Title
Bayesian High-Dimensional Regression for Change Point Analysis.

Journal
Statistics and Its Interface, 12(2)

ISSN
1938-7989

Authors
Datta, Abhirup
Zou, Hui
Banerjee, Sudipto

Publication Date
2019

DOI
10.4310/sii.2019.v12.n2.a6

Peer reviewed

HHS Public Access

Author manuscript

Stat Interface. Author manuscript; available in PMC 2019 September 20.

Published in final edited form as:

Stat Interface. 2019 ; 12(2): 253–264. doi:10.4310/SII.2019.v12.n2.a6.

Bayesian High-Dimensional Regression for Change Point Analysis

Abhirup Datta,

Department of Biostatistics, Johns Hopkins University, abhidatta@jhu.edu

Hui Zou,

Department of Statistics, University of Minnesota, zouxx019@umn.edu

Sudipto Banerjee

Department of Biostatistics, University of California, Los angeles, sudipto@ucla.edu

Abstract

In many econometrics applications, the dataset under investigation spans heterogeneous regimes that are more appropriately modeled using piece-wise components for each of the data segments separated by change-points. We consider using Bayesian high-dimensional shrinkage priors in a change point setting to understand segment-specific relationship between the response and the covariates. Covariate selection before and after each change point can identify possibly different sets of relevant covariates, while the fully Bayesian approach ensures posterior inference for the change points is also available. We demonstrate the flexibility of the approach for imposing different variable selection constraints like grouping or partial selection and discuss strategies to detect an unknown number of change points. Simulation experiments reveal that this simple approach delivers accurate variable selection, and inference on location of the change points, and substantially outperforms a frequentist lasso-based approach, uniformly across a wide range of scenarios. Application of our model to Minnesota house price dataset reveals change in the relationship between house and stock prices around the sub-prime mortgage crisis.

Keywords

Bayesian Inference; Change Point Detection; High-dimensional Regression; Markov Chain Monte Carlo; Minnesota House Price Data; Variable Selection

1. INTRODUCTION

Modern statistical modeling and inference continue to evolve and be molded by the emergence of complex datasets, where the dimension of each observation in a dataset substantially exceeds the size of the dataset. Largely due to recent advances in technology, such high-dimensional datasets are now ubiquitous in fields as diverse as genetics, economics, neuroscience, public health, imaging, and so on. One important objective of

SUPPLEMENTARY MATERIALS

Code to generate data and run the single and two change point high dimensional regression models are provided as supplementary materials.

high-dimensional data analysis is to segregate a small set of regressors, associated with the response of interest, from the large number of redundant ones. Penalized least square approaches like lasso [46], SCAD [18], Elastic Net [51], adaptive lasso [50] etc. are widely employed for high-dimensional regression analysis. Bayesian alternatives typically proceed by using hierarchical priors for the regression coefficients aimed at achieving variable selection. Common approaches include the spike-and-slab prior [24] or its variants [29, 35] or, more generally, scale mixture of normal priors, like Bayesian lasso [39] and horseshoe prior [11] among others.

Most of the aforementioned approaches assume a single underlying model from which the data is generated. Such homogeneity assumptions are often violated for time series data spanning different economic regimes [14, 48]. Such examples of structural change in relationship between variables is also common in other fields like climate change [41], DNA micro-array analysis [3] and so on. Change point models provide a convenient depiction of such complex relationships by splitting the data based on a threshold variable and using a homogeneous model for each segment. There is a burgeoning literature on Bayesian methodology addressing various change point problems [see, e.g., 10, 5, 33, 1, 47].

Changing linear regression models are a subclass of change point problems, where the linear model relating the response to the predictors varies over different segments of the data. Segmentation of the dataset is typically based on unknown change points of a threshold variable like time or age or some other contextual variable observed along with the data. In a low dimensional setting, [10] used Gibbs' sampling techniques for changing linear models to deliver fully Bayesian inference about the location of the change points and the regression coefficients for each segment.

When the set of possible predictors is large, an additional objective is to identify the (possibly different) sparse supports for each segment. Even when the sample size n is larger than the number of predictors p, in presence of one or more change points, the effective sample size for each segment may be much less than p, thereby necessitating shrinkage or variable selection based approaches.

This manuscript demonstrates how we can seamlessly exploit Bayesian variable selection techniques in a change point setup to simultaneously detect the location of the change points as well as to identify the true sparse support for each of the linear models. The Bayesian approach delivers full posterior inference on the change points, posterior selection probabilities for each variable for all segments and posterior predictive distributions for the response. Flexibility in the choice of variable selection priors offers the scope for structural variable selection tailored to specific data applications. For example, constraints like grouping the selection of a variable across all the segments can be easily achieved using group selection priors. Other constraints like partial selection within or between the segments can also be accommodated.

Classical penalized least square approaches mentioned earlier can also be used in a change point setup. By treating the unknown change points as additional tuning parameters, one can split the data using fixed values of these change points and use some penalized loss function

to achieve variable selection for each segment. For example, [30] uses lasso penalty to estimate the coefficients for each segment. Subsequent application of cross validation or model selection techniques will yield the optimal change points from a grid of possible values. However, a fully Bayesian approach has several advantages over this. Firstly, the grid search approach is computationally highly inefficient especially for more than one change points. On the other hand, a prior specification for the change points in the Bayesian model enables standard MCMC techniques to efficiently generate posterior samples. Moreover, in many real applications, change in association between variables can occur over a range of the threshold variable. Point estimates of change points obtained from classical approaches fail to accurately depict such scenarios. Bayesian credible intervals obtained from the posterior distributions provide a much more realistic quantification of the uncertainty associated with the location of the change points.

Numerical studies reveal that for a wide range of scenarios, our proposed methodology performs uniformly better than the frequentist approach both in terms of model selection and parameter estimation. We also demonstrate the applicability of our method for a macro-economic analysis of Minnesota house price index data. The results strongly favor our change point model over a homogeneous high-dimensional regression model.

The rest of the manuscript is organized as follows. The motivating dataset of Minnesota house price index time series is presented in Section 2. In Section 3 we present our methodology in details including extensions to unknown number of change points and alternate prior choices. Results from several simulated numerical studies are provided in Section 4. In Section 5 we present the details of the house price index data analysis using our change point methodology. We conclude in Section 6 with a brief review and pointers to future research.

2. MOTIVATING DATA

Economic datasets constitute a major domain of application of change point models. Many economic time series datasets may be collected over different political and financial regimes, thereby containing several change points with respect to the association with the predictors. For example, the relationship between house prices and macro-economic variables is often observed to exhibit differential trends over time. As noted in [2], the US stock market crash in the ‘Internet bubble burst’ of 2001–2002 was not accompanied by declining house prices whereas in the sub-prime mortgage crisis in 2007–2009, stocks and house prices witnessed simultaneous collapse.

Multivariate regression models have been used to understand the relationship between house price index (hpi) and macro economic variables in Ukraine [32], Sweden [45] and Malaysia [37]. These analyses often assume a single underlying time-homogeneous relationship between hpi and the explanatory variables, which may not be always appropriate. To illustrate, in Figure 1, we present the quarterly hpi time-series of Minnesota for a 24-year period from first quarter of 1991 to first quarter of 2015. We observe that there are two possible change points (marked by red dots) with respect to time — one around 2006–2008 where hpi starts to depreciate after reaching a peak and, later, one around 2012 where hpi

starts its revival. This only suggests possible change points in terms of the overall temporal trend for hpi and doesn't necessarily imply concurrent change points in terms of association with macroeconomic variables like stocks. To understand how hpi co-varied with stock prices, we also added the quarterly Dow Jones Industrial Average on the secondary y-axis. We see that there are roughly three segments (separated by dotted gray vertical lines) with distinct relationships between house and stock prices. Both indices were upward moving till the early 2000s when the stock market crashed but hpi continued to appreciate and there seem to be little correlation among the two series. Finally after the sub-prime mortgage crisis both indices showed strong co-movement, plummeting for a while and subsequently rallying. Hence Figure 1, it seems that a change-point regression model would be more suited to analyze Minnesota hpi data.

While the impact of macroeconomic variables on US house prices has been analyzed in the literature [12, 13] any relevant literature focusing on similar analysis at state level has eluded us. As housing markets are local in nature [22], a state level macro-analysis may reveal trends not reflected in a similar nationwide study. The state of Minnesota is home to 18 Fortune 500 companies and has the second highest number of Fortune 500 companies per capita. Furthermore, the Minneapolis-St. Paul metropolitan area hosts the highest number of Fortune 500 companies per capita among the 30 largest metropolitan areas in US. Our objective for this analysis was to understand association between real estate prices and local industries in Minnesota.

In this analysis, we use 22 macro-economic variables, including 16 Minnesota based stock prices, (detailed in Section 5) as covariates. Given the reach of modern high-dimensional methods to analyze datasets with million of covariates, this may not seem to be a high-dimensional problem. However, as mentioned earlier, changing linear regression estimates the regression coefficient vector for each of the data segments separated by change points. Hence, the total sample size n need not be less than the number of covariates p for requiring high-dimensional techniques in change point analysis. For example, as discussed above, the trends of house and stock prices displayed in Figure 1 insinuates that there may be a change point near 2008–2009 (the right gray vertical line), implying that sample size of rightmost segment of the data will be close to or less than 22. Hence, the although the number of time points ($n = 96$) was larger than the number of covariates ($p = 22$), if there is indeed a change point around 2008–09 (as the analysis will confirm in Section 5), this becomes a $p > n$ (or at least $p \approx n$) problem, and, shrinkage is essential to obtain reliable inference. This points out a general issue in regression analysis with change points. A change point towards the very end or very beginning, or two very proximal change points in the middle will result in small data segments, thereby, necessitating highdimensional techniques like shrinkage and variable selection.

3. BAYESIAN HIGH-DIMENSIONAL CHANGING LINEAR REGRESSION

We consider a traditional high-dimensional setup with the $n \times 1$ response vector $y = (y_1, y_2, \ldots, y_n)'$ and corresponding $n \times p$ covariate matrix $X = (x_1, x_2, \ldots, x_n)'$ where p can be larger than n. We further assume that for every observation y_i, we observe another quantitative variable t_i such that the association between y_i and x_j depends on the values of

t_i. In a linear regression setup, this dynamic relationship between the response y_i and the corresponding $p \times 1$ vector of covariates x_i can be expressed as $E(y_i \mid x_i, t_i) = x_i'\beta_k$ for all i such that $\tau_{k-1} < t_i < \tau_k$ where $\tau_0 < \tau_1 < \ldots < \tau_K < \tau_{K+1}$ where τ_0 and τ_{K+1} are two arbitrary constants such that $\tau_0 < t_i < \tau_{K+1}$ for all i. The change-points $\tau_1, \tau_2, \ldots, \tau_K$ are typically unknown while the number of change-points K may or may not be known depending on the application.

As the number of regressors (p) is large, our goal is to select the relevant variables for this regression. However, for this changing linear regression, the set of relevant regressors may depend on the value of the threshold variable t and variable selection procedures applied disregarding the dependence on t can lead to erroneous variable selection. Let S_k denotes the support of β_k where $s_k = |S_k|$ is typically much less than p. We intend to simultaneously detect the change-points τ_k and estimate S_k for all $k = 1, 2, \ldots, K$. We initially assume only one change-point τ i.e. $K = 1$. Extensions to more than one change points are discussed later in Section 3.2.

3.1 One Change Point Model

We assume a changing linear regression model

$$y_i = \begin{cases} x_i'\beta_1 + \epsilon_i \text{if } t_i \le \tau \\ x_i'\beta_2 + \epsilon_i \text{if } t_i > \tau \end{cases} \quad (1)$$

where β_1, β_2 are both sparse $p \times 1$ vectors such that $\beta_1 \neq \beta_2$ and $\epsilon_i \sim N(0, \sigma^2)$ denotes the independent and identically distributed noise. In order to accomplish variable selection both before and after the change point, we use spike-and-slab type shrinkage priors [24] for β_1 and β_2. To be specific, we assume $\beta_k | Z_k, \sigma^2 \sim N(0, \sigma^2 diag(\gamma_{1k} Z_k + \gamma_{0k}(1 - Z_k)))$ for $k = 1, 2$ where $Z_k = (Z_{k1}, Z_{k2}, \ldots, Z_{kp})'$ is a $p \times 1$ vector of zeros and ones. The hyper-parameters γ_{0k} and γ_{1k} are scalars chosen to be very small and very large, respectively. Under this prior specification, β_{kj}—the j^{th} component of β_k—is assigned a shrinking (concentrated around zero) prior if Z_{kj} equals 0 and a diffusion (flat) prior if $Z_{kj} = 1$. Zkj's are assumed to be apriori independent each following $Bernoulli(q_k)$. Hence q_k controls the prior model size for the k^{th} segment. The choices for the hyper-parameters γ_{0k}, γ_{1k} and q_k are discussed in Section 4. We assume a uniform prior for the change-point τ and a conjugate Inverse Gamma prior for the noise variance σ^2. The full Bayesian model is

$$\prod_{i:t_i \le \tau} N\left(y_i | x_i'\beta_1, \sigma^2\right) \times \prod_{i:t_i > \tau} N\left(y_i | x_i'\beta_2, \sigma^2\right) \times Unif(\tau | a_\tau, b_\tau) \times IG\left(\sigma^2 | a_\sigma, b_\sigma\right) \quad (2)$$
$$\times \prod_{k=1}^{2} \left(N\left(\beta_k | 0, \sigma^2 diag\left(\gamma_{1k} Z_k + \gamma_{0k}(1 - Z_k)\right)\right) \times \prod_{j=1}^{p} Bernoulli\left(Z_{kj} | q_k\right) \right).$$

We use Gibbs' sampler to obtain posterior samples of all the parameters. Let $\tau \mid \cdot$ denote the full-conditional distribution of τ in the Gibbs' sampler. We use similar notation to denote the other full conditionals. Let $U_1 = \{i \mid t_i \leq \tau\}$ and $U_2 = \{i \mid t_i > \tau\}$. For $k = 1, 2$, let Y_k and X_k denote the response vector and covariate matrix obtained by stacking up the observations corresponding to Uk. From the full likelihood in (2), we have

$\beta_k \mid \cdot \sim N\left(V_k X_k' Y_k, \sigma^2 V_k\right)$, where

$$V_k = (X_k' X_k + diag(\gamma_{1k} Z_k + \gamma_{0k}(1 - Z_k))^{-1})^{-1},$$

$$\sigma^2 \mid \cdot \sim IG(a_\sigma + n/2, b_\sigma + \frac{1}{2} \sum_{k=1}^{2} \|Y_k - X_k \beta_k\|^2),$$

$$p(\tau \mid \cdot) \propto \prod_{k=1}^{2} \prod_{i \in U_k} N(y_i \mid x_i' \beta_k, \sigma^2) \times Unif(\tau \mid a_\tau, b_\tau),$$

$Z_{kj} \mid \cdot \sim Bernoulli(\pi_{kj})$, where

$$\pi_{kj} = \frac{q_k \phi(\beta_{kj} / \sqrt{\sigma^2 \gamma_{1k}})}{q_k \phi(\beta_{kj} / \sqrt{\sigma^2 \gamma_{1k}}) + (1 - q_k) \phi(\beta_{kj} / \sqrt{\sigma^2 \gamma_{0k}})}$$

where $\phi(\cdot)$ denotes the density of standard normal distribution. We observe that the full conditionals of β_k, Z_{kj} and σ^2 follow conjugate distributions and are easily updated via the Gibbs' sampler. Only $p(\tau \mid \cdot)$ does not correspond to any standard likelihood and we use a Metropolis-Hastings random walk step within the Gibbs' sampler to update τ.

3.2 Multiple change points

A highlight of our approach is the ease with which the setup in Section 3.1 can be extended to multiple change points. This is important as although many methods exist for single change points, analyzing datasets with multiple change points are trickier and few methods generalize to this case. If we have K change points $\tau_1 < \ldots < \tau_K$, the joint likelihood in (2) can be generalized to

$$\prod_{k=1}^{K+1} \left(\prod_{i: \tau_{k-1} < t_i \leq \tau_k} N(y_i \mid x_i' \beta_k, \sigma^2) \times N(\beta_k \mid 0, \sigma^2 diag(\gamma_{1k} Z_k + \gamma_{0k}(1 - Z_k))) \right. \quad (3)$$
$$\left. \times \prod_{j=1}^{p} Bernoulli(Z_{kj} \mid q_k) \right) \times p(\tau_1, \tau_2, \ldots, \tau_K) \times IG(\sigma^2 \mid a_\sigma, b_\sigma).$$

To ensure identifiability of the change points, the prior $p(\tau_1, \tau_2, \ldots, \tau_K)$ should be supported on $\tau_1 < \tau_2 < \ldots < \tau_K$. The Gibbs' sampler remains essentially same as in Section 3 with the Metropolis random walk step now being used to update the entire change point vector $(\tau_1, \tau_2, \ldots, \tau_K)'$.

3.3 Determining the number of change points

Often in applications, the number of change points is unknown. In our fully Bayesian approach this can potentially be handled by adding a prior for the number of change points (K). Introducing this additional level of hierarchy comes with the caveat that different values of K yields parameter sub-spaces of different sizes and interpretations. [25] proposed the extremely general and powerful reversible jump MCMC (RJMCMC) sampler for sampling across multiple parameter spaces of variable dimensions. We can seamlessly adopt an RJMCMC joint sampler to obtain the posterior distribution for the number of change points. When naively implemented, RJMCMC experiences poor acceptance rates for transitions to parameter sub-spaces with different dimensionality. This leads to widely documented convergence issues [26, 19]. The problem will be exacerbated in our setup due to the high-dimensionality of the parameter spaces.

Several improvements and alternatives to RJMCMC have been proposed over the years including efficient proposal strategies to effectuate frequent cross-dimensional jumps [42, 7, 17, 20], product space search [9, 16] and parallel tempering [31]. All these approaches can be adapted in our setup to determine the number of change points. However, many of these approaches are accompanied by their own computational burden such as running several chains or apriori obtaining posterior distributions for each individual model before running the joint sampler. We concur with [27] and [28] that it is often expedient to use simpler model selection approaches based on individual models. Hence, popular Bayesian model comparison metrics like DIC [44] and posterior predictive loss [23] remains relevant to select the number of change points in our case. For example, if θ is the complete set of parameters associated with the model, for each K we can compute the DIC score

$$\text{DIC} = 2E(D(y \mid \theta) \mid y) - D(y \mid E(\theta \mid y)) = E(D(y \mid \theta) \mid y) + p_D \quad (4)$$

where $D(y \mid \theta)$ is the deviance function and $p_D = E(D(y \mid \theta) \mid y) - D(y \mid E(\theta \mid y))$ is interpreted as effective sample size. The DIC penalizes more complex models and is particularly appropriate for our change point problem, where higher number of change points will lead to overfitting. Parallel computing enables us to simultaneously run the MCMC sampler for different values of K. Subsequently, the optimal K is selected as the one yielding lowest DIC score.

All the methods for selecting the number of change points discussed here can be used in conjunction with our approach. It is prudent to predicate the choice on the nature of the application at hand and the computational resources available.

3.4 Alternate prior choices

We observe from Equation 2 that, conditional on the value of the change point τ, the joint likelihood can be decomposed into individual likelihoods for the regression before and after the change point along with the corresponding priors for the regression coefficients. This allows for a lot of flexibility in the choice of priors for the regression coefficients. One can also use other priors to achieve variable selection. For example, using Laplace (double exponential) priors for the β_k's will yield a Bayesian lasso [39] with change point detection. However, unlike spike-and-slab priors, Bayesian lasso does not perform model selection.

Additional information regarding grouping or structuring of the variables is often available in the context of variable selection. In the presence of a change point, additional constraints can specify grouped selection both within and/or between the β_k's. For example, in a single change point setup, it may be plausible that the set of relevant variables remain unchanged before and after the change point, with change occurring only with respect to the strength of association between y_i and x_i. Such additional structural constraints both within and across β_k's can easily be accommodated in our setup via a suitable choice of prior. To elucidate, we can rewrite (1) as $y_i = z_i(\tau)'\zeta + \epsilon_i$ where $z_i = \left(I(t_i \leq \tau)x_i', I(t_i > \tau)x_i'\right)'$ and $\zeta = \left(\beta_1', \beta_2'\right)'$. To incorporate the constraint that β_1 and β_2 share the same support, one can use a Bayesian group lasso [40] with M-Laplace priors on the groups $\zeta_j = (\beta_{1j}, \beta_{2j})'$ for $j = 1, 2, \ldots, p$. The M-Laplace prior

$$p(\zeta_j \mid \sigma^2, \lambda^2) \propto \frac{2\lambda^2}{\sigma^2}\exp\left(-\sqrt{\frac{2\lambda^2}{\sigma^2}}\|\zeta_j\|_2\right)$$

has a convenient two-step hierarchical specification:

$$\begin{aligned} \zeta_j \mid \eta_j &\overset{ind}{\sim} N(0, \sigma^2\eta_j I); \\ \eta_j \mid \lambda^2 &\overset{ind}{\sim} Gamma\left(3/2, \lambda^2\right); \lambda^2 \sim Gamma(r, s) \end{aligned} \tag{5}$$

The full conditional distributions of the parameters provided in [40] can now be used to implement the Gibbs' sampler with the additional Metropolis random walk step for updating the change point τ. Any other information like hierarchical selection or anti-hierarchical selection both within and between the β_k's can also be accommodated via suitable priors.

Often, in real data applications, prior knowledge dictates the inclusion of certain variables in the model and variable selection is sought only for the remaining variables. Such constraints can be easily achieved in our setup by using standard Gaussian prior for that specified subset and spike-and-slab prior for the remaining variables.

3.5 Variable selection after MCMC

When there are finitely many candidate models, Bayesian model selection typically proceeds by selecting the candidate model with the highest posterior probability. However, in our setup the regression coefficients are continuous. For variable selection, we use the median

probability model [4] which is computationally easy and is optimal in terms of prediction. To be specific, β_{kj} is included in the model if the posterior probability of $Z_{kj} = 1$ is greater than 0.5.

4. NUMERICAL STUDIES

We conducted numerical experiments to assess the performance of our method both for single and multiple change points. For assessing convergence and determining length of MCMC run for all the simulation scenarios, we generated data from one of the simulation settings in the single change point model and ran three MCMC chains of length 10,000 with different initial values. Satisfactory convergence was confirmed using a combination of trace plots, posterior density plots and Gelman-Rubin shrink factors. Similarly using three MCMC chains for the two change point model, satisfactory convergence was achieved in 30,000 iterations. Nonetheless, for all the simulations we were more conservative and ran a single chain of 100, 000 iterations.We discarded the first 50,000 as burn-in and used the subsequent 50, 000 samples for inference.

4.1 One change point

We assume $t_i = i$ and generate data from the model $y_i = N(x_i'\beta_1, \sigma^2)$ for $i \leq \tau$ and $y_i = N(x_i'\beta_2, \sigma^2)$ for $i > \tau$ where $\beta_1 = (3, 1.5, 0, 0, 2, 0, \ldots, 0)$ and $\beta_2 = -\beta_1$. The rows of X were independent and identically distributed normal random variables with zero mean and covariance Σ_X. Two structures were used for Σ_X — auto-regressive (AR) with $\Sigma_{X,ij} = 0.5^{|i-j|}$ and compound symmetry (CS) with $\Sigma_{X,ij} = 0.5 + 0.5I(i = j)$. The noise variance σ^2 was fixed at 1 and the sample size was chosen to be 200. Two different model sizes, $p = 250$ and $p = 500$, were used. The change point τ was chosen to vary among 50, 100 and 150. Since the sample size is 200, these three choices of τ correspond to a change point at the initial, middle or later portion of the data, respectively. For each combination of parameter choices, we generated 100 Monte Carlo replicates. We used three different models – the Bayesian change points models using the spike-and-slab prior (BSAS), the Bayesian Group lasso (BGL) prior [40], and a frequentist change point lasso (FL) which, for a given value of the change point, estimates the lasso-penalized regression coefficients for each segment. The optimal value of the change point for FL is selected using cross-validation. This is a variant of the method proposed in [30] and is much simpler to implement. The Bayesian group lasso model was used to investigate any possible benefits of using a grouped variable selection as it is known here that β_1 and β_2 has same support. The range of the uniform prior for τ was chosen to be (20,180) and a normal proposal density with tuning variance of 0.1 was used for the Metropolis update of τ. The prior for σ^2 was chosen to be $IG(2, 1)$. The hyper-parameters γ_{0k}, γ_{1k} and q_k were chosen as follows. Let τ_0 denote the initial estimate for τ. Then $n_1 = [\tau_0]$ and $n_2 = n - n_1$ denotes the initial sample sizes for the two segments. We used $\gamma_{0k} = \frac{\hat{\sigma}_k^2}{10n_k}$ and $\gamma_{1k} = \hat{\sigma}_k^2 \max\left(\frac{p^{2.1}}{100n_k}, \log n_k\right)$ where $\hat{\sigma}_k^2$ was the sample variance of Y_k for $k = 1, 2$. The hyper-parameters q_k were chosen such that the prior model sizes $\sum_{j=1}^{p} Z_{kj}$ were greater than $\min(p-1, \max(10, \log n_k))$ with probability 0.1. These choices of γ_{0k}, γ_{1k} and q_k were adapted from [35].

Let C_k and IC_k denote the number of true and false regressors respectively selected for the k^{th} segment of the data for $k = 1, 2$. As discussed earlier, we used a cutoff of 0.5 for the posterior probability of the binary Z_{kj}'s in the spike-and-slab model to select the variables. The Bayesian group lasso is devoid of such binary selection parameters and variable selection was based on the posterior credible intervals i.e β_{kj} was not selected if its posterior credible interval covered zero. In addition to the variable selection metrics, we also assess the three methods based on the coefficient estimates for the true predictors. Let $\beta_k[\mathcal{S}_k]$ denote the sub-vector of β_k corresponding to the true support $\mathcal{S}_k$. We use the Mean Squared Error $MSE_k = \|\beta_k[\mathcal{S}_k] - \hat{\beta}_k[\mathcal{S}_k]\|_2^2$ for $k = 1, 2$ where $\hat{\beta}_k$ denote the posterior estimate of β_k.

Table 1 presents the *C*, *IC* and *MSE* numbers along with estimates of τ for all the three models. The numbers are averages over 100 Monte Carlo simulations. We only present the numbers for the scenarios with $p = 500$. The analogous set of results for $p = 250$ were similar. Firstly, we observe that both Bayesian models estimate the change point τ with high accuracy while the frequentist lasso is slightly inaccurate in some instances. The *C* and *IC* numbers reveal that the spike-and-slab prior achieves perfect variable selection for all the scenarios while group lasso prior performs nearly as well. The selection accuracy for the group lasso was somewhat surprising given the crude variable selection technique used. The frequentist approach, although almost always accurately selecting the true set of regressors, tends to select a large number of uncorrelated regressors as indicated by the high *IC* numbers.

In terms of estimating the regression coefficients β_k, once again we observe that spike-and-slab prior stands out with uniformly lowest *MSE* numbers across all scenarios. The *MSE* for the frequentist method is significantly higher indicating that the estimates corresponding to the true predictors gets considerably shrunk. It is important to note that, for all models, MSE_1 tends to be higher when $\tau = 50.5$ while MSE_2 is higher when $\tau = 150.5$. This behavior is expected as for $\tau = 50.5$, sample size for estimating β_1 is effectively 50 while that for β_2 is 150. Overall, the spike-and-slab priors produced the most accurate variable selection, change point detection and estimation across all scenarios while the frequentist lasso consistently performs worst.

4.2 Two change points

We demonstrate the applicability of our method to multiple change points using a two change point setup. The three coefficient vectors are given by

$$\beta_1 = (3, 0, 0, \ldots, 0)', \beta_2 = (3, 1.5, 0, 0, \ldots, 0)',$$
$$\beta_3 = (3, 1.5, 0, 0, 2, 0, 0, \ldots, 0)'$$

The change in β_k's at each change point in this simulation study is much lesser than what was used in Section 4.1. Three pairs of values for the change points (τ_1, τ_2) are selected — (50,100), (50,150) and (100,150). Other specifications including sample size, model size and covariance of the predictors are kept unchanged from Section 4.1. We do not use the Bayesian group lasso here as the coefficient vectors for different segments do not share a

common support. Hence, we only present the results for the other two methods. Table 2 presents the C, IC and MSE numbers along with estimates of τ_k's for all the two models. The numbers are averages over 50 Monte Carlo simulations. Once again, we only present the numbers for the scenarios with $p = 500$ as the numbers for $p = 250$ were similar.

We observe from Table 2 that for the two change point model the frequentist approach becomes quite erratic in terms of estimating the change points. The Bayesian model remains much more accurate, except for the scenario with $\tau_1 = 100$ and $\tau_2 = 150$ where the estimates are slightly off. In general, the estimates of the change points are less accurate than those in Section 4.1. This is not surprising as the effective sample size per segment for the two-change point model is smaller. Also, the change in the regression coefficients are also of smaller magnitude, thereby making it harder to detect the change points accurately. Turning to variable selection, the Bayesian model identified the exact set of predictors across all scenarios almost in all the Monte Carlo simulations. The frequentist lasso, once again accurately identified the true predictors but produced a high false discovery rate. The MSE numbers for were also an order of magnitude lower for the Bayesian model.

5. MINNESOTA HOUSE PRICE INDEX DATA

All the data used in the analysis were publicly available. We use quarterly Minnesota hpi data published by the Federal Housing Finance Agency (FHFA). The macro-economic indices used as explanatory variables include national unemployment rate (unemp) and national consumer price indices (cpi). Instead of including a national stock index in the model like the S&P 500 or the Dow Jones Industrial Average, we use the stock prices of Minnesota based Fortune 500 companies. 14 out of the 18 Minnesota-based Fortune 500 companies were publicly traded since before 1991 and we include their stock prices in the regression model. Additionally, the list of top 10 employers in Minnesota include Wal-Mart Stores Inc. and Wells Fargo Bank Minnesota. Hence, the stock prices of these two companies are also included in the model. The 16 stocks used in total are listed in Table 3.

Financial indices often exhibit strong autocorrelation and consequently autoregressive components commonly feature in house price models [34]. Figure 2 plots the partial autocorrelation values of the hpi time series as a function of the lag. We observe that the index lagging one quarter behind (AR(1)) has very high correlation with the hpi time series but it quickly falls off beyond the first lag and all the subsequent lags have insignificant partial correlations. Hence, we include only the AR(1) term in the regression model.

Statistical analysis involving economic time series is often preceded by customary seasonality adjustment of the indices using standard techniques. It is well known that house price time series reveal a predictable and repetitive pattern with systematic highs in summer and lows in winter [36]. Consequently, publishers of popular house price indices like the FHFA or Standard and Poor's (Case-Shiller index) produce a version of their indices discounting this effect [21]. However, Minnesota is a land of extreme climates experiencing one of the widest range of temperatures in U.S. It is of interest to investigate if the impact of weather in Minnesota on its house prices extends beyond the routine pattern. Hence, we

include the state level quarterly average temperatures (temp) and precipitation (precip) in the model.

Our model for the k^{th} segment $\tau_{k-1} \leq t \leq \tau_k$ is given by:

$$hpi_t = \beta_k^{intercept} + \beta_k^{ar(1)} hpi_{t-1} + \beta_k^{cpi} \text{cpi}_t + \beta_k^{unemp} \text{unemp}_t + \beta_k^{temp} temp_t + \beta_k^{\text{precip}} \text{precip}_t + \beta_k^{\text{stocks}'} \text{stocks}_t + \epsilon_t \quad (6)$$

Here stocks_t denote the 16×1 vector formed by stacking up the stock prices at time t of the companies listed in Table 3 and β_k^{stocks} is the corresponding coefficient vector.

5.1 Results

We used data from the second quarter of 1991 to the second quarter to 2014 for model fitting. The first quarter data of 1991 was used for the AR(1) term, whereas the data for last two quarters of 2014 and first quarter of 2015 were held out for out-of-sample validation. Under the assumption of K change points, separate regression models are fit to each of the $K+1$ segments. Higher values of $K (\geq 3)$ implies that average sample size for each segment $(n/(K+1))$ becomes small and the estimates obtained may not be reliable. Hence, we restrict ourselves to $K = 0$, 1 and 2 and fit model (6) with spike-and-slab priors for the coefficient vectors in each segment. Note that for $K = 0$ i.e. no change point, the model is simply the traditional Bayesian high-dimensional regression model. The models for different values of K were assessed based on their in-sample DIC score and out-of-sample root mean square predictive error (RMSPE) [49]. Due to the presence of the autoregressive term, out-of-sample forecasts were obtained using one-step-ahead predictions.

The models for $K = 0$, 1 and 2 are denoted by BSAS_0, BSAS_1 and BSAS_2 respectively. For comparison, we also used the frequentist lasso with one change point, referred as FL_1. Additionally, to elucidate why low dimensional analysis is not suitable for this data, we also used two low dimensional models — a low dimensional one change point linear model (LM_1) which is similar to FL_1 but uses classical least squares to estimate the coefficients for each τ_1, and a low dimensional Bayesian linear model BLM_1 with one change point [similar to 10] with normal Inverse gamma (NIG) priors for $(\beta_1, \beta_2, \sigma^2)$ and uniform prior for τ_1. Since the size of this dataset was considerably smaller than what is considered in the simulation settings, it sufficed to run the MCMC for 10, 000 iterations discarding the first 5, 000 as burn-in.

Table 4 contains the DIC scores (only for the Bayesian models), RMSPE values and estimated change points for all the models. Both the DIC score and the RMSPE score for $K = 0$ were significantly worse than the scores for $K = 1$ and 2 justifying the use of a change point model. The single change point model detected change around late 2008- early 2009 which coincides with the sub-prime mortgage crisis. The two change point model detected

change points in mid 2006 and early 2011. The DIC score for the single change point model was substantially better. RMSPE scores for change point models for time series data only validate the accuracy of the models after the last change point. We observed that the RMSPE score were similar for $K = 1$ and 2 with the former turning out to be marginally better. The FL_1 model also estimated a change point around mid 2008 but the RMSPE score was higher than our Bayesian model.

The low dimensional models LM_1 and BLM_1 were also able to detect a change point in 2008. However, their model evaluation metrics were significantly worse. This is not surprising as a change point in mid 2008 leaves less than 25 observations to estimate a 22-dimensional vector β_2. In a low dimension approach like linear least squares and, to a lesser extent, in Bayesian linear model, this will lead least to unstable estimates. This elucidates that in-spite of $n = 93$ being sufficiently larger than $p = 22$, in presence of a change point, the location of the change point may warrant a regularized approach to ensure numerically stable analysis. Table 4 also provides the credible intervals for the estimated change points for the Bayesian models. The frequentist approach does not provide credible intervals of the change points.

We present the subsequent analysis only for the single change point model $BSAS_1$ as both in-sample and out-of-sample validations provide strongest evidence in favor of a single change point. Figure 3 plots the probability of selection for each of the regressors in model (6) before and after the change point. We observe that the set of variables selected by the median probability model differs across the two segments. The AR(1) index and precipitation are selected with high probabilities in both segments. However, the selection of stocks differ considerably on either side of the change point. We see that prior to change point in 2008, there was little correlation between hpi and stocks with only General Mills (GIS) having a posterior median probability close to 0.5 (0.498). Perhaps this is a reflection of the fact discussed earlier that stock prices and hpi did not exhibit comovements during the early 2000s. After the change point in 2008 the stocks of 3M (MMM), Medtronic (MDT) and Xcel Energy (XEL) are selected with high probability.

Table 5 presents the coefficient estimates for the variables selected either before or after the change point. We observe that the value for the coefficient corresponding to the AR(1) index drops significantly post change point indicating less autoregressive behavior after the change point. We also observe a positive association of hpi with precipitation. Since summer months witness significantly higher precipitation than winter, this merely corroborates the traditional ‘hot season cold season’ trend of house prices. A more interesting observation from Table 5 is the fact that this effect is much more pronounced after 2008 indicating more disparity between summer and winter house prices in the post-recession market.

Observe from Figure 3 that in presence of the stock prices of Minnesota based companies, national level macroeconomic indicators like the cpi or unemployment were not selected in the model. This perhaps provides evidence in support of the conjecture that hpi is strongly correlated with local macro-economics [22]. However, its worthwhile to point out that multivariate regression models, although a simple and powerful tool to determine correlation, rarely implies causality. Any confirmatory assessment of the change points

detected and the variables selected by our method would require further research. Nevertheless, the model evaluation metrics in Table 4 provide very strong evidence in favor of one or more change points thereby justifying the use of our methodology to analyze the data.

6. CONCLUSION

We have demonstrated the effectiveness of combining existing Bayesian shrinkage priors in a changing linear regression setup. The fully Bayesian approach offers several inferential advantages including quantifying uncertainty regarding the change points as well as variable selection for each segment. This framework is flexible to the choice of variable selection priors although the spike-and-slab prior empirically outperformed other competing choices. A wide range of constrained variable selections like grouping or partial selection can be seamlessly accomplished in this setup. The Bayesian method consistently and substantially outperformed the change point lasso for all the simulated and real data analyses. This, in fact, corroborates the findings by [35] in alternate settings not involving changing linear regression. The analysis of the Minnesota hpi data using our methodology revealed strong evidence for a potential change point with respect to the association of Minnesota house price with other macro-economic variables.

Many methods exist for detection of single change point in regression. These include long-established control chart-based tests for change points using arithmetic sums of residuals (CUSUM) [38, 8] and geometric sums (EWMA) [43]. However, we are not aware of any literature that extends these testing procedures to a high-dimensional regression setting. Also, we believe that instead of tests for change points that predicate decision-making on single p-values, a Bayesian approach providing adequate uncertainty information about the location of change points is probably more useful and realistic. Furthermore, these control-chart-based methods do not extend to multiple change points. Our method, as demonstrated through simulations and the hpi data analysis, works for both single and multiple change points, and outperforms the frequentist change point lasso [30] by significant margins for both these scenarios. The performance of our method for multiple change points is in particular very promising as few methodologies exist for multiple change points. If the number of change points is 3 or more, the cross-validation for the frequentist change point lasso will become exponentially more expensive. It will be much easier to use our Bayesian approach.

We have discussed several approaches for handling unknown number of change points. However, most of them comes with statutory warnings regarding computational requirements. More efficient models and algorithms for simultaneous detection of number of change points need to be researched. Other potential extensions include accommodating missing data, measurement errors or non-Gaussian responses in a high-dimensional changing regression setup. Extensions to change point detection in high-dimensional VAR models also need to be explored due to the extensive usage of VAR models in economics research [15, 6]. In a time series context, our work is restricted to detecting historical change points. Detecting future change points in high-dimensional time series is equally important to provide accurate predictions. We identify all these areas as directions for future research.

Also, implementing this methodology as a software package in a widely accessible platform like R would be a first priority.

Acknowledgments

*We thank the Editors, the Associate Editor and Referees for their suggestions that improved the article considerably. Sudipto Banerjee's work was supported, in part, by grants NIH/NIEHS 1R01ES027027–01, NSF IIS-1562303 and NSF DMS-1513654.

REFERENCES

[1]. Adams RP AND MacKay DJ (2007). Bayesian online change-point detection. arXiv preprint arXiv:0710.3742.

[2]. Ahamada I and Diaz Sanchez JL (2013). A Retrospective Analysis of the House Prices Macro-relationship in the United States. The World Bank: Policy Research Working Paper Series July, 6549 http://ideas.repec.org/p/wbk/wbrwps/6549.html.

[3]. Baladandayuthafani V, Ji Y, Talluri R, Nietu-Barajas LE, AND Morris JS (2010). Bayesian random segmentation models to identify shared copy number aberrations for array cgh data. Journal of the American Statistical Association 105, 492, 1358–1375. 10.1198/jasa.2010.ap09250. [PubMed: 21512611]

[4]. Barbieri MM AND Berger JO (2004). Optimal predictive model selection. Annals of Statistics 32, 870–897.

[5]. Barry D and Hartigan JA (1993). A bayesian analysis for change point problems. Journal of the American Statistical Association 88, 421, 309–319. http://www.jstor.org/stable/2290726.

[6]. Bernanke B. s., Boivin J, and Eliasz P (2004). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. Working Paper 10220, National Bureau of Economic Research. January http://www.nber.org/papers/w10220.

[7]. Brooks SP, Giudici P, and Roberts GO (2003). Efficient construction of reversible jump markov chain monte carlo proposal distributions. Journal of the Royal Statistical Society: Series B 65, 1, 3–39. 10.1111/1467-9868.03711.

[8]. Brown RL, Durbin J, and Evans JM (1975). Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society. Series B (Methodological), 149–192.

[9]. Carlin BP AND Chib s. (1995). Bayesian model choice via markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B 57, 473–484.

[10]. Carlin BP, Gelfand AE, and Smith AFM (1992). Hierarchical bayesian analysis of changepoint problems. Journal of the Royal Statistical Society: Series C 41, 2, 389—405. http://www.jstor.org/stable/2347570.

[11]. Carvalho CM, Polson NG, and Scott JG (2010). The horseshoe estimator for sparse signals. Biometrika 97, 2, 465–480.

[12]. Case KE, Shiller RJ, and Quigley JM (2001). Comparing wealth effects: The stock market versus the housing market Working Paper 8606, National Bureau of Economic Research 11 http://www.nber.org/papers/w8606.

[13]. Catte P, Girouard N, Price R, and Andre C (2004). Housing Markets, Wealth and the Business Cycle. OECD Working Paper, No 394.

[14]. Chen J and Gupta AK (1997). Testing and locating variance changepoints with application to stock prices. Journal of the American Statistical Association 92, 438, 739–747. http://www.jstor.org/stable/2965722.

[15]. Christiano LJ, Eichenbaum M, and Evans CL (2000). Monetary policy shocks: What have we learned and to what end? In Handbook of Macroeconomics, Taylor J and Woodford M, Eds. Elsevier, Amsterdam, Holland, 65–148. http://www.sciencedirect.com/science/article/pii/S1574004899010058.

[16]. Dellaportas P, Forster JJ, and Ntzoufras I (2002). On bayesian model and variable selection using mcmc. Statistics and Computing 12, 27–36.

[17]. Ehlers RS and Brooks s. P. (2008). Adaptive proposal construction for reversible jump mcmc. Scandinavian Journal of Statistics 35, 4, 677–690. 10.1111/j.1467-9469.2008.00606.x.

[18]. Fan J and Li R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.

[19]. Fan Y and Sisson, s. (2011). Reversible jump markov chain monte carlo In Handbook of Markov Chain Monte Carlo, Brooks SP, Gelman A, Jones GL, and Meng X-L, Eds. Chapman & Hall/CRC, Boca Raton, FL, 67–87.

[20]. Farr WM, Mandel I, and Stevens D (2015). An efficient interpolation technique for jump proposals in reversible-jump markov chain monte carlo calculations. Royal Society Open Science 2, 6.

[21]. Federal Housing Finance Agency. (2014). U.s. house price index report - 3q 2014 / september 2014. Technical report 9.

[22]. Garmaise MJ and Moskowitz TJ (2002). Confronting information asymmetries: Evidence from real estate markets Working Paper 8877, National Bureau of Economic Research 4 http://www.nber.org/papers/w8877.

[23]. Gelfand AE and Ghosh SK (1998). Model choice: A minimum posterior predictive loss approach. Biometrika 85, 1–11.

[24]. George EI and McCulloch RE (1993). Variable selection via gibbs sampling. Journal of the American Statistical Association 88, 423, 881–889. http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476353.

[25]. Green PJ (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika 82, 4, 711–732. http://biomet.oxfordjournals.org/content/82/4/711.abstract.

[26]. Green PJ and Hastie DI (2009). Reversible jump mcmc http://people.ee.duke.edu/lcarin/rjmcmc20090613.pdf.

[27]. Han C and Carlin BP (2001). Markov chain monte carlo methods for computing bayes factors. Journal of the American Statistical Association 96, 455, 1122–1132. 10.1198/016214501753208780.

[28]. Hastie DI and Green PJ (2012). Model choice using reversible jump markov chain monte carlo. Statistica Neerlandica 66, 3, 309–338. 10.1111/j.1467-9574.2012.00516.x.

[29]. Ishwaran H and Rao JS (2005). Spike and slab variable selection: Frequentist and bayesian strategies. The Annals of Statistics 33, 2 (04), 730–773. 10.1214/009053604000001147.

[30]. Lee S, Seo MH, AND Shin Y (2016). The lasso for high dimensional regression with a possible change point. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78, 1, 193–210. 10.1111/rssb.12108. [PubMed: 27656104]

[31]. Littenberg TB and Cornish NJ (2009). Bayesian approach to the detection problem in gravitational wave astronomy. Physical Review 80, 063007–1–063007–19.

[32]. Mavrodiy A (2005). Factor analysis of real estate prices. EERC MA thesis.

[33]. McCulloch RE AND Tsay R. s. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. Journal of the American Statistical Association 88, 423, 968–978. http://www.jstor.org/stable/2290788.

[34]. Nagaraja CH, Brown LD, and Zhao LH (2011). An autoregressive approach to house price modeling. The Annals of Applied Statistics 5, 1 (03), 124–149. 10.1214/10-AOAS380.

[35]. Narisetty NN AND He X (2014). Bayesian variable selection with shrinking and diffusing priors. The Annals of Statistics 42, 2 (04), 789–817. 10.1214/14-A0S1207.

[36]. Ngai LR AND Tenreyro s. (2014). Hot and cold seasons in the housing market. American Economic Review 104, 12, 3991–4026. http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.12.3991.

[37]. Ong TS AND Chang Y. s. (2013). Macroeconomic determinants of malaysian housing market. Journal of Human and Social Science Research 1, 119–127.

[38]. Page E. s. (1954). Continuous inspection schemes. Biometrika 41, 1/2, 100–115.

[39]. Park T and Casella G (2008). The bayesian lasso. Journal of the American Statistical Association 103, 482, 681–686. 10.1198/016214508000000337.

[40]. Raman s., Fuchs TJ, Wild PJ, Dahl E, and Roth V (2009). The bayesian group-lasso for analyzing contingency tables In Proceedings of the 26th Annual International Conference on Machine Learning. ICML '09. ACM, New York, NY, USA, 881–888. http://doi.acm.org/10.1145/1553374.1553487.

[41]. Reeves J, Chen J, Wang X, Lund R, and Lu Q (2007). A review and comparison of changepoint detection techniques for climate data. Journal of Applied Meteorology and Climatology 46, 6, 900–915.

[42]. Richardson S and Green PJ (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). Journal of the Royal Statistical Society: Series B 59, 4, 731–792. 10.1111/1467-9868.00095.

[43]. ROBERTS S (1959). Control chart tests based on geometric moving averages. Technometrics 1, 3, 239–250.

[44]. SPIEGELHALTER DJ, BEST NG, CARLIN BP, AND VAN DER LiNDE A (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B 64, 583–639.

[45]. Strömberg P, Hedman M, and Broberg M (2011). Forecasting the house price index in stockholm county 2011–2014: A multiple regression analysis of four influential macroeconomic variables. Bachelor's thesis from Mlardalen University / School of Sustainable Development of Society and Technology.

[46]. Tibshirani R (1994). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B 58, 267–288.

[47]. Turner R, Saatci Y, and Rasmussen C (2009). Adaptive Sequential Bayesian Change Point Detection.

[48]. Wang J AND Zivot E (2000). A bayesian time series model of multiple structural changes in level, trend, and variance. Journal of Business & Economic Statistics 18, 3, 374–386. http://amstat.tandfonline.com/doi/abs/10.1080/07350015.2000.10524878.

[49]. Yeniay O AND Goktas A (2002). A comparison of partial least squares regression with other prediction methods. Hacettepe Journal of Mathematics and Statistics 31, 99–111.

[50]. Zou H (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.

[51]. Zou H AND Hastie T (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B 67, 301–320.

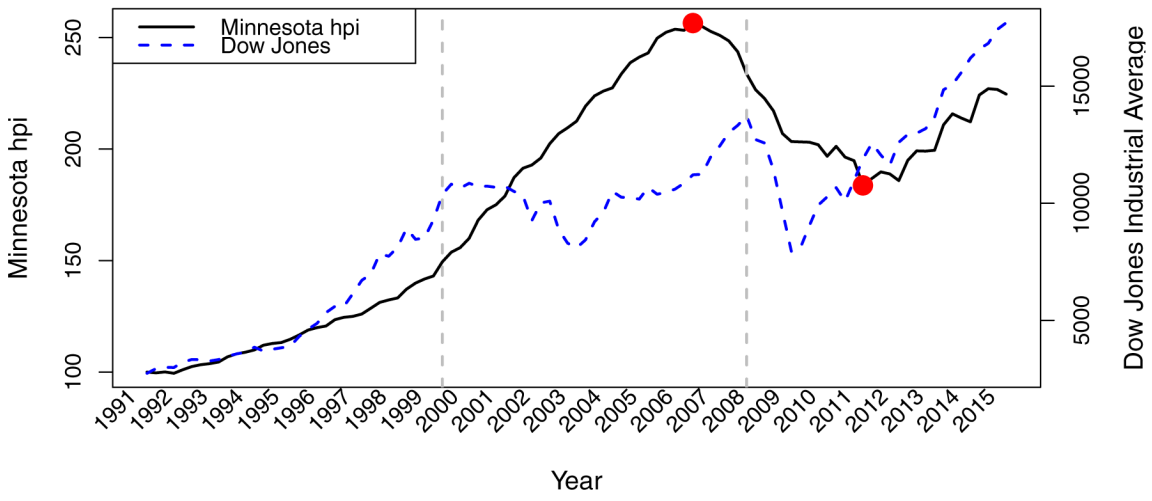

Figure 1.

Minnesota hpi time series (left y-axis) and Dow Jones Industrial Average (right y-axis). Red dots indicate eyeballed change-points in terms of temporal trend of hpi. Grey lines indicate eyeballed change-points with respect to relationship of the two series.

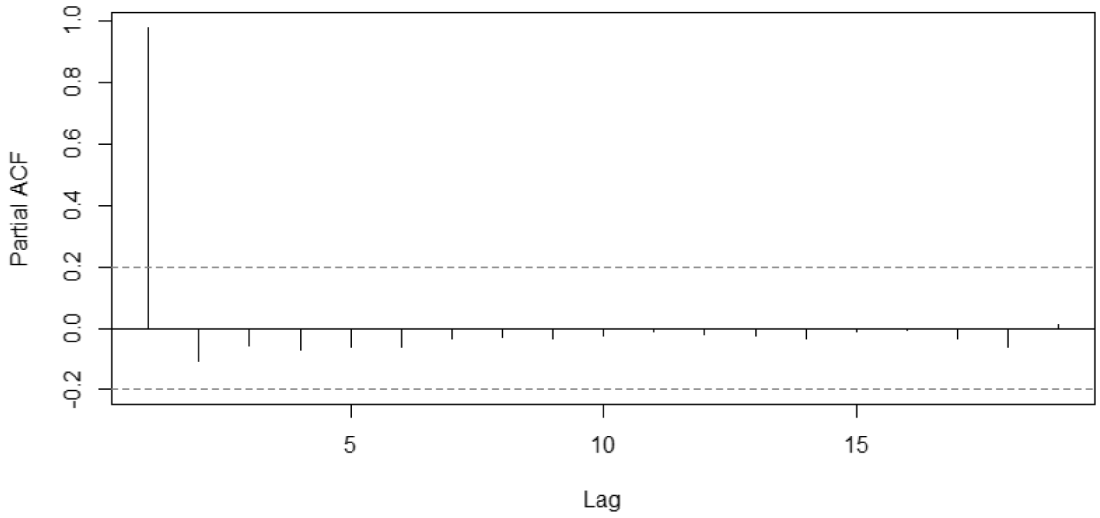

Figure 2.
Partial autocorrelation function for Minnesota hpi time series

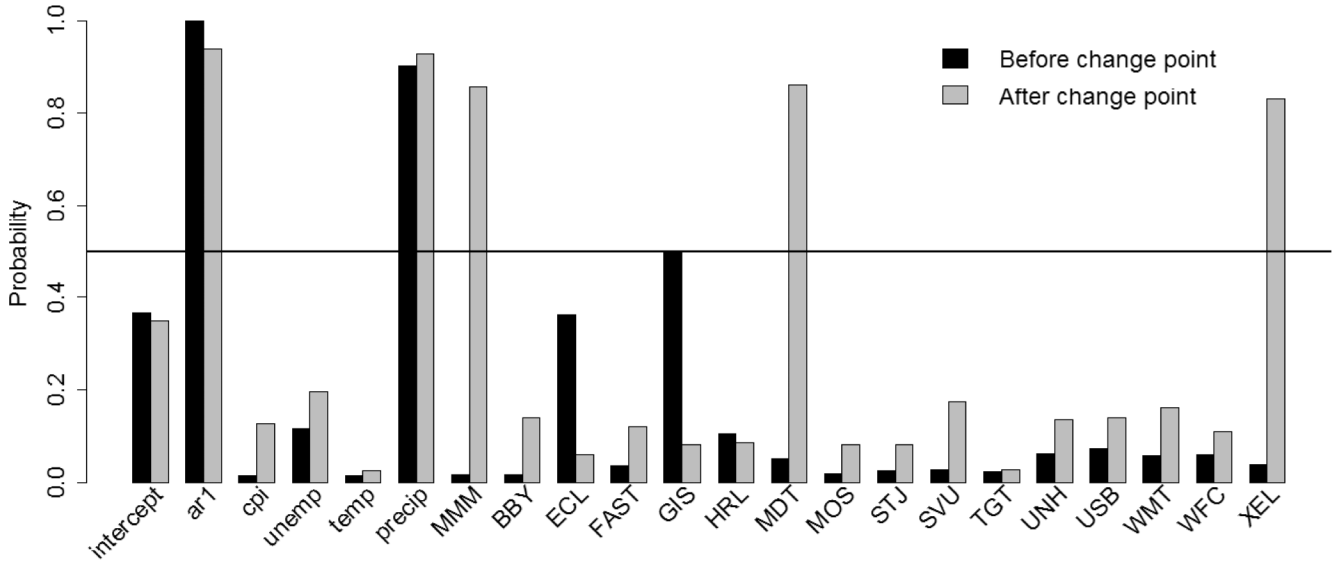

Figure 3.

Minnesota hpi analysis: Posterior median probabilities of variable selection using single change point model

Stat Interface. Author manuscript; available in PMC 2019 September 20.

Table 1.

Summary statistics for one change point model for p = 500

		$\tau = 50$						
	Model	$\hat{\tau}$	C_1	C_2	IC_1	IC_2	MSE_1	MSE_2
AR	BSAS	50.5	3.0	3.0	0.1	0.0	0.45	0.22
	FL	57.0	2.8	3.0	13.3	1.7	3.37	3.20
	BGL	50.5	3.0	3.0	0.0	0.0	1.39	0.54
CS	BSAS	50.5	3.0	3.0	0.0	0.0	0.55	0.27
	FL	51.1	2.6	3.0	18.2	22.7	3.39	3.09
	BGL	50.5	2.6	3.0	0.0	0.0	1.86	0.81
		$\tau = 100$						
	Model	$\hat{\tau}$	C_1	C_2	IC_1	IC_2	MSE_1	MSE_2
AR	BSAS	100.4	3.0	3.0	0.0	0.0	0.28	0.27
	FL	99.6	3.0	3.0	6.6	3.3	3.20	3.21
	BGL	100.4	3.0	3.0	0.0	0.0	0.77	0.75
CS	BSAS	100.4	3.0	3.0	0.0	0.0	0.34	0.32
	FL	100.5	3.0	3.0	21.6	20.9	3.17	3.16
	BGL	100.4	3.0	3.0	0.0	0.0	1.13	1.11
		$\tau = 150$						
	Model	$\hat{\tau}$	C_1	C_2	IC_1	IC_2	MSE_1	MSE_2
AR	BSAS	150.3	3.0	3.0	0.0	0.0	0.23	0.41
	FL	145.1	3.0	2.8	3.6	10.3	3.18	3.31
	BGL	150.4	3.0	3.0	0.0	0.0	0.53	1.34
CS	BSAS	150.2	3.0	3.0	0.0	0.1	0.28	0.55
	FL	149.8	3.0	2.8	23.9	16.5	3.11	3.34
	BGL	150.4	3.0	2.7	0.0	0.0	0.79	1.81

Table 2.

Summary statistics for two change point model for p = 500

		$\tau_1 = 50,\ \tau_2 = 100$										
	Model	$\hat{\tau}_1$	$\hat{\tau}_2$	C_1	C_2	C_3	IC_1	IC_2	IC_3	MSE_1	MSE_2	MSE_3
AR	BSAS	51.5	102.0	1	2.0	3.0	0.4	0.0	0.0	0.15	0.35	0.28
	FL	67.2	132.9	1	2.0	2.9	13.3	32.2	10.1	2.95	3.00	3.18
cs	BSAS	50.5	101.8	1	1.9	3.0	0.6	0.1	0.0	0.23	0.52	0.37
	FL	62.9	136.5	1	1.9	2.7	17.2	35.5	21.8	2.94	3.11	3.27
		$\tau_1 = 50,\ \tau_2 = 150$										
	Model	$\hat{\tau}_1$	$\hat{\tau}_2$	C_1	C_2	C_3	IC_1	IC_2	IC_3	MSE_1	MSE_2	MSE_3
AR	BSAS	51.6	150.3	1	2.0	3.0	0.1	0.0	0.0	0.15	0.23	0.45
	FL	75.1	146.7	1	2.0	2.7	14.4	30.0	7.8	2.93	3.11	2.99
CS	BSAS	51.1	150.9	1	2.0	2.8	0.2	0.0	0.0	0.21	0.28	0.69
	FL	60.4	143.9	1	1.9	2.5	17.4	40.9	22.2	2.89	3.15	3.19
		$\tau_1 = 100,\ \tau_2 = 150$										
	Model	$\hat{\tau}_1$	$\hat{\tau}_2$	C_1	C_2	C_3	IC_1	IC_2	IC_3	MSE_1	MSE_2	MSE_3
AR	BSAS	89.3	145.0	*1*	1.8	2.9	0.1	0.1	0.0	0.13	0.59	0.60
	FL	71.2	139.6	1	1.9	2.8	9.3	34.6	9.9	2.92	3.09	3.04
CS	BSAS	94.8	146.0	1	1.8	2.7	0.1	0.0	0.0	0.16	0.58	0.90
	FL	69.1	152.4	1	1.6	2.5	16.4	44.0	18.8	2.96	3.18	3.23

Table 3.

List of stocks used in Minnesota hpi analysis

Company Name	Ticker Symbol	Company Name	Ticker Symbol
3M Company	MMM	St. Jude Medical, Inc.	STJ
Best Buy Co., Inc.	BBY	SuperValu, Inc.	SVU
Ecolab, Inc.	ECL	Target Corporation	TGT
Fastenal Co.	FAST	UnitedHealth Group Inc.	UNH
General Mills, Inc.	GIS	U.S. Bancorp	USB
Hormel Foods Corporation	HRL	Wal-Mart Stores, Inc.	WMT
Medtronic Pic.	MDT	Wells Fargo & Company	WFC
Mosaic Company	MOS	Xcel Energy Inc.	XEL

Table 4.

Minnesota hpi analysis: DIC, RMSPE scores and estimated change points

Model	DIC	RMSPE	$\hat{\tau}_1$	$\hat{\tau}_2$
$BSAS_0$	287	4.72		
$BSAS_1$	250	2.17	2008Q4 (2008Q2, 2009Q1)	
$BSAS_2$	269	2.23	2006Q2 (2005Q4, 2007Q3)	2011Q1 (2010Q4, 2011Q2)
FL_1		2.71	2008Q2	
BLM_1	288	10.00	2008Q3 (2008Q2, 2008Q4)	
LM_1		23.31	2008Q2	

Table 5.

Minnesota hpi analysis: Posterior median for the selected coefficients. (* indicates that the variable was selected in that segment)

	Before τ	**After τ**			**Before τ**	**After τ**
AR(1)	1.034*	0.761*		GIS	−0.151*	−0.002
precip	0.964*	1.952*		MDT	0.071	1.214*
MMM	0.019	−0.573*		XEL	−0.005	1.279*