HOW GOOD ARE SPATIAL GLM'S? A SIMULATION STUDY

Roger G. Collins

Walter W. Stroup

Stephen D. Kachman

HOW GOOD ARE SPATIAL GLM'S? A SIMULATION STUDY.

by
Roger G. Collins, Walter W. Stroup, and Stephen D. Kachman
Department of Biometry
University of Nebraska-Lincoln

ABSTRACT

An area of increasing interest to agricultural and ecological researchers is the analysis of spatially correlated non-normal data. A generalized linear model (GLM) accounting for spatial covariance was presented by Gotway and Stroup (1997). Their method included approximate inference based on asymptotic distributions. A simulation study was conducted to assess the small sample behavior of their proposed estimates and test statistics. This study suggests that the spatial GLM yields unbiased estimates of treatment means and differences for binomial data, that the spatial GLM improves precision, as measured by MSE, and that the approximate F-statistic is acceptable for hypothesis testing.

1. INTRODUCTION

Spatial variability refers to the tendency of adjacent or nearby observations to be more alike than those that are farther apart. Spatial variability occurs in many agricultural and ecological experiments. In recent years, the need for methods to account or adjust for spatial variability has been increasingly appreciated by biological researchers. These methods typically involve modeling spatial correlation among the errors; however, standard spatial correlation methods assume normality. Agricultural and ecological researchers frequently do experiments or studies in which spatial variability is present **and** the response variable(s) of primary interest are non-normal, e.g. categorical or count data.

The **generalized linear model** (GLM) presented by Nelder and Wedderburn (1972) is a generalization of linear model theory to non-normal members of the exponential family. Wedderburn (1974), McCullagh (1983), and others extended the GLM approach to more general types of data using quasi-likelihood. Standard GLM's assume independent errors. Zeger and Liang (1986) proposed GLM's for serially correlated observations in repeated measures data.

Gotway and Stroup (1997) proposed a GLM for spatially correlated data. Their approach included approximate inference involving *ad hoc* adaptations of

Kansas State University

inference for standard GLM's. However, they did not address the small sample properties of these procedures. The purpose of this paper is to present the results of a simulation study to evaluate the behavior of estimates and test-statistics proposed by Gotway and Stroup.

2. REVIEW OF THE GOTWAY-STROUP METHOD

The standard GLM assumes an n x 1 vector of random variables, $\mathbf{y} = [y_1, y_2, \ldots, y_n]$, such that $E(y_i)=\mu_i$ and $Var(y_i) = v(\mu_i)a(\phi_i)$ where $v(\mu_i)$ is the variance function and $a(\phi_i)$ is the scale parameter. Thus, $E(\mathbf{y})=\mu$ and $Var(\mathbf{y})= V = V_\mu^{1/2} A V_\mu^{1/2}$, where $V_\mu^{1/2} = diag[sqrt(v(\mu_i)]$ and $A = diag[a(\phi_i)]$. The standard GLM is

$$\mu = h(X\beta)$$

where X is an n x p matrix of known constants, β is a p x 1 parameter vector and $h(\cdot)$ is the inverse link function.

The spatial GLM proposed by Gotway and Stroup generalizes V, so that $V(y) = V_\mu^{1/2} R(\alpha) V_\mu^{1/2}$, where $R(\alpha)$ is a "working spatial correlation matrix." Examples of working spatial correlation matrices include semivariogram models, such as the spherical and exponential, commonly used in geostatistics. For example, letting r_{ij} be the ij^{th} element of $R(\alpha)$, the spherical model is

$$r_{ij} = 1 - 0.5(d_{ij}/\alpha) + 1.5\,(d_{ij}/\alpha)^3 \text{ if } d_{ij} < \alpha, \text{ and}$$

$$r_{ij} = 0 \text{ if } d_{ij} > \alpha,$$

where d_{ij} is the distance between the i^{th} and j^{th} observations.

The parameters of the spatial GLM are estimated by solving the equation

$$X'WX\beta = X'Wy^*$$

where $y^* = X\beta + D^{-1}(y-\mu)$

$$D^{-1} = diag[\partial h^{-1}(\mu_i)/\partial\mu_i], \text{ and}$$

$$W=DV^{-1}D.$$

Approximate inference on estimable functions of the form $\mathbf{K}'\beta$ is based on the following asymptotic results:

1. $AV(\mathbf{K}'\beta) = \mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}$
2. Hence, for vector $\mathbf{k}$, the asymptotic standard error of $\mathbf{K}'\beta$ is
   $a.s.e.(\mathbf{K}'\beta) = sqrt[\mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{K}]$
3. The Wald statistic for testing H_0: $\mathbf{K}'\beta = 0$,
   $(\mathbf{K}'\beta)'[\mathbf{K}'(\mathbf{X}'\mathbf{W}\mathbf{X})\mathbf{K}]^{-1}(\mathbf{K}'\beta)$
   is asymptotically χ^2 with rank($\mathbf{K}$) degrees of freedom.

4. The Wald statistic divided by ϕ[rank(**K**)], where ϕ is the estimated scale parameter, is asymptotically distributed $F_{(\nu 1, \nu 2)}$ where ν_1 = rank(K) and ν_2 = error degrees of freedom. The test statistic Wald/{ϕ[rank(**K**)]} is hereafter referred to as the F-statistic for H_0: $\mathbf{K}'\beta = 0$.

The Wald and F-statistic produce similar results if ϕ=1, i.e. if A=I in Var(**y**) = $V_\mu^{1/2} A V_\mu^{1/2}$. However, if ϕ>1, the Wald statistic will over-reject H_0. In the spatial GLM, ϕ>1 is expected as a result of the contribution of spatial correlation, R(α), to Var(**y**) in addition to the variance function V_μ.

3. THE SIMULATION STUDY

The small sample properties of the spatial GLM were investigated using a simulated experiment consisting of 4 replications of 16 treatments. The treatments were laid out on an 8 x 8 grid so that each replication consisted of a 4 x 4 balanced lattice (see Figure 1). For each of the 64 experimental units, spatially correlated binomial data were generated as follows:

1. The working correlation matrix R(α) was determined. In these simulations, the range was set to 3, i.e. α=3.

2. Normal deviates were generated for each experimental unit. For the ijth experimental unit, corresponding to the jth replication (j=1, 2, 3, 4) of the ith treatment (i=1, 2, ..., 16), the random deviate is denoted z_{ij}. To simulate spatial correlation, the z_{ij} were generated so that $\mathbf{z} = [z_{11}, z_{12, ,,,,}, z_{16,4}] \sim MVN[0, R(\alpha)]$.

3. The probability of a "success" for the ijth experimental unit, denoted $\pi_{ij,}$ was determined as $\pi_{ij} = \Phi^{-1}(z_{ij} - \tau_i)$, where τ_i is the effect of the ith treatment. Note that τ_i sets π_{ij} through the inverse normal c.d.f. For example, τ_i=0 would imply that the expected π_{ij} for the ith treatment is 0.5. Several configurations of τ_i to represent different patterns of treatment effects were used. These are described below.

4. The number of "successes", denoted y_{ij}, out of n_{ij} binary observations on the ijth experimental unit were generated from a Bin(n_{ij}, π_{ij}) random number generator. For this simulation, n_{ij} was set to either 10 or 50.

The data were generated with the 17 different τ vectors that appear in Table 1. The τ vectors used represent a mix of equal treatment effects and unequal treatment effects. The equal treatment effect vectors represent a spectrum from low probability of a success (τ = -1.5) to high probability (τ = 1.5). The unequal

Kansas State University

treatment effect vectors represent some scenarios where treatment effects are relatively close and others where treatment effects are farther apart.

A Fortran program was written to generate the data and compute the analysis. The random number generators IGNBIN, to generate binomial random deviates, and AS 66, to evaluate the tail area of the standardized normal curve, that were used in the Fortran program were acquired from http://lib.stat.cmu.edu/apstat/. The data were analyzed using a standard GLM, i.e. where the range, α=0, and hence R(α)=I, and using a spatial GLM with range equal 3. The authors are aware of no programs available at the present time to perform this type of analysis. The current option is to write one's own code in Fortran, C++, or SAS IML. Fortran was chosen for its speed in performing this type of analysis.

Analyses using the standard [R(0)] and spatial [R(3)] GLM were compared using the following criteria:

1. The mean and MSE of the estimated π_i and four selected differences, π_i-$\pi_{i'}$: the four differences used were, Treatment 1 versus Treatment 2, Treatment 1 versus Treatment 5, Treatment 1 versus Treatment 10, and Treatment 1 versus Treatment 15.

2. The percent rejection rate of the Wald tests for overall equal means and four differences.

3. The percent rejection rate of the F tests for overall equal means and four differences. These four differences were chosen because they represent groups of direct and indirect comparisons for the balanced lattice design that was used.

Of particular interest was how well the quantiles of the estimates fit the expected asymptotic quantiles under theory. Theory predicts that the estimated π_i and π_i-$\pi_{i'}$ are distributed normally. Under H_o, the Wald statistic is asymptotically χ^2, and the Wald statistic adjusted for ϕ is asymptotically F. For the vectors with equal treatment means, the distributions were checked by comparing the observed 5^{th}, 25^{th}, 75^{th}, 95^{th} quantiles with the expected quantiles.

4. RESULTS

Table 2 gives the estimated MSE, bias, % rejection rate of Wald tests, % rejection rate of F tests, and the number of simulated experiments whose spatial GLM estimation algorithm converged for each combination at n_{ij}=10 level. MSE is consistently lower for the spatial GLM than for the standard GLM. Both procedures show negligible bias. As expected, the Wald statistic uncorrected for overdispersion is biased upward, resulting in excessive rejection rates. On the other hand, the F tests in the table show more reasonable rejection rates. There is a tendency for the spatial GLM to come closer than the standard GLM to the nominal 5% level when the null hypothesis is true. The spatial GLM algorithm was deemed to have converged if the convergence criteria was met after 20 iterations. Generally, this occurred over 95% of the time. Inspection of the cases that failed to converge indicate that convergence would have occurred if the iterations continued a few more rounds.

Table 3 gives the estimated MSE, bias, % rejection rate of Wald tests, % rejection rate of F tests, and the number of iterations that converged for each combination at n_{ij}=50 level. The results are generally similar to those observed when n_{ij}=10. The only difference between n_{ij}=50 and n_{ij}=10 is that the F tests seem to be closer to the nominal 5% level in the n_{ij}=50 table.

Table 4 gives the expected quantiles of the F distribution under H_o and the observed quantiles of the F-statistics for the mean vectors where treatment effects were equal. For both n_{ij}=10 and n_{ij}=50, the empirical distribution of the F-statistics show reasonable agreement with the expected quantiles.

5. DISCUSSION

1. The uncorrected Wald statistics are strongly biased upward, producing excessive rejection rates. In no case did the Wald statistic give reliable results.
2. Under H_o, the F tests using the spatial GLM appear to come close to the nominal 5% level when spatial error structure is included in the analysis.
3. In general, when H_o is false, including the spatial structure of the errors in the GLM results in more powerful F tests.
4. Finally, and most importantly it appears that the small sample behavior of the F tests are consistent with their expected behavior under asymptotic theory. This result gives credibility to the use of approximate F-statistics for hypothesis testing in spatial GLM's.
5. Taking (1) through (4) together, we strongly recommend use of the approximate F rather than the Wald statistic for hypothesis testing in spatial GLM's.
6. Spatial GLM appears to produce unbiased estimates of treatment means and differences.
7. MSE of estimates of π_i, π_i-$\pi_{i'}$ are lower for spatial GLM than for standard GLM.

Kansas State University

The main result of this study was to examine the small-sample properties of the spatial GLM proposed by Gotway and Stroup (1997). The results suggest that the estimates of treatment means and differences are unbiased and reduce MSE compared to standard GLM alternatives when spatial variation is present. More importantly, they suggest that the small-sample behavior of the approximate F-statistic is acceptable for use in hypothesis testing.

Although these results are promising, more work needs to be done. These results apply to cases where the response is binary, and the spatial correlation model and the range are known. In the future, we plan to investigate the behavior on the spatial GLM when the range is estimated and when the model is misspecified. Also, we plan to look at variety of designs and response variable distributions.

These results only compared the spatial GLM to a standard GLM. We also plan to compare the spatial GLM to non-GLM alternatives. In this simulation, for instance, we plan to compare the spatial GLM to the standard ANOVA for a lattice design with the percent of successes per experimental unit as the response variable and the arc-sine square root transformation, as would likely be standard practice for most researchers.

Finally, a main feature of the Gotway and Stroup (1997) article was a method for spatial prediction. The small sample properties of this method were not addressed in this study. We plan to address them in future studies.

Figure 1. Example Layout of Simulated 4x4 Lattice Experiment in 4 Replications.
Double-Lines demark complete replications.

Row\Column	1	2	3	4	5	6	7	8
1	Trt 4	Trt 2	Trt 8	Trt 7	Trt 9	Trt 13	Trt 8	Trt 16
2	Trt 3	Trt 1	Trt 6	Trt 5	Trt 5	Trt 1	Trt 12	Trt 4
3	Trt 15	Trt 16	Trt 9	Trt 10	Trt 11	Trt 3	Trt 10	Trt 6
4	Trt 14	Trt 13	Trt 12	Trt 11	Trt 7	Trt 15	Trt 2	Trt 14
5	Trt 2	Trt 5	Trt 6	Trt 11	Trt 2	Trt 13	Trt 9	Trt 6
6	Trt 15	Trt 12	Trt 1	Trt 16	Trt 11	Trt 8	Trt 15	Trt 4
7	Trt 7	Trt 10	Trt 14	Trt 9	Trt 10	Trt 3	Trt 14	Trt 1
8	Trt 13	Trt 4	Trt 3	Trt 8	Trt 16	Trt 5	Trt 7	Trt 12

Table 1. Vectors of treatment effects (τ_i). P{success} = $\Phi^{-1}(z_{ij}-\tau_i)$, where z~MVN(0,R)

	Trt 1	Trt 2	Trt 3	Trt 4	Trt 5	Trt 6	Trt 7	Trt 8	Trt 9	Trt 10	Trt 11	Trt 12	Trt 13	Trt 14	Trt 15	Trt 16
A	-1	-1	-1	-1	0	0	0	0	1	1	1	1	1	1	1	1
B	-1	-1	-1	-1	1	1	1	1	0	0	0	0	0	0	0	0
C	0	0	0	0	-1	-1	-1	-1	1	1	1	1	1	1	1	1
D	0	0	0	0	-0.5	-0.5	-0.5	-0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
E	-1	-1	-1	-1	-0.5	-0.5	-0.5	-0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
F	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	0
G	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1
H	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
I	-1	0	0	0	0.5	0.5	0.5	0.5	1	1	1	1	1	1	1	1
J	-1	0	0	0	1	1	1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
K	-1	-0.5	-0.5	-0.5	0	0	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
L	-1	-0.5	-0.5	-0.5	0	0	0	0	1	1	1	1	1	1	1	1
M	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5
N	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Q	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 2. Simulation results for standard vs. spatial GLM, nij=10.
Convergence, MSE of estimates, and % rejection of F statistics.

	Converged		MSE		Overall F % rejection		F diff 1 % rejection		F diff 2 % rejection		F diff 3 % rejection		F diff 4 % rejection	
Sets of τ_i	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM
n_{ij}=10	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3
A	1000	956	0.01466	0.01175	98.1%	100.0%	2.0%	3.7%	43.9%	61.4%	94.7%	98.7%	95.0%	99.0%
B	1000	963	0.01822	0.01499	90.4%	99.8%	0.7%	2.1%	93.5%	98.5%	35.1%	55.2%	36.0%	52.6%
C	1000	956	0.01409	0.01178	98.9%	100.0%	7.8%	11.3%	41.3%	56.4%	42.3%	61.1%	40.1%	60.5%
D	1000	992	0.01997	0.01551	34.9%	85.6%	3.5%	7.1%	10.6%	20.6%	9.4%	21.5%	9.2%	20.7%
E	1000	972	0.01750	0.01463	89.2%	99.7%	1.3%	2.9%	7.6%	16.8%	74.9%	90.9%	74.9%	88.7%
F	1000	966	0.01783	0.01417	59.5%	89.0%	0.5%	2.4%	0.5%	2.1%	37.6%	56.9%	35.6%	56.2%
G	1000	968	0.01184	0.01006	100.0%	100.0%	3.1%	5.5%	2.8%	5.0%	96.0%	99.1%	96.9%	99.8%
H	1000	987	0.01965	0.01533	59.8%	96.1%	2.8%	4.8%	2.8%	5.4%	39.8%	61.1%	37.7%	61.3%
I	1000	954	0.01530	0.01256	72.1%	96.8%	38.4%	60.3%	75.1%	89.8%	92.4%	99.0%	93.3%	99.1%
J	1000	959	0.01820	0.01403	52.2%	88.5%	35.4%	55.6%	90.6%	98.9%	71.6%	89.7%	72.6%	89.9%
K	1000	980	0.02002	0.01569	50.2%	94.6%	5.7%	13.9%	32.0%	53.0%	70.2%	89.8%	68.4%	91.4%
L	1000	961	0.01519	0.01269	94.2%	99.9%	10.3%	15.9%	42.0%	59.8%	95.5%	98.6%	95.7%	99.2%
M	1000	937	0.00502	0.00421	13.8%	8.5%	5.6%	6.4%	5.1%	7.2%	5.6%	7.5%	5.7%	6.5%
N	1000	953	0.00611	0.00506	13.5%	8.9%	4.9%	8.3%	5.3%	6.7%	5.0%	8.0%	5.4%	8.7%
O	1000	1000	0.02406	0.01757	0.5%	3.4%	2.9%	6.3%	3.0%	5.3%	2.2%	6.2%	2.8%	5.9%
P	1000	930	0.01079	0.00919	4.0%	3.8%	3.1%	5.8%	3.9%	6.3%	3.3%	6.2%	2.4%	5.1%
Q	1000	959	0.01023	0.00792	4.7%	3.5%	3.3%	5.7%	3.9%	5.3%	2.7%	5.9%	3.5%	6.3%

Table 2. (Continued) Simulation results for standard vs. spatial GLM, nij=10.
Bias of estimates, and % rejection of Wald Statistics.

	Bias		Overall Wald % rejection		Wald diff 1 % rejection		Wald diff 2 % rejection		Wald diff 3 % rejection		Wald diff 4 % rejection	
Sets of τ_i	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM
n_{ij}=10	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3
A	-0.00665	-0.00550	18.1%	34.4%	73.0%	89.7%	99.8%	100.0%	99.9%	99.9%	100.0%	100.0%
B	-0.00148	-0.00391	17.2%	33.1%	99.8%	100.0%	76.4%	88.6%	75.5%	88.8%	100.0%	100.0%
C	-0.00488	-0.00466	29.7%	43.8%	76.8%	86.9%	74.7%	90.1%	76.5%	87.6%	100.0%	100.0%
D	-0.00198	-0.00289	29.0%	38.2%	45.9%	62.4%	44.1%	63.2%	45.9%	61.4%	99.8%	100.0%
E	0.00267	0.00701	17.0%	34.2%	39.4%	56.3%	96.3%	99.0%	95.7%	99.4%	100.0%	100.0%
F	0.01028	0.01522	20.3%	35.5%	17.7%	33.7%	76.3%	90.1%	75.4%	90.6%	100.0%	100.0%
G	-0.00159	0.00100	16.2%	36.3%	14.2%	35.7%	99.7%	99.9%	99.8%	100.0%	100.0%	100.0%
H	-0.00121	0.00078	28.7%	40.2%	24.8%	40.2%	78.8%	92.2%	81.0%	90.9%	100.0%	100.0%
I	-0.01039	-0.01278	76.4%	88.6%	94.7%	99.1%	99.5%	99.9%	99.5%	100.0%	100.0%	100.0%
J	-0.00911	-0.01332	74.2%	89.1%	99.6%	100.0%	95.4%	99.3%	95.6%	99.6%	100.0%	100.0%
K	-0.00118	-0.00520	37.5%	55.7%	75.9%	89.5%	95.9%	99.3%	95.8%	99.1%	100.0%	100.0%
L	-0.00228	-0.00838	38.0%	57.9%	76.0%	88.7%	99.7%	100.0%	99.7%	100.0%	100.0%	100.0%
M	0.03482	0.03471	5.2%	15.5%	5.4%	15.9%	5.2%	16.2%	6.3%	16.2%	20.6%	44.3%
N	-0.03392	-0.03409	4.9%	16.7%	4.4%	14.1%	4.8%	15.2%	4.2%	15.5%	23.2%	43.9%
O	0.00568	-0.00023	30.5%	38.7%	31.7%	39.2%	28.4%	41.7%	27.4%	40.8%	94.7%	99.6%
P	0.02039	0.02440	17.8%	31.7%	16.5%	30.6%	16.0%	31.4%	16.5%	29.8%	64.1%	90.9%
Q	-0.02067	-0.02135	17.9%	30.1%	17.4%	30.4%	17.7%	31.8%	16.1%	32.4%	64.3%	90.8%

Table 3. Simulation results for standard vs. spatial GLM, nij=50.
Converged, MSE of estimates, % rejection of F statistics.

	Converged		MSE		Overall F % rejection		F diff 1 % rejection		F diff 2 % rejection		F diff 3 % rejection		F diff 4 % rejection	
Sets of τ_i	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM
n_{ij}=50	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3
A	1000	952	0.01366	0.00972	99.2%	100.0%	1.3%	3.9%	45.1%	71.4%	96.8%	100.0%	96.2%	99.9%
B	1000	965	0.01741	0.01230	91.7%	99.9%	0.8%	2.0%	95.3%	99.9%	40.6%	68.2%	39.7%	71.7%
C	1000	957	0.01335	0.00952	99.2%	100.0%	5.1%	8.8%	40.1%	68.2%	41.8%	69.6%	42.9%	69.4%
D	1000	997	0.01938	0.01348	39.0%	95.1%	2.6%	5.6%	10.0%	24.6%	11.1%	27.2%	12.3%	24.8%
E	1000	969	0.01689	0.01190	92.2%	100.0%	0.2%	2.2%	8.1%	16.2%	76.8%	96.9%	76.1%	96.2%
F	1000	945	0.01666	0.01151	67.4%	96.4%	0.4%	2.0%	0.3%	2.4%	39.3%	70.6%	39.9%	71.1%
G	1000	951	0.01064	0.00815	100.0%	100.0%	2.2%	4.6%	1.9%	3.7%	97.6%	99.5%	97.6%	100.0%
H	1000	998	0.01828	0.01268	68.8%	99.5%	1.6%	4.9%	2.8%	5.9%	45.4%	75.4%	44.9%	76.7%
I	1000	937	0.01479	0.01056	72.7%	99.5%	40.0%	70.2%	76.5%	96.9%	95.1%	99.9%	96.3%	99.8%
J	1000	954	0.01722	0.01135	54.7%	97.1%	36.9%	72.2%	95.0%	99.9%	75.5%	96.8%	75.7%	97.2%
K	1000	983	0.01947	0.01320	54.0%	98.3%	4.8%	15.8%	32.8%	67.0%	75.5%	96.8%	72.6%	95.8%
L	1000	946	0.01473	0.01062	95.3%	100.0%	8.3%	20.5%	43.0%	69.8%	95.5%	99.7%	96.5%	100.0%
M	1000	729	0.00373	0.00271	8.7%	1.9%	2.2%	5.9%	2.2%	4.9%	1.1%	5.2%	1.4%	6.4%
N	1000	792	0.00493	0.00354	10.1%	1.1%	1.9%	3.4%	3.1%	3.3%	1.8%	5.1%	2.4%	4.5%
O	1000	1000	0.02265	0.01501	0.2%	2.7%	2.6%	4.3%	2.9%	3.9%	2.4%	4.4%	1.6%	4.5%
P	1000	903	0.01019	0.00740	1.8%	1.2%	1.4%	4.4%	1.3%	4.4%	0.7%	4.7%	1.7%	4.1%
Q	1000	948	0.00956	0.00648	1.8%	0.9%	1.4%	4.0%	2.4%	2.8%	1.5%	3.6%	0.6%	3.9%

Table 3. (Continued) Simulation results for standard vs. spatial GLM, nij=50.
Bias of estimates, % rejection of Wald statistics.

	Bias		Overall Wald % rejection		Wald diff 1 % rejection		Wald diff 2 % rejection		Wald diff 3 % rejection		Wald diff 4 % rejection	
Sets of τ_i	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM	Standard GLM	Spatial GLM
n_{ij}=50	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3	range=0	range=3
A	-0.00155	-0.00050	100.0%	100.0%	56.5%	68.9%	95.0%	98.2%	100.0%	100.0%	100.0%	100.0%
B	-0.00150	-0.00260	100.0%	100.0%	57.0%	63.8%	99.9%	100.0%	94.9%	98.2%	95.1%	98.1%
C	0.00230	0.00113	100.0%	100.0%	63.4%	70.0%	94.7%	98.1%	93.7%	98.4%	94.1%	98.7%
D	-0.00235	0.00159	100.0%	100.0%	59.0%	68.1%	76.4%	86.3%	77.6%	86.6%	77.2%	86.8%
E	-0.00056	0.00075	100.0%	100.0%	57.6%	66.4%	72.2%	83.0%	99.8%	100.0%	99.8%	100.0%
F	0.00583	0.00379	100.0%	100.0%	60.4%	66.5%	58.1%	63.5%	93.9%	97.8%	94.8%	98.1%
G	0.00112	-0.00029	100.0%	100.0%	59.1%	68.7%	55.0%	69.8%	100.0%	100.0%	100.0%	100.0%
H	-0.00051	0.00064	100.0%	100.0%	62.7%	69.2%	62.2%	67.3%	96.1%	98.9%	95.5%	99.0%
I	-0.00442	-0.00533	100.0%	100.0%	94.1%	98.1%	99.9%	99.9%	100.0%	100.0%	100.0%	100.0%
J	-0.00144	0.00211	100.0%	100.0%	94.4%	98.0%	99.9%	100.0%	99.7%	100.0%	99.4%	100.0%
K	-0.00179	0.00182	100.0%	100.0%	73.9%	82.5%	94.6%	98.2%	99.7%	99.8%	99.6%	100.0%
L	-0.00157	-0.00357	100.0%	100.0%	74.7%	80.5%	94.4%	98.6%	100.0%	100.0%	100.0%	100.0%
M	0.00383	0.00653	98.9%	99.5%	44.4%	58.6%	45.8%	52.9%	44.0%	53.4%	44.3%	59.0%
N	-0.00500	-0.00660	98.5%	99.2%	48.4%	57.1%	46.4%	52.8%	45.4%	54.8%	46.0%	53.2%
O	0.00199	0.00048	100.0%	100.0%	61.0%	68.0%	63.6%	67.7%	60.9%	68.3%	62.1%	63.8%
P	0.00300	0.00680	100.0%	100.0%	57.7%	63.9%	56.2%	63.8%	58.1%	65.8%	58.6%	62.1%
Q	-0.00336	-0.00201	100.0%	100.0%	57.0%	63.3%	56.5%	63.4%	60.0%	65.1%	57.4%	65.2%

Table 4. Quantiles of the F Tests for vectors with equal treatment effects.

Spatial GLM	Overall F quantiles				Diff 1 F quantiles				Diff 2 F quantiles			
n_{ij}=10	5^{th}	25^{th}	75^{th}	95^{th}	5^{th}	25^{th}	75^{th}	95^{th}	5^{th}	25^{th}	75^{th}	95^{th}
Theoretical	0.458	0.722	1.282	1.880	0.004	0.103	1.356	4.043	0.004	0.103	1.356	4.043
M	0.565	0.848	1.464	2.098	0.008	0.147	1.621	4.464	0.005	0.146	1.529	4.696
N	0.564	0.842	1.449	2.249	0.004	0.150	1.887	5.190	0.003	0.135	1.720	4.821
O	0.492	0.761	1.280	1.748	0.003	0.106	1.507	4.349	0.005	0.096	1.505	4.233
P	0.515	0.750	1.222	1.756	0.006	0.167	1.754	4.497	0.005	0.137	1.680	4.260
Q	0.490	0.752	1.231	1.745	0.006	0.126	1.635	4.329	0.008	0.143	1.569	4.202

Spatial GLM	Diff 3 F quantiles				Diff 4 F quantiles			
n_{ij}=10	5^{th}	25^{th}	75^{th}	95^{th}	5^{th}	25^{th}	75^{th}	95^{th}
Theoretical	0.004	0.103	1.356	4.043	0.004	0.103	1.356	4.043
M	0.006	0.148	1.772	5.067	0.006	0.141	1.868	4.637
N	0.004	0.121	1.769	5.309	0.005	0.132	1.669	5.154
O	0.004	0.128	1.530	4.376	0.005	0.126	1.455	4.277
P	0.006	0.147	1.750	4.589	0.005	0.137	1.577	4.109
Q	0.005	0.119	1.718	4.306	0.006	0.165	1.762	4.629

Spatial GLM	Overall F quantiles				Diff 1 F quantiles				Diff 2 F quantiles			
n_{ij}=50	5^{th}	25^{th}	75^{th}	95^{th}	5^{th}	25^{th}	75^{th}	95^{th}	5^{th}	25^{th}	75^{th}	95^{th}
Theoretical	0.458	0.722	1.282	1.880	0.004	0.103	1.356	4.043	0.004	0.103	1.356	4.043
M	0.457	0.647	1.084	1.566	0.008	0.187	1.818	4.413	0.005	0.146	1.555	4.004
N	0.449	0.633	1.043	1.477	0.008	0.142	1.495	3.630	0.007	0.143	1.418	3.539
O	0.442	0.718	1.228	1.684	0.007	0.112	1.359	3.666	0.004	0.110	1.415	3.640
P	0.460	0.656	1.088	1.537	0.006	0.127	1.551	3.883	0.004	0.134	1.555	3.929
Q	0.430	0.662	1.065	1.516	0.007	0.129	1.608	3.794	0.007	0.129	1.540	3.550

Spatial GLM	Diff 3 F quantiles				Diff 4 F quantiles			
n_{ij}=50	5^{th}	25^{th}	75^{th}	95^{th}	5^{th}	25^{th}	75^{th}	95^{th}
Theoretical	0.004	0.103	1.356	4.043	0.004	0.103	1.356	4.043
M	0.003	0.119	1.718	4.083	0.009	0.171	1.687	4.463
N	0.012	0.169	1.636	4.044	0.006	0.159	1.603	3.735
O	0.006	0.122	1.468	3.896	0.003	0.096	1.240	3.715
P	0.006	0.141	1.564	3.864	0.003	0.119	1.590	3.722
Q	0.006	0.139	1.594	3.616	0.009	0.143	1.495	3.578

New Prairie Press
https://newprairiepress.org/agstatconference/1997/proceedings/13

Kansas State University

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis*. New York: John Wiley.

Gotway, C.A. and Stroup, W.W. (1997), "A generalized linear model approach to spatial data analysis and prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 157-178.

Liang, K., and Zeger, S.L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13-22.

Littell, R.C., Milliken, G.A., Stroup, W.W. and Wolfinger, R.D. (1996), *SAS System for Mixed Models*. SAS Institute, Inc., Cary, NC, 633p.

McCullagh, P. (1983), "Quasi-Likelihood Functions," *Annals of Statistics*, 11, 59-67.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, Second Edition, New York: Chapman & Hall.

Nelder, J.A. and Wedderburn, R.W.M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society*, Ser. A, 135, 370-384.

Wedderburn, R.W.M. (1974), "Quasi-likelihood Functions, Generalized Linear Models and the Gauss-Newton Method," *Biometrika*, 61, 439-447.

Zeger, S.L. and Laing, K. (1986), "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics* 42, 121-130.