.[2] Second, working with d_t facilitates the

[2]Maheu and Gordon did not consider the smoothed distribution of breaks and only focus on the filtered

modeling of regime durations directly which we discuss later.

Formally, define d_t as the duration of the most recent regime up to time t and $d_t \in \{1, \ldots, t\}$ by construction. If a break happens at time t, then $d_t = 1$. If $d_t = t$, then there is no break throughout the whole sample. Define $Y_{i,t} = (y_i, \ldots, y_t)$ for $1 \leq i \leq t$. If $i > t$, $Y_{i,t}$ is an empty set.

To form the predictive density for y_{t+1} conditional on duration d_{t+1}, we require the posterior density based on data $Y_{1,t}$. Let the data density of y_{t+1} given the model parameter θ and information set $Y_{1,t}$ be denoted as $p(y_{t+1} \mid \theta, Y_{1,t})$. There are two cases to consider. The first case is that the regime continues one more period while the second is a structural change and a new draw of the parameter θ occurs between t and $t+1$. If $p(\theta)$ is the prior for θ then conditional on duration d_{t+1} the posterior is

$$p(\theta|d_{t+1}, Y_{1,t}) \equiv p(\theta|Y_{t-d_{t+1}+2,t}) \propto \begin{cases} p(y_{t-d_{t+1}+2}, \ldots, y_t|\theta)p(\theta) & d_{t+1} > 1 \\ p(\theta) & d_{t+1} = 1 \end{cases} \tag{1}$$

The predictive density conditional on the duration is given by

$$p(y_{t+1} \mid d_{t+1}, Y_{1,t}) = \int p(y_{t+1} \mid \theta, Y_{1,t})p(\theta \mid d_{t+1}, Y_{1,t})d\theta \tag{2}$$

$$= \int p(y_{t+1} \mid \theta, Y_{1,t})p(\theta \mid Y_{t-d_{t+1}+2,t})d\theta \tag{3}$$

The second equality comes from the assumption that the data before a break point is uninformative to the regime after it. For example, if $d_{t+1} = 1$, $p(\theta \mid Y_{t-d_{t+1}+2,t})$ is equivalent to its prior $p(\theta)$. The data density conditions on $Y_{1,t}$ to allow for autoregressive models and other time series specifications.

The conditional distribution of $y_{t+1} \mid \theta, Y_{1,t}$ is a linear model with an i.i.d. normal error term. The prior is assumed a Normal-Gamma distribution, which is conjugate to the model. By conjugacy, the posterior distribution $\theta \mid Y_{t-d_{t+1}+2,t}$ is also Normal-Gamma. The predictive density $p(y_{t+1} \mid d_{t+1}, Y_{1,t})$ is a Student-t distribution if we integrate out θ. Conditional on d_t, the posterior distribution and the predictive density have analytic forms. If we assume a constant structural break probability $\pi \in (0, 1)$, the model with the conjugate prior can be written as follows:

$$\begin{aligned} d_t &= \begin{cases} d_{t-1} + 1 & \text{w.p. } 1 - \pi \\ 1 & \text{w.p. } \pi \end{cases} \\ (\beta_t, \sigma_t^{-2}) &\sim \mathbf{1}(d_t = 1)\mathbf{NG}(\underline{\beta}, \underline{H}^{-1}, \underline{\chi}/2, \underline{\nu}/2) + \mathbf{1}(d_t > 1)\delta_{(\beta_{t-1}, \sigma_{t-1}^{-2})} \\ y_t \mid \beta_t, \sigma_t, Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta_t, \sigma_t^2) \end{aligned} \tag{4}$$

The covariate x_t can include exogenous or lagged dependent variables. In this paper we consider $x_t = (1, y_{t-1}, \ldots, y_{t-q})'$, which is an AR(q) model in each regime. To emphasize that θ will change at each break point define $\theta_t \equiv (\beta_t, \sigma_t^2)$ as the collection of the parameters which characterize the data density at time t. If a break happens ($d_t = 1$), θ_t is drawn independently from the prior $\mathbf{NG}(\underline{\beta}, \underline{H}^{-1}, \underline{\chi}/2, \underline{\nu}/2)$, where $\mathbf{NG}$ represents a Normal-Gamma

distribution of change points.

distribution.[3] δ_θ represents a degenerate distribution at a mass point θ. If there is no break ($d_t > 1$), all parameters are the same as those in the previous period.

By conjugacy of the prior, the posterior distribution of the parameters which characterize the data density at time t is still Normal-Gamma conditional on the duration d_t.

$$\beta_t, \sigma_t^{-2} \mid d_t, Y_{1,t} \sim \mathbf{NG}(\hat{\beta}, \hat{H}^{-1}, \hat{\chi}/2, \hat{\nu}/2) \tag{5}$$

with

$$\hat{\beta} = \hat{H}^{-1}(\underline{H}\underline{\beta} + X'_{t-d_t+1,t}Y_{t-d_t+1,t}), \quad \hat{H} = \underline{H} + X'_{t-d_t+1,t}X_{t-d_t+1,t} \tag{6}$$

$$\hat{\chi} = \underline{\chi} + Y'_{t-d_t+1,t}Y_{t-d_t+1,t} + \underline{\beta}'\underline{H}\underline{\beta} - \hat{\beta}'\hat{H}\hat{\beta}, \quad \hat{\nu} = \underline{\nu} + d_t \tag{7}$$

where $X_{t-d_t+1,t} = (x_{t-d_t+1}, \ldots, x_t)'$.

If there is no break at time $t+1$, the new duration increase by 1 ($d_{t+1} = d_t + 1$) and the parameters which characterize the data dynamics stay the same ($\theta_{t+1} = \theta_t$) as the last period. The posterior distribution of θ_t is used to compute the predictive density.

$$p(y_{t+1} \mid d_{t+1} = d_t + 1, Y_{1,t}) \propto \left(1 + \frac{(y_t - x_t'\hat{\beta})^2}{\hat{\chi}(x_t'\hat{H}^{-1}x_t + 1)}\right)^{-\frac{(\hat{\nu}+1)}{2}} \tag{8}$$

which is the kernel of a Student-t distribution,

$$y_{t+1} \mid d_{t+1} = d_t + 1, Y_{1,t} \sim \mathbf{t}\left(x_t'\hat{\beta}, \frac{\hat{\chi}(x_t'\hat{H}^{-1}x_t + 1)}{\hat{\nu}}, \hat{\nu}\right). \tag{9}$$

For the special case of $d_{t+1} = 1$, a structural change happens at time $t+1$, so the data before $t+1$ is uninformative to the predictive density. In this case the posterior is replaced by the prior we obtain

$$y_{t+1} \mid d_{t+1} = 1, Y_{1,t} \sim \mathbf{t}\left(x_t'\underline{\beta}, \frac{\underline{\chi}(x_t'\underline{H}^{-1}x_t + 1)}{\underline{\nu}}, \underline{\nu}\right). \tag{10}$$

By integrating out the model parameters, the predictive density depends on the duration d_{t+1} and the past information $Y_{1,t}$. Now Chib's (1996) method to jointly sample the discrete latent variable from a hidden Markov model can be applied to sample $D_{1,T} = (d_1, \ldots, d_T)$ jointly.

The first step in sampling $D_{1,T}$ is the forward-filtering pass as follows.

1. At $t = 1$, the distribution of the duration is $p(d_1 = 1 \mid y_1) = 1$ by assumption.

[3] The precision (inverse of variance) σ_t^{-2} is drawn from a Gamma distribution $\mathbf{G}(\underline{\chi}/2, \underline{\nu}/2)$, where $\underline{\chi}/2$ is the multiplier and $\underline{\nu}/2$ is the degree of freedom. Its prior mean is $\frac{\underline{\nu}}{\underline{\chi}}$ and the prior variance is $\frac{2\underline{\nu}}{\underline{\chi}^2}$. It also implies the prior mean of the variance σ_t^2 is $\frac{\underline{\chi}}{\underline{\nu}-2}$. Conditional on the variance, the vector of the regression coefficients β_t is drawn from a multivariate normal distribution $\mathbf{N}(\underline{\beta}, \underline{H}^{-1}\sigma_t^2)$.

2. The forecasting step:

$$p(d_{t+1} = j \mid \pi, Y_{1,t}) = \begin{cases} p(d_t = j-1 \mid \pi, Y_{1,t})(1-\pi) & \text{for } j = 2, \cdots, t+1 \\ \pi & \text{for } j = 1 \end{cases}$$

3. The updating step:

$$p(d_{t+1} = j \mid \pi, Y_{1,t+1}) = \frac{p(y_{t+1} \mid d_{t+1} = j, Y_{1,t})p(d_{t+1} = j \mid \pi, Y_{1,t})}{p(y_{t+1} \mid \pi, Y_{1,t})}$$

   for $j = 1, \ldots, t+1$. The first term in the numerator on the right hand side is a Student-t distribution density function which we have derived using the conjugate prior. The second term is obtained from step 2. The denominator is the predictive likelihood given π and is computed by summing over all the values of the duration d_{t+1},

$$p(y_{t+1} \mid \pi, Y_{1,t}) = \sum_{j=1}^{t+1} p(y_{t+1} \mid d_{t+1} = j, Y_{1,t})p(d_{t+1} = j \mid \pi, Y_{1,t}). \tag{11}$$

4. Iterate over step 2 and 3 until the last period T.

Then, use the backward-sampling method to draw the vector of durations $D_{1,T} = (d_1 \ldots, d_T)$ jointly.

1. Sample the last period duration d_T from $d_T \mid \pi, Y_{1,T}$, which is obtained from the last iteration of the forward-filtering step.

2. If $d_t > 1$, then $d_{t-1} = d_t - 1$.

3. If $d_t = 1$, then sample d_{t-1} from the distribution $d_{t-1} \mid \pi, Y_{1,t-1}$. This is because $d_t = 1$ implies a structural change at time t. Hence, for any $\tau \geq t$, the data y_τ is in a new regime and uninformative to d_{t-1}. The distribution $d_{t-1} \mid d_t = 1, \pi, Y_{1,t-1}$ is equivalent to $d_{t-1} \mid d_t = 1, \pi, Y_{1,T}$.

4. Iterate step 2 and 3 until the first period $t = 1$.

Using the conjugate prior has several features. First, the computational burden is negligible compared to the original model with the non-conjugate priors. The computer memory required by the predictive likelihoods is $O(T^2)$, which is manageable for a sample size up to several thousands. The number of regimes is equal to the number of t such that $d_t = 1$ in the sample $D_{1,T}$. If K as the number of regimes implied by one sample of the vector of the durations $D_{1,T}$ from the posterior distribution, then $K = \sum_{t=1}^{T} \mathbf{1}(d_t = 1)$. The posterior distribution of $K-1$ is the distribution of the number of change-points. Finally, the posterior sampler is efficient based on Casella and Robert (1996), because the parameters $\Theta_{1,T} = \{\theta_t\}_{t=1}^{T}$ are integrated out.

2.1 Estimation and Inference

In the case of the constant break probability, the prior of the break probability π is assumed as a Beta distribution, $\mathbf{B}(\pi_a, \pi_b)$. Because the analytic conditional marginal likelihood $p(Y_{1,T} \mid \pi)$ exists, π can be sampled through a Metropolis-Hastings framework by integrating out the time-varying parameters $\Theta_{1,T}$ and the regime durations $D_{1,T}$.

For an efficient proposal sampling distribution, we exploit the information from the previous sample of the regime durations $D_{1,T}^{(i-1)}$ in the Markov chain. In other words, we use the known conditional distribution of π given the previous sample of $D_{1,T}^{(i-1)}$ as a tailored proposal distribution in a Metropolis-Hastings algorithm.

In the following, we sample from $p(\pi \mid Y_{1,T})$ first and then from $p(\Theta_{1,T}, D_{1,T} \mid \pi, Y_{1,T})$. This is equivalent to sampling from the joint posterior distribution $p(\pi, \Theta_{1,T}, D_{1,T} \mid Y_{1,T})$. The sampling steps are as follows.

1. Sample $\pi \mid Y_{1,T}$ from a proposal distribution:

   $$\pi^{(i)} \mid Y_{1,T} \sim \mathbf{Beta}(\pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})$$

   $K^{(i-1)}$ is the number of regimes implied from the previous sample of $D_{1,T}^{(i-1)}$. Accept $\pi^{(i)}$ with probability

   $$\min\left\{1, \frac{p(\pi^{(i)} \mid \pi_a, \pi_b)}{p(\pi^{(i-1)} \mid \pi_a, \pi_b)} \frac{p(Y_T \mid \pi^{(i)})}{p(Y_T \mid \pi^{(i-1)})} \frac{p(\pi^{(i-1)} \mid \pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})}{p(\pi^{(i)} \mid \pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})}\right\}$$

   If rejected, $\pi^{(i)}$ is set equal to $\pi^{(i-1)}$.

2. Sample $\Theta_{1,T}, D_{1,T} \mid \pi, Y_{1,T}$:

   (a) Sample $D_{1,T} \mid \pi, Y_{1,T}$ from the previously described forward-backward sampler. Calculate the number of regimes K and index the regimes by $1, \cdots, K$. Use an auxiliary variable s_t to represent the regime index at time t. Define $s_1 = 1$ and $s_t = 1$ for $t > 1$ until time τ with $d_\tau = 1$, which implies there is a break and the data is in a new regime. Then set $s_\tau = 2$ at this break point, and iterate until the last period with $s_T = K$.

   For example, if $D_{1,T} = (1, 2, 3, 1, 2, 1, 2, 3, 4)$, we can infer there are $K = 3$ regimes and the time series of regime indicators is $S_{1,T} = (s_1, \ldots, s_T) = (1, 1, 1, 2, 2, 3, 3, 3, 3)$. There is a one-to-one relationship between $D_{1,T}$ and $S_{1,T}$.

   (b) To sample $\Theta_{1,T} \mid D_{1,T}, \pi, Y_{1,T}$, we only need to sample K different sets of parameters because their values are constant in each regime. Define $\{\beta_i^*, \sigma_i^*\}$ as the distinct parameters which characterize the ith regime, where $i = 1, \ldots, K$.

   $$\beta_i^*, \sigma_i^{*-2} \sim \mathbf{NG}(\overline{\beta}_i, \overline{H}_i^{-1}, \overline{\chi}_i/2, \overline{\nu}_i/2)$$

   with

   $$\overline{\beta}_i = \overline{H}_i^{-1}(\underline{H}\underline{\beta} + X_i'Y_i), \quad \overline{H}_i = \underline{H} + X_i'X_i$$

$$\overline{\chi}_i = \underline{\chi} + Y_i'Y_i + \underline{\beta}'\underline{H}\underline{\beta} - \overline{\beta}_i'\overline{H}_i\overline{\beta}_i, \quad \overline{\nu}_i = \underline{\nu} + D_i$$

$X_i = (x_{t_0}, \ldots, x_{t_1})'$ and $Y_i = (y_{t_0}, \ldots, y_{t_1})'$, where $s_t = i$ if and only if $t_0 \le t \le t_1$. So, X_i and Y_i represent the data in the ith regime. $D_i = t_1 - t_0 + 1$ is the duration of the ith regime.

The Markov chain is run for $N_0 + N$ times and the first N_0 iterations are discarded as burn-in samples. The rest of the samples of the parameters $\left\{\pi^{(i)}, \Theta_{1,T}^{(i)}, D_{1,T}^{(i)}\right\}_{i=1}^{N}$ are used for inferences and forecasting as if they were drawn from the posterior distribution. For example, the posterior mean of the break probability is computed as the sample average of $\pi^{(i)}$ as $\hat{E}(\pi \mid Y_{1,T}) = \frac{1}{N}\sum_{i=1}^{N}\pi^{(i)}$. The posterior mean of the volatility at time t is $\hat{E}(\sigma_t^2 \mid Y_{1,T}) = \frac{1}{N}\sum_{i=1}^{N}\sigma_t^{2(i)}$. Similarly, we can estimate the predictive density for time $T+1$ by averaging over the MCMC draws of π using (11) or based on the draws of π and $D_{1,T}$ following

$$\hat{p}(y_{T+1} \mid Y_{1,T}) = \frac{1}{N}\sum_{i=1}^{N}\left\{p(y_{T+1} \mid d_{T+1} = d_T^{(i)} + 1, Y_{1,T})(1 - \pi^{(i)}) + p(y_{T+1} \mid d_{T+1} = 1, Y_{1,T})\pi^{(i)}\right\}. \quad (12)$$

This model has two crucial assumptions. One is the conjugate prior for the regime dependent parameters which characterize the conditional data density. The other is that the data before a break point is uninformative to the regime after it conditional on the time-invariant parameters. Both are necessary for the analytic form of the predictive density. If we do not use the conjugate prior, each predictive density $p(y_{t+1} \mid d_{t+1}, Y_{1,t})$ has to be estimated numerically. If the second assumption is violated, the data before the break can provide information to the regime after it, the duration d_t itself is not sufficient for the predictive density given the time-invariant parameters. For example, in Koop and Potter's (2007) model, in order to integrate out the parameters in the most recent regime, they need to know the whole sample path of the durations $D_{1,t} = (d_1, \ldots, d_t)$. However, since the vector of durations $D_{1,t}$ takes 2^t values in their model, it is computationally infeasible to calculate the predictive likelihood for every case, while in the new model it is feasible.

Because data prior to a break point may be useful in forecasting we next consider a hierarchical prior to exploit this but still maintain the computational feasibility of our approach.

3 Hierarchical Structural Break Model

In our model, forecasts immediately after a break are dominated by the prior and could be poor if the prior is at odds with the new parameter value of the data density. Of course as more data arrives the predictive density improves but this can take some time.

Pesaran et al. (2006) proposed to estimate the prior to improve forecasting by exploiting the information across regimes. This section introduces a hierarchical prior for the structural break model. This is computationally feasible only if using the conjugate prior as in the previous section. The model is referred as the hierarchical SB-LSV model: SB means structural break and LSV means that the level, the slope and the variance are subject to breaks. The model in the previous section is labelled as the non-hierarchical SB-LSV model.

The previous prior parameters $\underline{\beta}, \underline{H}, \underline{\chi}, \underline{\nu}$ are not fixed any more but given a prior distribution. The hierarchical SB-LSV model is the following:

$$
\begin{aligned}
\underline{\beta}, \underline{H} &\sim \mathbf{NW}(m_0, \tau_0^{-1}, A_0, a_0) \\
\underline{\chi} &\sim \mathbf{G}(d_0/2, c_0/2) \\
\underline{\nu} &\sim \mathbf{Exp}(\rho_0) \\
d_t &= \begin{cases} d_{t-1}+1 & \text{w.p. } 1-\pi \\ 1 & \text{w.p. } \pi \end{cases} \\
(\beta_t, \sigma_t^{-2}) &\sim \mathbf{1}(d_t = 1)\mathbf{NG}(\underline{\beta}, \underline{H}^{-1}, \underline{\chi}/2, \underline{\nu}/2) + \mathbf{1}(d_t > 1)\delta_{(\beta_{t-1}, \sigma_{t-1}^{-2})} \\
y_t \mid \beta_t, \sigma_t, Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta_t, \sigma_t^2)
\end{aligned}
\tag{13}
$$

The positive definite matrix $\underline{H}$ has a Wishart distribution $\mathbf{W}(A_0, a_0)$, where A_0 is a positive definite matrix and a_0 is a positive scalar. The prior mean of $\underline{H}$ is $a_0 A_0$. The prior variance of $\underline{H}_{ij}$ is $a_0(A_{ij}^2 + A_{ii}A_{jj})$, where subscript ij means the ith row and the jth column. $\underline{\beta} \mid \underline{H}$ is a multivariate Normal $\mathbf{N}(m_0, \tau_0^{-1}\underline{H}^{-1})$, where τ_0 is a positive scalar. $\underline{\chi}$ has a Gamma distribution with a prior mean of c_0/d_0 and a prior variance of $2c_0/d_0^2$. $\underline{\nu}$ has an Exponential distribution with both the prior mean and variance equal to ρ_0.

Conditional on the number of regimes K and the distinct parameter values $\{\beta_i^*, \sigma_i^*\}_{i=1}^K$, the posterior distribution of the hierarchical parameters $\underline{\beta}$ and $\underline{H}$ are still Normal-Wishart.

$$
\underline{\beta}, \underline{H} \mid \{\beta_i^*, \sigma_i^*\}_{i=1}^K \sim \mathbf{NW}(m_1, \tau_1^{-1}, A_1, a_1)
$$

with

$$
m_1 = \frac{1}{\tau_1}\left(\tau_0 m_0 + \sum_{i=1}^K \sigma_i^{*-2}\beta_i^*\right), \quad \tau_1 = \tau_0 + \sum_{i=1}^K \sigma_i^{*-2} \tag{14}
$$

$$
A_1 = \left(A_0^{-1} + \sum_{i=1}^K \sigma_i^{*-2}\beta_i^*\beta_i^{*'} + \tau_0 m_0 m_0' - \tau_1 m_1 m_1'\right)^{-1}, \quad a_1 = a_0 + K \tag{15}
$$

The posterior of $\underline{\chi} \mid \underline{\nu}, K, \{\sigma_i^*\}_{i=1}^K$ is a Gamma distribution,

$$
\underline{\chi} \mid \underline{\nu}, \{\sigma_i^*\}_{i=1}^K \sim \mathbf{G}(d_1/2, c_1/2) \tag{16}
$$

with $d_1 = d_0 + \sum_{i=1}^K \sigma_i^{*-2}$ and $c_1 = c_0 + K\underline{\nu}$. The posterior of $\underline{\nu} \mid \underline{\chi}, K, \{\sigma_i^*\}_{i=1}^K$ is

$$
p(\underline{\nu} \mid \underline{\chi}, K, \{\sigma_i^*\}_{i=1}^K) \propto \left(\frac{(\underline{\chi}/2)^{\underline{\nu}/2}}{\Gamma(\underline{\nu}/2)}\right)^K \left(\prod_{i=1}^K \sigma_i^{*-2}\right)^{\underline{\nu}/2} \exp\left\{-\frac{\underline{\nu}}{\rho_0}\right\}.
$$

which does not have a convenient form. It is sampled by a Metropolis-Hastings algorithm using a random walk as the proposal distribution.

Sampling from the posterior density of the break probability π and the hierarchical prior parameters follows the same approach used in non-hierarchical SB-LSV model. To implement the sampler, define $\Psi = (\pi, \underline{\beta}, \underline{H}, \underline{\chi}, \underline{\nu})$ as the collection of the break probability and the parameters of the hierarchical prior, which are all time-invariant. To obtain a good proposal density we base it on the previous iteration of the sampler and exploit known conditional posterior densities. Since the analytic form of the marginal likelihood $p(Y_{1,T} \mid \Psi)$ exists, the joint sampler draws Ψ from this proposal distribution and accepts the new draw with a probability implied by the Metropolis-Hastings algorithm.

Following this, sample the regime durations $D_{1,T}$ and the time-varying parameters $\Theta_{1,T}$ conditional on Ψ and the data $Y_{1,T}$. As in the previous specification we are sampling jointly from the full posterior $\Theta_{1,T}, D_{1,T}, \Psi | Y_{1,T}$ which results is a well mixing Markov chain. The details are in the appendix.

After discarding the burn-in samples, the rest of the sample is used to draw inferences from the posterior as in the non-hierarchical model. The predictive density, $p(y_{T+1} \mid Y_{1,T})$ is estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\left\{p(y_{T+1} \mid d_{T+1} = d_T^{(i)} + 1, \Psi^{(i)}, Y_{1,T})(1 - \pi^{(i)}) + p(y_{T+1} \mid d_{T+1} = 1, \Psi^{(i)}, Y_{1,T})\pi^{(i)}\right\}.$$

Alternatively, averaging over the predictive density expression in (11) can be used. This latter approach integrates out the durations.

4 Extensions

This section extends the model while preserving the two assumptions. The first extension allows the structural breaks only in the variance σ_t^2 or in the regression coefficients β_t. The second extension considers the duration dependent break probabilities.

4.1 Breaks in the Variance

The model with breaks only in the variance is referred as the hierarchical SB-V model. It assumes a time-invariant vector of the regression coefficients β. The time-varying variance σ_t^2 are drawn from a hierarchical prior. The model is

$$\begin{aligned}
\underline{\chi} &\sim \mathbf{G}(d_0/2, c_0/2) \\
\underline{\nu} &\sim \mathbf{Exp}(\rho_0) \\
\beta &\sim \mathbf{N}(\underline{\beta}, \underline{H}^{-1}) \\
d_t &= \begin{cases} d_{t-1} + 1 & \text{w.p. } 1 - \pi \\ 1 & \text{w.p. } \pi \end{cases} \\
\sigma_t^{-2} &\sim \mathbf{1}(d_t = 1)\mathbf{G}(\underline{\chi}/2, \underline{\nu}/2) + \mathbf{1}(d_t > 1)\delta_{\sigma_{t-1}^{-2}} \\
y_t \mid \beta, \sigma_t, Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta, \sigma_t^2).
\end{aligned} \tag{17}$$

The prior for the regression coefficients β is not modelled as hierarchical since it is constant across all regimes. The parameters of its prior $\underline{\beta}$ and $\underline{H}$ are fixed. On the other hand, the prior for the variance σ_t^2 is modelled as hierarchical to share the information across regimes. Since the regression coefficient β is the same in all regimes, the data before a break point is informative to the regime after it. So the duration of the most recent regime d_t is not sufficient for computing the posterior of the parameters in that regime. That is, $p(\theta_t \mid d_t, Y_{1,t}) \neq p(\theta_t \mid d_t, Y_{t-d_t+1,t})$. The predictive density $p(y_{t+1} \mid d_{t+1}, Y_{1,t})$ is not a Student-t distribution any more as in the non-hierarchical SB-LSV model.

However, $p(\theta_t \mid d_t, \beta, Y_{1,t}) = p(\theta_t \mid d_t, \beta, Y_{t-d_t+1,t})$ holds. Namely, conditional on β, if a break happens, the volatility is independently drawn from the hierarchical prior and the previous information is not useful for the current regime.

Meanwhile, conditional on β, the prior for the variance is conjugate. So the model can be estimated using the method similar to that in the hierarchical SB-LSV model. Specifically, define the collection of the time-invariant parameters as $\Psi = (\pi, \beta, \underline{\chi}, \underline{\nu})$. The posterior MCMC sampler first randomly draw $\Psi \mid Y_{1,T}$ using a tailored proposal distribution and accept it with the probability implied by a Metropolis-Hastings algorithm. Then, conditional on Ψ and the data $Y_{1,T}$, draw the regime durations $D_{1,T}$ and the time-varying parameters $\Theta_{1,T}$. In the hierarchical SB-V model, $\Theta_{1,T} = \{\sigma_t\}_{t=1}^{T}$, because the time-invariant regression coefficients $\beta \in \Psi$ are sampled in the first step. The details are in the appendix.

4.2 Breaks in the Regression Coefficients

We can also fix the variance σ^2 as time-invariant and only allow the regression coefficients to change over time. This model is named as the hierarchical SB-LS since the breaks only happen for the level and slopes. Conditional on the variance σ^2, the data before a break is not informative to the current regime. Also, the conjugate prior exists for the regression coefficient β_t in each regime. The hierarchical SB-LS model can be estimated as the hierarchical SB-LSV or SB-V model. The model is:

$$
\begin{aligned}
\underline{\beta}, \underline{H} &\sim \mathbf{NW}(m_0, \tau_0^{-1}, A_0, a_0) \\
\sigma^{-2} &\sim \mathbf{G}(\underline{\chi}/2, \underline{\nu}/2) \\
d_t &= \begin{cases} d_{t-1} + 1 & \text{w.p. } 1 - \pi \\ 1 & \text{w.p. } \pi \end{cases} \\
\beta_t &\sim \mathbf{1}(d_t = 1)\mathbf{N}(\underline{\beta}, \underline{H}^{-1}) + \mathbf{1}(d_t > 1)\delta_{\beta_{t-1}} \\
y_t \mid \beta_t, \sigma, Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta_t, \sigma^2)
\end{aligned}
\tag{18}
$$

The posterior sampler randomly draws the time-invariant parameter $\Psi = (\pi, \underline{\beta}, \underline{H}, \sigma)$ from its posterior distribution using a MCMC sampler. Then it samples the regime durations $D_{1,T}$ and the time varying parameters $\Theta_{1,T} = \{\beta_t\}_{t=1}^{T}$ conditional on the time-invariant parameter Ψ and the data $Y_{1,T}$. The details are in the appendix.

4.3 Duration Dependent Break Probability

Due to the analytic form for the predictive density even with duration dependent break probabilities, our approach continues to be computationally straightforward. Since modeling the duration dependent break probability is equivalent to modeling the duration, we assume a Poisson distribution for each regime.

The hazard rate represents the duration dependent break probabilities[4]. The Poisson distribution function is $P(\text{Duration} = d \mid \lambda) = e^{-\lambda}\frac{\lambda^{(d-1)}}{(d-1)!}$, where $d \geq 1$. The implied break probability is

$$\begin{aligned}\pi_j = P(d_{t+1} = 1 \mid d_t = j, \lambda) &= P(\text{Duration} = j \mid \text{Duration} \geq j, \lambda) \\ &= \frac{P(\text{Duration} = j \mid \lambda)}{P(\text{Duration} \geq j \mid \lambda)} \\ &= \frac{e^{-\lambda}\lambda^{(j-1)}}{(j-1)\gamma(j-1, \lambda)}\end{aligned}$$

where $\gamma(x, y)$ is the incomplete gamma functions with $\gamma(x, y) = \int_0^y t^{x-1}e^{-t}dt$. The no-break probability $P(d_{t+1} = j + 1 \mid d_t = j, \lambda)$ is simply $1 - P(d_{t+1} = 1 \mid d_t = j, \lambda)$.

Previously, the time-invariant structural break probability π is used in the forecasting step to compute $p(d_{t+1} = j \mid \pi, Y_{1,t})$ in order to construct the filtered probability $p(d_t = j \mid \pi, Y_{1,t})$ and the predictive density $p(y_{t+1} \mid \pi, Y_{1,t})$. If the break probability depends on the regime duration, $p(d_{t+1} = 1 \mid \lambda, d_t = j) = \pi_j$, then $p(d_{t+1} = j \mid \lambda, Y_{1,t})$ is calculated as

$$p(d_{t+1} = j \mid \lambda, Y_{1,t}) = \begin{cases} p(d_t = j - 1 \mid \lambda, Y_{1,t})(1 - \pi_{j-1}) & \text{for } j = 2, \cdots, t + 1 \\ \sum_{k=1}^{t} p(d_t = k \mid \lambda, Y_{1,t})\pi_k & \text{for } j = 1. \end{cases}$$

The updating step of the forward filtering procedure and the backward sampling procedure are not affected. Conditional on the durations $D_{1,T}$, the posterior of the parameters which characterize each regime are not changed either. So the estimation is still computationally straightforward.

The priors for the other parameters are set the same as the hierarchical SB-LSV model. This extension is labelled as the hierarchical DDSB-LSV model, where DD means duration dependent.

To estimate the hierarchical DDSB-LSV model, notice that the set of the time-invariant parameters Ψ now is $(\lambda, \underline{\beta}, \underline{H}, \underline{\chi}, \underline{\nu})$. The posterior sampler draws Ψ from its posterior distribution by a Metropolis-Hastings sampler. Then the time-varying parameters $\Theta_{1,T}$ and the regime durations $D_{1,T}$ are sampled conditional on the time-invariant parameter Ψ and the data $Y_{1,T}$. This is still a joint sampler as in the hierarchical SB-LSV with the time-invariant break probability. Details are in the appendix.

5 Application to Canadian Inflation

The model is applied to Canadian quarterly inflation time series to investigate its dynamic instability and forecast performance. The data is constructed from the quarterly CPI, which is downloaded from CANSIM[5]. The quarterly inflation rate is calculated as the log difference

[4]In general, any hazard function in the survival analysis can be applied to model the duration.

[5]TABLE NUMBER: 3800003. TABLE TITLE: GROSS DOMESTIC PRODUCT (GDP) INDEXES. Data Sources: IMDB (Integrated Meta Data Base) Numbers: 1901 - NATIONAL INCOME AND EXPENDITURE ACCOUNTS. SERIES TITLE: CANADA; IMPLICIT PRICE INDEXES 2002=100; PERSONAL EXPENDITURE ON CONSUMER GOODS AND SERVICES SERIES FREQUENCY: Quarterly

of the CPI data and scaled by 100. It starts from 1961Q1 and ends at 2009Q4 with 196 observations in total. The summary statistics are in Table 1.

The hierarchical models used are SB-LSV, SB-V, SB-LS and DDSB-LSV models. Two non-hierarchical SB-LSV models are also applied, one estimates the break probability π and the other fixes $\pi = 0.01$. Linear autoregressive models are used as benchmarks for model comparison and a GARCH specification to capture heteroskedasticity. For all the structural break models, we assume that the explanatory variables in each regime include an intercept and the one-period lag of the dependent variable. So the data follows an AR(1) process in the each regime.

The prior of the hierarchical SB-LSV model is:

$$\pi \sim \mathbf{B}(1,9), \;\; \underline{H} \sim \mathbf{W}(0.2I,5), \;\; \underline{\beta} \mid \underline{H} \sim \mathbf{N}(0,\underline{H}^{-1}), \;\; \underline{\chi} \sim \mathbf{G}(2,2), \;\; \underline{\nu} \sim \mathbf{Exp}(2)$$

This prior is informative but covers a wide range of empirically realistic values. The prior mean of the break probability is $E(\pi) = 0.1$, which implies infrequent breaks. The inverse of the variance in each regime is drawn from a Gamma distribution, which has a degree of freedom centered at 2 and a multiplier centered at 1.

The prior and the posterior summary of the parameters are in Table 2. The posterior mean of the structural change probability π is 0.04, which is less than its prior mean of 0.1. The posterior mean implies an average duration of 6 years and 1 quarter. The 95% density interval is narrower than that of the prior. Although the posterior mean of $\underline{H}$ is similar to the prior the density intervals are tighter than the prior intervals. On the other hand, the prior and the posterior mean for the intercept $\underline{\beta}_0$ are 0 and 0.82, respectively. And the posterior 95% density interval of $\underline{\beta}_0$ does not cover 0. After a structural break the new intercept tends to be positive. The expected value of $\underline{\chi}$ does not change from the prior to posterior but its density interval shrinks, which implies that the data confirms the prior assumption. Lastly, for $\underline{\nu}$ there is a significant rightward shift in the posterior.

The posterior means of the regression coefficients $E(\beta_t \mid Y_{1,T})$, the standard deviations $E(\sigma_t \mid Y_{1,T})$ and the structural change probabilities $p(d_t = 1 \mid Y_{1,T})$ for $t = 1, \ldots, T$, are plotted in Figure 1.[6] The top panel is the data. The second panel plots the break probabilities over time. The middle panel plots the intercept $\beta_{t,0}$ over time. The AR(1) coefficient $\beta_{t,1}$ is plotted in the fourth panel and is labelled as persistence. The standard deviation σ_t is in the bottom panel. From the plot of the break probabilities, we can visually identify 4 major breaks in the inflation process. The first is in the mid-60's, which is featured by an increase of the inflation level. The second is in the early 70's, which is associated with oil crisis and characterized by an increase of the persistence and the volatility. In the mid 80's, a structural change happened by decreasing both the persistence and the volatility, which is consistent with the great moderation. The last break happened in the early 90's, which featured decreases in both the inflation level and its volatility and coincides with the introduction of an inflation target by the Bank of Canada. Figured 1 shows that each break brings different dynamic patterns to the inflation process.

The non-hierarchical SB-LSV model fixes the parameters of the priors at $\underline{\beta} = (0,0)'$, $\underline{H} = I_2$, $\underline{\chi} = 1$, $\underline{\nu} = 2$, which are the prior means of the hierarchical SB-LSV model. The break

[6] At time $t = 1$, we plot $p(d_t = 1 \mid Y_{1,T}) = 0$, because it is in the first regime by construction and not marked as a change-point in this paper.

probability π has the same prior as that of the hierarchical model, which is $\mathbf{B}(1, 9)$. The posterior mean of π equals to 0.01. Its 95% density interval is $(0.002, 0.029)$. The non-hierarchical SB-LSV model implies a longer regime duration than the hierarchical SB-LSV model. The posterior means of the time varying parameters $E(\beta_t \mid Y_{1,T})$, $E(\sigma_t \mid Y_{1,T})$ and the probabilities of breaks, $p(d_t = 1 \mid Y_{1,T})$, are plotted in Figure 2. Each panel has the same interpretation as in Figure 1. There are three points worth noticing. First, there is only one spike in the break probabilities, which is in the early 90's. It captures the decrease of the variance and is consistent with the last change-point identified by the hierarchical SB-LSV model. Second, although it is not visually identifiable in the second panel, from the middle and the fourth panel, we can observe a gradual increase of the persistence and a decrease of the intercept between the mid 60's and the early 70's. Lastly, the non-hierarchical SB-LSV model fails to identify the great moderation in the mid 80's.

As one alternative to the time-invariant break probability, the duration is modeled as a Poisson distribution to fit the inflation dynamics. The prior of the duration parameter λ is assumed as an exponential distribution with a mean of 50. The other priors are set as the same as that of the hierarchical SB-LSV model. For simplicity, the first period is assumed to be the first period of its regime. Table 3 shows the posterior summary of the parameters. The posterior of summaries of $\underline{\beta}$ are similar to that in the hierarchical SB-LSV model, but $\underline{\chi}$ and $\underline{\nu}$ are different. The estimates of λ implies one regime lasts about 7 years and a quarter, which is comparable to the length of 6 years and a quarter implied by the hierarchical SB-LSV model. Figure 3 shows that the change-points identified by the duration dependent model is consistent with Figure 1. The dynamic patterns of these two figures are similar except for the last 10 years. The hierarchical DDSB-LSV model shows some structural change uncertainty around the year 2000, after which the volatility increased. Although the smoothed parameters for the hierarchical DDSB-LSV model are similar to that of the hierarchical SB-LSV model, the later model comparison below shows that the Poisson duration is strongly rejected. This is attributed to the fact that the duration dependent break probability implied by the Poisson distribution is very small if the regime duration is short. For example, if the duration parameter λ equals to the posterior mean of 28.9, the break probability $p(d_{t+1} = 1 \mid d_t)$ is less than $1.0\mathrm{e}^{-5}$ if the duration $d_t < 10$. This feature causes the model to learn regime changes slower than a constant break probability model.

For the hierarchical SB-V model, which only allows breaks in the variance, the prior of the time-invariant regression coefficient vector is $\beta \sim N(0, I)$. Its mean and the precision matrix are the prior means in the hierarchical SB-LSV model. The priors of π, $\underline{\chi}$ and $\underline{\nu}$ are the same as the hierarchical SB-LSV model. The posterior summary is in Table 4. The most prominent feature is that the posterior mean of the break probability π is 0.16, which is much higher than that of the hierarchical or the non-hierarchical SB-LSV model. By fixing β over the sample more breaks in σ_t are required to accommodate the data.

The frequent change of volatility is shown in Figure 4. The middle panel is the break probability, from which we can observe that the process is characterized by many breaks in the variance. This frequent break pattern is similar to the ARCH effects in Engle (1982; 1983). The bottom panel plots the posterior means of the standard deviations $E(\sigma_t \mid Y_T)$. Although we can see some episodes such as from the mid 80's to the early 90's are more stable, there is no general pattern about the volatility evolution. In practise, it is not desirable to have too frequent structural changes, which implies that less data can be used to estimate the most recent regime. The frequent break pattern of Canadian inflation estimated by the hierarchical SB-V model reflects model misspecification, because the more general hierarchical SB-LSV model nests the hierarchical SB-V model and it does not find as many breaks as the latter one does.

On the other hand, the hierarchical SB-LS model allows the breaks to happen only in the regression coefficients and keeps the variance constant. The prior of the inverse of the variance is $\sigma^{-2} \sim \mathbf{G}(1, 0.5)$. The values of the multiplier and the degree of freedom in this prior are the means implied by the prior for the hierarchical SB-LSV model. The priors for π, $\underline{\beta}$ and $\underline{H}$ are set the same as that of the hierarchical SB-LSV model. The posterior summary is in Table 5. The posterior for the break probability is similar to that of the hierarchical SB-LSV model. Figure 5 plots the posterior means of the regression coefficients and the probabilities of breaks. Surprisingly, the hierarchical SB-LS model locates the same change-points as the hierarchical SB-LSV model does in Figure 1.

5.1 Model Comparison

Some questions are raised from the above results. Are changes in volatility important for the Canada inflation series? Is the great moderation a feature of data? Can a duration dependent break probability improve the out-of-sample forecasting? Which of the specification provide the best fit to the data? Which model provides the best out-of-sample forecasts?

We use the log-marginal likelihoods to answer these questions. Let $\mathcal{M}_i$ denote a particular model. The marginal likelihood for $\mathcal{M}_i$ is defined as

$$p(Y_{1,T} \mid \mathcal{M}_i) = \prod_{t=1}^{T} p(y_t \mid Y_{1,t-1}, \mathcal{M}_i). \tag{19}$$

This decomposition shows that the marginal likelihood is intrinsically the comparison based on the out-of-sample forecasts[7], which automatically penalizes the over-parameterized model. An improvement on the marginal likelihood implies better forecasting ability over the whole sample.

The log-marginal likelihood is calculated as $\sum_{t=1}^{T} \log p(y_t \mid Y_{1,t-1}, \mathcal{M}_i)$. The one-period predictive likelihood $p(y_t \mid Y_{1,t-1}, \mathcal{M}_i)$ is calculated by using the data up to $t-1$ to estimate the model and plugging the value of y_t into the predictive density function. The first period is simply to use the prior as the posterior estimates.

Kass and Raftery (1995) propose to compare the model $\mathcal{M}_i$ and $\mathcal{M}_j$ by the log Bayes factors $\log(BF_{ij})$, where $BF_{ij} = \frac{p(Y_{1,T}|\mathcal{M}_i)}{p(Y_{1,T}|\mathcal{M}_j)}$ is the ratio of the marginal likelihoods. A positive value of $\log(BF_{ij})$ supports model $\mathcal{M}_i$ against $\mathcal{M}_j$. Quantitatively, Kass and Raftery (1995) suggest the results barely worth a mention for $0 \leq \log(BF_{ij}) < 1$; positive for $1 \leq \log(BF_{ij}) < 3$; strong for $3 \leq \log(BF_{ij}) < 5$; and very strong for $\log(BF_{ij}) \geq 5$.

Table 6 shows the log marginal likelihoods of different models. The autoregressive models,

$$y_t \mid \beta, \sigma, Y_{1,t-1} \sim \mathbf{N}(\beta_0 + \beta_1 y_{t-1} + \ldots + \beta_q y_{t-q}, \sigma^2) \tag{20}$$

[7] The marginal likelihood is a sequence of one-period ahead predictive likelihoods each of which have parameter uncertainty integrated out based on the respective posterior using data $Y_{1,t-1}$

are applied as benchmarks. The prior is set as Normal-Gamma $(\beta, \sigma^{-2}) \sim \mathbf{NG}(\underline{\beta}, \underline{H}^{-1}, \underline{\chi}/2, \underline{\nu}/2)$. The parameters $\underline{\beta} = \underline{0}_{(q+1)\times 1}, \underline{H} = I_{q+1}, \underline{\chi} = 1, \underline{\nu} = 2$. If $q = 1$, it is an AR(1) process and the values are the same as in the non-hierarchical SB-LSV model. An AR(2)-GARCH(1,1) model is also included to check that the structural break model is doing more than capturing neglected heteroskedasticity. This model combines (20) with the conditional variance specified as $\sigma_t^2 = \eta_0 + \eta_1(y_{t-1} - \beta_0 - \cdots - \beta_2 y_{t-3})^2 + \eta_2\sigma_{t-1}^2$.

The hierarchical SB-V, the hierarchical DDSB-LSV, the non-hierarchical models and the AR(1) model perform the worst and have log marginal likelihoods less than -155. The duration dependent break probability is not appropriate for the Canada inflation dynamics in the application. The AR(2) and the AR(3) model improve the performance by adding more lags.

The hierarchical SB-LS model has a log marginal likelihood of -140.4, which is larger than that of the AR(2) and the AR(3) model by $-140.4-(-144.2) = 3.8$ and $-140.4-(-144.7) = 4.3$, respectively. Maintaining the AR(1) dynamics but allowing the breaks in the regression coefficients improves the marginal likelihood more than adding extra lags. The optimal choice is the hierarchical SB-LSV model with the log marginal likelihood of -122.5, which dominates all other models strongly. The value of including the hierarchical prior in the SB-LSV model is substantial. The log-Bayes factor for this model versus the non- hierarchical break model is -122.5-(-158.4)=35.9.

The AR(2)-GARCH(1,1) model improves upon the homoskedastic AR(2) but is still strongly dominated by the hierarchical SB-LSV specification.

For a robustness check, we estimate each of the break models by assuming an AR(2) or AR(3) in each regime. For the hierarchical SB-LSV model, the log marginal likelihoods are -126.3 and -129.7 for the AR(2) or AR(3) case, which are less than that with the basic AR(1) model. The largest log marginal likelihood of the rest of the break models using an AR(2) or AR(3) specification is -144.4. The optimal model is still the hierarchical SB-LSV with AR(1) process in each regime. Hence, after controlling for the structural breaks, adding extra number of lags does not improve the marginal likelihood or forecasting in terms of the predictive likelihoods.

Based on this we conclude that the breaks in the regression and variance parameters shown in Figure 1 are significant. None of the partial break models match this model with breaks in both moments. In addition, the hierarchical prior provides a large improvement in the log-marginal likelihood.

The last column of Table 6 report the the root mean squared error(RMSE) of the predictive mean, computed as $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - E(y_t \mid Y_{1,t-1}, \mathcal{M}_i))^2}$. RMSE comparison is consistent with the results implied by the marginal likelihood. Modeling the structural changes in inflation translate into better out-of-sample point forecasts. For instance, there is about a 6% reduction in the RMSE in moving from the AR(2) model to the hierarchical SB-LSV model.

5.2 Inflation Targeting

In February 1991, the Bank of Canada and the Government of Canada issued a joint statement setting out a target path for inflation reduction, which is measured by the change of

12-month CPI index excluding food , energy and temporary effect of indirect taxes. The target is 3% by the end of 1992, 2.5% by the middle of 1994, and 2% by the end of 1995 with a range of $\pm 1\%$. In December 1993, the $1\% - 3\%$ plan was extended to the end of 1998. In 1998, it was further extended to 2001. In May 2001, it was extended to the end of 2006.[8] In 2006, it was extended to the end of 2011.[9]

We further investigate whether a linear model is sufficient to describe inflation dynamics by taking into account this important policy change. Two sample periods are used. The first one is from 1991Q2 till the end (2009Q4) and represents the whole period of the inflation targeting policy. The second one is from 1994Q1 till the end to show a more homogeneous policy regime, in which the inflation target range is $1\% - 3\%$. The linear models only use the subsample data and the data that are necessary to calculate the first period predictive density. For example, to calculate the predictive density at 1991Q2 from an AR(2) model in the first subsample, the data starts at 1990Q4, which is the two-period lag of 1991Q2. On the other hand, the hierarchical SB-LSV model uses the data from the first period(1961Q1), because it can automatically learn about structural change.

The model comparison is based on the predictive likelihood. Geweke and Amisano (2010) find the interpretation of the predictive likelihood is equivalent to the marginal likelihood if the initial data set is viewed as a training sample to form the priors. The priors are assumed the same as the previous model comparison in Table 6.

Table 7 shows the model comparison between linear models and the hierarchical SB-LSV model. The log-predictive likelihood is included along with the RMSE of the predictive mean. The last row of each subpanel is the naive forecast that the inflation next period is 2% annually.[10]

The two subsamples have the same implication for density forecasts as the full sample results. First, the SB-LSV model is still strongly supported by the predictive likelihood. So the learning of the hierarchical structure improves forecasting even after a well recognized break point. Second, the linear models outperforms our approach in point forecasts. We can see that the new monetary policy is associated with a simple forecasting rule, which produces good first moment forecasts. On the other hand, a nonlinear model can play an important role if incorporating higher moments is a concern to the policy makers.

5.3 Subjective Forecast

The econometrician is likely to have less information than a policy researcher does in the Bank of Canada. Exogenous information can be modeled in our framework with a simple revision. We consider 3 extra pieces of information in a revised version of the hierarchical SB-LSV model:

1. The inflation on 1991Q1 experiences a temporary increase after the introduction of GST (Goods and Services Tax). We assume its mean as 2%, which is consistent with some conjectures of the policy researchers before 1991.

[8] See Freedman (2001)

[9] See *Renewal of the Inflation-Control Target: Background Information* by the Bank of Canada, Nov 2006.

[10] The quarterly rate used in the calculation is $100 \times \frac{\log(1.02)}{4}$.

2. A deterministic structural change happens on 1991Q1, because of the introduction of the inflation targeting policy.

3. The dynamics of inflation follows a simple 2% rule starting from 1994Q1.

In order to encapsulate this extra information, the data is transformed and the model is revised as follows. First, construct $\tilde{y}_{1991Q1} = y_{1991Q1} - 100\log(1.02)$ to replace y_{1991Q1} by removing the expected inflation change on 1991Q1 with the introduction of the GST. Second, to impose a deterministic break at time $t = 1991Q1$, set $p(d_{1991Q1} = 1 \mid Y_{1961Q1,1991Q1}) = 1$ in the forward filtering step when sampling the regime allocation. Lastly, construct $\tilde{y}_t = y_t - 100\frac{\log(1.02)}{4}$ and $\tilde{x}_t = 0$ to replace y_t and x_t in Equation (4) for $t \geq 1994Q1$.

After the data transformation, we set the rest of $\tilde{y}_t$ and $\tilde{x}_t$ equal to y_t and x_t. We can simply apply our model to the transformed data and compute the predictive densities and means. Since $\tilde{y}_t$ and $\tilde{x}_t$ are the same as y_t and x_t for $t < 1991Q1$, we will focus on the predictive likelihood starting from the GST introduction in 1991Q1. The first panel of Table 8 includes these results while the second and third panel of the table correspond to the first and second panel of Table 7. The differences in the log-predictive likelihoods for the AR(1) in the first and second panel show the 1991Q1 observation to be very influential.

The top panel of Table 8 shows the revised hierarchical SB-LSV model outperforms the hierarchical SB-LSV model (the best model in Table 6) and the AR(1) model (the best linear model in Table 7). The log of the predictive Bayes factor is 3.0 in favor of the revised model and the predictive mean forecasts improve by 11% (from 0.430 to 0.382).

Even if we ignore the 1991Q1 outlier, the revised hierarchical SB-LSV model is still close to the original hierarchical SB-LSV model in density forecasts (second and third panel of Table 8). Meanwhile, the revised hierarchical SB-LSV model always provides better predictive mean forecasts than the original hierarchical SB-LSV model but cannot match the 2% rule between 1991Q1 and 1993Q4.

5.4 Exogenous Predictors

Are there other exogenous predictors useful in Canadian inflation forecasting? An anonymous attendant in the workshop on nowcasting and short-term forecasting at the Bank of Canada in 2011 suggested that the commodity price and the exchange rate (CAD/USD) are helpful predictors for the Canadian inflation.

The monthly commodity price data is from CANSIM (v52673496) and converted to continuously compounded quarterly growth rate. The monthly exchange rate data is from CITIBASE and transformed to quarterly exchange rate by taking the simple average of 3 months. The latest starting time of these data is 1972M1 (commodity price). All time series are truncated to have the same length from 1972Q1 to 2009Q4.

The exogenous variables augment the regressors x_t in Equation (4). For example, in a hierarchical SB-LSV model with AR(1) process in each regime, $x_t = (1, y_{t-1}, x_{\text{comm},t-1}, x_{\text{ex},t-1})$, where x_{comm} and x_{ex} are the commodity price growth rate and the exchange rate. We use the data at period $t-1$ to forecast the inflation y_t.

For the above model, the informative priors assume the 3rd and the 4th element on the diagonal of matrix A_0 as $A_{0,3,3} = \frac{1}{5}\frac{\sigma^2_{\text{infl}}}{\sigma^2_{\text{comm}}}$ and $A_{0,4,4} = \frac{1}{5}\frac{\sigma^2_{\text{infl}}}{\sigma^2_{\text{ex}}}$. $\sigma^2_{\text{infl}}, \sigma^2_{\text{comm}}$ and σ^2_{ex} are the sample variance of the inflation, the commodity price growth rate and the exchange rate. Since our approach is hierarchical, the hierarchical parameters are estimated. Meanwhile, we performed robustness checks and the results were not changed significantly with different priors. For example, if A_0 is set as proportional to an identity matrix as that of the original hierarchical SB-LSV model, our results are only different in the first decimal place for the marginal likelihood. The rest of the priors are the same as that of the original hierarchical SB-LSV model.

Table 9 shows the model comparison results after adding the exogenous variables. AR(p)+X means the regressors include p lags of y_t and the two exogenous variables. The top panel shows that the best linear model is AR(3)+X based on the marginal likelihood and AR(2)+X based on the predictive means. So adding exogenous variables is helpful for forecasting the inflation in the linear framework.

On the other hand, the lower panel of Table 9 shows that after controlling for structural instability, adding extra exogenous variables is strongly rejected by the original hierarchical SB-LSV model. Table 9 also shows that the structural change approach strongly dominates the linear models.

5.5 Computational Speed and Numeric Efficiency

To consider computational speed and numerical efficiency the Koop and Potter's (2007) model is compared with the hierarchical SB-LSV model. We assume each model has an AR(1) process in each regime, which is the optimal model in Table 6. These two approaches are not directly comparable in general since they assume slightly different data dynamics.

Both methods are applied to the whole sample of the Canadian inflation time series in the application. Each posterior sampler draws 6000 random samples and the first 1000 are discarded as burn-in samples. The CPU time used in Koop and Potter's (2007) model and our model are 1.1e^9 and 1.4e^8, respectively.[11] The relative numeric efficiency[12] (RNE) for the mean of the number of regimes implied by Koop and Potter's (2007) model is 49.2. Our approach is more efficient since the RNE is only 0.49. If we combine the computational time and the relative numeric efficiency together, Koop and Potter's (2007) model requires about $\frac{1.1e^9}{1.4^8}\frac{49.2}{0.49} \approx 789$ times more computation time than our approach in order to achieve the same numeric efficiency for estimating the posterior mean of the number of regimes.

6 Conclusion

This paper builds on existing structural change models to provide an improved approach to estimate and forecast time series with multiple change-points. This methodology obtains the analytic form of the predictive density by taking advantage of the conjugate prior for the

[11] Wall clock time is about 2 minutes for our model while it is between 15-20 minutes for the Koop and Potter's (2007) model.

[12] This is the ratio of the variance of the mean of a parameter relative to the variance from an iid sequence computed as $(1 + 2\sum_{i=1}^{\tau} \frac{\tau-i}{\tau}\hat{\rho}_i)$ where $\tau = 1000$ and $\hat{\rho}_i$ is the i-th sample autocorrelation computed from the posterior sample.

parameters that characterize each regime. The prior is modeled as hierarchical to exploit the information across regimes to improve forecast.

We discuss how to allow for breaks in the variance, the regression coefficients or both. Duration dependent break probabilities can be used and one extension assumes the regime duration has a Poisson distribution.

A new Markov Chain Monte Carlo sampler is introduced to draw the parameters from the posterior distribution efficiently. Each of the models sample the parameters jointly as one block which results in very good mixing.

We apply the model to a Canada inflation data. The best model is the hierarchical model which allows the breaks in the regression coefficients and the variance simultaneously. Modeling break results in improvements in density forecasts and point forecasts. We discuss the importance of inflation targets introduced in the 1990s and investigate if forecasts can be improved after this policy change.

Acknowledgements

We thank Gary Koop, Tom McCurdy and seminar participants at the Bank of Canada and CFE'11 for helpful comments. We thank the Social Sciences and Humanities Research Council of Canada for financial support.

7 Appendix

7.1 Hierarchical SB-LSV Model

1. Sampling $\pi^{(i)}, \underline{\beta}^{(i)}, \underline{H}^{(i)}, \underline{\chi}^{(i)}, \underline{\nu}^{(i)} \mid Y_T$ from the following proposal distribution.

   (a) Sample $\pi^{(i)} \mid K^{(i-1)} \sim \mathbf{B}(\pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})$ as the non-hierarchical model.

   (b) Sample $\underline{H}^{(i)} \mid \{\beta_k^{(i-1)}, \sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{W}(A_1, a_1)$

   (c) Sample $\underline{\beta}^{(i)} \mid \underline{H}^{(i)}, \{\beta_k^{(i-1)}, \sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{N}(m_1, (\tau_1 \underline{H}^{(i)})^{-1})$

   (d) Sample $\underline{\chi}^{(i)} \mid \underline{\nu}^{(i-1)}, \{\sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{G}(d_1/2, c_1/2)$

   (e) Sample $\underline{\nu}^{(i)} \mid \underline{\nu}^{(i-1)} \sim \mathbf{G}\left(\frac{\zeta}{\underline{\nu}^{(i-1)}}, \zeta\right)$

   with

$$
\begin{aligned}
m_1 &= \frac{1}{\tau_1}\left(\tau_0 m_0 + \sum_{i=1}^K \sigma_i^{-2}\beta_i\right) \\
\tau_1 &= \tau_0 + \sum_{i=1}^K \sigma_i^{-2} \\
A_1 &= \left(A_0^{-1} + \sum_{i=1}^K \sigma_i^{-2}\beta_i\beta_i' + \tau_0 m_0 m_0' - \tau_1 m_1 m_1'\right)^{-1} \\
a_1 &= a_0 + K \\
d_1 &= d_0 + \sum_{i=1}^K \sigma_i^{-2} \\
c_1 &= c_0 + K\underline{\nu}^{(i-1)}
\end{aligned}
$$

   Accept the whole set $\Psi^{(i)} = (\pi^{(i)}, \underline{\beta}^{(i)}, \underline{H}^{(i)}, \underline{\chi}^{(i)}, \underline{\nu}^{(i)})$ with probability

$$
\min\left\{1, \frac{p(\Psi^{(i)})}{p(\Psi^{(i-1)})}\frac{p(Y_T \mid \Psi^{(i)})}{p(Y_T \mid \Psi^{(i-1)})}\frac{p_{\text{prop}}(\Psi^{(i-1)})}{p_{\text{prop}}(\Psi^{(i)})}\right\}
$$

   where $p(\Psi)$ is the prior density and $p_{\text{prop}}(\Psi)$ is the proposal density.

2. Sample $\{s_t, \beta_t, \sigma_t\}_{t=1}^T \mid \Psi$ as the non-hierarchical structural break model.

7.2 Hierarchical SB-V Model

The predictive likelihood is computed as:

$$
p(y_t \mid s_t, Y_{t-1}, \beta) \propto \left(1 + \frac{(y_t - x_t'\beta)^2}{\hat{\chi}}\right)^{-\frac{(\hat{\nu}+1)}{2}}
$$

or

$$y_t \mid s_t, Y_{t-1}, \beta \sim t\left(x_t'\beta, \frac{\hat{\chi}}{\hat{\nu}}, \hat{\nu}\right)$$

with the mean $x_t'\beta$ and the variance $\frac{\hat{\chi}}{\hat{\nu}-2}$, where

$$\begin{aligned}
\hat{\chi} &= \underline{\chi} + E'_{t-s_t+1,t-1}E_{t-s_t+1,t-1} \\
\hat{\nu} &= \underline{\nu} + s_t - 1
\end{aligned}$$

$E_{t-s_t+1,t-1} = (e_{t-s_t+1}, \ldots, e_{t-1})'$ is the residual vector with $e_t = y_t - x_t'\beta$. The posterior sampling scheme is the following:

1. Sampling $\pi^{(i)}, \beta^{(i)}, \underline{\chi}^{(i)}, \underline{\nu}^{(i)} \mid Y_T$ from the following proposal distribution.
   (a) Sample $\pi^{(i)} \mid K^{(i-1)} \sim \mathbf{B}(\pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})$ as the non-hierarchical model.
   (b) Sample $\beta^{(i)} \mid \{\sigma_k^{(i-1)}\}_{k=1}^K, S_T \sim \mathbf{N}(\overline{\beta}, \overline{H}^{-1})$
   (c) Sample $\underline{\chi}^{(i)} \mid \underline{\nu}^{(i-1)}, \{\sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{G}(d_1/2, c_1/2)$
   (d) Sample $\underline{\nu}^{(i)} \mid \underline{\nu}^{(i-1)} \sim \mathbf{G}\left(\frac{\zeta}{\underline{\nu}^{(i-1)}}, \zeta\right)$

   with

$$\begin{aligned}
\overline{\beta} &= \overline{H}^{-1}\left(\underline{H}\,\underline{\beta} + \sum_{t=1}^T \frac{x_t y_t}{\sigma_t^2}\right) \\
\overline{H} &= \underline{H} + \sum_{t=1}^T \frac{x_t x_t'}{\sigma_t^2} \\
d_1 &= d_0 + \sum_{i=1}^K \sigma_i^{-2} \\
c_1 &= c_0 + K\underline{\nu}^{(i-1)}
\end{aligned}$$

   Accept the whole set $\Psi^{(i)} = (\pi^{(i)}, \beta^{(i)}, \underline{\chi}^{(i)}, \underline{\nu}^{(i)})$ with probability

$$\min\left\{1, \frac{p(\Psi^{(i)})}{p(\Psi^{(i-1)})} \frac{p(Y_T \mid \Psi^{(i)})}{p(Y_T \mid \Psi^{(i-1)})} \frac{p_{\text{prop}}(\Psi^{(i-1)})}{p_{\text{prop}}(\Psi^{(i)})}\right\}$$

   where $p(\Psi)$ is the prior density and $p_{\text{prop}}(\Psi)$ is the proposal density.

2. Sample $\{s_t, \sigma_t\}_{t=1}^T \mid \Psi$ similar to the non-hierarchical structural break model.

7.3 Hierarchical SB-LS Model

The predictive likelihood of $y_t \mid s_t, Y_{t-1}, \sigma$ is

$$y_t \mid s_t, Y_{t-1}, \sigma \sim N(x_t'\hat{\beta}, x_t'\hat{H}^{-1}x_t + \sigma^2)$$

where $\hat{\beta} = \hat{H}^{-1}(\underline{H}\underline{\beta} + \sigma^{-2}X'_{t-s_t+1,t-1}Y_{t-s_t+1,t-1})$ and $\hat{H} = \underline{H} + \sigma^{-2}X'_{t-s_t+1,t-1}X_{t-s_t+1,t-1}$. The posterior sampler is

1. Sampling $\pi^{(i)}, \underline{\beta}^{(i)}, \underline{H}^{(i)}, \underline{\sigma}^{(i)} \mid Y_T$ from the following proposal distribution.

   (a) Sample $\pi^{(i)} \mid K^{(i-1)} \sim \mathbf{B}(\pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})$ as the non-hierarchical model.

   (b) Sample $\underline{H}^{(i)} \mid \{\beta_k^{(i-1)}\}_{k=1}^K \sim \mathbf{W}(A_1, a_1)$

   (c) Sample $\underline{\beta}^{(i)} \mid \underline{H}^{(i)}, \{\beta_k^{(i-1)}\}_{k=1}^K \sim \mathbf{N}(m_1, (\tau_1\underline{H}^{(i)})^{-1})$

   (d) Sample $\sigma^{-2(i)} \mid \{\beta_k^{(i-1)}\}_{k=1}^K, S_T \sim G(\chi_1/2, \nu_1/2)$

   with

$$\begin{aligned}
m_1 &= \frac{1}{\tau_1}\left(\tau_0 m_0 + \sum_{i=1}^K \beta_i\right)\\
\tau_1 &= \tau_0 + K\\
A_1 &= \left(A_0^{-1} + \sum_{i=1}^K \beta_i\beta_i' + \tau_0 m_0 m_0' - \tau_1 m_1 m_1'\right)^{-1}\\
a_1 &= a_0 + K\\
\chi_1 &= \chi_0 + \sum_{t=1}^T (y_t - x_t\beta_t)^2\\
\nu_1 &= \nu_0 + T
\end{aligned}$$

   Accept the whole set $\Psi^{(i)} = (\pi^{(i)}, \underline{\beta}^{(i)}, \underline{H}^{(i)}, \sigma^{(i)})$ with probability

$$\min\left\{1, \frac{p(\Psi^{(i)})}{p(\Psi^{(i-1)})}\frac{p(Y_T \mid \Psi^{(i)})}{p(Y_T \mid \Psi^{(i-1)})}\frac{p_{\text{prop}}(\Psi^{(i-1)})}{p_{\text{prop}}(\Psi^{(i)})}\right\}$$

   where $p(\Psi)$ is the prior density and $p_{\text{prop}}(\Psi)$ is the proposal density.

2. Sample $\{s_t, \beta_t\}_{t=1}^T \mid \Psi$ similar the non-hierarchical structural break model.

7.4 Hierarchical DDSB-LSV Model

1. Sampling $\lambda^{(i)}, \underline{\beta}^{(i)}, \underline{H}^{(i)}, \underline{\chi}^{(i)}, \underline{\nu}^{(i)} \mid Y_T$ from the following proposal distribution.

   (a) Sample $\lambda^{(i)}$ by a random walk proposal distribution

(b) Sample $\underline{H}^{(i)} \mid \{\beta_k^{(i-1)}, \sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{W}(A_1, a_1)$

(c) Sample $\underline{\beta}^{(i)} \mid \underline{H}^{(i)}, \{\beta_k^{(i-1)}, \sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{N}(m_1, (\tau_1 \underline{H}^{(i)})^{-1})$

(d) Sample $\underline{\chi}^{(i)} \mid \underline{\nu}^{(i-1)}, \{\sigma_k^{(i-1)}\}_{k=1}^K \sim \mathbf{G}(d_1/2, c_1/2)$

(e) Sample $\underline{\nu}^{(i)} \mid \underline{\nu}^{(i-1)} \sim \mathbf{G}(\frac{\zeta}{\underline{\nu}^{(i-1)}}, \zeta)$

with

$$
\begin{aligned}
m_1 &= \frac{1}{\tau_1}\left(\tau_0 m_0 + \sum_{i=1}^K \sigma_i^{-2}\beta_i\right) \\
\tau_1 &= \tau_0 + \sum_{i=1}^K \sigma_i^{-2} \\
A_1 &= \left(A_0^{-1} + \sum_{i=1}^K \sigma_i^{-2}\beta_i\beta_i' + \tau_0 m_0 m_0' - \tau_1 m_1 m_1'\right)^{-1} \\
a_1 &= a_0 + K \\
d_1 &= d_0 + \sum_{i=1}^K \sigma_i^{-2} \\
c_1 &= c_0 + K\underline{\nu}^{(i-1)}
\end{aligned}
$$

Accept the whole set $\Psi^{(i)} = (\lambda^{(i)}, \underline{\beta}^{(i)}, \underline{H}^{(i)}, \underline{\chi}^{(i)}, \underline{\nu}^{(i)})$ with probability

$$
\min\left\{1, \frac{p(\Psi^{(i)})}{p(\Psi^{(i-1)})}\frac{p(Y_T \mid \Psi^{(i)})}{p(Y_T \mid \Psi^{(i-1)})}\frac{p_{\text{prop}}(\Psi^{(i-1)})}{p_{\text{prop}}(\Psi^{(i)})}\right\}
$$

where $p(\Psi)$ is the prior density and $p_{\text{prop}}(\Psi)$ is the proposal density.

2. Sample $\{s_t, \beta_t, \sigma_t\}_{t=1}^T \mid \Psi$ as the non-hierarchical structural break model.

Table 1: Summary statistics of Canada inflation

Mean	1.01
Min	-0.54
Max	3.12
Variance	0.69
Skewness	0.83
Excess Kurtosis	0.09

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 2: Posterior summary of the hierarchical SB-LSV model

	Prior		Posterior		
	Mean	0.95DI	Mean	Sd	0.95 DI
π	0.1	(0.003, 0.34)	0.04	0.02	(0.01, 0.07)
$\underline{\beta}_0$	0.0	(-3.08, 3.08)	0.82	0.20	(0.45, 1.23)
$\underline{\beta}_1$	0.0	(-3.08, 3.08)	-0.05	0.17	(-0.40, 0.29)
$\underline{H}_{00}$	1.0	(0.16, 2.52)	1.01	0.45	(0.35, 2.08)
$\underline{H}_{01}$	0.0	(-0.91, 0.91)	-0.01	0.37	(-0.77, 0.70)
$\underline{H}_{11}$	1.0	(0.16, 2.52)	1.30	0.58	(0.45, 2.68)
$\underline{\chi}$	1.0	(0.12, 2.79)	1.01	0.37	(0.43, 1.87)
$\underline{\nu}$	2.0	(0.05, 7.38)	6.04	2.27	(2.45, 11.1)

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 3: Posterior summary of the hierarchical DDSB-LSV model

	Prior		Posterior		
	Mean	0.95DI	Mean	Sd	0.95 DI
λ	50.0	(1.27, 184.4)	28.9	7.32	(14.70, 44.22)
$\underline{\beta}_0$	0.0	(-3.08, 3.08)	0.74	0.16	(0.44, 1.08)
$\underline{\beta}_1$	0.0	(-3.08, 3.08)	-0.03	0.15	(-0.33, 0.28)
$\underline{H}_{00}$	1.0	(0.16, 2.52)	1.00	0.35	(0.43, 1.80)
$\underline{H}_{01}$	0.0	(-0.91, 0.91)	-0.06	0.28	(-0.61, 0.48)
$\underline{H}_{11}$	1.0	(0.16, 2.52)	1.19	0.39	(0.56, 2.11)
$\underline{\chi}$	1.0	(0.12, 2.79)	3.86	1.76	(1.22, 8.26)
$\underline{\nu}$	2.0	(0.05, 7.38)	26.1	12.2	(8.03, 56.7)

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 4: Posterior summary of the hierarchical SB-V model

	Prior		Posterior		
	Mean	0.95DI	Mean	Sd	0.95 DI
π	0.1	(0.003, 0.34)	0.16	0.07	(0.05, 0.32)
β_0	0.0	(-1.96, 1.96)	0.16	0.05	(0.09, 0.28)
β_1	0.0	(-1.96, 1.96)	-0.05	0.17	(-0.40, 0.29)
$\underline{\chi}$	1.0	(0.12, 2.79)	0.96	0.42	(0.38, 1.99)
$\underline{\nu}$	2.0	(0.05, 7.38)	4.65	1.59	(2.22, 8.42)

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 5: Posterior summary of the hierarchical SB-LS model

	Prior		Posterior		
	Mean	0.95DI	Mean	Sd	0.95 DI
π	0.1	(0.003, 0.34)	0.03	0.01	(0.01, 0.07)
σ^2	-	(0.14, 19.7)	0.17	0.02	(0.14, 0.21)
$\underline{\beta}_0$	0.0	(-3.08, 3.08)	0.53	0.36	(-0.16, 1.21)
$\underline{\beta}_1$	0.0	(-3.08, 3.08)	-0.27	0.32	(-0.93, 0.33)
$\underline{H}_{00}$	1.0	(0.16, 2.52)	1.63	0.67	(0.62, 3.22)
$\underline{H}_{01}$	0.0	(-0.91, 0.91)	0.12	0.53	(-0.98, 1.14)
$\underline{H}_{11}$	1.0	(0.16, 2.52)	2.10	0.89	(0.76, 4.42)

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 6: Model comparison

	Log-marginal likelihood	RMSE
Hierarchical SB-LSV	**-122.5**	**0.457**
Hierarchical SB-V	-159.8	0.543
Hierarchical SB-LS	-140.4	0.471
Hierarchical DDSB-LSV	-155.6	0.479
Non-hierarchical SB-LSV	-158.4	0.514
Non-hierarchical SB-LSV($\pi = 0.01$)	-156.6	0.517
AR(1)	-160.8	0.541
AR(2)	-144.2	0.486
AR(2)-GARCH(1,1)	-141.8	0.497
AR(3)	-144.7	0.489

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 7: Model comparison after inflation targeting

	Log-predictive likelihood	RMSE
From the first quarter after inflation targeting(1991Q2-2009Q4)		
Hierarchical SB-LSV	**-36.4**	0.410
AR(1)	-46.5	0.365
AR(2)	-47.8	0.371
AR(3)	-48.2	0.377
2% target		**0.353**
From the first quarter of $1\% - 3\%$ target (1994Q1-2009Q4)		
Hierarchical SB-LSV	**-26.6**	0.367
AR(1)	-40.5	**0.346**
AR(2)	-41.4	0.352
AR(3)	-42.0	0.355
2% target		0.352

Canada quarterly inflation rate from 1961Q1-2009Q4. The SB-LSV model uses all data while the linear models only use part of the data that is associated with the inflation targeting period. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 8: Subjective forecasting

	Log-predictive likelihood	RMSE
From the first quarter of inflation targeting(1991Q1-2009Q4)		
Hierarchical SB-LSV	-40.2	0.430
Revised Hierarchical SB-LSV	**-37.2**	0.382
AR(1)	-53.1	0.481
2% target		**0.353**
From the first quarter after inflation targeting(1991Q2-2009Q4)		
Hierarchical SB-LSV	-36.4	0.410
Revised Hierarchical SB-LSV	**-35.7**	0.368
AR(1)	-46.5	0.365
2% target		**0.353**
From the first quarter of $1\% - 3\%$ target (1994Q1-2009Q4)		
Hierarchical SB-LSV	**-26.6**	0.367
Revised Hierarchical SB-LSV	-26.9	0.352
AR(1)	-40.5	**0.346**
2% target		0.352

Canada quarterly inflation rate from 1961Q1-2009Q4. There are 196 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901

Table 9: Model comparison: Commodity price and exchange rate as predictors

	Marginal likelihood	RMSE
AR(1)	-125.5	0.530
AR(2)	-120.6	0.542
AR(3)	-120.0	0.530
AR(4)	-120.1	0.531
AR(1)+X	-129.8	0.575
AR(2)+X	-118.3	0.502
AR(3)+X	-118.2	0.504
AR(4)+X	-118.3	0.508
Hierarchical SB-LSV	**-101.8**	**0.466**
Hierarchical SB-LSV+X	-109.7	0.527

Canada quarterly inflation rate from 1972Q2-2009Q4. There are 151 observations in total. The data is scaled by 100 to represent quarterly percentage change. Data Sources: IMDB (Integrated Meta Data Base) TABLE NUMBER: 3800003. Numbers: 1901. The monthly commodity price data is downloaded from CANSIM(v52673496). The monthly exchange rate data is downloaded from CITIBASE.

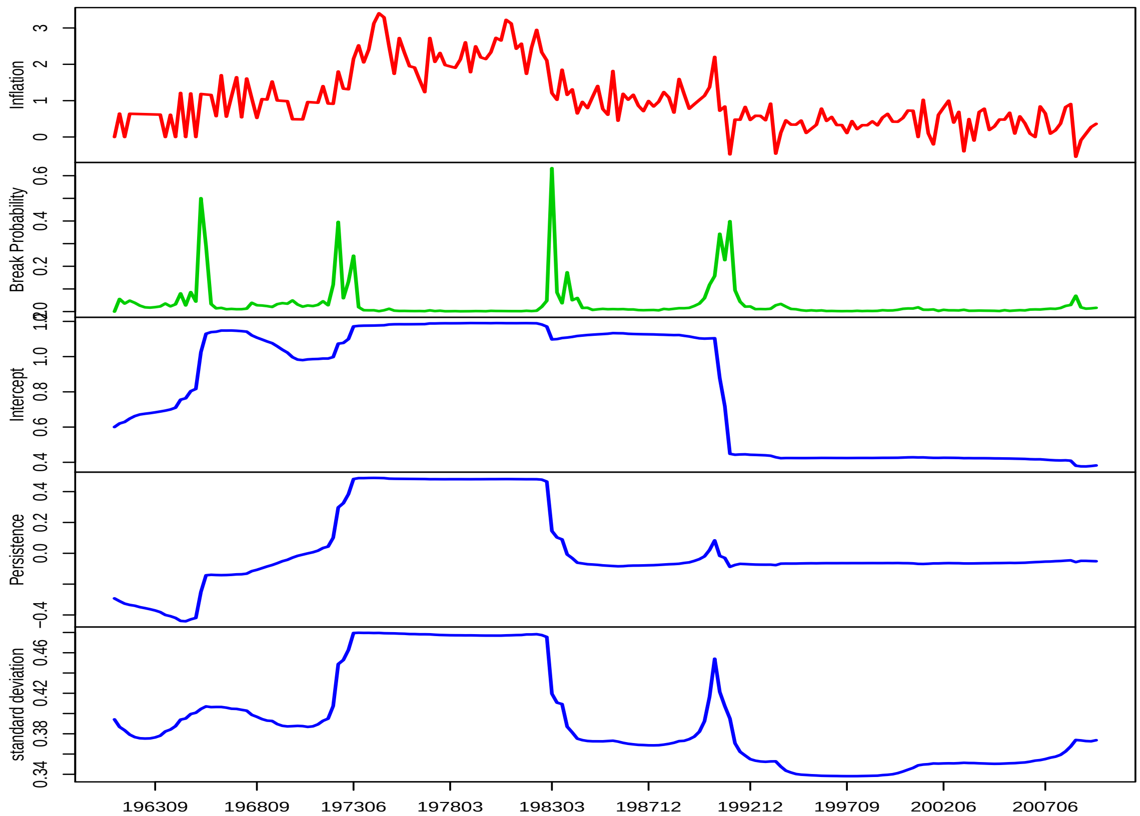

Figure 1: Posterior mean of the regression coefficients, the standard deviations and the break probabilities from the hierarchical SB-LSV model applied to a Canada quarterly inflation series from 1961Q1-2009Q4.

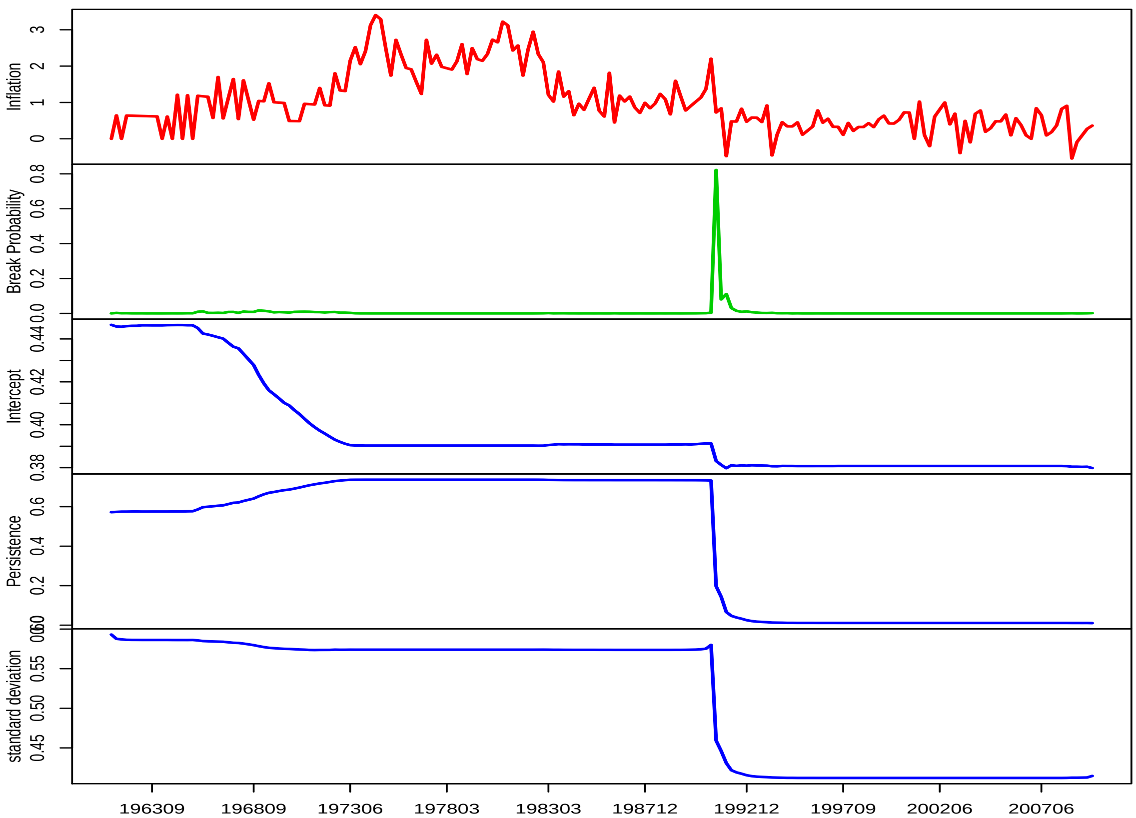

Figure 2: Posterior mean of the regression coefficients, the standard deviations and the break probabilities from the non-hierarchical SB-LSV model applied to a Canada quarterly inflation series from 1961Q1-2009Q4.

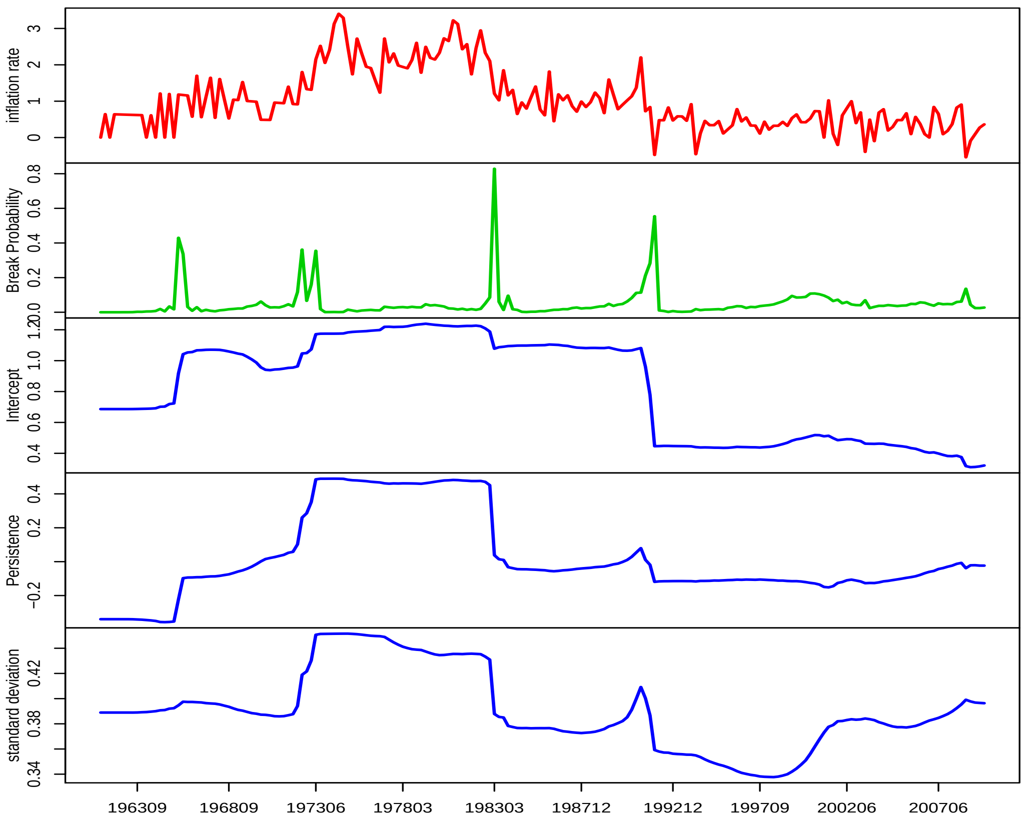

Figure 3: Posterior mean of the regression coefficients, the standard deviations and the break probabilities from the hierarchical DDSB-LSV model applied to a Canada quarterly inflation series from 1961Q1-2009Q4.

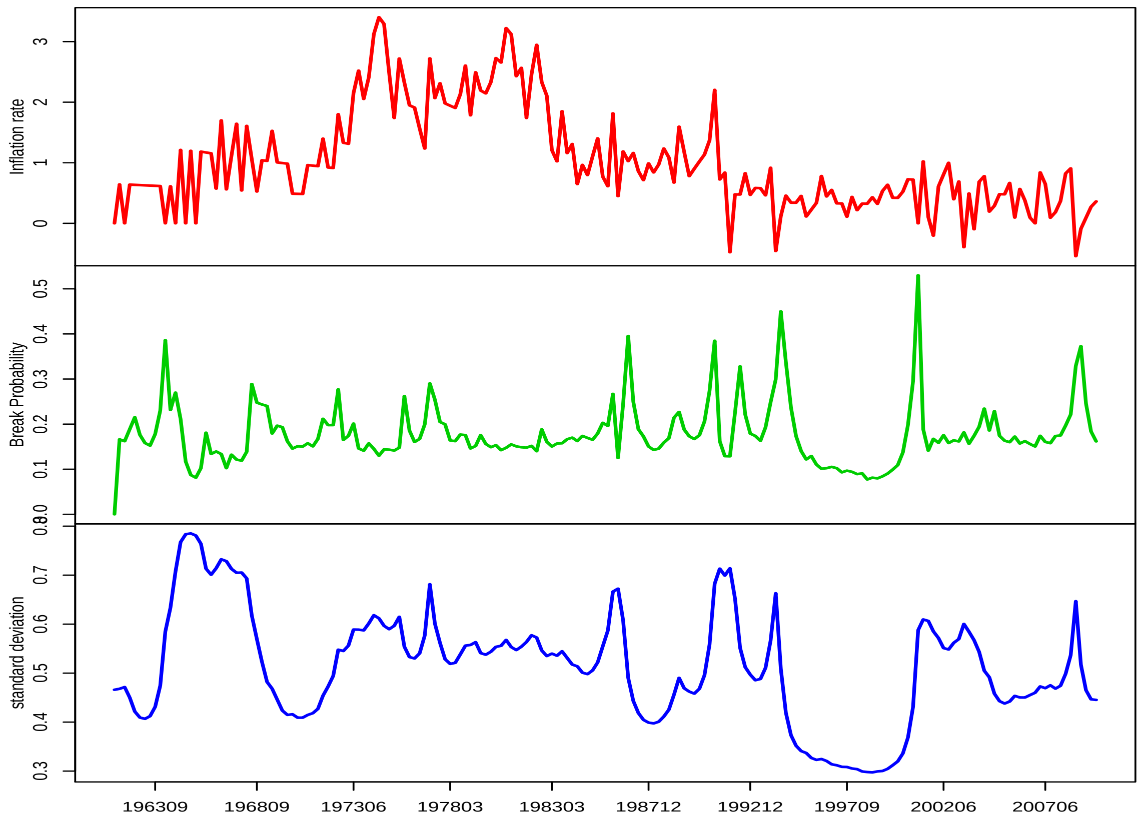

Figure 4: Posterior mean of the standard deviations and the break probabilities from the hierarchical SB-V model applied to a Canada quarterly inflation series from 1961Q1-2009Q4.

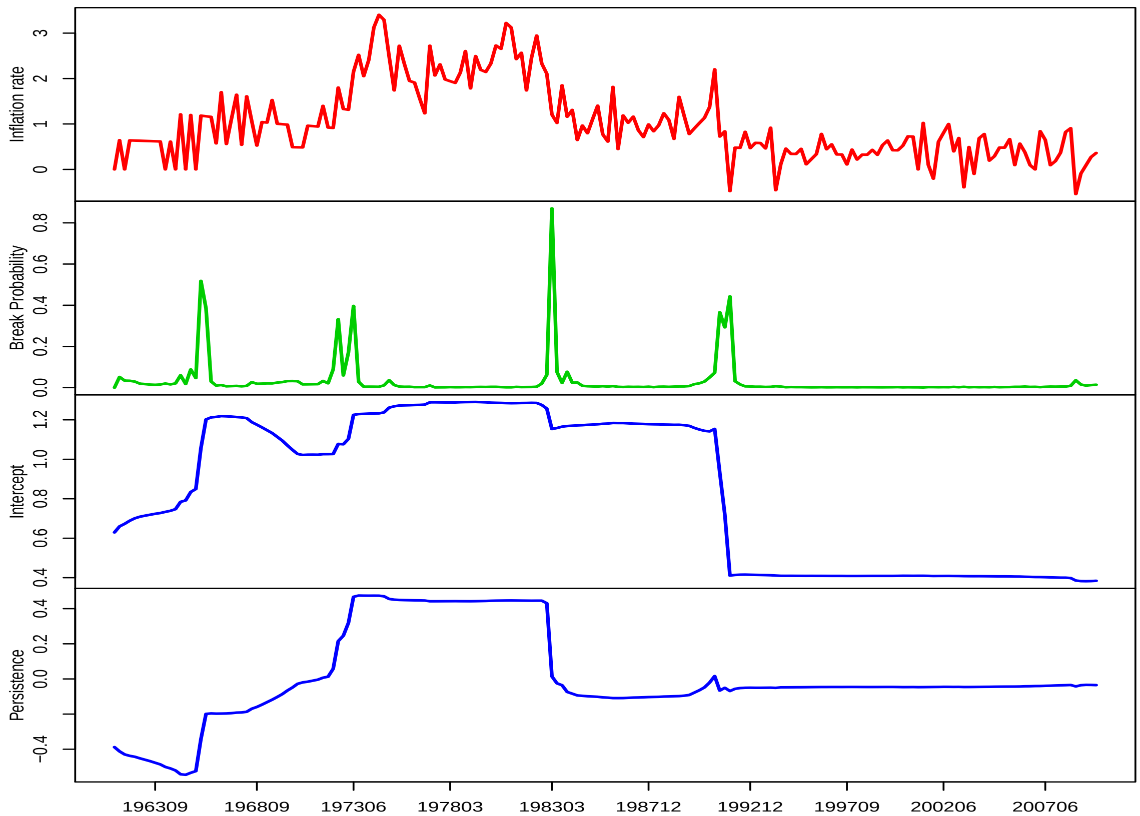

Figure 5: Posterior mean of the regression coefficients and the break probabilities from the hierarchical SB-LS model applied to a Canada quarterly inflation series from 1961Q1-2009Q4.

References

Casella, G. and Robert, C.P. Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1): 81, 1996.

Chib, S. Calculating posterior distributions and modal estimates in Markov mixture models* 1. *Journal of Econometrics*, 75(1):79–97, 1996.

Clark, Todd E. and McCracken, Michael W. Averaging forecasts from vars with uncertain instabilities. *Journal of Applied Econometrics*, 25(1):5–29, 2010.

Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4): 987–1007, 1982.

Engle, R.F. Estimates of the Variance of US Inflation Based upon the ARCH Model. *Journal of Money, Credit and Banking*, 15(3):286–301, 1983.

Freedman, Charles. Monetary policy formulation: The process in canada. *Business Economics*, pages 52–56, 2001.

Geweke, J. and Amisano, G. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 2010.

Geweke, John and Jiang, Yu. Inference and prediction in a multiple-structural-break model. *Journal of Econometrics*, 163(2):172 – 185, 2011.

Giordani, Paolo, Kohn, Robert, and van Dijk, Dick. A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, 137(1):112–133, 2007.

Kass, R.E. and Raftery, A.E. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

Koop, G. and Potter, S.M. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74(3):763–789, 2007.

Liu, Chun and Maheu, John M. Are there structural breaks in realized volatility? *Journal of Financial Econometrics*, 6(3):326–360, 2008.

Maheu, J.M. and Gordon, S. Learning, forecasting and structural breaks. *Journal of Applied Econometrics*, 23(5):553–583, 2008.

Maheu, J.M. and McCurdy, T.H. How useful are historical data for forecasting the long-run equity return distribution? *Journal of Business and Economic Statistics*, 27(1):95–112, 2009.

Pesaran, M.H., Pettenuzzo, D., and Timmermann, A. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73(4):1057–1084, 2006.

Stock, J.H. and Watson, M.W. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30, 1996.

Wang, J. and Zivot, E. A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business & Economic Statistics*, 18(3):374–386, 2000.