A New Structural Break Model with Application to Canadian Inflation Forecasting

John M. Maheu* Yong Song†

May 2013

Abstract

This paper develops an efficient approach to modelling and forecasting 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 modelled 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 a Canadian inflation 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: DeGroote School of Business, McMaster University, 1280 Main Street West, Hamilton, ON, Canada, L8S 4M4, 1-905-525-9140 x26198 and RCEA, Italy, maheujm@mcmaster.ca

†Economics Discipline Group, University of Technology, Sydney and RCEA, Italy, ysong1@gmail.com

1 Introduction

This paper develops an efficient Bayesian approach to modelling and forecasting time series data with an unknown number of change-points. The approach simplifies structural break analysis and reduces the computational burden relative to existing approaches in the literature. A conjugate prior is modeled as hierarchical to exploit information across regimes. Regime duration can be inferred from a fixed structural change probability or modelled as a Poisson distribution. Compared to existing time series models of Canadian inflation, including alternative structural break models, our specification produces superior density forecasts and point forecasts.

Accounting for structural instability in macroeconomic and financial time series models 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 provide a change-point model suitable for out-of-sample forecasting with the attractive features of the previous approaches but which is computationally less demanding. Parameters in each regime are drawn independently from a hierarchical prior. This allows for learning about the structural change process and its affect on model parameters and is convenient for computation. We introduce a new Markov chain Monte Carlo (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 better than that of a regular Gibbs sampling scheme as in Chib (1998). Lastly, different types of break dynamics including having breaks in the variance, the regression coefficients or both are nested in this framework.

We extend Maheu and Gordon (2008) and Maheu and McCurdy (2009) in four directions. First, a conjugate prior for the parameters that characterize each regime is adopted. Conditional on this prior and the time-invariant parameters, the predictive density has a closed form, which reduces the computational burden compared to Maheu and Gordon (2008).[1] Second, a hierarchical structure for the conjugate prior is introduced to allow pooling of information across regimes, as in Pesaran et al. (2006). Third, we show how to model the regime duration as a Poisson distribution, which implies duration dependent break probabilities. Lastly, we show how to produce the smoothed distribution of the change-points.

Koop and Potter (2007) also model regime durations but 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. Koop and Potter (2007)

[1] Maheu and Gordon (2008) assume a conditional conjugate prior and use Gibbs sampling to compute the predictive density. The computational benefits of our approach require a conjugate prior and simplified structural break process compared with other models in the literature.

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 increases computation time.

Different versions of our model are applied to a Canadian 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 the usefulness of the data prior to this date in forecasting after 1991. We also show that incorporating exogenous subjective information from policy changes into our model can further improve forecasts.

The log-predictive likelihood is used as the criterion 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 parametrization. A sub-sample analysis is consistent with the results from the full sample. We also show how to incorporate exogenous information or variables in our framework for out-of-sample forecasting. A posterior analysis based on the optimal model identifies 4 major change-points in the Canadian inflation dynamics. The duration dependent break probability is not a significant feature of the data.

The paper is organized as follows. Section 2 introduces the model and a MCMC 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, only in the regression coefficients or independent breaks in both. A duration dependent break probability is also modeled by assuming a Poisson distribution for the regime durations. Section 5 applies the model to a Canadian 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. We discuss how to compute the posterior density of θ for each regime as well as the predictive density. The following subsection 2.1 then gives specifics and shows how to integrate out all possible structural break points (regimes) to form predictions.

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 parameter 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 modeling of regime durations directly, which we discuss later.

[2]Maheu and Gordon did not consider the smoothed distribution of breaks and only focused on the filtered distribution of change points.

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 for one more period while the second case is the occurrence of a structural change, with a new draw of the parameter θ occurring 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 for 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.

We assume a constant structural break probability $\pi \in (0, 1)$. 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 that characterize the data density at time t. The full specification we focus on is the following.

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

The discrete measure concentrated at the mass point $(\beta_{t-1}, \sigma_{t-1}^{-2})$ is denoted as $\delta_{(\beta_{t-1}, \sigma_{t-1}^{-2})}$. 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. 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.[3] If there is no break ($d_t > 1$), all parameters are the same as those in the previous period.

[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 rate and $\underline{\nu}/2$ is the degree of freedom. Its prior mean is $\frac{\underline{\nu}}{\underline{\chi}}$ and prior variance is $\frac{2\underline{\nu}}{\underline{\chi}^2}$. It also implies that 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 $\mathbf{N}(\underline{\beta}, \underline{H}^{-1}\sigma_t^2)$.

2.1 Estimation and Inference

Due to the conjugacy of the prior, the posterior distribution of the parameters that 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 increases 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 in the last period.

The posterior distribution of θ_t given d_t is used to compute the predictive density for y_{t+1}

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

which is a Student-t distribution. 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 for the predictive density. In this case the posterior is replaced by the prior and we obtain the following predictive density,

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

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.

To sample $D_{1,T}$ a forward-filtering pass is made followed by a backward-sampling method. The forward-filtering pass is conducted as follows.

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

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 density function which we have derived above. 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},

$$p(y_{t+1} \mid \pi, Y_{1,t}) = \sum_{j=1}^{t+1} p(y_{t+1} \mid d_{t+1} = j, Y_{1,t}) p(d_{t+1} = j \mid \pi, Y_{1,t}). \tag{10}$$

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

Following this, the backward-sampling method samples the vector of durations $D_{1,T} = (d_1 \dots, d_T)$ jointly as follows.

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 for 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 steps 2 and 3 until the first period $t = 1$.

Using the conjugate prior has several advantages. First, the computational burden is negligible compared to the approach of Maheu and Gordon (2008) with non-conjugate priors.[4] 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 is 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 according to Casella and Robert (1996), because the parameters $\Theta_{1,T} = \{\theta_t\}_{t=1}^T$ are integrated out which improves the accuracy of posterior estimates.

In the case of the constant break probability, the prior of the break probability π is specified 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.

[4] Not all beliefs can be represented conveniently with the conjugate prior. For instance, stationary conditions are difficult to impose in the hierarchical model of Section 3.

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

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

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

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

and otherwise set $\pi^{(i)}$ 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 that there is a break and the data are in a new regime. Then set $s_\tau = 2$ at this break point, and iterate until the last period with $s_T = K$.[5]

   (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 $i = 1, \dots, K$. Then

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

   with

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

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

   and $X_i = (x_{t_0}, \dots, x_{t_1})'$ and $Y_i = (y_{t_0}, \dots, y_{t_1})'$, 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. The duration of the ith regime is $D_i = t_1 - t_0 + 1$.

The Markov chain is run for $N_0 + N$ iterations and the first N_0 iterations are discarded as burn-in iterations. The rest of the parameters draws, $\left\{\pi^{(i)}, \Theta_{1,T}^{(i)}, D_{1,T}^{(i)}\right\}_{i=1}^N$, are used for posterior inference and forecasting. 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

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

density for time $T+1$ by averaging the MCMC draws of π using (10) or based on the draws of π and $D_{1,T}$ following

$$\hat{p}(y_{T+1} \mid Y_{1,T}) = \frac{1}{N}\sum_{i=1}^{N}\left\{p(y_{T+1} \mid d_{T+1} = d_T^{(i)} + 1, Y_{1,T})(1-\pi^{(i)}) + p(y_{T+1} \mid d_{T+1} = 1, Y_{1,T})\pi^{(i)}\right\}. \quad (11)$$

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 for 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 provide information for the regime after it, then the duration d_t itself is not sufficient for an analytic 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.[6] 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 the conjugate prior is used 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:

$$\begin{aligned}
&\pi \sim \mathbf{B}(\pi_a, \pi_b),\ \underline{\beta}, \underline{H} \sim \mathbf{NW}(m_0, \tau_0^{-1}, A_0, a_0),\ \underline{\chi} \sim \mathbf{G}(d_0/2, c_0/2),\\
&\underline{\nu} \sim \mathbf{Exp}(\rho_0),\\
&d_t = \begin{cases} d_{t-1}+1 & \text{w.p. } 1-\pi,\\ 1 & \text{w.p. } \pi,\end{cases}\\
&(\beta_t, \sigma_t^{-2}) \sim \mathbf{1}(d_t = 1)\mathbf{NG}(\underline{\beta}, \underline{H}^{-1}, \underline{\chi}/2, \underline{\nu}/2) + \mathbf{1}(d_t > 1)\delta_{(\beta_{t-1}, \sigma_{t-1}^{-2})},\\
&y_t \mid \beta_t, \sigma_t, Y_{1,t-1} \sim \mathbf{N}(x_t'\beta_t, \sigma_t^2).
\end{aligned} \quad (12)$$

[6] Recall that after a break occurs ($d_t = 1$) a new parameter (β_t, σ_t^{-2}) is drawn from the prior distribution in (4).

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

$$m_1 = \frac{1}{\tau_1}\left(\tau_0 m_0 + \sum_{i=1}^K \sigma_i^{*-2}\beta_i^*\right), \quad \tau_1 = \tau_0 + \sum_{i=1}^K \sigma_i^{*-2}, \tag{13}$$

$$A_1 = \left(A_0^{-1} + \sum_{i=1}^K \sigma_i^{*-2}\beta_i^*\beta_i^{*\prime} + \tau_0 m_0 m_0' - \tau_1 m_1 m_1'\right)^{-1}, \quad a_1 = a_0 + K. \tag{14}$$

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

$$\underline{\chi} \mid \underline{\nu}, \{\sigma_i^*\}_{i=1}^K \sim \mathbf{G}(d_1/2, c_1/2), \tag{15}$$

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 chain 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 the 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 π 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, sample jointly from the full posterior $\Theta_{1,T}, D_{1,T}, \Psi | Y_{1,T}$, which results in a well-mixing Markov chain. The details are given in Section A.1.

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 (10) can be used. This latter approach integrates out the durations.

4 Extensions

This section extends the model while preserving the two assumptions: the conjugate prior and the conditional independence between the parameters in each regime. Up to this point we have assumed that breaks affect both the conditional mean and variance at the same time. The extensions allow for only breaks in the variance, only breaks in the regression coefficients or independent breaks in both. Another extension allows for 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 is drawn from a hierarchical prior. The model is

$$\begin{aligned}
\pi &\sim \mathbf{B}(\pi_a,\pi_b),\ \underline{\chi}\sim\mathbf{G}(d_0/2,c_0/2),\ \underline{\nu}\sim\mathbf{Exp}(\rho_0),\ \beta\sim\mathbf{N}(\underline{\beta},\underline{H}^{-1}),\\
d_t &= \begin{cases} d_{t-1}+1 & \text{w.p. } 1-\pi,\\ 1 & \text{w.p. } \pi,\end{cases}\\
\sigma_t^{-2} &\sim \mathbf{1}(d_t=1)\mathbf{G}(\underline{\chi}/2,\underline{\nu}/2)+\mathbf{1}(d_t>1)\delta_{\sigma_{t-1}^{-2}},\\
y_t\mid\beta,\sigma_t,Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta,\sigma_t^2).
\end{aligned}\tag{16}$$

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, $p(\theta_t\mid d_t,\beta,Y_{1,t})=p(\theta_t\mid d_t,\beta,Y_{t-d_t+1,t})$ still 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 has the following steps. First, randomly draw $\Psi \mid Y_{1,T}$ using a tailored proposal distribution and accept it with the probability implied by the Metropolis-Hastings algorithm. Second, 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 Section A.2.

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 labelled 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:

$$
\begin{aligned}
\pi &\sim \mathbf{B}(\pi_a, \pi_b),\ \underline{\beta}, \underline{H} \sim \mathbf{NW}(m_0, \tau_0^{-1}, A_0, a_0),\ \sigma^{-2} \sim \mathbf{G}(\underline{\chi}/2, \underline{\nu}/2), \\
d_t &= \begin{cases} d_{t-1}+1 & \text{w.p. } 1-\pi, \\ 1 & \text{w.p. } \pi, \end{cases} \\
\beta_t &\sim \mathbf{1}(d_t = 1)\mathbf{N}(\underline{\beta}, \underline{H}^{-1}) + \mathbf{1}(d_t > 1)\delta_{\beta_{t-1}}, \\
y_t \mid \beta_t, \sigma, Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta_t, \sigma^2).
\end{aligned} \tag{17}
$$

The posterior sampler consists of the following steps. First, randomly draw the time-invariant parameter $\Psi = (\pi, \underline{\beta}, \underline{H}, \sigma)$ from its posterior distribution using a MCMC sampler. Second, sample 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 given in Section A.3

4.3 Independent Breaks in Regression Coefficients and Variance

It is possible that breaks in the regression coefficients are independent of breaks in the variance. This can be considered in our framework but all the parameters cannot be sampled in one single block as before. Nevertheless, we are still able to use the durations as the state variable conditional on the time-invariant parameters, except that we have two sets of duration variables, one for the regression coefficients and the other for the variance, and each set is sampled conditional on the other one. The model is the following.

$$
\begin{aligned}
&\pi_\beta \sim \mathbf{B}(\pi_{\beta,a}, \pi_{\beta,b}),\ \pi_\sigma \sim \mathbf{B}(\pi_{\sigma,a}, \pi_{\sigma,b}), \\
&\underline{\beta}, \underline{H} \sim \mathbf{NW}(m_0, \tau_0^{-1}, A_0, a_0),\ \underline{\chi} \sim \mathbf{G}(d_0/2, c_0/2),\ \underline{\nu} \sim \mathbf{Exp}(\rho_0), \\
&d_{\beta,t} = \begin{cases} d_{\beta,t-1}+1 & \text{w.p. } 1-\pi_\beta, \\ 1 & \text{w.p. } \pi_\beta, \end{cases} \\
&\beta_t \sim \mathbf{1}(d_{\beta,t} = 1)\mathbf{N}(\underline{\beta}, \underline{H}^{-1}) + \mathbf{1}(d_{\beta,t} > 1)\delta_{\beta_{t-1}},
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
d_{\sigma,t} &= \begin{cases} d_{\sigma,t-1}+1 & \text{w.p. } 1-\pi_\sigma, \\ 1 & \text{w.p. } \pi_\sigma, \end{cases} \\
\sigma_t^{-2} &\sim \mathbf{1}(d_{\sigma,t}=1)\mathbf{G}(\underline{\chi}/2, \underline{\nu}/2) + \mathbf{1}(d_{\sigma,t}>1)\delta_{\sigma_{t-1}^{-2}}, \\
y_t \mid \beta_t, \sigma_t, Y_{1,t-1} &\sim \mathbf{N}(x_t'\beta_t, \sigma_t^2).
\end{aligned}
$$

The breaks of β_t and σ_t are independent. We use $d_{\beta,t}$ and $d_{\sigma,t}$ to represent the duration for the regression coefficients β_t and the volatility σ_t, respectively. This model is labeled as hierarchical SB-LS-V model. The posterior sampling algorithm is in Section A.4.

4.4 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[7]. The Poisson distribution function is $P(\text{Duration} = d \mid \lambda) = e^{-\lambda}\frac{\lambda^{(d-1)}}{(d-1)!}$, where $d \geq 1$ and the duration is a discrete count variable. 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 and follows the previous discussion.

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.

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

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 Section A.5.

5 Application to Canadian Inflation

The model is applied to Canadian quarterly inflation. The data is constructed from the quarterly CPI, which is downloaded from CANSIM[8]. The quarterly inflation rate is calculated as the log difference of the CPI data and scaled by 100. It starts from 1961Q1 and ends at 2012Q2 with 206 observations. The top panel of Figure 1 plots the data.

The hierarchical models used are SB-LSV, SB-V, SB-LS, DDSB-LSV and SB-LS-V models. Two non-hierarchical SB-LSV models are also applied, one estimates the break probability π and the other fixes $\pi = 0.01$. For all the structural break models, we assume that each regime has an AR(q) representation and estimate $q = 1, 2$ and 3 for each model.

For comparison we include Koop and Potter's (2007) model (KP) of structural change. The KP model links the regression coefficients and the log-variances of adjacent regimes through a random walk process. For instance, if a break occurs the regime parameter changes according to $\theta_t = \theta_{t-1} + \varepsilon_t$, where ε_t is a normal innovation. In our case, θ_t is independent of θ_{t-1} conditional on the hierarchical prior.

The final comparison specifications which assume no-breaks include homoskedastic linear autoregressive (AR) models and AR-GARCH specifications to capture heteroskedasticity.

5.1 Priors

The prior of the hierarchical SB-LSV model is:

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

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 degrees of freedom parameter centered at 2 and a rate centered at 1.

The non-hierarchical SB-LSV model fixes the parameters of the priors at $\underline{\beta} = (0, \cdots, 0)', \underline{H} = I, \underline{\chi} = 1, \underline{\nu} = 2$, which are the prior means of the hierarchical SB-LSV model. The break probability π has the same prior as that of the hierarchical model, which is $\mathbf{B}(1, 9)$.

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 specified as exponential with a mean equal to 50. The other priors are the same as for the

[8]Table number: 3800003; table title: GDP indexes; data sources: IMDB (Integrated Meta Data Base) numbers: 1901; series title: Canada; implicit price indexes 2002=100; personal expenditure on consumer goods and services series; frequency: quarterly.

hierarchical SB-LSV model. For simplicity, the first period $t = 1$ is assumed to be the first period of its regime.

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 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 in the hierarchical SB-LSV model.

For the hierarchical SB-LS model, the prior of the inverse of the variance is $\sigma^{-2} \sim \mathbf{G}(0.5, 1)$. The values of the rate and the degrees 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.

For the hierarchical SB-LS-V model, the priors of the break probabilities are $\pi_\beta \sim \mathbf{B}(1, 9)$ and $\pi_\sigma \sim \mathbf{B}(1, 9)$. The other priors are the same as in the hierarchical SB-LSV model.

For the KP model, the prior is set the same as in Koop and Potter (2007). Their prior is informative but covers a wide range of reasonable parameter values. It reflects an assumption of small differences between parameters of adjacent regimes. We do not repeat their notations here in order to avoid any confusion.

Several time invariant autoregressive models are estimated as benchmarks:

$$y_t \mid \beta, \sigma, Y_{1,t-1} \sim \mathbf{N}(\beta_0 + \beta_1 y_{t-1} + \ldots + \beta_q y_{t-q}, \sigma^2), \tag{19}$$

with various q. 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 $\underline{\beta} = (0, \cdots, 0), \underline{H} = I, \underline{\chi} = 1, \underline{\nu} = 2$, which are the prior mean from the hierarchical SB-LSV model.

Autoregressive models coupled with GARCH (AR(q)-GARCH(1,1)) with no breaks are also included to check that the structural break model is doing more than capturing neglected heteroskedasticity. This model combines an AR structure for the conditional mean with the conditional variance specified as $\sigma_t^2 = \eta_0 + \eta_1(y_{t-1} - \beta_0 - \cdots - \beta_q y_{t-q})^2 + \eta_2 \sigma_{t-1}^2$ and assumes normal innovations. The prior for the regression coefficients are $\beta \sim \mathbf{N}(0, I)$, which is the same as for the hierarchical SB-V model. The priors for the volatility coefficients are $\eta_0 \sim \mathbf{N}(0, 1)\mathbf{1}(\eta_0 > 0)$, $\eta_1 \sim \mathbf{N}(0.05, 1)\mathbf{1}(\eta_1 \geq 0)$ and $\eta_2 \sim \mathbf{N}(0.9, 1)\mathbf{1}(\eta_2 \geq 0)$ with stationary restriction $\eta_1 + \eta_2 < 1$.

5.2 Forecast Performance

In this section we compare models using the full sample of data and consider density forecasts using the marginal likelihood and point forecasts using the predictive mean.

Let $\mathcal{M}_i$ denote a particular model. The marginal likelihood for $\mathcal{M}_i$ is defined as

$$p(Y_{1,T} \mid \mathcal{M}_i) = \prod_{t=1}^{T} p(y_t \mid Y_{1,t-1}, \mathcal{M}_i). \tag{20}$$

This decomposition shows that the marginal likelihood is intrinsically the out-of-sample density forecast record[9] of a model. It automatically penalizes the over-parametrized models

[9] 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 $Y_{1,t-1}$

since parameter uncertainty is integrated out. An increase in the marginal likelihood implies better density forecasts over the sample. Kass and Raftery (1995) propose to compare the models $\mathcal{M}_i$ 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.[10]

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 uses the prior as the posterior estimates.

We also report the root mean squared forecasting error(RMSFE), which is computed as $\text{RMSFE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}$, where $\hat{y}_t$ is the predictive mean of y_t. As in the predictive likelihoods, only data up to and including $t-1$ is used to estimate the predictive mean for y_t. A two-sided Diebold and Mariano (1995) (DM) test is included to assess the statistical significance of model forecast errors $e_t = \hat{y}_t - y_t$, assuming a quadratic loss function. Since the DM test is pairwise, we compare the optimal model implied by the marginal likelihoods, which is the hierarchical SB-LSV AR(2) from Table 1, to other specifications. We also report the mean absolute scaled error proposed by Hyndman and Koehler (2006), which is computed as $\frac{\frac{1}{T}\sum_{t=1}^{T}|e_t|}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}$, and label it as HK-MASE. A smaller value of HK-MASE means better point forecasts.

Table 1 reports the log-marginal likelihoods, log-Bayes factors, RMSFE and p-values for the DM test and the HK-MASE statistic. The Log-Bayes factors and DM tests are for the optimal model (hierarchical SB-LSV AR(2)) against every other model.

According to Table 1 the best model is the hierarchical SB-LSV AR(2) model based on all three measures which gauge the accuracy of density forecasts (Log ML) and point forecasts (RMSFE, HK-MASE). In terms of the marginal likelihoods, the hierarchical SB-LSV AR(2) model is strongly favored by the data. The second best model, which is the hierarchical SB-LSV AR(3) model, has a marginal likelihood of -137.5 , which is 7.3 ($137.5 - 130.8 = 7.3$) less than the optimal model. If we look at models other than the hierarchical SB-LSV models, the smallest log-Bayes factor between the hierarchical SB-LSV AR(2) model and its competitors is larger than 14. This shows that the hierarchical SB-LSV specification dominates other linear and nonlinear models. The hierarchical SB-LSV AR(2) model has the smallest RMSFE and is often significantly better than the alternatives. This model also produces the best forecasts in terms of the HK-MASE statistic.

The hierarchical structure is very important as the nonhierarchical SB-LSV versions have much lower marginal likelihoods. The hierarchical model allows for learning where a new parameter is most likely to be after a break has occurred. We found no evidence of structural change from the nonhierarchical SB-LSV models and these models are generally inferior to the benchmark AR specifications. The hierarchical SB-LSV AR(2) model is substantially

[10] 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 $\log(BF_{ij}) \geq 5$.

better than the KP model in terms of the marginal likelihoods. The log-Bayes factors are 14 or more in favour of our parametrization versus the KP-AP(p), $p = 1, 2, 3$, while the point forecasts are similar.

All the extensions such as the hierarchical SB-V, the hierarchical SB-LS, the hierarchical DDSB-LSV and the hierarchical SB-LS-V model, are not supported by Canadian inflation data. The performance of these extensions is comparable to, or even worse, than the linear models. Among them, the hierarchical SB-LS-V model performs the worst and clearly indicates that breaks in the regression coefficient and variance occur at the same time. Finally, the AR(2)-GARCH(1,1) model improves upon the homoskedastic AR(2) and is better than many of the extensions, but it is still strongly dominated by all the hierarchical SB-LSV specifications.

5.3 Sub-sample Forecast Performance

As a robustness check, Table 2 reports forecast results for various sub-samples using the best models found in Table 1. The data before 1977Q4 is used as a training sample and the predictive likelihood, RMSFE and HK-MASE are displayed based on the rest of the out-of-sample data.[11] From the top panel, the hierarchical SB-LSV AR(2) model is still the optimal model and strongly supported by the data based on the predictive likelihood. The hierarchical DDSB-LSV AR(1) has the smallest RMSFE and HK-MASE. We further decompose the sub-sample into two parts, before and after inflation targeting (1991Q1) began. The middle and the bottom panel of Table 2 show that the conclusion from the top panel is not affected by different sub-samples.

5.4 Forecasts in the Presence of 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 the temporary effect of indirect taxes. The target was 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.[12] In 2006, it was extended to the end of 2011.[13]

Perhaps our structural break model is capturing nothing more than publicly announced policy changes. We further investigate whether a linear model is sufficient to describe inflation dynamics by taking into account these important policy changes. Two sample periods are used. The first one is from 1991Q2 until the end (2012Q2) and represents the whole period of the inflation targeting policy. The second one is from 1994Q1 until the end to show a more homogeneous policy regime, in which the inflation target range is $1\% - 3\%$. The linear models only use the sub-sample data and the data that are necessary to calculate the first period predictive density. For example, to calculate the predictive density at 1991Q2 from

[11] 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.

[12] See Freedman (2001)

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

an AR(2) model in the first sub-sample, the data start 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 priors are assumed the same as the previous model comparison in Table 1.

Table 3 shows the comparison between linear models and the hierarchical SB-LSV model. The log-predictive likelihood is included along with the RMSFE, the DM test p-values and the HK-MASE. The last row of each sub-panel is the naive forecast that the inflation next period is 2% annually.[14]

The two sub-samples have the same implication for density forecasts as the full sample results. The hierarchical SB-LSV model is still strongly supported by the predictive likelihood. The hierarchical structure improves forecasts even after a well recognized break point. On the other hand, the linear models outperform our approach in point forecasting, although not significantly. It is not surprising that the best point forecasting approach is the 2% rule or the AR(1) model, as the Bank of Canada quickly established credibility for its inflation policy in the market.

5.5 Subjective Forecasts

An advantage of our approach is that subjective information can easily be introduced into the model to produce forecasts. Exogenous information can be modeled in our framework with some simple revisions. We consider 3 extra pieces of information for constructing a revised version of the hierarchical SB-LSV model and label the model as the subjective hierarchical SB-LSV model. The information is the following.

1. Inflation in 1991Q1 experiences a temporary increase after the introduction of the GST (Goods and Services Tax). We assume that the increase is 2%, which is consistent with some conjectures from policy researchers before 1991.

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

3. The dynamics of inflation follow a simple 2% rule starting from 1994Q1 and therefore the forecast is always 2%.

In order to incorporate this extra information, the data are 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 set $\tilde{x}_t = 0$ to replace y_t and x_t in Equation (12) for $t \geq 1994Q1$.

After the data transformation, set the rest of $\tilde{y}_t$ and $\tilde{x}_t$ equal to y_t and x_t, respectively. We can simply apply the revised model to the transformed data and compute the predictive densities and means.[15] 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 top panel

[14]The quarterly rate used in the calculation is $100 \times \frac{\log(1.02)}{4}$.
[15]All of these adjustments can be done in real time.

of Table 4 shows these results while the middle and bottom panel of the table correspond to Table 3. The difference in the log-predictive likelihoods for the AR(1) in the top and middle panel ($58.1 - 51.1 = 7$) shows that the 1991Q1 observation is influential.

The top panel of Table 4 shows the subjective hierarchical SB-LSV model outperforms the hierarchical SB-LSV model (the best model in Table 1 and Table 2) and the AR(1) model (the best linear model in Table 3). The subjective model improves the log-predictive likelihood of the hierarchical SB-LSV model by 7.8 (from -45.3 to -37.5) and the predictive mean by 7% (from 0.42 to 0.39).

Even if we ignore the 1991Q1 outlier, the subjective hierarchical SB-LSV model is still better than the original hierarchical SB-LSV model in density forecasts (middle and bottom panel of Table 4). Furthermore, the subjective hierarchical SB-LSV model always provides better point forecasts than the original hierarchical SB-LSV model but cannot match the 2% rule between 1991Q1 and 1993Q4. In summary, it is straightforward to incorporate subjective information into our model which can lead to improved forecasts.

5.6 Exogenous Predictors

Are there other exogenous predictors useful in Canadian inflation forecasting? We consider unemployment and industrial production, following Stock and Watson (1999), who forecast inflation through a Phillips curve or a generalized Phillips curve by using real sector variables.

The industrial production data is from Statistics Canada while the unemployment data is from International Monetary Fund (IMF). The growth rate of industrial production is computed as the first difference of the logarithmic values. The unemployment is computed as the simple average from the monthly data. All time series are truncated to have the same length from 1976Q2 to 2012Q2.

The exogenous variables augment the regressors x_t in Equation (12) and their coefficients are subject to breaks. For example, in a hierarchical SB-LSV model with AR(1) process in each regime, we can have $x_t = (1, y_{t-1}, x_{\mathrm{Um},t-1})$, where $x_{\mathrm{Um},t-1}$ is the unemployment. We use the data at period $t-1$ to forecast the inflation y_t. The regressor $x_{\mathrm{Um},t-1}$ can be replaced by $x_{IP,t-1}$ to reflect the real sector's influence or both can be included. We can also add more lags of the exogenous variable to the model.

We use an informative prior similar to the Minnesota prior to the exogenous predictors' coefficients. For instance, including one lag of the unemployment rate, the 3rd element on the diagonal of A_0 is $\frac{1}{5}\frac{\sigma^2_{\mathrm{Um}}}{\sigma^2_{\mathrm{infl}}}$ where σ^2_{infl} and σ^2_{Um} are the sample variance of inflation and unemployment and makes the prior independent of data scaling.[16] This prior is used for additional lags of unemployment and similarly for lags of industrial production. The rest of the priors are the same as that of the original hierarchical SB-LSV model. We have performed robustness checks for this prior and the results are qualitatively the same. For example, if we set the prior the same as that of the hierarchical SB-LSV model, our results are only different in the first decimal place for the marginal likelihood.

Table 5 shows the model comparison results after adding the exogenous variables. The model label Hie-SB-LSV AR(1) + IP(1), means that the regressors are the intercept, the first lag of inflation and the first lag of industrial production in each regime of the hierarchical

[16] Recall from (12) that the prior mean of $\underline{H}$ is $a_0 A_0$.

SB-LSV model. The other models' structure can be inferred from their names in a similar manner. Our finding is consistent with Stock and Watson (1999) in the sense that including real sector variables improves inflation density forecasts. The hierarchical SB-LSV AR(1) + IP(1) model has the largest marginal likelihood and the performance of the point forecasts is close to the hierarchical SB-LSV AR(2).

5.7 Computational Speed and Numeric Efficiency

An advantage of our approach is computational speed and numerical sampling efficiency. To consider these, Koop and Potter's (2007) model is compared with the hierarchical SB-LSV model. We assume each model has an AR(2) process in each regime, which is the optimal model in Table 1. These two approaches are not directly comparable in general since they assume 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[17] used in Koop and Potter's (2007) model and our model are $1.1e^9$ and $1.4e^8$, respectively. 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. The relative numeric efficiency (RNE) is the ratio of the variance of the sample mean of a vector of draws relative to the variance from an iid sequence and is 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. Larger values indicate less efficient sampling. The RNE for the posterior mean of the number of regimes implied by the Koop and Potter (2007) model is 49.2. Our approach is more efficient since the RNE is only 0.49. If we consider the computational time and the RNE 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 computational time than our approach in order to achieve the same numeric efficiency for estimating the posterior mean of the number of regimes.

To investigate the source of improvement we estimate the SB-LSV model using a Gibbs sampler. We sample the durations conditional on $\Theta_{1,T}$ and then sample $\Theta_{1,T}$ conditional on the durations. To evaluate the efficiency of an algorithm we compute the effective sample size (ESS) which is the number of effectively independent draws from the posterior distribution that the Markov chain is equivalent to. The *nominal* ESS is calculated as $R\left[1 + 2\sum_{i=1}^{\tau} \frac{\tau-i}{\tau}\hat{\rho}_i\right]^{-1}$ and represents the true posterior sample size after accounting for autocorrelation in the chain. A sampler with a larger *nominal* ESS will result in a more accurate estimate of the posterior quantity of interest. The ESS is the *nominal* ESS normalized for CPU run time, which is computed as ESS/S, where S is the seconds of CPU run time.

Table 6 reports on the accuracy of the posterior mean using the ESS for our new joint sampling approach and a regular Gibbs sampler for the SB-LSV model. The joint sampling is much more efficient based on this criterion. For example, for the number of regimes K, the new method is more than 100 times as efficient as the Gibbs sampler. This is due to the fact that K is likely to be highly correlated in the Gibbs sampler but our posterior simulation

[17]CPU time is a measure of the time taken for a specific program to run, while wall clock time can encompass the time for a specific program as well as other unrelated processes running on the computer.

method samples the whole set of parameters as a single block.

5.8 Posterior Analysis

In this section, we discuss posterior estimates for the hierarchical SB-LSV AR(2) model. Table 7 shows the prior and the posterior summary of the parameters. The posterior mean of the structural change probability π is 0.03, which is less than its prior mean of 0.1. The posterior mean implies an average duration of 8 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 0.73, respectively. The posterior 95% density interval of $\underline{\beta}_0$ does not cover 0, which means that 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 confirm 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[18] along with 0.90 density intervals. 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 persistence, which is the sum of AR coefficients $\beta_{t,1}$ and $\beta_{t,2}$ is plotted in the fourth panel. 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 the oil crisis and characterized by an increase of the persistence and the volatility. In the mid 80's, a structural change decreased both the persistence and the volatility, which is consistent with the great moderation. The last break happened in the early 90's, which decreased 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 induces different dynamic patterns in the inflation process.

6 Conclusion

This paper builds on existing structural change models to provide an improved approach to estimating and forecasting 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 parameters that characterize each regime. The prior is modeled as hierarchical to exploit the information across regimes to improve forecasts.

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

[18] 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.

one block which results in very good mixing and improved accuracy of posterior estimates.

We apply the model to Canadian inflation data. The best model is the hierarchical model which allows the breaks in the regression coefficients and the variance to occur simultaneously. We discuss the importance of inflation targets introduced in the 1990s and investigate if forecasts can be improved after this policy change. We identify 4 major change-points in the Canadian inflation dynamics. Modeling breaks results in improvements in density forecasts and point forecasts.

A Appendix

The appendix provides the additional details of posterior simulation for the following models.

A.1 Hierarchical SB-LSV Model

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

   (a) Sample $\pi^{(i)} \mid K^{(i-1)} \sim \mathbf{B}(\pi_a + K^{(i-1)} - 1, \pi_b + T - K^{(i-1)})$ as in 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 $\{d_t, \beta_t, \sigma_t\}_{t=1}^T \mid \Psi$ as in the non-hierarchical structural break model.

A.2 Hierarchical SB-V Model

The predictive likelihood is:

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

or

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

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

$$\hat{\chi} = \underline{\chi} + E_{t-d_t+1,t-1}'E_{t-d_t+1,t-1}, \ \hat{\nu} = \underline{\nu} + d_t - 1$$

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

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

$$\overline{\beta} = \overline{H}^{-1}\left(\underline{H}\underline{\beta} + \sum_{t=1}^T \frac{x_t y_t}{\sigma_t^2}\right),$$

$$\overline{H} = \underline{H} + \sum_{t=1}^T \frac{x_t x_t'}{\sigma_t^2},$$

$$d_1 = d_0 + \sum_{i=1}^K \sigma_i^{-2},$$

$$c_1 = c_0 + K\underline{\nu}^{(i-1)}.$$

   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 $\{d_t, \sigma_t\}_{t=1}^T \mid \Psi$ similarly to the non-hierarchical structural break model.

A.3 Hierarchical SB-LS Model

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

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

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

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

   (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 in 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 $\{d_t, \beta_t\}_{t=1}^T \mid \Psi$ similarly to the non-hierarchical structural break model.

A.4 Hierarchical SB-LS-V Model

1. Sample $D_{\beta,1,T}, \{\beta_t\}_{t=1}^T \mid \underline{\beta}, \underline{H}, Y_{1,T}, \{\sigma_t\}_{t=1}^T$. The conditional posterior distribution of $\beta_t \mid d_{\beta,t}, \underline{\beta}, \underline{H}, Y_{1,t}, \{\sigma_t\}_{t=1}^T$ is a normal distribution $\mathbf{N}(\hat{\beta}, \hat{H})$, where $\hat{\beta} = \hat{H}^{-1}(\underline{H}\underline{\beta} +$

$\sum_{\tau=t-d_{\beta,t}+1}^{t} \frac{x_\tau y_\tau}{\sigma_\tau^2})$ and $\hat{H} = \underline{H} + \sum_{\tau=t-d_{\beta,t}+1}^{t} \frac{x_\tau x'_\tau}{\sigma_\tau^2})$. After integrating out β_t, the predictive distribution of y_{t+1} is a normal distribution,

$$y_{t+1} \mid d_{\beta,t+1} = d_{\beta,t} + 1, \underline{\beta}, \underline{H}, Y_{1,t}, \{\sigma_t\}_{t=1}^T \sim \mathbf{N}(x'_{t+1}\hat{\beta}, x'_{t+1}\hat{H}^{-1}x_{t+1} + \sigma_{t+1}^2).$$

Otherwise, if $d_{\beta,t+1} = 1$, use $\underline{\beta}$ and $\underline{H}$ to replace $\hat{\beta}$ and $\hat{H}$.

The filtered probability of $p(d_{\beta,t} \mid Y_{1,t}, \{\sigma_t\}_{t=1}^T)$ and the sampler of $D_{\beta,1,T} \mid Y_{1,t}, \{\sigma_t\}_{t=1}^T$ is computed in a similar way to the sampler discussed in Section 2.

Each distinct β_i^* is sampled from a normal distribution $\mathbf{N}(\overline{\beta}_i, \overline{H}_i^{-1})$, where $\overline{\beta}_i = \overline{H}_i^{-1}(\underline{H}\underline{\beta} + \sum_{s_{\beta,t}=i} \frac{x_t y_t}{\sigma_t^2})$ and $\overline{H}_i = \underline{H} + \sum_{s_{\beta,t}=i} \frac{x_t x'_t}{\sigma_t^2}$. The regime indicator $s_{\beta,t}$ has similar interpretation as s_t in Section 2.1. There is a one-to-one relationship between the set $S_{\beta,1,T} = (s_{\beta,1}, \ldots, s_{\beta,T})$ and $D_{\beta,1,T}$.

2. Sample $D_{\sigma,1,T}, \{\sigma_t\}_{t=1}^T \mid \underline{\chi}, \underline{\nu}, Y_T, \{\beta_t\}_{t=1}^T$. The conditional posterior distribution of $\sigma_t^{-2} \mid d_{\sigma,t}, \underline{\chi}, \underline{\nu}, Y_{1,t}, \{\beta_t\}_{t=1}^T$ is a gamma distribution $\mathbf{G}(\frac{\hat{\chi}}{2}, \frac{\hat{\nu}}{2})$, where $\hat{\nu} = \underline{\nu} + d_{\sigma,t}$, $\hat{\chi} = \underline{\chi} + E'_{t-d_{\sigma,t}+1,t}E_{t-d_{\sigma,t}+1,t}$, $E_{t-d_{\sigma,t}+1,t} = (e_{t-d_{\sigma,t}+1}, \ldots, e_t)'$ and $e_t = y_t - x'_t\beta_t$. After integrating out σ_t, the predictive distribution of y_{t+1} is a Student-t distribution,

$$y_{t+1} \mid d_{\sigma,t+1} = d_{\sigma,t} + 1, \underline{\chi}, \underline{\nu}, Y_{1,t}, \{\beta_t\}_{t=1}^T \sim \mathbf{t}\left(x'_{t+1}\beta_{t+1}, \frac{\hat{\chi}}{\hat{\nu}}, \hat{\nu}\right),$$

with density proportional to $\left(1 + \frac{(y_{t+1} - x'_{t+1}\beta_{t+1})^2}{\hat{\chi}}\right)^{-\frac{\hat{\nu}+1}{2}}$. Otherwise, if $d_{\sigma,t+1} = 1$, use $\underline{\chi}$ and $\underline{\nu}$ to replace $\hat{\chi}$ and $\hat{\nu}$. Each distinct σ_i^* is sampled based on a gamma distribution as $\sigma_i^{*-2} \sim \mathbf{G}(\frac{\overline{\chi}}{2}, \frac{\overline{\nu}}{2})$, where $\overline{\chi} = \underline{\chi} + \sum_{s_{\sigma,t}=i} e_t^2$, $\overline{\nu} = \underline{\nu} + d_{\sigma,i}^*$, and $d_{\sigma,i}^*$ is the duration of regime i. The regime indicator $s_{\sigma,t}$ has similar interpretation as s_t in Section 2.1. There is a one-to-one relationship between the set $S_{\sigma,1,T} = (s_{\sigma,1}, \ldots, s_{\sigma,T})$ and $D_{\sigma,1,T}$.

3. Sample $\pi_\beta, \pi_\sigma \mid D_\beta, D_\sigma$. Let K_β and K_σ represent the number of distinct β_i^*'s and σ_i^*'s, then the break probabilities π_β and π_σ are sampled from beta distributions, $\mathbf{B}(\pi_{\beta,a} + K_\beta - 1, \pi_{\beta,b} + T - K_\beta)$ and $\mathbf{B}(\pi_{\sigma,a} + K_\sigma - 1, \pi_{\sigma,b} + T - K_\sigma)$, respectively.

4. Sample $\underline{\beta}, \underline{H} \mid \{\beta_i^*\}_{i=1}^{K_\beta}$. The conditional posterior distribution is $\underline{H} \mid \{\beta_i^*\}_{i=1}^{K_\beta} \sim \mathbf{W}(A_1, a_0)$ and $\underline{\beta} \mid \underline{H}, \{\beta_i^*\}_{i=1}^{K_\beta} \sim \mathbf{N}(m_1, (\tau_1 H_1)^{-1})$, where $A_1 = (A_0^{-1} + \sum \beta_i^* \beta_i^* + \tau_0 m_0 m'_0 - \tau_1 m_1 m'_1)^{-1}$, $a_1 = a_0 + K_\beta$, $\tau_1 = \tau_0 + K_\beta$ and $m_1 = \tau_1^{-1}(\tau_0 m_0 + \sum \beta_i^*)$.

5. Sample $\underline{\chi} \mid \underline{\nu}, \{\sigma_i^*\}_{i=1}^{K_\sigma}$. The notation K_σ is the number of distinct σ_i^*'s. The conditional posterior is $\underline{\chi} \mid \underline{\nu}, \{\sigma_i^*\}_{i=1}^{K_\sigma} \sim \mathbf{G}(\frac{d_1}{2}, \frac{c_1}{2})$, where $d_1 = d_0 + \sum \sigma_i^{*-2}$ and $c_1 = c_0 + K_\sigma \underline{\nu}$.

6. Sample $\underline{\nu} \mid \underline{\chi}, \{\sigma_i^*\}_{i=1}^{K_\sigma}$. Use the proposal distribution $\underline{\nu}^{(i)} \sim \mathbf{G}\left(\frac{\zeta}{\underline{\nu}^{(i-1)}}, \zeta\right)$ and accept with the Metropolis-Hastings method. The conditional posterior density is proportional to $\exp(-\frac{\underline{\nu}}{\rho_0}) \left(\frac{(\frac{\underline{\chi}}{2})^{\frac{\underline{\nu}}{2}}}{\Gamma(\frac{\underline{\nu}}{2})}\right)^{K_\sigma} (\prod \sigma_i^{*-2})^{\underline{\nu}/2}$.

A.5 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 distributions.

   (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 $\{d_t, \beta_t, \sigma_t\}_{t=1}^T \mid \Psi$ as done in the non-hierarchical structural break model.

Table 1: Forecast Performance: 1961Q1 – 2012Q2

	Log ML	Log BF	RMSFE	DM p-value	HK-MASE
Hie SB-LSV AR(1)	-138.8	8.0	0.47	0.16	0.87
Hie SB-LSV AR(2)	**-130.8**		**0.46**		**0.84**
Hie SB-LSV AR(3)	-137.5	6.7	0.47	0.24	0.85
KP AR(1)	-145.1	14.3	0.48	0.16	0.87
KP AR(2)	-154.4	23.6	0.50**	0.02	0.92
KP AR(3)	-157.2	26.4	0.53***	0.00	0.96
Hie SB-LS-V AR(1)	-180.2	49.4	0.54***	0.00	0.99
Hie SB-LS-V AR(2)	-205.7	74.9	0.55***	0.00	1.00
Hie SB-LS-V AR(3)	-199.3	68.5	0.53***	0.00	0.97
Hie SB-V AR(1)	-162.0	31.2	0.54***	0.00	0.98
Hie SB-V AR(2)	-145.2	14.4	0.49**	0.03	0.89
Hie SB-V AR(3)	-145.0	14.2	0.49**	0.04	0.89
Hie SB-LS AR(1)	-145.0	14.2	0.47	0.27	0.85
Hie SB-LS AR(2)	-150.6	19.8	0.47	0.11	0.86
Hie SB-LS AR(3)	-155.5	24.7	0.48*	0.06	0.87
Hie DDSB-LSV AR(1)	-156.8	26.0	0.47	0.19	0.86
Hie DDSB-LSV AR(2)	-161.4	30.4	0.47	0.28	0.86
Hie DDSB-LSV AR(3)	-167.0	36.2	0.48	0.16	0.87
Nonhie SB-LSV AR(1)	-163.2	32.4	0.51***	0.00	0.94
Nonhie SB-LSV AR(2)	-154.4	23.6	0.49**	0.04	0.89
Nonhie SB-LSV AR(3)	-154.5	23.7	0.49**	0.04	0.89
Nonhie SB-LSV($\pi = 0.01$) AR(1)	-161.4	30.6	0.51***	0.00	0.94
Nonhie SB-LSV($\pi = 0.01$) AR(2)	-152.6	21.8	0.48*	0.06	0.88
Nonhie SB-LSV($\pi = 0.01$) AR(3)	-152.9	22.1	0.48*	0.07	0.88
AR(1)	-168.1	37.3	0.54***	0.00	0.98
AR(2)	-151.5	20.7	0.49**	0.03	0.90
AR(3)	-151.4	20.6	0.49*	0.05	0.89
AR(1)-GARCH(1,1)	-168.6	37.8	0.58***	0.00	1.03
AR(2)-GARCH(1,1)	-148.3	17.5	0.50***	0.01	0.91
AR(3)-GARCH(1,1)	-150.0	19.2	0.51***	0.00	0.94

DM is a Diebold-Mariano test for squared error loss. The notations ***, ** and * mean that the test is significant at 1%, 5% and 10% level. We compare the hierarchical (Hie) SB-LSV AR(2) model with the other models. Log ML is the log-marginal likelihood, Log BF is the associated log-Bayes factor of the model with the SB-LSV AR(2) model, RMSFE is the root mean square forecast error and HK-MASE is the mean absolute scaled error from Hyndman and Koehler.

Table 2: Sub-sample Forecast Performance

From 1978Q1-2012Q2

	Log PL	Log BF	RMSFE	DM p-value	HK-MASE
Hie SB-LSV AR(2)	**-76.1**		0.43		0.84
KP AR(1)	-84.5	9.3	0.47	0.11	0.88
Hie SB-LS-V AR(1)	-121.8	45.7	0.52***	0.00	1.04
Hie SB-V AR(3)	-87.2	11.1	0.46*	0.06	0.90
Hie SB-LS AR(1)	-83.3	7.2	0.42	0.54	0.82
Hie DDSB-LSV AR(1)	-85.1	9.0	**0.42**	0.39	**0.81**
Nonhie SB-LSV AR(2)	-93.7	17.6	0.46	0.15	0.90
AR(3)	-91.6	15.5	0.46*	0.08	0.90
AR(2)-GARCH(1,1)	-89.9	13.8	0.47**	0.05	0.93

From 1978Q1-1990Q4

	Log PL	Log BF	RMSFE	DM p-value	HK-MASE
Hie SB-LSV AR(2)	**-30.9**		0.44		0.89
KP AR(1)	-33.5	2.6	0.45	0.89	0.86
Hie SB-LS-V AR(1)	-53.6	22.7	0.51*	0.08	1.00
Hie SB-V AR(3)	-32.8	1.9	0.45	0.92	0.86
Hie SB-LS AR(1)	-34.9	4.0	0.44	0.69	0.85
Hie DDSB-LSV AR(1)	-35.6	4.7	**0.43***	0.09	**0.83**
Nonhie SB-LSV AR(2)	-35.4	4.5	0.45	0.93	0.88
AR(3)	-34.3	3.4	0.44	0.94	0.86
AR(2)-GARCH(1,1)	-32.5	1.6	0.45	0.80	0.89

From 1991Q1-2012Q2

	Log PL	Log BF	RMSFE	DM p-value	HK-MASE
Hie SB-LSV AR(2)	**-45.3**		0.42		0.82
KP AR(1)	-51.0	5.7	0.48*	0.10	0.91
Hie SB-LS-V AR(1)	-68.3	23.0	0.52***	0.00	1.08
Hie SB-V AR(3)	-54.4	9.1	0.46**	0.04	0.94
Hie SB-LS AR(1)	-48.4	3.1	0.41	0.64	0.80
Hie DDSB-LSV AR(1)	-49.5	4.2	**0.42**	0.98	**0.81**
Nonhie SB-LSV AR(2)	-58.3	13.0	0.46	0.13	0.92
AR(3)	-57.3	12.0	0.47*	0.04	0.93
AR(2)-GARCH(1,1)	-57.4	12.1	0.48*	0.04	0.98

DM is a Diebold-Mariano test for squared error loss. The notations ***, ** and * mean that the test is significant at 1%, 5% and 10% level. We compare the hierarchical (Hie) SB-LSV AR(2) model with the other models. Log PL is the log-predictive likelihood, Log BF is the associated log-Bayes factor of the model with the SB-LSV AR(2) model, RMSFE is the root mean square forecast error and HK-MASE is the mean absolute scaled error from Hyndman and Koehler.

Table 3: Forecasts in the Presence of Inflation Targeting

	Log PL	Log BF	RMSFE	DM p-value	HK-MASE
From the first quarter after inflation targeting (1991Q2-2012Q2)					
Hie SB-LSV AR(2)	**-41.8**		0.40		0.82
AR(1)	-51.1	9.3	0.36	0.15	0.78
AR(2)	-52.4	10.6	0.37	0.28	0.81
AR(3)	-53.0	11.2	0.38	0.31	0.81
2% target			**0.36***	0.10	**0.75**
From the first quarter of 1% – 3% target (1994Q1-2012Q2)					
Hie SB-LSV AR(2)	**-32.1**		0.37		0.82
AR(1)	-45.2	13.1	**0.35**	0.50	0.84
AR(2)	-45.8	13.7	0.35	0.56	0.86
AR(3)	-46.3	14.2	0.35	0.63	0.86
2% target			0.35	0.41	**0.81**

DM is a Diebold-Mariano test for squared error loss. The notations ***, ** and * mean that the test is significant at 1%, 5% and 10% level. We compare the hierarchical (Hie) SB-LSV AR(2) model with the other models. Log PL is the log-predictive likelihood, Log BF is the associated log-Bayes factor of the model with the SB-LSV AR(2) model, RMSFE is the root mean square forecast error and HK-MASE is the mean absolute scaled error from Hyndman and Koehler.

Table 4: Subjective Forecasting

	Log PL	Log BF	RMSFE	DM p-value	HK-MASE
From the first quarter of inflation targeting(1991Q1-2012Q2)					
Subjective Hie SB-LSV AR(2)	**-37.5**		**0.39**		0.80
Hie SB-LSV AR(2)	-45.3	7.8	0.42	0.40	0.82
AR(1)	-58.1	20.6	0.47	0.11	0.84
AR(2)	-59.4	20.9	0.48*	0.07	0.89
AR(3)	-60.4	22.9	0.50**	0.03	0.93
2% target			0.40	0.80	**0.77**
From the first quarter after inflation targeting(1991Q2-2012Q2)					
Subjective Hie SB-LSV AR(2)	**-35.6**		0.37		0.77
Hie SB-LSV	-41.8	6.2	0.40	0.22	0.82
AR(1)	-51.1	15.5	0.36	0.90	0.78
AR(2)	-52.4	16.8	0.37	0.76	0.81
AR(3)	-53.0	17.4	0.38	0.51	0.81
2% target			**0.36**	0.40	**0.75**
From the first quarter of $1\% - 3\%$ target (1994Q1-2012Q2)					
Subjective Hie SB-LSV AR(2)	**-28.3**		0.35		**0.81**
Hie SB-LSV	-32.1	3.8	0.37	0.41	0.82
AR(1)	-45.2	16.9	**0.35**	0.80	0.84
AR(2)	-45.8	17.5	0.35	0.89	0.86
AR(3)	-46.3	18.0	0.35	0.97	0.86
2% target			0.35	1.00	**0.81**

DM is a Diebold-Mariano test for squared error loss. The notations ***, ** and * mean that the test is significant at 1%, 5% and 10% level. We compare the hierarchical (Hie) SB-LSV AR(2) model with the other models. Log PL is the log-predictive likelihood, Log BF is the associated log-Bayes factor of the model with the SB-LSV AR(2) model, RMSFE is the root mean square forecast error and HK-MASE is the mean absolute scaled error from Hyndman and Koehler.

Table 5: Forecasting with Umemployment or Industrial Production as Predictors

	Log ML	Log BF	RMSFE	DM p-value	HK-MASE
Hie SB-LSV AR(1)	-88.4	3.5	0.45	0.65	0.85
Hie SB-LSV AR(2)	-87.2	2.3	**0.44**		**0.83**
Hie SB-LSV AR(3)	-93.3	8.4	0.48***	0.01	0.92
Hie SB-LSV AR(1) + IP(1)	**-84.9**		0.45	0.41	0.85
Hie SB-LSV AR(1) + Um(1)	-90.8	5.9	0.46	0.23	0.86
Hie SB-LSV AR(1) + IP(2)	-86.0	1.1	0.46	0.30	0.84
Hie SB-LSV AR(1) + Um(2)	-96.3	11.4	0.49**	0.03	0.91
Hie SB-LSV AR(1) + IP(3)	-87.7	2.8	0.47*	0.09	0.88
Hie SB-LSV AR(1) + Um(3)	-103.4	18.5	0.49**	0.02	0.94
Hie SB-LSV AR(2) + IP(1)	-88.5	3.6	0.46	0.19	0.85
Hie SB-LSV AR(2) + Um(1)	-93.9	9.0	0.46*	0.04	0.88
Hie SB-LSV AR(2) + Um(1) & IP(1)	-96.7	11.8	0.49***	0.00	0.94
Hie SB-LSV AR(2) + Um(2) & IP(2)	-104.6	19.7	0.50***	0.00	0.99

DM is a Diebold-Mariano test for squared error loss. The notations ***, ** and * mean that the test is significant at 1%, 5% and 10% level. We compare the hierarchical (Hie) SB-LSV AR(2) model with the other models. Log PL is the log-predictive likelihood, Log BF is the associated log-Bayes factor of the model with the SB-LSV AR(2) model, RMSFE is the root mean square forecast error and HK-MASE is the mean absolute scaled error from Hyndman and Koehler.

Table 6: Comparison of Posterior Sampling Efficiency

	nominal ESS		ESS	
	Gibbs Sampler	Joint Sampler	Gibbs Sampler	Joint Sampler
π	287.4	1250	0.43	10.2
$\underline{\beta}_0$	326.8	2000	0.49	16.3
$\underline{\beta}_1$	211.9	1678	0.32	13.6
$\underline{\beta}_2$	6172	2994	9.30	24.3
$\underline{H}_{00}$	1519	1845	2.29	15.0
$\underline{H}_{01}$	1510	1000	2.27	8.13
$\underline{H}_{02}$	10570	4717	15.9	38.3
$\underline{H}_{11}$	4347	2941	6.55	23.9
$\underline{H}_{12}$	1582	4167	2.38	33.9
$\underline{H}_{22}$	3703	3571	5.58	29.0
$\underline{\chi}$	78.9	151	0.12	1.23
$\underline{\nu}$	106.6	119	0.16	0.97
K	75.2	1613	0.11	13.1

The *nominal* ESS is calculated as $R\left[1+2\sum_{i=1}^{\tau}\frac{\tau-i}{\tau}\hat{\rho}_i\right]^{-1}$ where R is the number of posterior samples and $\hat{\rho}_i$ is the sample autocorrelation. The ESS is the nominal ESS normalized for CPU run time, which is computed as ESS/S, where S is the seconds of CPU run time.

Table 7: Posterior Summary of the Hierarchical SB-LSV AR(2) Model

	Prior		Posterior		
	Mean	0.95DI	Mean	Sd	0.95 DI
π	0.1	(0.003, 0.34)	0.03	0.01	(0.01, 0.06)
$\underline{\beta}_0$	0.0	(-3.08, 3.08)	0.73	0.21	(0.32, 1.14)
$\underline{\beta}_1$	0.0	(-3.08, 3.08)	-0.07	0.18	(-0.42, 0.29)
$\underline{\beta}_2$	0.0	(-3.08, 3.08)	-0.09	0.17	(-0.26, 0.45)
$\underline{H}_{00}$	1.0	(0.16, 2.52)	1.06	0.38	(0.46, 1.94)
$\underline{H}_{01}$	0.0	(-0.91, 0.91)	-0.02	0.27	(-0.56, 0.50)
$\underline{H}_{02}$	0.0	(-0.91, 0.91)	0.04	0.28	(-0.55, 0.60)
$\underline{H}_{11}$	1.0	(0.16, 2.52)	1.16	0.40	(0.51, 2.05)
$\underline{H}_{12}$	0.0	(-0.91, 0.91)	0.06	0.29	(-0.51, 0.62)
$\underline{H}_{22}$	1.0	(0.16, 2.52)	1.25	0.43	(0.57, 2.23)
$\underline{\chi}$	1.0	(0.12, 2.79)	0.93	0.46	(0.30, 2.09)
$\underline{\nu}$	2.0	(0.05, 7.38)	5.37	3.01	(1.30, 12.7)

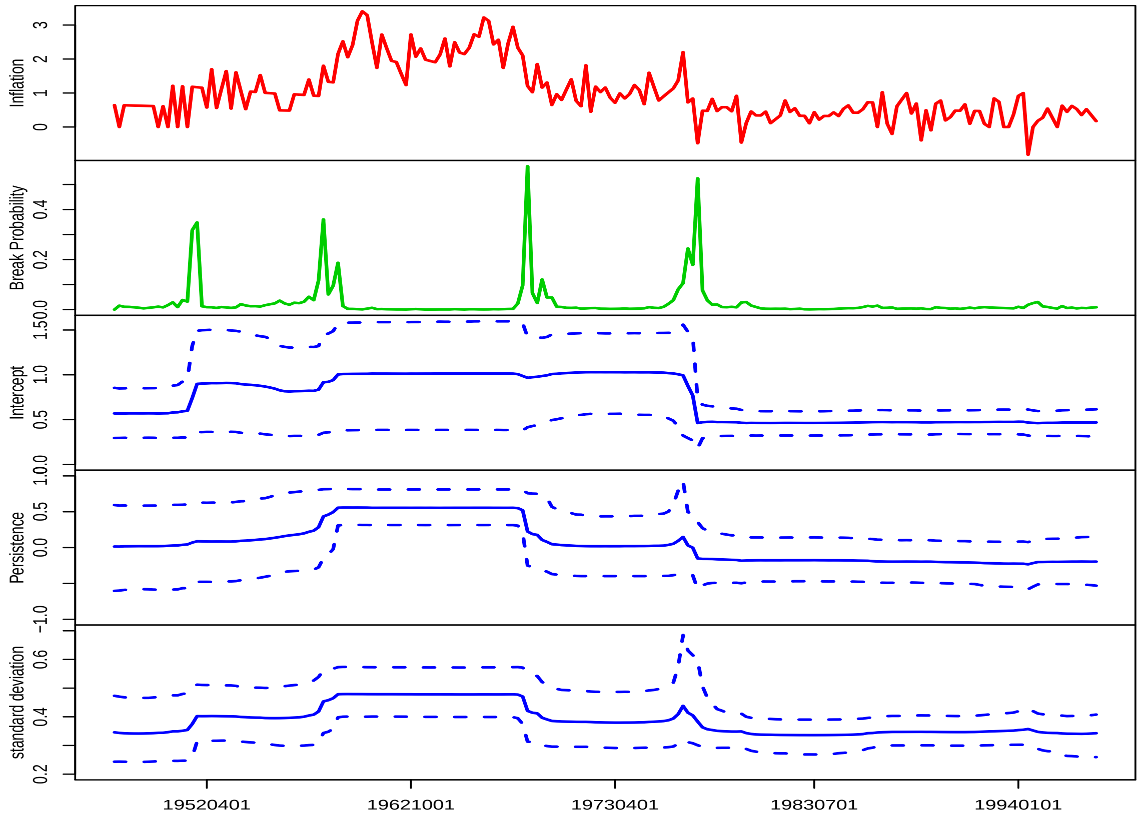

Figure 1: The top panel display quarterly Canadian inflation over 1961Q1-2012Q2. The second panel shows the posterior probability of a structural break followed by the posterior mean (solid line) of the intercept, the sum of the autoregression coefficients, and the standard deviation for the hierarchical SB-LSV model with an AR(2) structure. The dashed lines show the 0.90 density intervals.

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. ISSN 0304-4076.

Chib S. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2):221–241, 1998. ISSN 0304-4076.

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

Diebold F.X. and Mariano R.S. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–265, 1995.

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

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

Hyndman R.J. and Koehler A.B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

Kass R.E. and Raftery A.E. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995. ISSN 0162-1459.

Koop G. and Potter S.M. Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, 74(3):763–789, 2007. ISSN 1467-937X.

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. ISSN 1099-1255.

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. ISSN 1467-937X.

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. ISSN 0735-0015.

Stock J.H. and Watson M.W. Forecasting inflation. *Journal of Monetary Economics*, 44(2): 293–335, 1999.

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. ISSN 0735-0015.