Supplementary Material for

"Prediction of General Small Area Parameters for Unit-Level Count Data"

In this supplement, we present extended simulation output. The primary purpose of this section of the supplement is to address the secondary objectives of the simulation study. The secondary objectives are described in the first paragraph of Section 4 of the main document. We also use this supplement to present output that we did not have space to include in the main document. In Section 1, we present the root mean square error and bias of the predictors in the absolute (not relative) scale. In Section 2, we present output using the `R` function `glmer`. Sections 3-5 concern the secondary objectives of the simulation. In section 3, we evaluate the choice of T for the Monte Carlo SIR algorithm used to compute the predictors for the GLMM model. In Section 4, we compare the computing times of the Gam-Pois and GLMM procedures. In Section 5, we present results for $L = 100$.

1 Raw (not relative) RMSE and RB

In the main manuscript, we present the relative root mean square errors and relative biases of the predictors. In this supplement, we present the absolute versions of those quantities. Specifically, we present the Monte Carlo (MC) root mean square error and the MC bias. Let $\hat{\theta}_i^{(m)}$ and $\theta_i^{(m)}$, respectively, denote a predictor and true parameter obtained in MC simulation m. We define the MC RMSE for area i as

$$RMSE_i = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\theta}_i^{(m)} - \theta_i^{(m)})^2}.$$

We define the MC bias for area i as

$$Bias_i = \frac{1}{M}\sum_{m=1}^{M}(\hat{\theta}_i^{(m)} - \theta_i^{(m)}).$$

We report the average RMSE and average absolute bias defined, respectively, as

$$\text{RMSE} = \frac{1}{D}\sum_{i=1}^{D} RMSE_i \tag{S1}$$

and

$$\text{Bias} = \frac{1}{D}\sum_{i=1}^{D} \mid Bias_i \mid . \tag{S2}$$

Tables 1 and 3 report the Bias and RMSE for the simulations in which the gamma-Poisson model is the true model. Tables 2 and 4 report the Bias and RMSE for the simulations in which the GLMM is the true model. For Tables 1 and 2, $L = 100$, and for Tables 3 and 4, $L = 1000$. The MC sample size is $M = 5000$ when $L = 100$ and is $M = 500$ when $L = 1000$.

Table 1: RMSE and Bias when gamma-Poisson is true model. $L = 100$

		Mean		Med.		IQR	
Predictor	α	RMSE	Bias	RMSE	Bias	RMSE	Bias
Gam-Pois	5.000	1.191	0.013	0.836	0.009	1.292	0.015
Gam-Pois-Alt	5.000	1.185	0.013				
GLMM	5.000	1.196	0.018	0.839	0.015	1.296	0.016
PI	5.000	1.191	0.032	0.857	0.135	1.511	0.682
Direct	5.000	4.198	0.052	3.280	0.627	5.266	1.128
Gam-Pois	0.500	0.378	0.004	0.306	0.003	0.519	0.006
Gam-Pois-Alt	0.500	0.376	0.004				
GLMM	0.500	0.385	0.019	0.309	0.008	0.524	0.008
PI	0.500	0.383	0.008	0.337	0.125	0.635	0.213
Direct	0.500	0.740	0.008	0.671	0.105	1.031	0.143

Table 2: RMSE and Bias when GLMM is true model. $L = 100$

Predictor	σ_b^2	Mean RMSE	Mean Bias	Med. RMSE	Med. Bias	IQR RMSE	IQR Bias
Gam-Pois	0.500	0.756	0.008	0.750	0.008	0.717	0.012
Gam-Pois-Alt	0.500	0.753	0.008				
GLMM	0.500	0.747	0.015	0.743	0.015	0.712	0.010
PI	0.500	0.743	0.009	0.755	0.122	1.219	0.970
Direct	0.500	1.197	0.012	1.387	0.124	1.841	0.705
Gam-Pois	1.500	1.020	0.012	0.994	0.012	0.973	0.020
Gam-Pois-Alt	1.500	1.016	0.012				
GLMM	1.500	1.008	0.013	0.985	0.013	0.965	0.013
PI	1.500	1.003	0.016	1.007	0.135	1.401	0.933
Direct	1.500	2.653	0.030	2.901	0.193	3.935	0.973

Table 3: RMSE and Bias when gamma-Poisson model is true model. $L = 1000$

Predictor	σ_b^2	Mean RMSE	Mean Bias	Med. RMSE	Med. Bias	IQR RMSE	IQR Bias
Gam-Pois	5.000	1.174	0.045	0.822	0.032	1.278	0.050
Gam-Pois-Alt	5.000	1.174	0.045				
GLMM	5.000	1.181	0.048	0.825	0.036	1.282	0.051
PI	5.000	1.180	0.048	0.846	0.131	1.504	0.680
Direct	5.000	4.164	0.140	3.225	0.632	5.141	1.108
Gam-Pois	0.500	0.372	0.013	0.302	0.011	0.514	0.017
Gam-Pois-Alt	0.500	0.372	0.013				
GLMM	0.500	0.380	0.019	0.305	0.013	0.518	0.019
PI	0.500	0.379	0.015	0.334	0.125	0.633	0.212
Direct	0.500	0.724	0.027	0.663	0.106	1.023	0.139

Table 4: RMSE and Bias when GLMM is true model. $L = 1000$

		Mean		Med.		IQR	
Predictor	σ_b^2	RMSE	Bias	RMSE	Bias	RMSE	Bias
Gam-Pois	0.500	0.753	0.025	0.750	0.026	0.712	0.028
Gam-Pois-Alt	0.500	0.752	0.025				
GLMM	0.500	0.743	0.026	0.741	0.027	0.705	0.028
PI	0.500	0.743	0.025	0.757	0.121	1.213	0.967
Direct	0.500	1.192	0.043	1.387	0.122	1.821	0.696
Gam-Pois	1.500	1.010	0.031	0.984	0.032	0.961	0.033
Gam-Pois-Alt	1.500	1.009	0.031				
GLMM	1.500	0.999	0.032	0.975	0.032	0.954	0.033
PI	1.500	0.999	0.031	1.002	0.139	1.394	0.927
Direct	1.500	2.678	0.099	2.942	0.199	3.824	0.961

The conclusions based on the absolute bias and RMSE are the same as the conclusions based on the relative measures reported in the main document. When the gamma-Poisson model is true, the Gam-Pois-Alt predictor is most efficient for the mean, and the Gam-Pois predictor is more efficient than the GLMM predictor. The GLMM predictor is more efficient than the Gam-Pois predictor when the GLMM model is true. The PI predictor is more efficient than the GLMM predictor for the mean when $L = 100$, but the PI predictor is inefficient for the median and for the IQR. Increasing L improves the efficiency of the Gam-Pois and GLMM predictors. The loss of efficiency from incorrect use of the GLMM or Gam-Pois predictor when the other model is true is negligible relative to the loss of efficiency from the use of the direct estimator. The bias is a small fraction of the RMSE, indicating that the variance is more important than the bias.

2 Output Using GLMER

We repeat the simulation, where we use the `R` function `glmer` to calculate the GLMM and PI predictors. The %RRMSE and %RB based on `glmer` are provided in Tables 5 and 6 below.

The use of `glmer` improves the GLMM and PI predictors slightly. Overall, the conclusions based on `glmer` are the same as the results obtained from the IRLS algorithm presented in the main document.

Table 5: Relative bias and relative root mean square error of the alternative predictors when the gamma-Poisson model is the true model. $L = 100$.

		Mean		Med		IQR	
	α	RRMSE	RB	RRMSE	RB	RRMSE	RB
Gam-Pois	5.000	17.530	0.189	20.507	0.214	19.527	0.195
Gam-Pois-Alt	5.000	17.445	0.189				
GLMM	5.000	17.616	0.187	20.580	0.213	19.590	0.195
PI	5.000	17.532	0.290	21.036	3.311	22.931	10.478
Direct	5.000	61.233	0.695	80.507	15.357	79.954	16.836
Gam-Pois	0.500	55.765	0.582	109.058	1.298	64.523	0.776
Gam-Pois-Alt	0.500	55.501	0.580				
GLMM	0.500	56.890	0.650	110.213	3.695	65.272	3.413
PI	0.500	56.792	3.554	118.525	39.865	78.630	28.455
Direct	0.500	107.558	1.121	235.484	36.186	126.890	17.814

Table 6: Relative bias and relative root mean square error of the alternative predictors when the Poisson-GLMM model is the true model. $L = 100$.

		Mean		Med		IQR	
	α	RRMSE	RB	RRMSE	RB	RRMSE	RB
Gam-Pois	0.500	24.552	0.258	28.902	0.320	25.219	0.379
Gam-Pois-Alt	0.500	24.442	0.259				
GLMM	0.500	24.254	0.258	28.608	0.309	25.019	0.283
PI	0.500	24.146	0.432	29.033	4.385	43.018	34.345
Direct	0.500	38.861	0.426	53.270	4.769	64.620	24.709
Gam-Pois	1.500	20.107	0.222	22.858	0.262	24.267	0.508
Gam-Pois-Alt	1.500	20.024	0.222				
GLMM	1.500	19.834	0.222	22.621	0.248	24.071	0.275
PI	1.500	19.741	0.250	23.101	3.190	35.105	23.540
Direct	1.500	51.370	0.586	65.581	4.547	96.024	24.215

3 Evaluation of the choice of T in the SIR algorithm

We evaluate the choice of T for the sampling importance resampling algorithm. We simplify the output and only present results for the GLMM model and procedure. Table 7 compares the RRMSE and RB of the alternative predictors. The RRMSE and RB are defined in (22) and (23) of the main document. Table 8 contains the corresponding absolute measures, defined as RMSE and Bias in (S1) and (S2) of Section 1.1 of this supplement. We use "GLMM, T = X" to denote the GLMM predictor with $T = X$. We consider $T = 200$ and $T = 2000$. For all simulations in this section, we set $L = 100$.

The main conclusion from Table 7 and Table 8 is that the effect of increasing T is minimal. For each combination of σ_b^2 and small area parameter of interest, the RRMSE based on $T = 200$ is nearly the same as the RRMSE based on $T = 2000$. That said, it is counter-intuitive that the RRMSE for $T = 2000$ tends to exceed the RRMSE for $T = 200$. We conjecture that this counter-intuitive result is due to the bias of the IRLS estimators of the fixed parameters. To support the conjecture, we repeated the comparison of $T = 200$

to $T = 2000$ using the true values of the fixed parameters instead of the estimated values. When we use the true parameters, we find that the RRMSE for $T = 200$ tends to exceed the RRMSE for $T = 2000$, as expected. We therefore think that the bias of the estimators of the fixed parameters is the main factor that leads to the slight increase in RRMSE when T increases from 200 to 2000.

In our discussion of the comparison of $T = 200$ to $T = 2000$, we have focused on the relative root mean square error. The conclusions based on the raw root mean square error are the same as the conclusions based on the relative root mean square error. The raw root mean square errors and biases for $T = 200$ are the same as those for $T = 2000$ through the third decimal place. We focus on mean square error instead of bias because the contribution from the bias to the overall mean square error of the predictor is negligible.

In summary, increasing T leads to a slight increase in the RRMSE and RB of the predictor. The evaluation of the choice of T supports the value of $T = 200$, used in the main manuscript.

Table 7: Relative bias and relative root mean square error of the alternative predictors when the GLMM is the true model.

		Mean		Med		IQR	
	σ_b^2	RRMSE	RB	RRMSE	RB	RRMSE	RB
GLMM, $T = 200$	1.500	19.946	0.448	22.763	0.479	24.578	0.639
GLMM, $T = 2000$	1.500	19.953	0.442	22.768	0.487	24.596	0.645
GLMM, $T = 200$	0.500	24.477	0.669	28.949	0.790	25.197	0.659
GLMM, $T = 2000$	0.500	24.479	0.671	28.940	0.799	25.200	0.672

Table 8: Absolute bias and root mean square error of the alternative predictors when the GLMM is the true model.

		Mean		Med		IQR	
	σ_b^2	RMSE	Bias	RMSE	Bias	RMSE	Bias
GLMM, $T = 200$	1.500	1.007	0.022	0.986	0.021	0.982	0.026
GLMM, $T = 2000$	1.500	1.008	0.022	0.986	0.021	0.983	0.026
GLMM, $T = 200$	0.500	0.746	0.020	0.743	0.020	0.708	0.018
GLMM, $T = 2000$	0.500	0.746	0.020	0.743	0.020	0.708	0.019

4 Comparison of Computing Times of Alternative Procedures

We compare the computing time of the GLMM procedure to the computing time of the Gam-Pois procedure. When implementing the GLMM procedure we split the calculations for the sampling importance resampling procedure onto three cores that are run in parallel. Table 9 contains the computing times for the four simulation models and the two procedures. The computing time of the Gam-Pois procedure is approximately half the computing time of the GLMM procedure. The comparison of the computing times of the two procedures supports the conclusion that a benefit of the gamma-Poisson model is computational simplicity.

Table 9: Computing times (seconds) of two procedures for four simulation models.

Model	Gam-Pois	GLMM
GLMM, $\sigma_b^2 = 0.5$	4.28	8.35
GLMM, $\sigma_b^2 = 1.5$	3.98	8.91
Gamma-Poisson, $\alpha = 5$	3.83	8.58
Gamma-Poisson, $\alpha = 0.5$	4.08	9.16

5 Results for $L = 100$

Table 10: %RB and %RRMSE of alternative predictors when the true model is the gamma-Poisson model and $L = 100$

		Mean		Med.		IQR	
	α	%RRMSE	%RB	%RRMSE	%RB	%RRMSE	%RB
Gam-Pois	5.000	17.544	0.192	20.538	0.231	19.559	0.232
Gam-Pois-Alt	5.000	17.459	0.192				
GLMM	5.000	17.621	0.268	20.606	0.378	19.614	0.248
PI	5.000	17.541	0.475	21.058	3.324	23.018	10.639
Direct	5.000	61.162	0.758	80.339	15.371	79.867	16.787
Gam-Pois	0.500	55.560	0.641	108.581	1.200	64.293	0.775
Gam-Pois-Alt	0.500	55.313	0.646				
GLMM	0.500	56.687	2.774	109.498	2.934	64.926	0.938
PI	0.500	56.311	1.122	119.499	44.742	78.893	26.475
Direct	0.500	107.935	1.206	237.138	37.307	127.339	17.519

Table 11: %RB and %RRMSE of alternative predictors when the true model is the Poisson-GLMM model and $L = 100$

		Mean		Med.		IQR	
	σ_b^2	%RRMSE	%RB	%RRMSE	%RB	%RRMSE	%RB
Gam-Pois	0.500	24.592	0.278	28.913	0.315	25.248	0.437
Gam-Pois-Alt	0.500	24.482	0.276				
GLMM	0.500	24.307	0.491	28.633	0.567	25.060	0.362
PI	0.500	24.185	0.311	29.123	4.702	43.103	34.373
Direct	0.500	38.882	0.402	53.443	4.773	64.769	24.815
Gam-Pois	1.500	20.129	0.243	22.884	0.289	24.332	0.495
Gam-Pois-Alt	1.500	20.032	0.245				
GLMM	1.500	19.880	0.263	22.672	0.294	24.130	0.317
PI	1.500	19.779	0.309	23.167	3.107	35.245	23.626
Direct	1.500	52.235	0.598	66.716	4.448	98.095	24.274