Seeking Effective Adjustments for Effective Areas

Presenter: Xiao-Li Meng Yang Chen, Xufei Wang

Joint work with Vinay Kashyap, Herman Marshall, David van Dyk, Matteo Guainazzi, Paul Plucinsky

March 1, 2016

Recap of the Problem

Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang

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.

Original 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

$$\log C_{ij} = \log A_i + \log F_j = B_i + G_j, \tag{1}$$

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).

Basic Model – Observation Level

Hierarchical regression model:

$$y_{ij} = \log(c_{ij}) = \alpha_{ij} + B_i + G_j + \epsilon_{ij}, \tag{2}$$

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.

Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang

Complications with Real Data

A multiplicative factor due to pile-up

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

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

Z_{ij} is an observed constant and

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

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

Model Fitting: identifiability assumptions

Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang

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 three major assumptions which are practically reasonable.

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

For model fitting, we calculate the maximum a posteriori estimation (MAP) for each model.

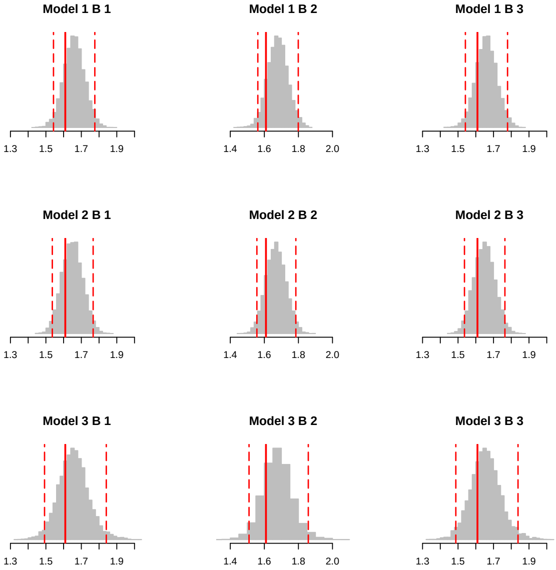
Besides, we also obtain the full posterior by Gibbs sampling and Hamiltonian Monte Carlo (HMC).

Simulation Experiment 1

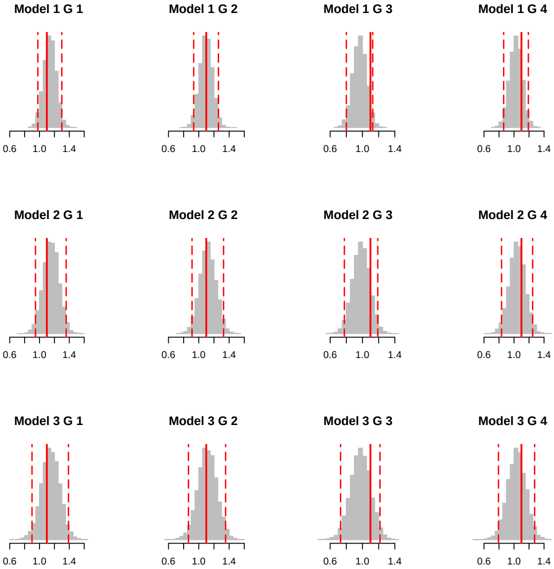
Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang

- Number of instruments: N = 3.
- Number of Sources: M = 100.
- True values: $B_i = \log(5) = 1.61$, $G_j = \log(3) = 1.10$.
- Variances: $\sigma_{ij} = 0.1$; $\tau_i = 0.1$; $1 \leq i \leq N$; $1 \leq j \leq M$.

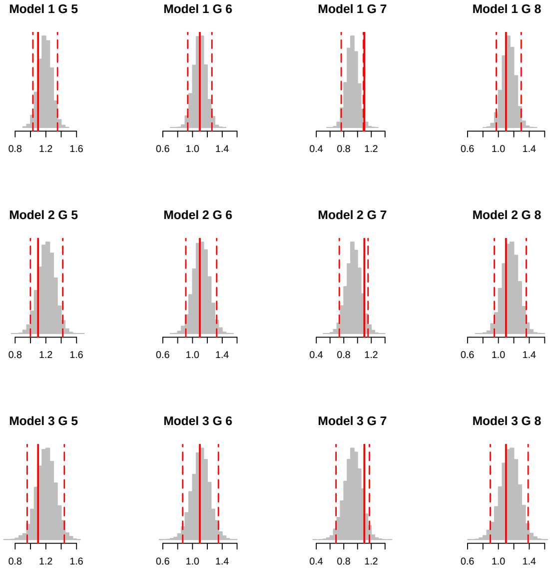
N = 3, M = 100, Effective Area (log)



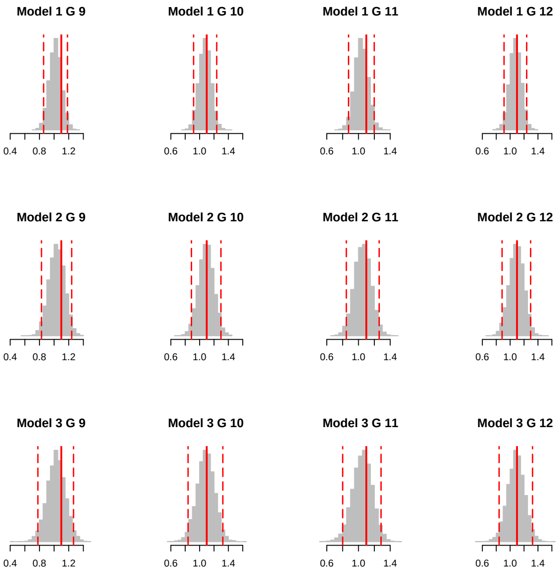
N = 3, M = 100, Flux (log)



N = 3, M = 100, Flux (log)

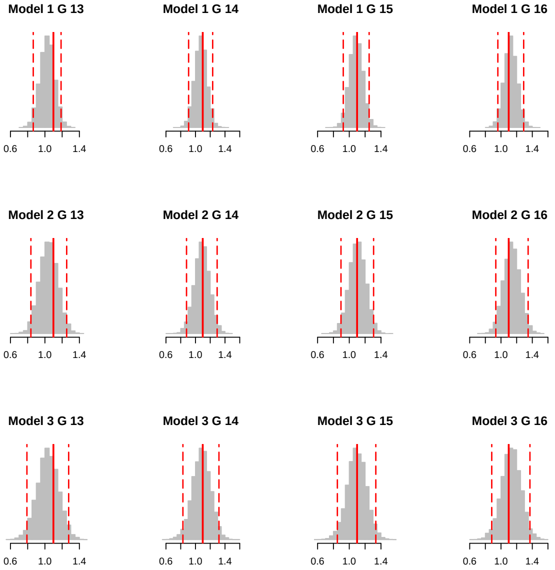


N = 3, M = 100, Flux (log)



N = 3, M = 100, Flux (log)

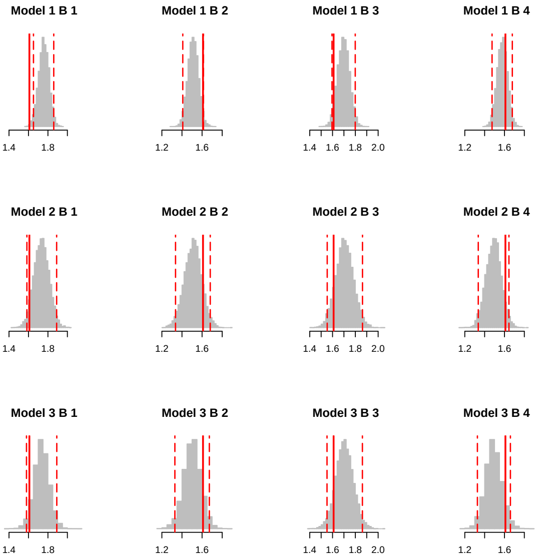
Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang



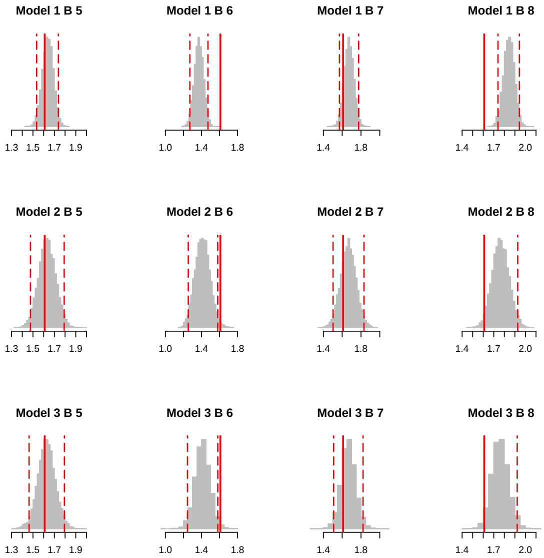
Simulation Experiment 2

- Number of instruments: N = 13.
- Number of Sources: M = 5.
- True values: $B_i = \log(5) = 1.61$, $G_j = \log(3) = 1.10$.
- Variances: $\sigma_{ij} = 0.1$; $\tau_i = 0.1$; $1 \leq i \leq N$; $1 \leq j \leq M$.

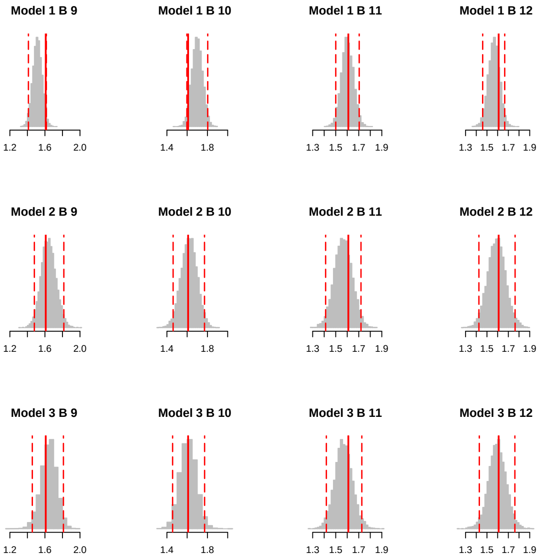
N = 13, M = 5, Effective Area (log)



N = 13, M = 5, Effective Area (log)

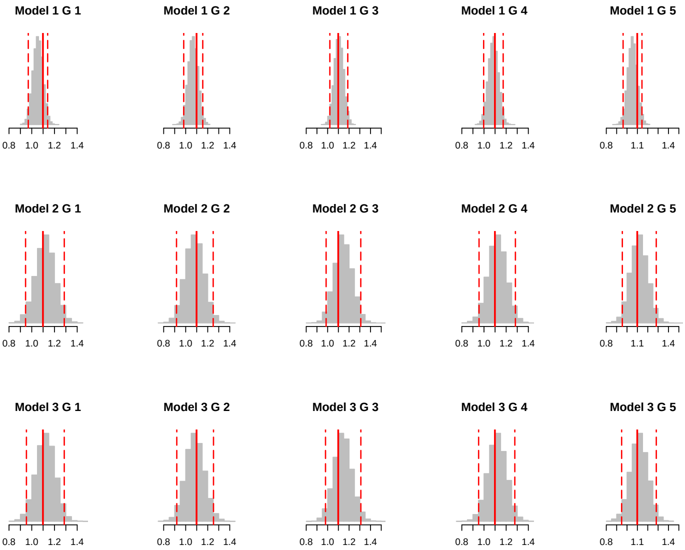


N = 13, M = 5, Effective Area (log)



Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang

N = 13, M = 5, Flux (log)

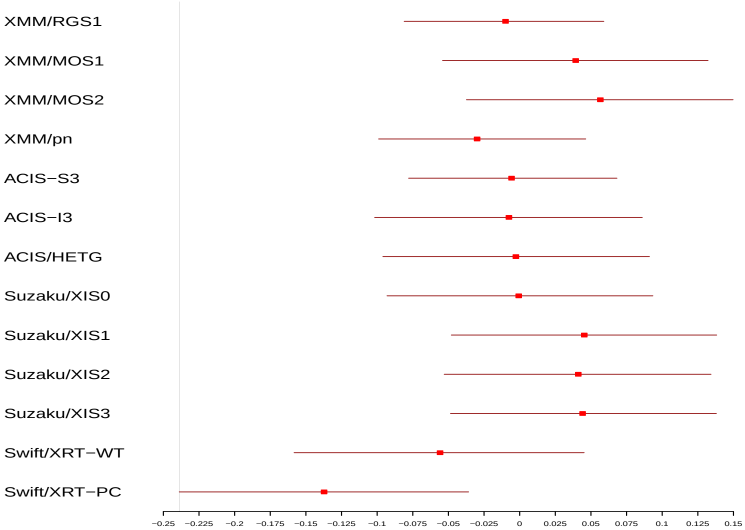


Real Data 1 (E0102 Data)

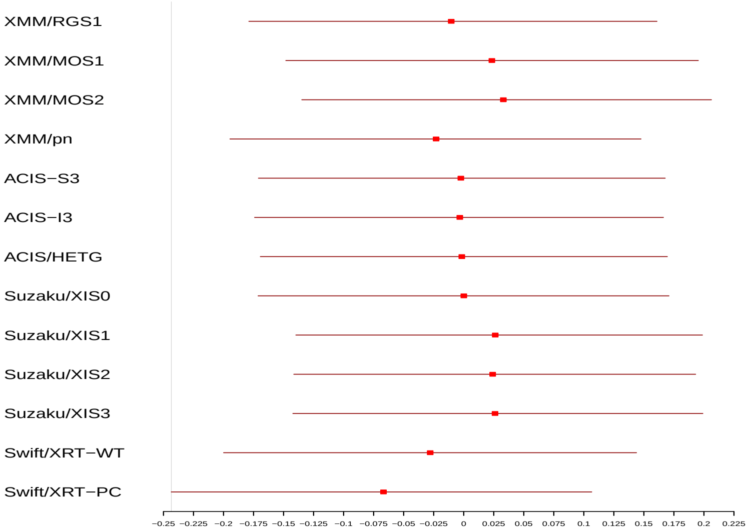
- Data Provided by: Paul Plucinsky, Vinay Kashyap.
- Number of instruments: N = 13.
- Number of Sources: M = 5.
- Source Names: 'const', 'O7', 'O8', 'Ne9', 'Ne10'.
- Instrument Names:

  'XMM/RGS1', 'XMM/MOS1', 'XMM/MOS2', 'XMM/pn',
  'ACIS-S3', 'ACIS-I3', 'ACIS/HETG',
  'Suzaku/XIS0', 'Suzaku/XIS1', 'Suzaku/XIS2', 'Suzaku/XIS3',
  'Swift/XRT-WT', 'Swift/XRT-PC'.

E0102 data Results (N = 13, M = 5, known variance)



E0102 data Results (N = 13, M = 5, unknown variance)

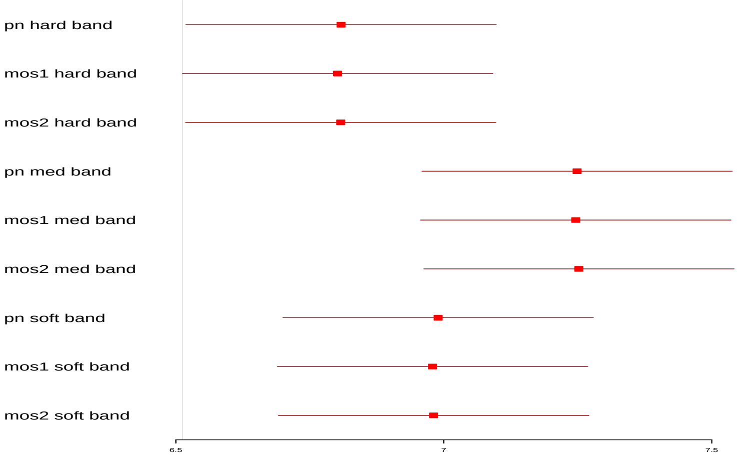


Real Data 2 (2XMM Data)

- Data Provided by: Herman Marshall & Matteo Guainazzi.
- Number of instruments: N = 3. 'pn', 'mos1', 'mos2'.
- Number of Sources: M = 35 (hard band); M = 39 (medium band); M = 34 (soft band).
- Source Names (hard band):

  RXJ0944.5+0357, HolmbergIX, 4C06.41, 1127-145, NGC4278, LBQS1228+1116, MS1229.2+6430, XCOMAE, XCOMAE, ESO323-G077, PKSB1334-127, NGC5252, PG1407+265, RBS1423, CenX-4, UZLIB, RXJ0136.9-3510, NGC6251, MS0205.7+3509, NGC7172, M31NN1, NGC1313, XComae, XComae, XComae, NGC5204X-1, NGC5204X-1, GRB080411, RXJ0228-40, PKS0237-23, RBS1055, V410Tau, V410Tau, VB50, 1E0919+515.

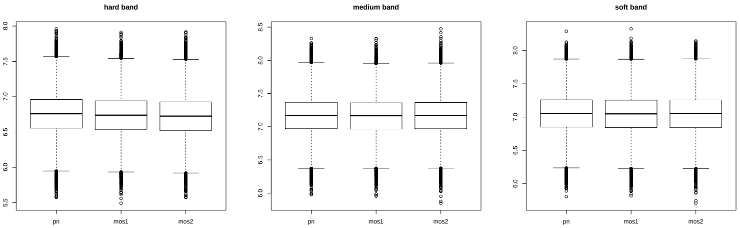
2XMM Data Results (Hard, Medium, Soft Band)



2XMM Data Results (Hard, Medium, Soft Band)

Astrostat
24/28

Presenter:
Xiao-Li Meng
Yang Chen,
Xufei Wang

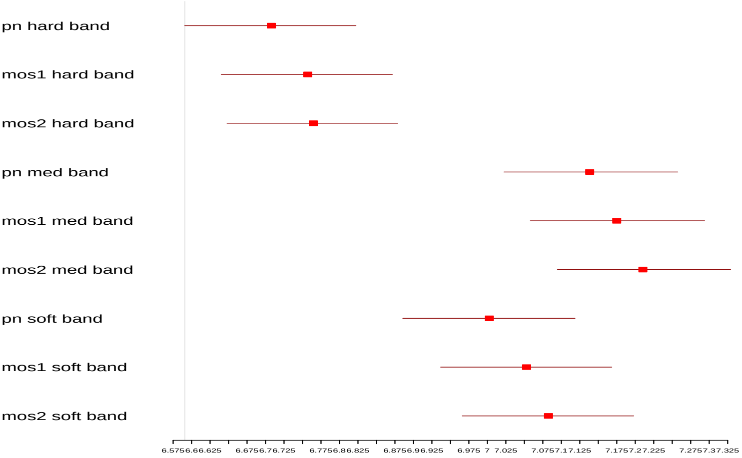


Real Data 3

- Data Provided by: Herman Marshall & Matteo Guainazzi.
- Number of instruments: $N = 3$. 'pn', 'mos1', 'mos2'.
- Number of Sources: $M = 94$ (hard band); $M = 103$ (medium band); $M = 108$ (soft band).
- Source Names (hard band):

  21 unique ones (total 94): 3C111, PKS2155-304, 3C120, 1H1219+301, H1426+428, 3C273, MKN501, PKS0558-504, 4U0543-31, Ark120, NGC526A, EXO0748-676, 1H0414+009, TON1388, PKS0548-322, 1ES1101-232, H2356-309, H1426+484, PG1116+215, Mkn501, 1ES1553+11.3.

Preliminary Results (Real Data 3)



Ongoing and Future Work

- Real data 3 by Herman Marshall & Matteo Guainazzi.
- Robustness to 'Outliers'.
- Poisson Model – observations are counts.
- Sensitivity of 'Priors'.

Questions?

Contact:

- Yang Chen: yangchen@fas.harvard.edu
- Xufei Wang: xufeiwang@fas.harvard.edu
- Xiaoli Meng: meng@stat.harvard.edu