A Tables of Data Description and Prior Influence

Tables 3 and 4 give summaries of the data used in Sections 4.1, 4.2 & 4.3.

Lines (Sources)	He-like OVII	H-like OVIII	He-like Ne IX	H-like Ne X
Spectrum	21.805Å	18.969Å	13.447Å	12.135Å

Telescopes	*Chandra*	XMM-*Newton*	*Suzaku*	*Swift*
Detectors (Instruments)	HETG, ACIS-S	RGS, EPIC-MOS, EPIC-pn	XIS	XRT

Table 3: Summary of E0102 data. The first table gives the sources for two data sets, Highly ionized Oxygen and Neon. The second table gives instruments for both data sets.

Observatory	XMM-*Newton* European Photon Imaging Cameras (EPIC)					
Detectors (Instruments)	EPIC-pn (pn), EPIC-MOS (MOS1 & MOS2)					
Data Acronym	2XMM			XCAL		
Energy Band	Hard	Medium	Soft	Hard	Medium	Soft
Energy (keV)	2.5-10.0	1.5-2.5	0.5-1.5	2.5-10.0	1.5-2.5	0.5-1.5
No. Sources	41	41	42	94	103	108

Table 4: Summary of 2XMM data and XCAL data. The number of instruments is $N = 3$ (pn, MOS1, MOS2) and the number of sources (M) is given in the last row for the six data sets, three from different energy bands of 2XMM data and XCAL data respectively.

B Details of Fitting the Log-Normal Model

The following three MCMC algorithms are used for our posterior sampling.

1. **Standard Gibbs Sampler:** iterates the following three sets of conditional distributions, all easily derived from (2.5):

   (a) Conditioning on $\boldsymbol{G}$ and $\boldsymbol{\sigma}^2$, sample B_i independently for $i = 1, \ldots, N$ from
   $$\mathcal{N}\left(\frac{b_i/\tau_i^2 + \sum_{j\in J_i}(y_{ij} + 0.5\sigma_i^2 - G_j)/\sigma_i^2}{1/\tau_i^2 + \sum_{j\in J_i} 1/\sigma_i^2},\ \frac{1}{1/\tau_i^2 + \sum_{j\in J_i} 1/\sigma_i^2}\right).$$

   (b) Conditioning on $\boldsymbol{B}$ and $\boldsymbol{\sigma}^2$, sample G_j independently for $1 \le j \le M$ from
   $$\mathcal{N}\left(\frac{\sum_{i\in I_j}(y_{ij} + 0.5\sigma_i^2 - B_i)/\sigma_i^2}{\sum_{i\in I_j} 1/\sigma_i^2},\ \frac{1}{\sum_{i\in I_j} 1/\sigma_i^2}\right).$$

   (c) Conditioning on $\boldsymbol{B}$ and $\boldsymbol{G}$, sample σ_i^2 independently for $i = 1, \ldots, N$ from
   $$\sigma_i^{-|J_i|-2-2\alpha} \exp\left\{-\frac{1}{2}\frac{\sum_{j\in J_i}(y_{ij} - B_i - G_j)^2 + 2\beta}{\sigma_i^2} - \frac{|J_i|\sigma_i^2}{8}\right\}$$
   via the Metropolis-Hastings algorithm using a simple random walk proposal (Gaussian proposal) on the log-scale, i.e., $\log(\sigma_i^2)$.

2. **Block Gibbs Sampler**: same as above except replace the two conditional steps (1a) and (1b) by a joint draw of $\{\boldsymbol{B}, \boldsymbol{G}\}$ from $(N+M)$-dimensional Gaussian with mean $\boldsymbol{\Omega(\sigma^2)}^{-1}\boldsymbol{\gamma(\sigma^2)}$ and covariance matrix $\boldsymbol{\Omega(\sigma^2)}^{-1}$; see (2.7) and (2.8) in Section 2.2.

Instrument	Oxygen		Neon	
	$\tau = 0.025$	$\tau = 0.05$	$\tau = 0.025$	$\tau = 0.05$
RGS1	0.570	0.205	0.063	0.016
MOS1	0.279	0.077	0.075	0.019
MOS2	0.355	0.065	0.077	0.017
pn	0.250	0.041	0.620	0.218
ACIS-S3	0.218	0.040	0.270	0.088
ACIS-I3	0.906	0.640	0.099	0.026
HETG	0.648	0.341	0.129	0.034
XIS0	0.180	0.051	0.069	0.018
XIS1	0.298	0.078	0.071	0.019
XIS2	0.463	0.140	0.063	0.016
XIS3	0.772	0.364	0.062	0.018
XRT-WT	0.726	0.278	0.154	0.026
XRT-PC	0.934	0.235	0.906	0.017

Table 5: Proportion of prior influence, as defined by $1-W_i$ (of (2.11)), for E0102 data in Section 4.1.

3. **Hamiltonian Monte Carlo (HMC)**: samples the entire vector $\boldsymbol{\theta} = \{B_i, G_j, \sigma_i^2\}$ through the non-U-turn HMC sampler (Hoffman and Gelman, 2014), implemented with the `STAN` package. Here we give a brief description of HMC; see Neal (2011) for more details. Let $\pi(\boldsymbol{\theta})$ denote the (unnormalized) joint posterior $\boldsymbol{\theta}$, as given by (2.5). Define potential energy as $U(\boldsymbol{\theta}) = -\log \pi(\boldsymbol{\theta})$ and kinetic energy as $k(\boldsymbol{p}) = \boldsymbol{p}^\top \mathcal{M}^{-1} \boldsymbol{p}$, where $\mathcal{M}$ is a symmetric positive-definite matrix, thus the total energy is $H(\boldsymbol{\theta}, \boldsymbol{p}) = U(\boldsymbol{\theta}) + k(\boldsymbol{p})$. We can obtain samples of $\pi(\boldsymbol{\theta})$ by sampling from the target density $\exp[-H(\boldsymbol{\theta}, \boldsymbol{p})] \propto \pi(\boldsymbol{\theta}) \exp(-\boldsymbol{p}^\top \mathcal{M}^{-1} \boldsymbol{p})$, which is essentially a data-augmentation technique (Tanner and Wong, 1987). By defining the potential energy and kinetic energy, we can propose MCMC moves according to the Hamiltonian dynamics, which explores the parameter space more efficiently by taking bigger and less correlated moves, as opposed to random walk Metropolis-Hastings or a Gibbs sampler. In practice, we use the leapfrog move to approximate the Hamiltonian dynamics. Due

Data Name	$\tau_i = 0.025$			$\tau_i = 0.05$		
	pn	mos1	mos2	pn	mos1	mos2
hard band 2XMM	0.093	0.075	0.082	0.025	0.020	0.022
medium band 2XMM	0.250	0.216	0.222	0.076	0.065	0.067
soft band 2XMM	0.093	0.075	0.069	0.025	0.020	0.018
hard band XCAL	0.010	0.019	0.031	0.003	0.005	0.008
medium band XCAL	0.023	0.016	0.028	0.006	0.004	0.007
soft band XCAL	0.021	0.011	0.007	0.005	0.003	0.002

Table 6: Proportion of prior influence for data used in the analysis in Sections 4.2 and 4.3.

to the energy-preserving property of Hamiltonian dynamics, the acceptance rate of the resulting HMC is approximately 1. It is not exactly 1 because we use the (discretized) leapfrog moves to approximate (continuous) Hamiltonian dynamics. The tuning parameters of the HMC algorithm include the covariance matrix $\mathcal{M}$, the leapfrog step size ϵ, and the number of leapfrog steps L. These are all self-tuned in the `STAN` package.

We compare the performance of these three algorithms using auto-correlation plots of the posterior samples and the effective sample size, in both the simulated and real data examples. Not surprisingly, the Gibbs sampler converges very slowly relative to the other two algorithms. We are able to cross check our results by comparing the samples obtained with the block Gibbs sampler and HMC – they give practically the same posterior distributions.

C Proprieties of the Posterior Distribution

C.1 Propriety of Posterior

Theorem 1. *Under the prior specifications for* $\{B_i, G_j, \sigma_i^2 : 1 \leq i \leq N, 1 \leq j \leq M\}$ *given in (2.4), the posterior is proper if each source is measured by at least one instrument, i.e.,* $|I_j| \geq 1$ *for all* $1 \leq j \leq M$.

Proof. We prove the propriety of the posterior by first integrating out the G_j first, then the B_i,

and finally the σ_i^2. By (2.5), $p(\boldsymbol{B}, \boldsymbol{G}, \boldsymbol{\sigma}^2 | \boldsymbol{D}, \boldsymbol{\tau}^2)$ is proportional to

$$\prod_{i=1}^{N} \sigma_i^{-|J_i|-2-2\alpha} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{M} \sum_{i \in I_j} \sigma_i^{-2} (y'_{ij} - B_i - G_j)^2 - \sum_{i=1}^{N} \left[\frac{(b_i - B_i)^2}{2\tau_i^2} + \frac{\beta}{\sigma_i^2} \right] \right\}. \tag{C.1}$$

Now for each $1 \leq j \leq M$, if we define a random index $\mathcal{I}$ on I_j such that $\Pr(\mathcal{I} = i) \propto \sigma_i^{-2}$, then

$$\frac{\sum_{i \in I_j} \sigma_i^{-2} (y'_{ij} - B_i - G_j)^2}{\sum_{i \in I_j} \sigma_i^{-2}} = \mathrm{E}\left[y'_{\mathcal{I}j} - B_{\mathcal{I}} - G_j \right]^2 \geq \left[\mathrm{E}(y'_{\mathcal{I}j} - B_{\mathcal{I}}) - G_j \right]^2. \tag{C.2}$$

Therefore, the first term in the exponential part of (C.1) is less than $-0.5 \left(\sum_{i \in I_j} \sigma_i^{-2} \right) (G_j - C_j)^2$, where $C_j = \mathrm{E}(y'_{\mathcal{I}j} - B_{\mathcal{I}})$ is free of G_j. The property of Normal density (for G_j) then yields

$$\int p(\boldsymbol{B}, \boldsymbol{G}, \boldsymbol{\sigma}^2 | \boldsymbol{D}, \boldsymbol{\tau}^2) \, d\boldsymbol{G} \leq C^* \prod_{i=1}^{N} \sigma_i^{-|J_i|-2-2\alpha} \prod_{j=1}^{M} \left[\sum_{i \in I_j} \sigma_i^{-2} \right]^{-1/2} \exp\left\{ -\sum_{i=1}^{N} \left[\frac{(b_i - B_i)^2}{2\tau_i^2} + \frac{\beta}{\sigma_i^2} \right] \right\}$$

where C^* is a constant that depends only on $\boldsymbol{D}, \boldsymbol{\tau^2}$. Integrating out $\boldsymbol{B}$ then gives

$$\int \int p(\boldsymbol{B}, \boldsymbol{G}, \boldsymbol{\sigma}^2 | \boldsymbol{D}, \boldsymbol{\tau}^2) \, d\boldsymbol{G} \, d\boldsymbol{B} \leq C^{**} \prod_{i=1}^{N} \sigma_i^{-|J_i|-2-2\alpha} \prod_{j=1}^{M} \left[\sum_{i \in I_j} \sigma_i^{-2} \right]^{-1/2} \exp\left\{ -\sum_{i=1}^{N} \frac{\beta}{\sigma_i^2} \right\} \tag{C.3}$$

where C^{**} is a constant that depends only on $\boldsymbol{D}, \boldsymbol{\tau^2}$. Since I_j is non-empty, it is meaningful to invoke the well-known harmonic-geometric mean inequality to obtain that

$$\prod_{j=1}^{M} \left[\sum_{i \in I_j} \sigma_i^{-2} \right]^{-1/2} \leq \prod_{j=1}^{M} |I_j|^{-1/2} \left[\prod_{i \in I_j} \sigma_i \right]^{1/|I_j|} \leq \prod_{i=1}^{N} \sigma_i^{\sum_{j \in J_i} |I_j|^{-1}}. \tag{C.4}$$

Inequalities (C.3) and (C.4) together imply that the unnormalized $p(\boldsymbol{\sigma}^2 | \boldsymbol{D}, \boldsymbol{\tau}^2)$ is dominated above by a constant times $\prod_{i=1}^{N} p_i(\sigma_i^2)$, where $p_i(x)$ is the density of the inverse Gamma distribution with shape parameter $\alpha_i = \alpha + [|J_i| - \sum_{j \in J_i} |I_j|^{-1}]/2$ and scale parameter β. Because $|I_j| \geq 1$, we have $\alpha_i \geq \alpha$. Hence as long as the hyperparameter $\alpha > 0$, which is always chosen to be so, p_i is a proper density. Consequently, $p(\boldsymbol{\sigma}^2 | \boldsymbol{D}, \boldsymbol{\tau}^2)$ is a proper density after renormalization.

□

C.2 Identifiability

When τ_i^2 is large, the likelihood information for estimating B_i (i.e., from c_{ij}) dominates the prior information (i.e., from b_i). In the extreme case of $\tau_i^2 = \infty$, the model is not identifiable because for

fixed variances, $\{B_i, G_j\}$ and $\{B_i+\delta, G_j-\delta\}$ yield the same posterior densities for $\{\boldsymbol{B}, \boldsymbol{G}\}$ for any constant δ. Let $\lambda_{\max}$ and $\lambda_{\min}$ be the maximum and minimum eigenvalues of $\boldsymbol{\Omega}(\boldsymbol{\sigma^2})$, as defined in Section 2.2. Taking $u = (\mathbf{1}_N, \mathbf{1}_M)^\top$ and $v = (\mathbf{1}_N, -\mathbf{1}_M)^\top$, the *condition number* of $\boldsymbol{\Omega}(\boldsymbol{\sigma^2})$ is

$$\frac{\lambda_{\max}}{\lambda_{\min}} \geq \frac{u^\top \boldsymbol{\Omega}(\boldsymbol{\sigma^2})\ u}{v^\top \boldsymbol{\Omega}(\boldsymbol{\sigma^2})\ v} = 1 + \frac{4\sum_{i=1}^N |J_i|\sigma_i^{-2}}{\sum_{i=1}^N \tau_i^{-2}}, \tag{C.5}$$

where $\mathbf{1}_n$ denotes an $n \times 1$ vector of ones. As a consequence, when $\{\tau_i^2\}$ are generally larger than $\{\sigma_i^2\}$, the ratio in (C.5) can be large, and the posterior contours, determined by $\boldsymbol{\Omega}$, are elongated in one direction and narrow in another. This provides a guideline that $\{\tau_i^2\}$ should not be set too large relative to $\{\sigma_i^2\}$ in practice, because large $\{\tau_i^2\}$ can lead to near model non-identifiability and consequently more costly computation. A computationally cheaper way of dealing with possible model non-identifiability is to set one of the $\{B_i\}$ equal to a fixed value, which is equivalent to setting the corresponding $\tau_i = 0$. We experiment with this computationally cheap strategy in our empirical evaluations, and find that it does not alter the results in substantive ways, but the resulting estimators for the Effective Areas are relative to some (arbitrarily) chosen values instead of in absolute terms/magnitudes.

D Derivation of Conditional Covariance Matrix

In this section, we give detailed derivations of $\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)$ when all instruments measure all sources. In this case, W_i defined in (2.11) becomes $W_i = \frac{M\sigma_i^{-2}}{M\sigma_i^{-2}+\tau_i^{-2}}$, $1 \leq i \leq N$. Define $\tilde{\sigma}^2 = \left(N^{-1}\sum_{i=1}^N \sigma_i^{-2}\right)^{-1}$.

Let $\boldsymbol{A}$ be the $(N+M)\times(N+M)$ diagonal matrix with diagonal elements equal to those of $\boldsymbol{\Omega}(\sigma^2)$. Let $\boldsymbol{U}$ be an $(N+M)\times 2$ matrix such that $U_{i,1} = \sigma_i^{-2}$, $U_{i,2} = 0$ for $i = 1, \ldots, N$, and $U_{j+N,1} = 0$, $U_{j+N,2} = 1$ for $j = 1, \ldots, M$. Let $\boldsymbol{C}$ be a 2×2 matrix such that $C_{i,j} = I_{i\neq j}$ $(i, j = 1, 2)$. Then $\boldsymbol{\Omega}(\boldsymbol{\sigma}^2) = \boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{U}^\top$. By the Woodbury matrix identity, we have

$$\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2) = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\left(\boldsymbol{C} + \boldsymbol{U}^\top\boldsymbol{A}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^\top\boldsymbol{A}^{-1}, \tag{D.1}$$

where $\boldsymbol{A}^{-1}$ is a diagonal matrix with diagonal elements

$$\left(\left\{W_i\sigma_i^2/M\right\}_{1\leq i\leq N}, \left\{\tilde{\sigma}^2/N\right\}_{1\leq j\leq M}\right).$$

Therefore, we can derive the inverse of 2×2 matrix $\boldsymbol{C} + \boldsymbol{U}^\top \boldsymbol{A}^{-1} \boldsymbol{U}$ as

$$\left(\boldsymbol{C} + \boldsymbol{U}^\top \boldsymbol{A}^{-1} \boldsymbol{U}\right)^{-1} = \begin{pmatrix} \frac{\sum_{i=1}^N W_i \sigma_i^{-2}}{M} & 1 \\ 1 & \frac{M}{N}\tilde{\sigma}^2 \end{pmatrix}^{-1} = -\frac{\sum_{i=1}^N \sigma_i^{-2}}{\sum_{i=1}^N W_i \tau_i^{-2}} \begin{pmatrix} \frac{M^2 \tilde{\sigma}^2}{N} & -M \\ -M & \sum_{i=1}^N W_i \sigma_i^{-2} \end{pmatrix}.$$

Further, let $\boldsymbol{W}$ be the $N \times 1$ column vector with ith element W_i, then we have

$$\boldsymbol{A}^{-1} \boldsymbol{U} = \begin{pmatrix} \boldsymbol{W}/M & 0_{N \times 1} \\ 0_{M \times 1} & \tilde{\sigma}^2/N \; 1_{M \times 1} \end{pmatrix}.$$

Consequently, $\boldsymbol{A}^{-1} \boldsymbol{U} \left(\boldsymbol{C} + \boldsymbol{U}^\top \boldsymbol{A}^{-1} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^\top \boldsymbol{A}^{-1}$ is equal to

$$-\left(\sum_{i=1}^N \tau_i^{-2} W_i\right)^{-1} \begin{pmatrix} \boldsymbol{W}\boldsymbol{W}^\top & -\boldsymbol{W} 1_{1 \times M} \\ -1_{M \times 1} \boldsymbol{W}^\top & [N^{-1}\tilde{\sigma}^2] \sum_{i=1}^N W_i \sigma_i^{-2} \; 1_{M \times M} \end{pmatrix}.$$

Finally, we arrive at the closed-form expression for $\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)$:

$$\begin{aligned} \left(\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)\right)_{i,i} &= \frac{1}{M\sigma_i^{-2} + \tau_i^{-2}} \left\{1 + \frac{M\sigma_i^{-2} W_i}{\sum_{u=1}^N \tau_u^{-2} W_u}\right\}, \\ \left(\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)\right)_{i,j+N} &= \left(\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)\right)_{j+N,i} = -W_i \left(\sum_{u=1}^N \tau_u^{-2} W_u\right)^{-1}, \\ \left(\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)\right)_{j+N,j+N} &= \left(\sum_{u=1}^N \sigma_u^{-2}\right)^{-1} \left\{1 + \frac{\sum_{i=1}^N W_i \; \sigma_i^{-2}}{\sum_{u=1}^N W_u \; \tau_u^{-2}}\right\}. \end{aligned}$$

E Likelihood Method

E.1 MLEs and Their Asymptotic Variances

Note that the variance-covariance matrix of the MLEs $\{\hat{\boldsymbol{B}}, \hat{\boldsymbol{G}}\}$ is in fact $\boldsymbol{\Omega}^{-1}(\boldsymbol{\sigma}^2)$ as defined in (2.8). Therefore, we have the following proposition.

Proposition 1. *If all detectors measure all objects, i.e., $J_i = \{1, \ldots, M\}$, $I_j = \{1, \ldots, N\}$ and $\{\sigma_i^2, \tau_i^2\}$ are known constants, then the variances of $\{\hat{B}_i\}$, $\{\hat{G}_j\}$ are given by*

$$\mathrm{Var}(\hat{G}_j) = \left[\sum_{i=1}^N \sigma_i^{-2}\right]^{-1} \mathcal{S}_G, \quad \mathrm{Var}(\hat{B}_i) = \left[M\sigma_i^{-2} + \tau_i^{-2}\right]^{-1} \mathcal{S}_B^{(i)}, \tag{E.1}$$

where the inflation factors $\mathcal{S}_G, \{\mathcal{S}_B^{(i)}\}$ *are given by*

$$\mathcal{S}_G = 1 + \frac{\sum_{i=1}^N \sigma_i^{-2} W_i}{\sum_{i=1}^N \tau_i^{-2} W_i}, \quad \mathcal{S}_B^{(i)} = 1 + \frac{M\sigma_i^{-2} W_i}{\sum_{u=1}^N \tau_u^{-2} W_u}.$$

Moreover, we have $\mathrm{Cov}(\hat{B}_i, \hat{G}_j) = -W_i \left[\sum_{k=1}^N \tau_k^{-2} W_k\right]^{-1}$.

Remark 1. *Under the additive model,* B_i *and* G_j *are negatively correlated for all* i, j*. The asymptotic variances of* $\hat{B}_i$ *and* $\hat{G}_j$ *can be written as*

$$\mathrm{Var}(\hat{G}_j) = \mathrm{Var}(\tilde{G}_j)\mathcal{S}_G, \quad \mathrm{Var}(\hat{B}_i) = \mathrm{Var}(\tilde{B}_i)\mathcal{S}_B^{(i)},$$

where $\mathrm{Var}(\tilde{G}_j) = \left[\sum_{i=1}^N \sigma_i^{-2}\right]^{-1}$ *is the inverse precision, i.e., asymptotic covariance, of* $\hat{G}_j$ *when the* B_i *are known constants;* $\mathrm{Var}(\tilde{B}_i) = [M\sigma_i^{-2} + \tau_i^{-2}]^{-1}$ *is the inverse precision, i.e., asymptotic covariance, of* $\hat{B}_i$ *when the* G_j *are known constants. The inflation factors* $\mathcal{S}_G$ *and* $\mathcal{S}_B^{(i)}$ *adjust for the fact that none of the* B_i *or the* G_j *are known.*

Proposition 1 directly yields the following asymptotic results as $N, M \to \infty$.

Corollary 1. *If* $\{\sigma_i/\tau_i\}$ *are uniformly bounded from below and above by finite positive constants, and* $\sum_{i=1}^N \sigma_i^{-2}/N$ *converges to a positive constant as* $N \to \infty$*, then for all* i, j*, as* $N, M \to \infty$*,*

$$\mathrm{Var}(\hat{G}_j) = O(N^{-1}),\ \mathrm{Var}(\hat{B}_i) = O(N^{-1} + M^{-1}),\ \mathrm{Cov}(\hat{B}_i, \hat{G}_j) = -O(N^{-1}).$$

Specifically, when $\tau = \tau_1 = \cdots = \tau_N$ *and* $\sigma = \sigma_1 = \cdots = \sigma_N$*,* (E.1) *simplifies to*

$$\mathrm{Var}(\hat{G}_j) = \frac{\sigma^2}{N},\ \mathrm{Var}(\hat{B}_i) = \frac{1}{M\sigma^{-2} + \tau^{-2}}\left(1 + \frac{M\sigma^{-2}}{N\tau^{-2}}\right),\ \mathrm{Cov}(\hat{B}_i, \hat{G}_j) = -\frac{\tau^2}{N}.$$

Remark 2. *The results above show that the asymptotic variances for* $\{B_i\}$ *and* $\{G_j\}$ *are not 'exchangeable' (i.e., switching* $\boldsymbol{B}$ *and* $\boldsymbol{G}$ *and correspondingly* N *and* M*), mainly for three reasons: first, for each* B_i *we assign an informative prior* $\mathcal{N}(b_i, \tau_i^2)$ *whereas for each* G_j *we assign a flat prior on the real line; second, for each instrument* i*, besides* B_i*, we also need to estimate* σ_i^2*; last, the measurement uncertainty depends only on the instrument but not on the sources (recall that* $\sigma_{ij}^2 = \sigma_i^2$ *for all* i, j*).*

E.2 Goodness-of-fit

We now give a goodness-of-fit test statistic for the random-effect regression model. Under the model (2.4), we have the following normalized residual sum of squares:

$$T(\boldsymbol{B}, \boldsymbol{G}) := \sum_{i=1}^{N} \frac{(b_i - B_i)^2}{\tau_i^2} + \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left(y'_{ij} - B_i - G_j\right)^2}{\sigma_i^2}. \tag{E.2}$$

We see this sum of squares has two parts. The first part involves $\{b_i\}$ only, measuring how good the prior means are relative to the prior variances $\{\tau_i\}$. The second part depends on $\{y_{ij}\}$ only, and it will allow us to access how good the fitted $\boldsymbol{B}, \boldsymbol{G}$ are relative to the sampling variances $\boldsymbol{\sigma}^2$. Here we put them together as an overall model check, but one can certainly use them individually if one wants to check the prior distribution and likelihood model separately.

Theorem 2. *When the variances σ_i^2, τ_i^2 are known and we insert the MLEs of B_i and G_j into (E.2), we obtain $T(\hat{\boldsymbol{B}}, \hat{\boldsymbol{G}}) \sim \chi^2_{NM-M}$.*

Proof. This conclusion regarding χ^2 distribution follows from standard results on residual sum of squares of linear regression with Gaussian error. To figure out the correct degrees of freedom, we have $(NM + N)$ independent observations in total, but with $N + M$ parameters. Therefore, the degrees of freedom for the residual sum of squares is $NM - M$. □

With unknown variances we do not have a closed-form distribution of T as defined in formula (E.2). Heuristically, we invoke the standard large-sample arguments and to continuously use the χ^2 approximation, but reduce the degrees of freedom to $MN - M - N$ to count for the number of estimated variance parameter $\{\sigma_i^2\}$. The resulting p-values of the fitted data in Sections 4.2 and 4.3 are not significant.

F More Simulation Results Under Misspecified Models

In Simulations IV and V, we generate data as $c_{ij} = \lambda_{ij} X_{ij}$, where $X_{ij} \sim \text{Poisson}(A_i F_j)$, and independently $\lambda_{ij} \sim \text{Uniform}[0.8, 1.2]$ for Simulation IV and $\lambda_{ij} \sim \text{Uniform}[0.4, 1.6]$ for Simulation V. In Simulations VI and VII, we generate data from $c_{ij} \sim \text{Poisson}(\lambda_{ij} A_i F_j)$, where the λ_{ij} are randomly generated from the uniform distribution on $[0.8, 1.2]$. The other parameters are set to be

the same as in Simulation II except that $\beta = 0.01$ for these simulations. Simulations VI and VII resemble the cases where the true model is Poisson and the estimation of T_{ij} is volatile, whereas Simulations IV and V resemble the cases that happen in practice, where the photon counts are multiplied by an adjustment factor, such as $\hat{T}_{ij}^{-1}$, as with the data pre-processing step for the XCAL data.

Figure 10 gives the results of Simulations IV and V. Figure 11 gives the results of Simulation VI with smaller counts ($B_i = 1$ and $G_j = 3$) and VII with larger counts ($B_i = 5$ and $G_j = 3$) under this scenario. It shows with large Poisson counts, controlling the uncertainty in the multiplicative constant can possibly lead to reasonably good results. Thus, even with compounded model misspecification, the log-Normal model is able to provide reasonable, though not as precise, results, as compared with the correctly-specified case. However, when the misspecified "known constant" is highly variable, the fit result is not as satisfactory; plugging in a "guesstimated" σ_i in this case can give disastrously optimistic but biased results.

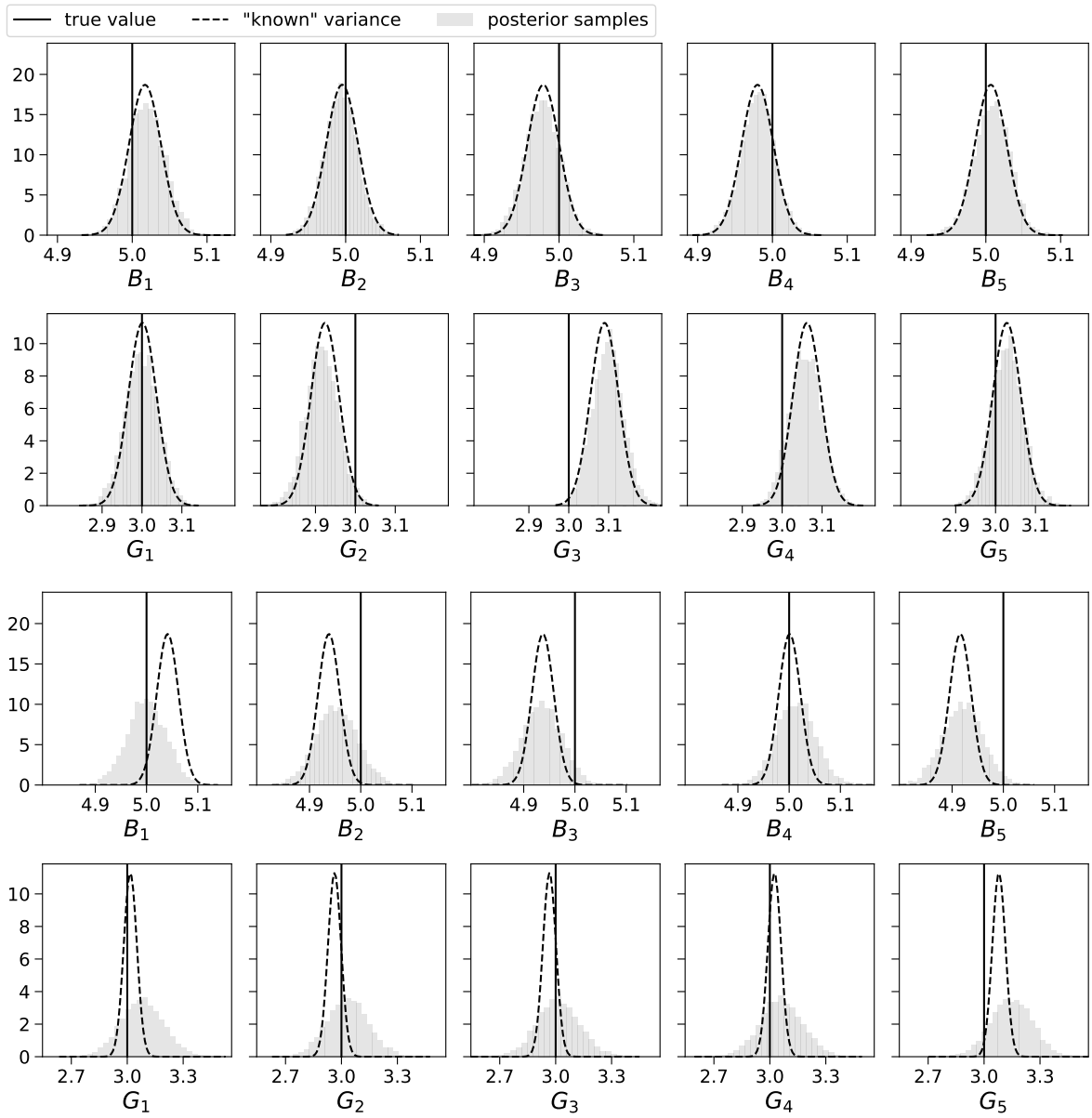

Figure 10: Simulations IV (rows 1 & 2) and V (rows 3 & 4). The gray shades are the posterior distributions of $\{B_i\}_{i=1}^5$ (rows 1 & 3) and $\{G_j\}_{j=1}^5$ (rows 2 & 4) fitted with unknown variances. The solid vertical black lines denote the true values. The black dashed density curves on top of the histograms denote the true posterior densities of $\{B_i\}$ and the $\{G_j\}$ with 'known' variances $\sigma_i^2 = 0.1^2$.

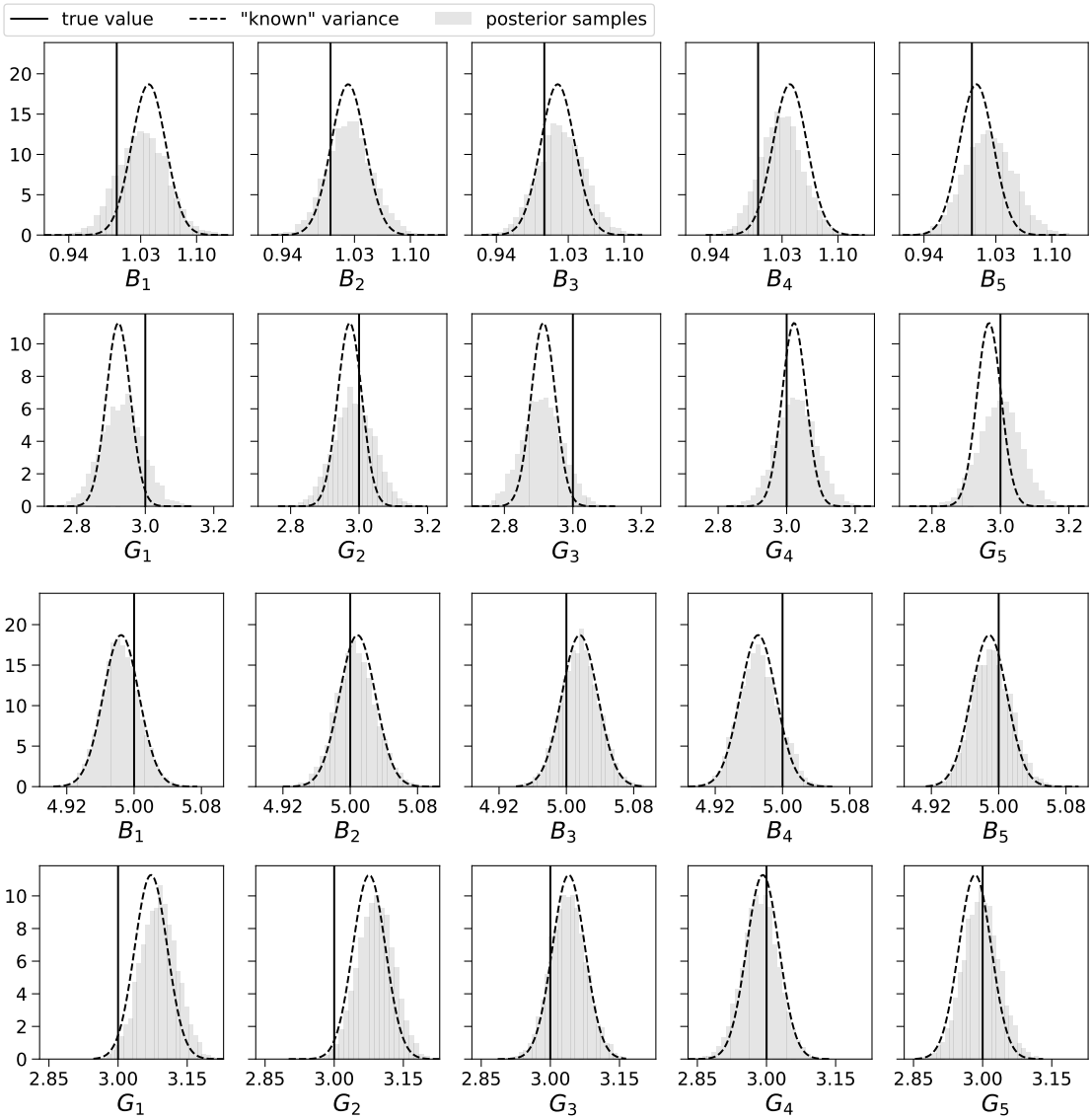

Figure 11: Same as Figure 10 but with Simulations VI (rows 1 & 2) and VII (rows 3 & 4).