Bayesian Estimation of Changepoints in a Partially Observed Latent Process Poisson Model

Chigozie Utazi
(Joint work with Dr. Peter Neal)

School of Mathematics
The University of Manchester

March, 2012

Outline

Background

Two classes of models exist for time series data involving Poisson counts:

- Observation-driven models: Lagged values of observed counts included in the mean function.
  Example: The INAR(p) model,
  $X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \epsilon_t$, where "$\circ$" denotes an operator, e.g. $\alpha \circ X \sim$ Binomial(X, α) and $\epsilon_t \sim$ *iid* Po(λ).
- Parameter-driven models: A latent process governs the mean function.
  Example: Zeger's (1988) model,
  $X_t | Y_t \sim$ Po(exp$(\mathbf{z}_t'\boldsymbol{\beta} + Y_t)$), where $E(\exp(Y_t)) = 1$.

Some features of parameter-driven models

- A stochastic model is postulated for the latent process (An extension of the Poisson regression model).
- The latent process accounts for overdispersion and autocorrelation in the model.
- Easy to interpret and derive model properties, but difficult to estimate.
- The model provides a framework for exchange of dynamics between the count process and the underlying latent process.

Motivation

$$X_t|Y_t, \mathbf{z}_t \sim \text{Po}(\exp(f(\mathbf{z}_t) + \omega Y_t)),$$

$$\text{where} \quad Y_t = \alpha Y_{t-1} + e_t \quad \text{and} \quad e_t \sim N(0, \sigma^2 = 1/\tau).$$

Process of interest: $\{Y_t, t = 1, \ldots, n\}$

- If $y_1, y_2, \ldots, y_n$ are fully observed, the x_t's are uninformative.
- If $y_1, y_2, \ldots, y_n$ are partially observed, the x_t's provide additional information (provided $\omega \neq 0$).

Motivating Example:

Temporal analysis of air pollution and health:

- Estimating the association between some health outcomes and air pollution;
- Estimating the parameters of a partially observed pollution variable;
- Detection of changes in one or both variables.

Our Contribution

- Model Formulation
- Parameter Estimation Procedure
- Changepoint Estimation

The Basic Latent Process Poisson Model

Model Specification

Observed Counts: $X_1, X_2, \ldots, X_n$

Latent Variable: $Y_1, Y_2, \ldots, Y_n$

Covariates: $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$, $\mathbf{z}_i = (1, z_{i,1}, z_{i,2}, \ldots, z_{i,p-1})$

The Model:

$$X_t | Y_t, \mathbf{z}_t \sim \text{Po}(\exp(\mathbf{z}_t' \boldsymbol{\beta} + \omega Y_t)), \tag{1}$$

where $Y_t = \alpha Y_{t-1} + e_t$, $e_t \sim N(0, \sigma^2 = 1/\tau)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$. To ensure stationarity in the latent process, it is assumed that $|\alpha| < 1$.

- Note that $\omega = 0 \Rightarrow$ a Poisson regression model.

Bayesian Estimation of Parameters

Let $\boldsymbol{\theta} = (\omega, \boldsymbol{\beta}, \alpha, \tau)$ denote the vector of parameters of the model and $\pi(\boldsymbol{\theta})$ denote its joint prior distribution. The **likelihood function** (conditional on Y_1)is given by

$$L(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}, \mathbf{z}) = \prod_{t=1}^{n} \frac{\exp(\mathbf{z}_t'\boldsymbol{\beta} x_t + \omega y_t x_t - e^{\mathbf{z}_t'\boldsymbol{\beta} + \omega y_t})}{x_t!} \times \prod_{t=2}^{n} \frac{\tau^{1/2}}{\sqrt{2\pi}} \exp(-\frac{\tau}{2}(y_t - \alpha y_{t-1})^2). \quad (2)$$

Priors: $\beta_i \sim N(u_1, 1/v_1), \quad i = 0, \ldots, p-1$, $\alpha \sim U(-1, 1)$, $\omega \sim N(u_2, 1/v_2)$ and $\tau \sim \text{Gamma}(a, b)$. The parameters are assumed to be apriori independent.

Given our choice of priors, the posterior distribution of $\boldsymbol{\theta}$ can now be written as:

$$\pi(\boldsymbol{\theta}|data) \propto \prod_{t=1}^{n} \text{exp}(\mathbf{z}_t'\boldsymbol{\beta} x_t + \omega y_t x_t - e^{\mathbf{z}_t'\boldsymbol{\beta}+\omega y_t}) \times \prod_{t=2}^{n} \tau^{1/2}\text{exp}(-\frac{\tau}{2}(y_t - \alpha y_{t-1})^2) \times e^{-\frac{v_1}{2}\sum_{i=0}^{p-1}(\beta_i - u_1)^2} \times e^{-\frac{v_2}{2}(\omega - u_2)^2} \times \tau^{a-1}e^{-b\tau}. \quad (3)$$

Conditional posterior distributions of the parameters:

$$\pi(\beta_i|\boldsymbol{\theta}_{-\beta_i},\boldsymbol{x},\boldsymbol{y},\mathbf{z}) \propto \exp\left(\beta_i\sum_{t=1}^{n} z_{ti}x_t - \sum_{t=1}^{n} e^{\mathbf{z}_t'\boldsymbol{\beta}+\omega y_t} - \frac{v_1}{2}(\beta_i^2 - 2u_1\beta_i)\right); \quad (4)$$

$$\pi(\omega|\boldsymbol{\theta}_{-\omega},\boldsymbol{x},\boldsymbol{y},\mathbf{z}) \propto \exp\left(\sum_{t=1}^{n}\omega y_t x_t - \sum_{t=1}^{n} e^{\mathbf{z}_t'\boldsymbol{\beta}+\omega y_t} - \frac{v_2}{2}(\omega^2 - 2u_2\omega)\right) \quad (5)$$

$$\pi(\alpha|\boldsymbol{\theta}_{-\alpha},\boldsymbol{x},\boldsymbol{y},\mathbf{z}) \sim N\left(\frac{\tau\sum_{t=2}^{n} y_t y_{t-1}}{\tau\sum_{t=2}^{n} y_{t-1}^2}, \frac{1}{\tau\sum_{t=2}^{n} y_{t-1}^2}\right), I(|\alpha|<1); \quad (6)$$

$$\pi(\tau|\boldsymbol{\theta}_{-\tau},\boldsymbol{x},\boldsymbol{y},\mathbf{z}) \sim \text{Gamma}\left(a + \frac{(n-1)}{2}, \frac{\sum_{t=2}^{n}(y_t - \alpha y_{t-1})^2}{2} + b\right). \quad (7)$$

Estimation of missing values in the latent process

Noting that by Markov property, $P(y_t|\boldsymbol{y}_{-t}) \propto P(y_t|y_{t-1})P(y_{t+1}|y_t)$, the conditional posterior distribution of y_t for $t = 2, \ldots, n-1$ can easily be derived from eqn (3) as

$$\pi(y_t|\boldsymbol{y}_{-t}, \boldsymbol{x}, \mathbf{z}, \boldsymbol{\theta}) \propto \exp(-\frac{\tau}{2}(y_t - \alpha y_{t-1})^2) \times \exp(-\frac{\tau}{2}(y_{t+1} - \alpha y_t)^2) \times \exp(\omega y_t x_t - e^{\mathbf{z}_t'\boldsymbol{\beta}+\omega y_t}). \quad (8)$$

We use the independent sampler to update the missing values.

Specifically, we use the Gaussian proposal density

$$q(y_t|\boldsymbol{y}_{-t},\theta) \sim N\left(\frac{\alpha(y_{t-1}+y_{t+1})}{1+\alpha^2}, \frac{1}{\tau(1+\alpha^2)}\right) \tag{9}$$

with acceptance probability

$$\alpha(y_t \rightarrow y_t') = min\left(1, \frac{\exp(\omega y_t' x_t - e^{\omega y_t' + \mathbf{z}_t'\boldsymbol{\beta}})}{\exp(\omega y_t x_t - e^{\omega y_t + \mathbf{z}_t'\boldsymbol{\beta}})}\right), \tag{10}$$

where y_t' denotes the proposed value of y_t. The proposal density given in equation 9 was determined to yield the best estimates based on pilot runs. The **proposal density** for Y_n is $N(\alpha y_{n-1}, 1/\tau)$.

MCMC Algorithm for the Basic Model

- Initialize the parameters and the missing values in $\boldsymbol{y}$,
- Update β,
- Update ω,
- Update α,
- Update τ,
- Update missing $\boldsymbol{y}$ values,
- Repeat steps 2-6 until a desired number of iterations is reached.

Simulation experiments using the Basic Model

The simulation studies were designed to examine how the model performs and compares with the AR(1) model under the following conditions:

- Different patterns of missingness in the latent process,
- Varying values of ω,
- High, moderate and low autocorrelation in the latent process.

Does the inclusion of X_t in the model lead to any improvement in parameter estimation in Y_t?

Simulation study using the Basic Model

The hyperparameters were chosen as follows:

$$\begin{aligned}\beta_i &\sim N(0,1)\\ \omega &\sim N(0.2,1/5)\\ \alpha &\sim U(-1,1)\\ \tau &\sim \text{Gamma}(1,1)\end{aligned}$$

Initial guesses for the missing values in y were drawn from $N(0,1)$.

Table: Comparing the Basic Model with an AR(1) Model (Regularly Missing Data)

		AR(1)[1]		LPPM[2]	
Amount of Missingness	Parameter (True value)	Posterior Mean (Std. Dev.)	95% Credible Interval	Posterior Mean (Std. Dev.)	95% Credible Interval
90%	$\beta_0 = 0.5$	-	-	0.5344 (0.1105)	(0.3178,0.7510)
	$\omega = 0.5$	-	-	0.5303 (0.0471)	(0.4379,0.6226)
	$\alpha = 0.5$	-0.0013 (0.3354)	(-0.6587,0.6561) -	0.4158 (0.1765)	(0.0700,0.7617)
	$\tau = 4.0$	3.1093 (0.6783)	(1.7798,4.4388)	3.6290 (0.9201)	(1.8256,5.4324)
75%	$\beta_0 = 0.5$	-	-	0.5288 (0.0826)	(0.3669,0.6907)
	$\omega = 0.5$	-	-	0.5372 (0.0435)	(0.4520,0.6225)
	$\alpha = 0.5$	-0.0011 (0.3312)	(-0.6502,0.6480) -	0.5177 (0.0930)	(0.3354,0.7000)
	$\tau = 4.0$	2.5635 (0.3599)	(1.8581,3.2689)	3.4915 (0.5859)	(2.3432,4.6398)

[1] AR(1) - First Order Autoregressive Model

[2] LPPM - Basic Latent Process Poisson Model

Sample size=400, No of iterations=100,000, Burn-in=10,000

Table: Comparing the Basic Model with an AR(1) Model (Data Missing at Random)

		AR(1)[1]		LPPM[2]	
Amount of Missingness	Parameter (True value)	Posterior Mean (Std. Dev.)	95% Credible Interval	Posterior Mean (Std. Dev.)	95% Credible Interval
90%	$\beta_0 = 0.5$	-	-	0.5125 (0.1215)	(0.2742,0.7517)
	$\omega = 0.5$	-	-	0.4829 (0.0475)	(0.3899,0.5759)
	$\alpha = 0.5$	0.1582 (0.2476)	(-0.3271,0.6435) -	0.3004 (0.1470)	(0.0124,0.5885)
	$\tau = 4.0$	3.0239 (0.6970)	(1.7798,4.4388)	3.1808 (0.6735)	(1.8608,4.5008)
75%	$\beta_0 = 0.5$	-	-	0.5129 (0.0977)	(0.3215,0.7043)
	$\omega = 0.5$	-	-	0.4553 (0.0473)	(0.3626,0.5480)
	$\alpha = 0.5$	0.2535 (0.1443)	(-0.0293,0.5363) -	0.5058 (0.0985)	(0.3129,0.6988)
	$\tau = 4.0$	3.7969 (0.5318)	(2.7546,4.8393)	4.1633 (0.6584)	(2.8728,5.4538)

[1] AR(1) - First Order Autoregressive Model

[2] LPPM - Basic Latent Process Poisson Model

Sample size=400, No of iterations=100,000, Burn-in=10,000

Table: Comparing the Basic Model with an AR(1) Model (Varying ω, 90% Data Missing [1])

Parameter (True value)	$\omega = 0.2$ AR(1)[2] Post. Mean (Std. Dev.)	$\omega = 0.2$ LPPM[2] Post. Mean (Std. Dev.)	$\omega = 0.5$ AR(1) Post. Mean (Std. Dev.)	$\omega = 0.5$ LPPM Post. Mean (Std. Dev.)	$\omega = 0.8$ AR(1) Post. Mean (Std. Dev.)	$\omega = 0.8$ LPPM Post. Mean (Std. Dev.)
ω	-	0.0365 (0.0947)	-	0.5147 (0.1213)	-	0.8057 (0.1218)
$\beta_0 = 0.5$	-	0.4811 (0.0402)	-	0.5288 (0.0476)	-	0.5997 (0.0476)
$\alpha = 0.5$	0.0079 (0.3695)	-0.0280 (0.3412)	-0.0025 (0.3166)	0.4045 (0.1907)	0.0066 (0.3159)	0.4163 (0.1334)
$\tau = 4.0$	2.6813 (0.5822)	3.0578 (0.7938)	3.1134 (0.6725)	3.6055 (0.9323)	4.2152 (0.9292)	4.2959 (0.9493)

[2] AR(1) - First Order Autoregressive Model LPPM - Basic Latent Process Poisson Model

[1] Data missing regularly, Sample size=400, No of iterations=100,000, Burn-in=10,000

Table: A Comparison between the Basic Model and the AR(1) Model (Varying ω, 75% Data Missing [1])

Parameter (True value)	$\omega = 0.2$ AR(1)[2] Post. Mean (Std. Dev.)	$\omega = 0.2$ LPPM[2] Post. Mean (Std. Dev.)	$\omega = 0.5$ AR(1) Post. Mean (Std. Dev.)	$\omega = 0.5$ LPPM Post. Mean (Std. Dev.)	$\omega = 0.8$ AR(1) Post. Mean (Std. Dev.)	$\omega = 0.8$ LPPM Post. Mean (Std. Dev.)
ω	-	0.0409 (0.0810)	-	0.5290 (0.0821)	-	0.7943 (0.0980)
$\beta_0 = 0.5$	-	0.4823 (0.0395)	-	0.5376 (0.0433)	-	0.5856 (0.0466)
$\alpha = 0.5$	-0.0036 (0.2573)	-0.0296 (0.2655)	0.0103 (0.2778)	0.5196 (0.0940)	-0.0116 (0.4120)	0.4507 (0.0974)
$\tau = 4.0$	3.1294 (0.4391)	3.4249 (0.6062)	2.5687 (0.3510)	3.5088 (0.5955)	4.1138 (0.5764)	4.5210 (0.7304)

[2] AR(1) - First Order Autoregressive Model LPPM - Basic Latent Process Poisson Model

[1] Data missing regularly, Sample size=400, No of iterations=100,000, Burn-in=10,000

Some Comments:

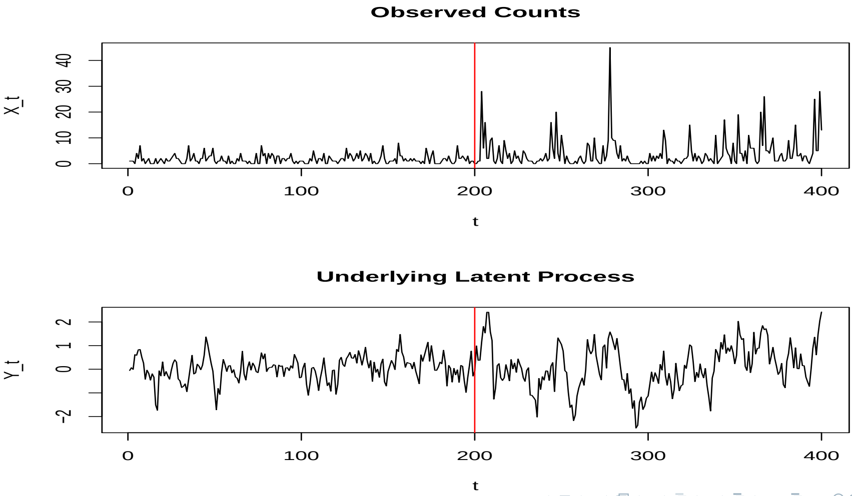
-Varying the sample size has no effect on parameter estimation.

- It was also observed from simulation experiments (though not presented) using $\alpha = 0.2, 0.5, 0.8$ that estimation of parameters in the AR(1) model did not improve even with very high correlation in Y.

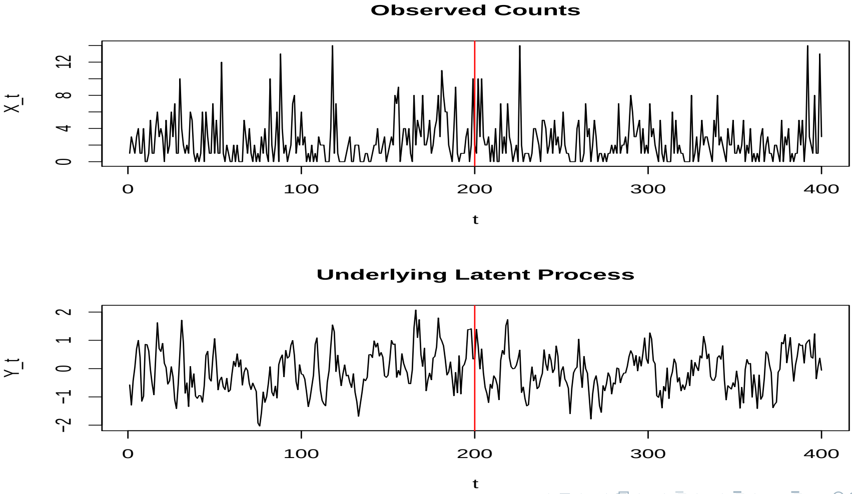
- In all, even though the AR(1) model is seen to be competing with our Basic LPPM in the estimation of τ, the latter yields estimates that are closer to the true parameter values.

Latent Process Poisson Model with a Changepoint

An illustration (Very clear changepoint)



An illustration (Changepoint quite unclear)



An illustration - Missing values (Any changepoint?)

The Model with a changepoint in all the parameters

Notation: X_t - Observed Counts, Y_t - Partially-Observed Latent Process, $\mathbf{z}_t$ -Covariate Vector and m - Unknown Changepoint

The Model:

$$X_t|Y_t, \mathbf{z}_t \sim \begin{cases} \text{Po}(\exp(\mathbf{z}_t'\boldsymbol{\beta}_1 + \omega_1 Y_t)) \quad \text{for} \quad 1 \leq t \leq m; \\ \qquad Y_t = \alpha_1 Y_{t-1} + e_{1,t} \quad \text{and} \quad e_{1,t} \sim N(0, 1/\tau_1); \\ \\ \text{Po}(\exp(\mathbf{z}_t'\boldsymbol{\beta}_2 + \omega_2 Y_t)) \quad \text{for} \quad m+1 \leq t \leq n; \\ \qquad Y_t = \alpha_2 Y_{t-1} + e_{2,t} \quad \text{and} \quad e_{2,t} \sim N(0, 1/\tau_2); \end{cases} \tag{11}$$

where $\boldsymbol{\beta}_1 = (\beta_{1,0}, \beta_{1,1}, \ldots, \beta_{1,p-1})$ and $\boldsymbol{\beta}_2 = (\beta_{2,0}, \beta_{2,1}, \ldots, \beta_{2,p-1})$ are regression coefficients. We assume that $-1 < \alpha_1, \alpha_2 < 1$ to ensure stationarity in the latent process.

Parameter Estimation

Priors: Same as given previously for the basic model. That is, Normal priors for $\beta_{1,i}$, $\beta_{2,i}$,ω_1, ω_2; Uniform priors for α_1, α_2; and Gamma priors for τ_1, τ_2. For m, we use a Discrete Uniform$(2, n-1)$ prior.

Likelihood:

$$L_c \propto \prod_{t=1}^{m} \exp(\mathbf{z}_t' \boldsymbol{\beta}_1 x_t + \omega_1 y_t x_t - e^{\mathbf{z}_t' \boldsymbol{\beta}_1 + \omega_1 y_t}) \times \prod_{t=m+1}^{n} \exp(\mathbf{z}_t' \boldsymbol{\beta}_2 x_t + \omega_2 y_t x_t - e^{\mathbf{z}_t' \boldsymbol{\beta}_2 + \omega_2 y_t}) \times \prod_{t=2}^{m} \tau_1^{1/2} e^{-\frac{\tau_1}{2}(y_t - \alpha_1 y_{t-1})^2} \times \prod_{t=m+1}^{n} \tau_2^{1/2} e^{-\frac{\tau_2}{2}(y_t - \alpha_2 y_{t-1})^2} \quad (12)$$

Joint Posterior Distribution of the parameters

Let $\boldsymbol{\theta}_c = (\omega_1, \omega_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \alpha_1, \alpha_2, \tau_1, \tau_2, m)$ denote the vector of parameters to be estimated. The **joint posterior distribution** is

$$\begin{aligned}
\pi(\theta_c|data) \propto & \prod_{t=1}^{m} \exp(\mathbf{z}_t' \boldsymbol{\beta}_1 x_t + \omega_1 y_t x_t - e^{\mathbf{z}_t' \boldsymbol{\beta}_1 + \omega_1 y_t}) \\
& \times \prod_{t=m+1}^{n} \exp(\mathbf{z}_t' \boldsymbol{\beta}_2 x_t + \omega_2 y_t x_t - e^{\mathbf{z}_t' \boldsymbol{\beta}_2 + \omega_2 y_t}) \times \\
& \times \prod_{t=2}^{m} \tau_1^{1/2} e^{-\frac{\tau_1}{2}(y_t - \alpha_1 y_{t-1})^2} \times \prod_{t=m+1}^{n} \tau_2^{1/2} e^{-\frac{\tau_2}{2}(y_t - \alpha_2 y_{t-1})^2} \\
& \times e^{-\frac{v_1}{2} \sum_{i=1}^{2} \sum_{j=0}^{p-1} (\beta_{i,j} - u_1)^2} \times e^{-\frac{v_2}{2} \sum_{i=1}^{2} (\omega_i - u_2)^2} \\
& \times (\tau_1 \tau_2)^{a-1} e^{-b(\tau_1 + \tau_2)}. \qquad (13)
\end{aligned}$$

Conditional Posterior Distribution of m

Let $t_1 = 1(2)$, $t_2 = m$ when $i = 1$ and $t_1 = m+1$, $t_2 = n$ when $i = 2$.

$$p(m|\theta_{-m}, \boldsymbol{x}, \boldsymbol{y}, \mathbf{z}) = \frac{p(\theta_{-m}, m; \boldsymbol{x}, \boldsymbol{y}, \mathbf{z})}{\sum_{j=1}^{n} p(\theta_{-m}, j; \boldsymbol{x}, \boldsymbol{y}, \mathbf{z})}; \qquad (14)$$

where

$$p(\theta_{-m}, m; \boldsymbol{x}, \boldsymbol{y}, \mathbf{z}) \propto e^{\sum_{t=1}^{m}(\mathbf{z}_t'\boldsymbol{\beta}_1 x_t + \omega_1 y_t x_t - e^{\mathbf{z}_t'\boldsymbol{\beta}_1 + \omega_1 y_t})} \times e^{\sum_{t=m+1}^{n}(\mathbf{z}_t'\boldsymbol{\beta}_2 x_t + \omega_2 y_t x_t - e^{\mathbf{z}_t'\boldsymbol{\beta}_2 + \omega_2 y_t})} \times \tau_1^{\frac{m-1}{2}} \tau_2^{\frac{n-m}{2}} \times e^{-\frac{\tau_1}{2}\sum_{t=2}^{m}(y_t - \alpha_1 y_{t-1})^2} \times e^{-\frac{\tau_2}{2}\sum_{t=m+1}^{n}(y_t - \alpha_2 y_{t-1})^2}.$$

Note: The full conditionals of other parameters have the same form as those of the Basic Model.

Estimating the Missing Values in the Latent Process

- Missing values in Y occurring before and after m are treated as given in the basic model.

- Y_n is also treated as given in the basic model.

- A special case arises when y_{m+1} is missing; for which the proposal density is $q(y_t \to y_t') \sim N\left(\frac{\tau_1\alpha_1 y_{t-1}+\tau_2\alpha_2 y_{t+1}}{\tau_1+\tau_2\alpha_2^2}, \frac{1}{\tau_1+\tau_2\alpha_2^2}\right)$ with acceptance probability $min\left(1, \frac{\exp(\omega_2 y_t' x_t - e^{\omega_2 y_t' + \mathbf{z}_t'\boldsymbol{\beta}_2})}{\exp(\omega_2 y_t x_t - e^{\omega_2 y_t + \mathbf{z}_t'\boldsymbol{\beta}_2})}\right)$.

Model Selection using RJMCMC

- We use the RJMCMC algorithm to choose between the basic latent process Poisson model (now Model 1)and the Latent process poisson model with a changepoint (now referred to as Model 2).
- In other words, we seek to answer the question: Is there any evidence of a changepoint?
- Model 1 parameters: $\theta = (\boldsymbol{\beta}, \omega, \alpha, \tau)$
- Model 2 parameters: $\theta_c = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \omega_1, \omega_2, \alpha_1, \alpha_2, \tau_1, \tau_2, m)$

Jump functions: Model 1$\rightarrow$ Model 2

- Generate the auxiliary variables (parameters): $a_i (i = 0, \ldots, p-1)$, b, c, d; each from $N(0, \sigma^2)$ and $k \sim$ Discrete Uniform$(2, n-1)$.
- Obtain the parameters of the proposed model (Model 2) as follows:

$$
\begin{aligned}
\beta_{1,i} &= \beta_i + (1-(m/n))a_i & \beta_{2,i} &= \beta_i - (m/n)a_i \\
\omega_1 &= \omega + (1-(m/n))b & \omega_2 &= \omega - (m/n)b \\
\alpha_1 &= \alpha + (1-(m/n))c & \alpha_2 &= \alpha - (m/n)c \\
\tau_1 &= \tau + (1-(m/n))d & \tau_2 &= \tau - (m/n)d \\
m &= k
\end{aligned} \tag{15}
$$

- We have used the idea of moment-matching (see Green, 1995) and weights obtained using m and n in defining the jump functions for the parameters.

Jump functions: Model 2→ Model 1

- Obtain the parameters of the proposed model (Model 1) as follows:

$$\begin{aligned} \beta_i &= (m/n)\beta_{1,i} + (1-(m/n))\beta_{2,i} & a_i &= \beta_{1,i} - \beta_{2,i} \\ \omega &= (m/n)\omega_1 + (1-(m/n))\omega_2 & b &= \omega_1 - \omega_2 \\ \alpha &= (m/n)\alpha_1 + (1-(m/n))\alpha_2 & c &= \alpha_1 - \alpha_2 \\ \tau &= (m/n)\tau_1 + (1-(m/n))\tau_2 & d &= \tau_1 - \tau_2 \\ k &= m \end{aligned} \quad (16)$$

- The new parameters are simply weighted averages of the corresponding parameters in Model 2 and are obtained by reversing the functions given previously.

Jacobian:Model 1→Model 2

$$J_{1\to 2} = \begin{pmatrix} \frac{\partial \beta_{1,0}}{\partial \beta_0} & \frac{\partial \beta_{2,0}}{\partial \beta_0} & \cdots & \frac{\partial \tau_2}{\partial \beta_0} & \frac{\partial m}{\partial \beta_0} \\ \frac{\partial \beta_{1,0}}{\partial a_0} & \frac{\partial \beta_{2,0}}{\partial a_0} & \cdots & \frac{\partial \tau_2}{\partial a_0} & \frac{\partial m}{\partial a_0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \beta_{1,0}}{\partial d} & \frac{\partial \beta_{2,0}}{\partial d} & \cdots & \frac{\partial \tau_2}{\partial d} & \frac{\partial m}{\partial d} \\ \frac{\partial \beta_{1,0}}{\partial k} & \frac{\partial \beta_{2,0}}{\partial k} & \cdots & \frac{\partial \tau_2}{\partial k} & \frac{\partial m}{\partial k} \end{pmatrix}. \qquad (17)$$

- The first $2p+6$ rows and columns form a block diagonal matrix with diagonal entries of the form: $\left[\frac{1}{(1-\frac{m}{n})} \; -\frac{1}{(\frac{m}{n})} \right]$.
- The last row and column have 1 as the diagonal entry and zeros elsewhere.
- The determinant is $|J_{1\to 2}| = 1^{(p+3)} \times 1 = 1$.
- The Jacobian for the reverse move is also 1.

Acceptance Probability (Model 1→Model2)

$$A_{1\to 2}(\theta,\theta_c) = \text{min}(1, \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \times |J_{1\to 2}|). \quad (18)$$

Likelihood ratio:

$$\frac{\exp(\sum_{t=1}^{m}(\mathbf{z}_t'\boldsymbol{\beta}_1 x_t + \omega_1 x_t y_t - e^{\mathbf{z}_t'\boldsymbol{\beta}_1+\omega_1 y_t}))\times\exp(\sum_{t=m+1}^{n}(\mathbf{z}_t'\boldsymbol{\beta}_2 x_t + \omega_2 x_t y_t - e^{\mathbf{z}_t'\boldsymbol{\beta}_2+\omega_2 y_t}))\times \tau_1^{\frac{m-1}{2}}\times e^{-\frac{\tau_1}{2}\sum_{t=2}^{m}(y_t-\alpha_1 y_{t-1})^2}\times\tau_2^{\frac{n-m}{2}}\times e^{-\frac{\tau_2}{2}\sum_{t=m+1}^{n}(y_t-\alpha_2 y_{t-1})^2}}{\exp(\sum_{t=1}^{n}(\mathbf{z}_t'\boldsymbol{\beta} x_t + \omega x_t y_t - e^{\mathbf{z}_t'\boldsymbol{\beta}+\omega y_t}))\times\tau^{\frac{n-1}{2}}\times e^{-\frac{\tau}{2}\sum_{t=2}^{n}(y_t-\alpha y_{t-1})^2}} \quad (19)$$

Prior ratio:

$$\frac{b^a (v_1)^{p/2}\sqrt{v_2} e^{-\frac{v_1}{2}\sum_{i=0}^{p-1}(\beta_{1,i}-u_1)^2} e^{-\frac{v_1}{2}\sum_{i=0}^{p-1}(\beta_{2,i}-u_1)^2} e^{-\frac{v_2}{2}\sum_{i=1}^{2}(\omega_i-u_2)^2} (\tau_1\tau_2)^{a-1} e^{-b(\tau_1+\tau_2)}}{2\Gamma(a)(2\pi)^{\frac{1+p}{2}}(n-2)e^{-\frac{v_1}{2}\sum_{i=0}^{p-1}(\beta-u_1)^2} e^{-\frac{v_2}{2}(\omega-u_2)^2}\tau^{a-1}e^{-b\tau}} \tag{20}$$

Proposal ratio:

$$\frac{(n-2)(2\pi)^{\frac{p+3}{2}}\sigma^{3+p}}{e^{-\frac{1}{2\sigma^2}(\sum_{i=0}^{p-1} a_i^2+b^2+c^2+d^2)}} \tag{21}$$

Note: The acceptance probability for the reverse move i.e. Model 2 $\rightarrow$ Model 1 is $A_{2\rightarrow 1}(\theta_c,\theta) = A_{1\rightarrow 2}^{-1}(\theta,\theta_c)$.

The Reversible Jump MCMC Algorithm

- Initialize all the parameters and the missing $\boldsymbol{y}$ values.
- If the current model is 1:
  a. Update $\boldsymbol{\beta}$, ω, α and τ
  b. Update the missing values in $\boldsymbol{y}$
- If the current model is 2:
  a. Update $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, ω_1, ω_2, α_1, α_2, τ_1 and τ_2
  b. Update m
  c. Update the missing values in $\boldsymbol{y}$
- Propose model switching as follows:

- If model=1:
  - Draw **a**, b, c, d and k .
  - Obtain the new parameters $\boldsymbol{\beta}_1'$, $\boldsymbol{\beta}_2'$, ω_1', ω_2', α_1', α_2', τ_1', τ_2' and m'.
  - Calculate $A_{1\rightarrow 2}$. If the move is accepted, switch to model 2 and set the model parameters equal to the proposed values as given above. Otherwise, remain in model 1.
- If model=2:
  - Obtain $\boldsymbol{\beta}'$, a, ω', b, α', c, τ', d and k.
  - Calculate $A_{2\rightarrow 1}$. If the move is accepted, switch to model 1 and set the model parameters equal to the proposed values as given above. Otherwise, remain in model 2.
- Repeat 2-4 for a desired number of iterations.

Simulation Experiments Using the Changepoint Model

The simulation experiments were designed to examine the performance of the model under the following conditions:

- Changepoints occuring at different positions in the data,
- Varying proportions of missingness in Y,
- Model selection and parameter estimation when there is no changepoint.

Table: Results of a simulation study investigating the estimation of changepoints at different positions(90% data missing in Y, Sample size=400)[1]

Parameter (True value)	m=100 Posterior Mean	m=100 Std. Deviation	m=200 Posterior Mean	m=200 Std. Deviation	m=300 Posterior Mean	m=300 Std. Deviation
$\beta_{1,0} = 0.6$	0.6858	0.1088	0.5304	0.0997	0.6394	0.1930
$\beta_{1,1} = 0.6$	0.5498	0.0747	0.6856	0.0692	0.5653	0.2930
$\omega_1 = 0.8$	0.7647	0.0837	0.7869	0.2606	0.8063	0.1671
$\alpha_1 = 0.8$	0.7010	0.1119	0.7366	0.1678	0.7660	0.2057
$\tau_1 = 2.0$	1.8210	0.3995	2.0108	0.5110	1.7416	0.4661
m	102.5919	10.4274	230.0713	27.8400	300.9430	9.4761
$\beta_{2,0} = 0.3$	0.3267	0.0537	0.3461	0.0731	0.3004	0.0559
$\beta_{2,1} = 0.3$	0.2927	0.0481	0.3228	0.0758	0.2548	0.0507
$\omega_2 = 0.5$	0.4191	0.1128	0.4107	0.2606	0.4522	0.2060
$\alpha_2 = 0.5$	0.4082	0.2323	0.4256	0.3280	0.4566	0.2446
$\tau_2 = 4.0$	3.7637	0.8721	3.0180	0.9922	3.2474	0.5464

[1] Number of iterations=100,000; Burn-in=10,000; Data assumed to be missing systematically; Algorithm initialized in Model 1; Percentage time spent in Model 2 after burn-in=100

Table: Results of a simulation study investigating the estimation of changepoints at different positions(75% data missing in Y, Sample size=400)[1]

Parameter (True value)	m=100 Posterior Mean	m=100 Std. Deviation	m=200 Posterior Mean	m=200 Std. Deviation	m=300 Posterior Mean	m=300 Std. Deviation
$\beta_{1,0} = 0.6$	0.6099	0.1053	0.6372	0.0825	0.6137	0.0777
$\beta_{1,1} = 0.6$	0.6207	0.0603	0.6865	0.0490	0.6987	0.1089
$\omega_1 = 0.8$	0.7562	0.0688	0.8369	0.0461	0.7873	0.0684
$\alpha_1 = 0.8$	0.7886	0.0668	0.8314	0.0454	0.7645	0.0781
$\tau_1 = 2.0$	2.4192	0.5388	1.7114	0.2725	1.6462	0.3796
m	117.7047	9.9526	200.2734	7.4865	310.9739	24.6959
$\beta_{2,0} = 0.3$	0.2112	0.0625	0.3412	0.0678	0.3513	0.0574
$\beta_{2,1} = 0.3$	0.2999	0.0529	0.3286	0.0550	0.3004	0.0492
$\omega_2 = 0.5$	0.5170	0.1334	0.4647	0.1395	0.4175	0.1474
$\alpha_2 = 0.5$	0.4208	0.1414	0.4423	0.2370	0.4081	0.1876
$\tau_2 = 4.0$	3.8594	0.7681	2.8856	0.7518	2.8935	0.6477

[1]Number of iterations=100,000; Burn-in=10,000; Data assumed to be missing systematically; Algorithm initialized in Model 1; Percentage time spent in Model 2 after burn-in=100

Figure: Histograms of the posterior distributions of m (90% data missing)

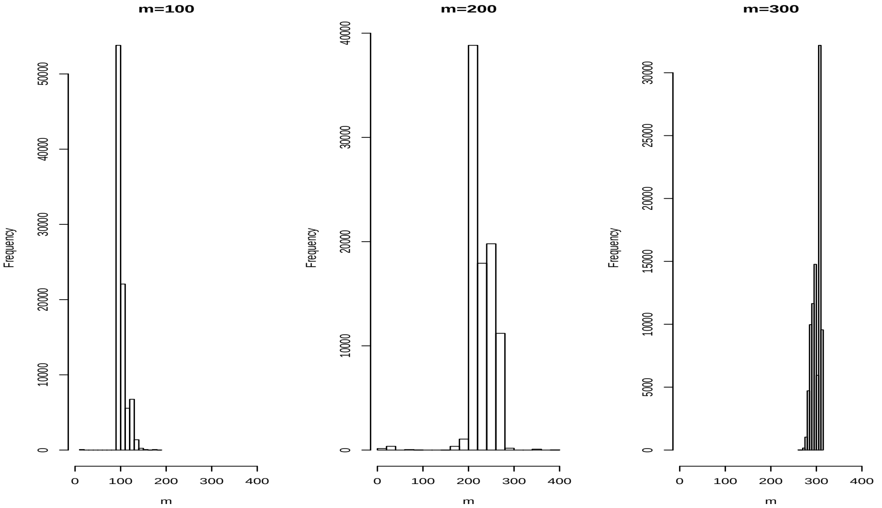


Figure: Histograms of the posterior distributions of m (75% data missing)

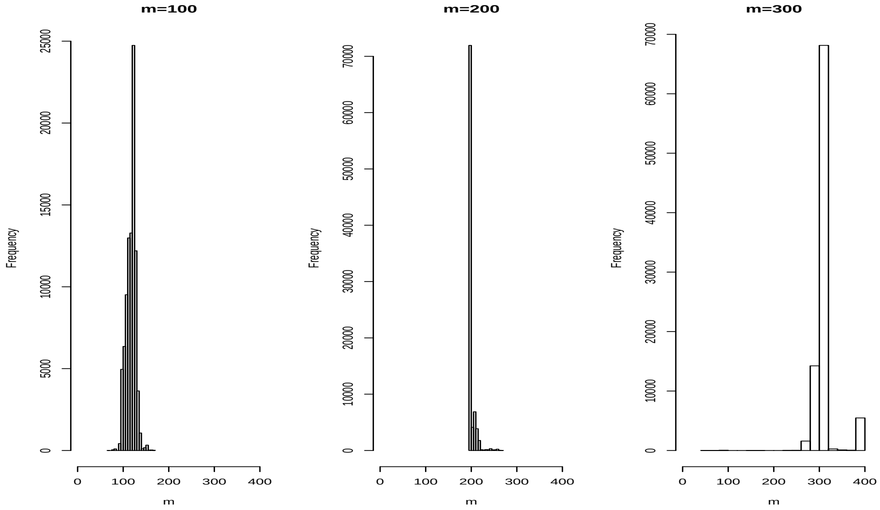


Table: Model and Parameter estimation in the absence of a changepoint using RJMCMC[1]

Amount of Missingness	Parameter (True value)	Posterior Mean	Posterior Std. Dev.	95% Credible Interval
	True Model=1	Post. prob.=0.986		
90%[2]	$\beta_0 = 0.3$	0.3259	0.0541	(0.2199,0.4319)
	$\beta_1 = 0.5$	0.4706	0.0445	(0.3835,0.5578)
	$\omega = 0.5$	0.4805	0.1811	(0.1256,0.8354)
	$\alpha = 0.5$	0.3522	0.2330	(-0.1045,0.8089)
	$\tau = 4.0$	3.5274	0.9153	(1.7335,5.3213)
	True Model=1	Post. prob.=0.995		
75%[3]	$\beta_0 = 0.3$	0.3082	0.0503	(0.2095,0.4068)
	$\beta_1 = 0.5$	0.4690	0.0434	(0.3840,0.5540)
	$\omega = 0.5$	0.5448	0.0977	(0.3533,0.7363)
	$\alpha = 0.5$	0.4134	0.1470	(0.1253,0.7015)
	$\tau = 4.0$	3.7221	0.6507	(2.4467,4.9974)

[1] Sample size=400, No of iterations=100,000, Burn-in=10,000

[2] Percent time spent in Model 1 after burn-in = 98.6%

[3] Percent time spent in Model 1 after burn-in = 99.5%

Extensions

- All or a subset of the covariates changing
- Changepoint in the latent process only
- Multiple Changepoints
- Extension to spatio-temporal models

References

1. Chen, K. S. and Ledolter J. (1995). Monte Carlo EM Estimation for Time Series Model Involving Counts. *Journal of the American Statistical Association*, Vol. 90, No. 429, pp. 242-252
2. Davis, R. A., Dunsmuir W. T. and Wang, Y. (1999). Modelling Time Series of Count Data. In Asymptotics, Nonparametrics and Time Series, Ed. S. Ghosh, pp. 63-114. New York: Marcel Dekker.
3. Davis, R. A., Dunsmuir W. T. and Wang, Y. (2000). On autocorrelation in a Poisson Regression Model. *Biometrika*, **87**, 3, pp. 491-505.
4. Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, **82**,4, pp. 711-32
5. Samet, J. M., Zeger, S. L., Dominici, F., Curriero, F., Coursac, I., Dockery, D. W., Schwartz, J. AND Zanobetti, A. (2000). The National Morbidity, Mortality, and Air Pollution Study, Part II: Morbidity and Mortality from Air Pollution in the United States. Cambridge, MA: Health Effects Institute.
6. Zeger, S. L. (1988). A Regression Model for Times Series of Counts. *Biometrika*, **75**,4, pp.621-9.

Thanks