Seeking Effective Adjustments for Effective Areas
Project Update 02/16/16

Presenter: Xufei Wang

Joint work with Yang Chen, Xiao-Li Meng, Vinay Kashyap, Herman Marshall, David van Dyk, Matteo Guainazzi

February 16, 2016
Recap of the Problem

Problem: Systematic errors in comparing effective areas.

Notations:
- Instruments \( \{1 \leq i \leq N\} \) with attributes \( \{A_i, 1 \leq i \leq N\} \).
- Sources \( \{1 \leq j \leq M\} \) with fluxes \( \{F_j, 1 \leq j \leq M\} \).
- Photon Counts \( \{C_{ij} = A_i F_j, 1 \leq i \leq N, 1 \leq j \leq M\} \) obtained from measuring flux \( F_j \) using effective area \( A_i \).

Questions:
1. How to adjust \( \{A_i, 1 \leq i \leq N\} \) such that \( \{C_{ij}/A_i, 1 \leq i \leq N\} \), the estimated \( F_j \) using observed values, agree with \( F_j \) within statistical uncertainty?
2. How to estimate the systematic error on the \( A_i \)'s?
Basic Model – Estimand Level

log-scale linear additive model

We start by noting a trivial fact that $C_{ij} = A_i F_j$ is mathematically equivalent to

$$\log C_{ij} = \log A_i + \log F_j = B_i + G_j,$$

where $B_i = \log A_i, \ G_j = \log F_j$.

However, this relationship holds at the *estimand* level, not at the *estimator/observation* level.

- Upper case: estimand $(A_i, F_j, B_i, G_j)$.
- Lower case: estimators / observations $(c_{ij}, a_i, b_i)$. 
Hierarchical regression model:

\[
y_{ij} = \log(c_{ij}) = \alpha_{ij} + B_i + G_j + \epsilon_{ij},
\]

(2)

where \(\epsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)\) independently; \(i \in \{1, \ldots, N\}\); \(j \in J_i = \{1 \leq j \leq M : c_{ij} \text{ is observed}\}\).

**Half-variance Correction:**

\(\alpha_{ij} = -0.5\sigma_{ij}^2\) is necessary to guarantee

\[E(c_{ij}) = C_{ij} = \exp(B_i + G_j) = A_i F_j.\]

**Priors:**

The prior for \(G_j\) is flat in \(\mathbb{R}\).

The prior for \(B_i\) is a Gaussian \(\mathcal{N}(b_i, \tau_i^2)\). \(b_i = \log a_i\) is known.
Hierarchical regression model likelihood function:

Let $D$ be our observed data $\{y_{ij}, b_i; 1 \leq i \leq N, j \in J_i\}$ and $y'_{ij} = y_{ij} + 0.5\sigma^2_{ij}$; \(\theta = \{B_i, G_j, i \in I, j \in J\}\) our estimand, i.e. parameters of interest; and \(\psi = \{\sigma^2_{ij}, \tau^2_i, i \in I, j \in J_i\}\), the nuisance parameters. We also denote \(I_j\) the collection of all \(i\)’s such that \(J_i\) covers \(j\).

The probability density of our data $D$ given $\theta$ and $\psi$ is

$$L(D|\theta, \psi) \propto \prod_{i=1}^{N} \prod_{j \in J_i} \left[ \frac{1}{\sigma_{ij}} e^{-\frac{(y'_{ij} - B_i - G_j)^2}{2\sigma^2_{ij}}} \right] \prod_{i=1}^{N} \left[ \frac{1}{\tau_i} e^{-\frac{(b_i - B_i)^2}{2\tau^2_i}} \right].$$
Complications with Real Data

A multiplicative factor due to pile-up

Let $Z_{ij}$ be the constant adjusting for the pile-up effect.

$$C_{ij} = Z_{ij}A_iF_j = Z_{ij} \exp(B_i + G_j).$$

(1) $Z_{ij} = z_{ij}$ is an observed constant.

$$y_{ij} = \log(c_{ij}) - \log(Z_{ij}) = \alpha_{ij} + B_i + G_j + \epsilon_{ij}.$$  

We only need to replace $y_{ij} = \log(c_{ij})$ with $\log(c_{ij}/Z_{ij})$.

(2) $Z_{ij}$ is observed with uncertainty. $Z_{ij}$ is a latent variable and the observations are $\log(z_{ij}) \sim \mathcal{N}(\log(Z_{ij}), \lambda^2)$.

$$\log(c_{ij}) - \log(z_{ij}) = \alpha_{ij} + B_i + G_j + \tilde{\epsilon}_{ij},$$

where $\text{Var}(\tilde{\epsilon}_{ij}) = \text{Var}(\epsilon_{ij}) + \lambda^2$. 
To estimate the $B_i$’s and $G_j$’s using observed data, we need to make assumptions on the variances to make sure the model is identifiable. Next, we will be focusing on two major assumptions which are practically reasonable.

1. **Known variance**: $\sigma_{ij}^2$ and $\tau_i^2$ are known constants.
2. **Unknown instrumental variance**: the noise term $\epsilon_{ij}$ only depends on the instrument-wise noise, i.e. $\sigma_{ij}^2 = \omega_i^2$; $\tau_i^2 = \tau^2$ for $1 \leq i \leq N$ is unknown.

**Remark**: The likelihood is unbounded in (2).
Model Fitting: MAP for known variances

\[ \sigma_{ij}^2 \text{ and } \tau_i^2 \text{ are known constants} \]

Maximum a posteriori (MAP): The \( B_i \)'s and \( G_j \)'s adopts the following form as shrinkage estimators.

\[
\hat{B}_i = \frac{b_i / \tau_i^2 + \sum_{j \in J_i} (y'_{ij} - \hat{G}_j) / \sigma_{ij}^2}{1 / \tau_i^2 + \sum_{j \in J_i} 1 / \sigma_{ij}^2};
\]

\[
\hat{G}_j = \frac{\sum_{i \in l_j} (y'_{ij} - \hat{B}_i) / \sigma_{ij}^2}{\sum_{i \in l_j} 1 / \sigma_{ij}^2}.\]

Asymptotic variances for MAP estimators: inverse of observed/expected Fisher information matrix.
Model Fitting: MAP for known variances

Lemma

When all instruments measure all sources and \( \{\sigma_{ij}^2 = \omega_i^2, \tau_i^2\}_{1 \leq i \leq N; 1 \leq j \leq M} \) are known constants:

\[
\hat{\text{Var}}(\hat{G}_j) = \left[ \sum_{i=1}^{N} \omega_i^{-2} \right]^{-1} S_G, \quad \hat{\text{Var}}(\hat{B}_i) = \left[ M\omega_i^{-2} + \tau_i^{-2} \right]^{-1} S_B^{(i)};
\]

where the shrinkage factors \( S_G, \{S_B^{(i)}\}_{1 \leq i \leq N} \) are given by

\[
S_G = \frac{\sum_{i=1}^{N} \omega_i^{-2} - (M - 1) \sum_{i=1}^{N} \omega_i^{-4} [M\omega_i^{-2} + \tau_i^{-2}]^{-1}}{\sum_{i=1}^{N} \omega_i^{-2} - M \sum_{i=1}^{N} \omega_i^{-4} [M\omega_i^{-2} + \tau_i^{-2}]^{-1}},
\]

\[
S_B^{(i)} = \frac{\sum_{u=1}^{N} \omega_u^{-2} - M \sum_{u \neq i} \omega_u^{-4} [M\omega_u^{-2} + \tau_u^{-2}]^{-1}}{\sum_{u=1}^{N} \omega_u^{-2} - M \sum_{u=1}^{N} \omega_u^{-4} [M\omega_u^{-2} + \tau_u^{-2}]^{-1}}.
\]
Model Fitting: MCMC for known variances

When $\sigma_{ij}^2$’s and $\tau_i^2$’s are known, iterate the following:

(a) For $1 \leq i \leq N$, sample $B_i$ from
\[
\mathcal{N}\left(\frac{b_i}{\tau_i^2} + \sum_{j \in J_i} (y'_{ij} - G_j)/\sigma_{ij}^2, \frac{1}{1/\tau_i^2 + \sum_{j \in J_i} 1/\sigma_{ij}^2}\right).
\]

(b) For $1 \leq j \leq M$, sample $G_j$ from
\[
\mathcal{N}\left(\frac{\sum_{i \in I_j} (y'_{ij} - B_i)/\sigma_{ij}^2}{\sum_{i \in I_j} 1/\sigma_{ij}^2}, \frac{1}{\sum_{i \in I_j} 1/\sigma_{ij}^2}\right).
\]

Alternative: Hamiltonian Monte Carlo algorithm.
Model Fitting: MAP for unknown variances

If $\sigma_{ij}^2 = \omega_i^2$ where $\{\omega_i^2\}_{1 \leq i \leq N}$ are unknown, $\{\tau_i^2\}_{1 \leq i \leq N}$ are known, we need an extra equation to update MAP estimators.

$$\omega_i^2 = 2 \sqrt{1 + \sum_{j \in J_i} (y_{ij} - B_i - G_j)^2 / |J_i|} - 2. \quad (3)$$

Furthermore, if $\tau_i^2 = \tau^2$ for $1 \leq i \leq N$ is unknown, we have an extra equation given by $\tau^2 = \sum_{i=1}^{N} (B_i - b_i)^2 / N$.

Again, the asymptotic variances are given by inverting the expected/observed Fisher information matrix.
If all instruments measure all sources and the priors for $\omega_i^2$ are flat;
Var($\hat{B}_i$) = $[\tau_i^{-2} + \frac{2M\omega_i^{-2}}{\omega_i^2+2}]^{-1}$ $\mathcal{R}_B^{(i)}$,
Var($\hat{G}_j$) = $[\sum_{i=1}^{N} \omega_i^{-2}]^{-1}$ $\mathcal{R}_G$,
Var($\hat{\omega}_i^2$) = $[\frac{M}{4} \frac{\tau_i^{-2}\omega_i^{-2}}{M\omega_i^{-2}+\tau_i^{-2}} + \frac{M}{2} \omega_i^{-4}]^{-1}$ $\mathcal{R}_\omega^{(i)}$.

The shrinkage factors $\{\mathcal{R}_B^{(i)}, \mathcal{R}_\omega^{(i)}\}_{1 \leq i \leq N}$, $\mathcal{R}_G$ are given by

\[
\mathcal{R}_B^{(i)} = \frac{\sum_{i=1}^{N} \omega_i^{-2}(\omega_i^2 + 2)^{-1} - 2M \sum_{k \neq i} \beta_k}{\sum_{i=1}^{N} \omega_i^{-2}(\omega_i^2 + 2)^{-1} - 2M \sum_{k=1}^{N} \beta_k};
\]

\[
\mathcal{R}_\omega^{(i)} = \frac{\sum_{u=1}^{N} \omega_u^{-2} \tau_u^{-2} - \sum_{u \neq i} \left[ 2\omega_u^{-4} + \frac{\omega_u^{-2} \tau_u^{-2}}{M \omega_u^{-2} + \tau_u^{-2}} \right]^{-1} \left[ \frac{\tau_u^{-2} \omega_u^{-2}}{M \omega_u^{-2} + \tau_u^{-2}} \right]^2}{\sum_{u=1}^{N} \omega_u^{-2} \tau_u^{-2} - \sum_{u=1}^{N} \left[ 2\omega_u^{-4} + \frac{\omega_u^{-2} \tau_u^{-2}}{M \omega_u^{-2} + \tau_u^{-2}} \right]^{-1} \left[ \frac{\tau_u^{-2} \omega_u^{-2}}{M \omega_u^{-2} + \tau_u^{-2}} \right]^2};
\]

\[
\mathcal{R}_G = \frac{M \sum_{i=1}^{N} \omega_i^{-2} - (M - 1) \left( \sum_{i=1}^{N} (\omega_i^2 + 2)^{-1} + 4M \sum_{k=1}^{N} \beta_k \right)}{M \sum_{i=1}^{N} \omega_i^{-2} - M \left( \sum_{i=1}^{N} (\omega_i^2 + 2)^{-1} + 4M \sum_{k=1}^{N} \beta_k \right)};
\]

\[
\beta_k = \omega_k^{-4} (\omega_k^2 + 2)^{-1} \left[ (\omega_k^2 + 2) \tau_k^{-2} + 2M \omega_k^{-2} \right]^{-1}, \ 1 \leq k \leq N.
\]
Because the log-likelihood is unbounded, it causes trouble when calculating the MAP with flat prior on $\omega^2_i$. In this way, we can add conjugate priors (inverse-gamma($\alpha, \beta$)) on $\omega^2_i$.

The update of $B_i$ and $G_j$ keeps the same. The update of $\omega_i$ is

$$\omega^2_i = 2\sqrt{\left[1 + \frac{2\alpha + 2}{|J_i|}\right]^2 + \frac{2\beta + S_i}{|J_i|}} - 2 \left[1 + \frac{2\alpha + 2}{|J_i|}\right].$$

where $S_i = \sum_{j \in J_i} (y_{ij} - B_i - G_j)^2$.

This update has a lower bound for $\omega^2_i$, which avoids the unboundness of the posterior likelihood on the boundary.
Model Fitting: MCMC for unknown variances

1. If \( \{ \sigma_{ij}^2 = \omega_i^2; j \in J_i \}_{1 \leq i \leq N} \) are unknown, \( \tau_i \)'s are known. We set independent \( \text{Inv} - \chi^2(\nu_\omega, s_\omega^2) \) priors for \( \omega_i^2 \). The Gibbs sampling iterates steps (a) (b) (c) till convergence.

2. If \( \{ \sigma_{ij}^2 = \omega_i^2, \tau_i^2 = \tau^2; j \in J_i \}_{1 \leq i \leq N} \) are unknown. We set \( \text{Inv} - \chi^2(\nu_\tau, s_\tau^2) \) prior for \( \tau^2 \). The Gibbs sampling iterates steps (a), (b), (c) and (d) till convergence.

(a) and (b), updates for \( B_i, G_j \), same as in known variances.
(c) Update \( \{ \omega_i^2 \} \) one-at-a-time using the Metropolis-Hastings.
(d) Sample \( \tau^2 \sim \text{Inv} - \chi^2(\nu_\tau + N, \nu_\tau s_\tau^2 + \sum_{i=1}^{N} (b_i - B_i)^2) \).

Alternative: Hamiltonian Monte Carlo algorithm.
Demonstration with Simulation Results

First, we simulate data with the fitting model and perform MAP calculation, MCMC and HMC.

Figure: When $\sigma^2_{ij}$, $\tau^2_i$ are known.
Figure: When $\sigma_{ij}^2 = \omega_i^2$ is unknown and $\tau_i^2$ is known.
Demonstration with Simulation Results

Figure: When $\sigma_{ij}^2 = \omega_i^2$ and $\tau_i^2 = \tau^2$ are unknown.
**Discussion:** In fact, HMC is not robust for this model. With different step sizes and leapfrog steps, HMC can generate some crazy results, especially for model 3. This might be because the derivative could be very large sometimes, and the posterior is very huge when \( \omega \) is small.

**Figure:** HMC result with different step size and leapfrog steps.
Real Data Results

In the real dataset, we have three instruments observing more than 100 sources. The observed fluxes are very huge, as well as the pile-up effect.

**Figure:** Histograms of $\log(C)$, $\log(Z)$ and $\log(C/Z)$. 
For our model fitting, neither MCMC nor HMC could get a converging chain.

Figure: Fitting real data with unknown $\omega_i^2$ and $\tau_i^2$. 
Figure: Fitting real data with unknown $\omega_i^2$ and $\tau_i^2$. 
Figure: Fitting real data with unknown $\omega_i^2$ and $\tau_i^2$. 
We also try our model fitting with another smaller data set ($N=5$, $M=13$). MCMC works for model 1 and model 2, while HMC still have troubles for robustness.

**Figure:** Fitting the smaller real data with known $\omega_i^2$ and $\tau_i^2$. 
**Figure:** Fitting the smaller real data with unknown $\omega_i^2$ and known $\tau_i^2$. 
Figure: Fitting the smaller real data with unknown $\omega_i^2$ and $\tau_i^2$. 
Discussions about Poisson Model

- Original scale versus log-scale.
- Choice of Priors.
Hierarchical Poisson Model (log scale)

Considering the fact that the observations $c_{ij}$ are actually counts, it is more natural to define the following Poisson model.

$$c_{ij} \sim \text{Poisson}(Z_{ij} \exp(B_i + G_j)),$$

independently for $j \in J_i$, $1 \leq i \leq N$. The prior for $G_j$ is flat. The prior for $B_i$ is $\mathcal{N}(b_i, \tau_i^2)$. When $z_{ij} = Z_{ij}$,

$$l(\xi, \theta|D) = \sum_{i=1}^{N} \sum_{j \in J_i} \left[ c_{i,j}(B_i + G_j) - z_{ij}e^{B_i+G_j} \right] - \sum_{i=1}^{N} \left[ \frac{(b_i - B_i)^2}{2\tau_i^2} \right],$$

where $\xi = (\tau_1^2, \ldots, \tau_N^2)$, $\theta = (B_1, \ldots, B_N; G_1, \ldots, G_M)$.

**Remark:** It is crucial to have the ‘prior part’ with $b_i$’s, otherwise this model is not identifiable. This can easily be seen from the degeneracy of the Fisher information matrix of $\{B_i\}_{1 \leq i \leq N}$ and $\{G_j\}_{1 \leq j \leq M}$ when the term with $b_i$’s is absent in the likelihood function in Equation ??.
Do we still need to work on the log-scale?

NO.
Poisson Model: Introduction

Considering the fact that the observations $c_{ij}$ are actually counts, it is more natural to define the following Poisson model.

$$c_{ij} \sim \text{Poisson}(Z_{ij}A_iF_j),$$

independently for $j \in J_i$, $1 \leq i \leq N$. $A_i \sim D_A(a_i, \tau_i^2)$.

Parameters: \( \xi = (\tau_1^2, \ldots, \tau_N^2), \ \theta = (A_1, \ldots, A_N; F_1, \ldots, F_M) \).

Assume that $Z_{ij}$, the multiplicative factor due to pile-up, is observed with independent noise: $z_{ij} \sim D_Z(Z_{ij})$.

Special case: $Z_{ij} = z_{ij}$, i.e. observed without uncertainty.

**Question:** What is $D_A$ and $D_Z$?
(1) $z_{ij} = Z_{ij}$, $A_i \sim \text{Gamma}(\nu, \frac{a_i}{\nu})$, log-likelihood $l(\xi, \theta|D)$ is

$$
\sum_{i=1}^{N} \sum_{j \in J_i} [c_{i,j}(\log A_i + \log F_j) - z_{ij} A_i F_j] + \sum_{i=1}^{N} (\nu - 1) \log A_i - \frac{\nu}{a_i} A_i.
$$

Setting score functions to zero gives the following iterative formula for calculating MLE:

$$
A_i = \frac{\nu - 1 + \sum_{j \in J_i} c_{ij}}{\nu / a_i + \sum_{j \in J_i} z_{ij} F_j}, \quad F_j = \frac{\sum_{i \in I_j} c_{ij}}{\sum_{i \in I_j} z_{ij} A_i}.
$$

The Gibbs sampling goes as follows:

$$
A_i|F_1, \ldots, F_M \sim \text{Gamma}\left(\nu + \sum_{j \in J_i} c_{ij}, [\nu / a_i + \sum_{j \in J_i} z_{ij} F_j]^{-1}\right),
$$

$$
F_j|A_1, \ldots, A_N \sim \text{Gamma}\left(\sum_{i \in I_j} c_{ij} + 1, \left[\sum_{i \in I_j} z_{ij} A_i\right]^{-1}\right).
$$
Poisson Model: Model Fitting (2)

(2) \( z_{ij} = Z_{ij}, \ A_i \sim \mathcal{N}(a_i, \tau^2_i) \), log-likelihood \( l(\xi, \theta|D) \) is

\[
\sum_{i=1}^{N} \sum_{j \in J_i} [c_{i,j} \left( \log A_i + \log F_j \right) - z_{ij} A_i F_j] - \sum_{i=1}^{N} \left[ \frac{(a_i - A_i)^2}{2 \tau^2_i} \right] - \frac{N}{2} \log(\tau^2_i).
\]

Setting the score functions to zero gives the following iterative formula:

\[
A_i = \frac{a_i - \tau^2_i \sum_{j \in J_i} z_{ij} F_j + \sqrt{(a_i - \tau^2_i \sum_{j \in J_i} z_{ij} F_j)^2 + 4 \tau^2_i \sum_{j \in J_i} c_{ij}}}{2};
\]

\[
F_j = \frac{\sum_{i \in I_j} c_{ij}}{\sum_{i \in I_j} z_{ij} A_i}.
\]

If \( \tau^2_i = \tau^2 \) is unknown, then we also need \( \tau^2 = \sum_{i=1}^{N} (a_i - A_i)^2 / N \).
Poisson Model: Model Fitting (2)

The MCMC sampling goes as follows:

- Update $A_1, \ldots, A_N$ using the Metropolis-Hastings algorithm.
- Update
  \[
  F_j | A_1, \ldots, A_N \sim \text{Gamma} \left( \sum_{i \in I_j} c_{ij} + 1, \left[ \sum_{i \in I_j} z_{ij} A_i \right]^{-1} \right).
  \]
- If $\tau_i^2 = \tau^2$ is unknown, update
  \[
  \tau^2 \sim \text{Inv} - \chi^2_N \left( \sum_{i=1}^{N} (a_i - A_i)^2 / N \right).
  \]
(3) $z_{ij} = Z_{ij}$, $\log A_i \sim \mathcal{N}(b_i, \tau_i^2)$, log-likelihood $l(\xi, \theta|D)$ is

$$
\sum_{i=1}^{N} \sum_{j \in J_i} [c_{i,j}(\log A_i + \log F_j) - z_{ij}A_iF_j] - \sum_{i=1}^{N} \left[ \frac{(b_i - \log A_i)^2}{2\tau_i^2} \right] - \sum_{i=1}^{N} \log A_i.
$$

The MCMC sampling goes as follows:

- Update $A_1, \ldots, A_N$ using the Metropolis-Hastings.
- Update
  $$
  F_j|A_1, \ldots, A_N \sim \text{Gamma} \left( \sum_{i \in I_j} c_{ij} + 1, \left[ \sum_{i \in I_j} z_{ij}A_i \right]^{-1} \right).
  $$
- If $\tau_i^2 = \tau^2$ is unknown, update
  $$
  \tau^2 \sim \text{Inv - } \chi^2_N \left( \sum_{i=1}^{N} (a_i - A_i)^2 / N \right).
  $$
Poisson Model: Model Fitting (4)

\[(4) \quad Z_{ij} \sim \text{Gamma}(\nu_z, z_{ij}/\nu_z), \quad A_i \sim \text{Gamma}(\nu_a, a_i/\nu_a), \]

\[
l(\xi, \theta, Z|D) = \sum_{i=1}^{N} \sum_{j \in J_i} c_{i,j} \log(A_i F_j) + \sum_{i=1}^{N} (\nu_a - 1) \log(A_i) - \sum_{i=1}^{N} \frac{\nu_a}{a_i} A_i \]
\[
+ \sum_{i=1}^{N} \sum_{j \in J_i} (\nu_z - 1 + c_{ij}) \log(Z_{ij}) - Z_{ij} \left(\frac{\nu_z}{Z_{ij}} + A_i F_j\right).\]

EM algorithm: the E-step relies on the conditional distribution
\[Z_{ij}|A_i, F_j \sim \text{Gamma}(c_{ij} + \nu_z, (\nu_z/z_{ij} + A_i F_j)^{-1});\]
thus Optimizing this Q-function over \(A_i, F_j\) gives the new \(A_i, F_j\)'s:

\[
A_i = \frac{\nu_a - 1 + \sum_{j \in J_i} c_{ij}}{\frac{\nu_a}{a_i} + \sum_{j \in J_i} F_j \frac{\nu_z + c_{ij}}{\nu_z z_{ij}^{-1} + A_i^{\text{old}} F_j^{\text{old}}}}, \quad F_j = \frac{\sum_{i \in I_j} c_{ij}}{\sum_{i \in I_j} A_i \frac{\nu_z + c_{ij}}{\nu_z z_{ij}^{-1} + A_i^{\text{old}} F_j^{\text{old}}}}.
\]
The Gibbs sampling goes as follows:

\[ Z_{ij} \sim \text{Gamma} \left( \nu_z + c_{ij}, \left[ \frac{\nu_z}{Z_{ij}} + A_i F_j \right]^{-1} \right), \]

\[ A_i \sim \text{Gamma} \left( \nu_a + \sum_{j \in J_i} c_{ij}, \left[ \frac{\nu_a}{a_i} + \sum_{j \in J_i} Z_{ij} F_j \right]^{-1} \right), \]

\[ G_j \sim \text{Gamma} \left( \sum_{i \in I_j} c_{ij}, \left[ \sum_{i \in I_j} \frac{Z_{ij} A_i}{Z_{ij}} \right]^{-1} \right). \]
(5) $\log z_{ij} \sim \mathcal{N}(\log Z_{ij}, \lambda^2)$, the log-likelihood function is

$$l(\xi, \theta, Z | D) = \sum_{i=1}^{N} \sum_{j \in J_i} [c_{i,j} \log(Z_{ij}A_iF_j) - Z_{ij}A_iF_j] - \sum_{i=1}^{N} \left[ \frac{(b_i - \log A_i)^2}{2\tau_i^2} \right]$$

$$- \sum_{i=1}^{N} \sum_{j \in J_i} \frac{\log(\lambda^2)}{2} - \sum_{i=1}^{N} \sum_{j \in J_i} \frac{(\log(z_{ij}) - \log(Z_{ij}))^2}{2\lambda^2}.$$ 

Remark: in this case, the latent variables $Z_{ij}$ are not easy to integrate out, neither does it have a nice form for Gibbs update.
We only plot the HMC result for model 1. Pure MCMC has some problem for it generates very large results. In fact, the results of HMC relies on the choice of $\nu$, that is to say the prior for $A_i$ very much.

Figure: HMC results for Poisson model.
For log-normal model, how could we improve the MCMC? For example, how can we choose HMC step size and leapfrog steps to gain a robust result?

For the real data, do we need and truncate because the range is so wide right now?

For Poisson model, which model assumption shall we choose?