# Online Appendix

The Fifty American States in Space and Time:
Applying Conditionally Autoregressive Models

Joshua L. Jackson & James E. Monogan III

This appendix serves several purposes. Section 1 describes to the practical researcher how to estimate conditional autoregressive (CAR) models in OpenBUGS or in R. Most importantly, this section offers a brief tutorial for OpenBUGS, for researchers who are not familiar with the software and wish to borrow our approach or replicate our findings. Section 2 describes an alternative algorithm for estimating a Bayesian conditional autoregressive (CAR) model. Section 3 presents details of how the Monte Carlo experiments were conducted. Section 4 presents an additional Monte Carlo analysis for the case when the dependent variable is dichotomous, calling for a logistic regression. Section 5 offers an applied example of how to use a Bayesian hierarchical CAR prior for frailty terms in an event history analysis, with the example data being time until a state adopts a lottery. Section 6 offers another applied example of a Bayesian hierarchical CAR model: a panel count model of the number of Clean Air Act enforcement actions taken by the 50 states over the years 2001-2009. Section 7 estimates the *Statehouse Democracy* model with an alternative spatial weights matrix, using ideological rather than geographical proximity. Finally, Section 8 details where our data come from. Replication information for this project is available at: http://dx.doi.org/10.7910/DVN/ADFBT7.

# 1 OpenBUGS and R Tutorial

## 1.1 Using OpenBUGS to Estimate a CAR Model

OpenBUGS (or the similar program WinBUGS) allows a researcher to implement a CAR error structure with the `car.normal` command. OpenBUGS makes Bayesian estimation using a weights matrix a fairly easy thing to do, but we realize that not all researchers are familiar with the program. Here, we offer a tutorial on OpenBUGS to show interested researchers how to use the software, and to facilitate replication of these results for those not familiar with OpenBUGS. Below is the OpenBUGS script for replicating the *Statehouse Democracy* model in Erikson, Wright and McIver (1993: 129). This is also available for download from our Dataverse page.

```
#MODEL
model {
for (i in 1:N) {
#Distributional assumption of each outcome
aveelite[i] ~ dnorm(muAveelite[i],tau.h[1])
pid[i] ~ dnorm(muPid[i],tau.h[2])
demstr[i] ~ dnorm(muDemstr[i],tau.h[3])
legideol[i] ~ dnorm(muLegideol[i],tau.h[4])
policyli[i] ~ dnorm(muPolicyli[i],tau.h[5])

#No constants: All variables are standardized
muAveelite[i] <- beta[1]*ideol[i]
muPid[i] <- beta[2]*ideol[i]+ beta[3]*aveelite[i]
muDemstr[i] <- beta[4]*ideol[i]+beta[5]*aveelite[i]+beta[6]*pid[i]
muLegideol[i] <- beta[7]*aveelite[i]+beta[8]*demstr[i]
muPolicyli[i] <- beta[9]*ideol[i]+beta[10]*demstr[i]+beta[11]*legideol[i]

#Dump variables not used in model
irrelevant[i]<-ID[i]
}


#Priors on coefficients are vague, conjugate
for (i in 1:11) {beta[i] ~ dnorm(0.0,0.001)}

#Bernardinelli et al (1995) hyperprior assuming 46 states would be
# tau.h[i] ~dgamma(4.18291,2.04522)
# tau.c~dgamma(1.0,1.0)
#For even variance between models, we lower the precision by half.
#Doing so gives an expected error variance of 0.9671807 in each model.
for (i in 1:5){
tau.h[i] ~dgamma(1.05731,1.02261)
}

# marginal SD of heterogeneity effects
for (i in 1:5){
sd.h[i] <- 1 / sqrt(tau.h[i])
}


}

#INITIAL VALUES
list(beta  = c(0,0,0,0,0,0,0,0,0,0,0),
```

```
  tau.h = c(1,1,1,1,1) )
list(beta  = c(.76,.83,-1.01,.15,-.24,.78,1.05,.63,.44,-.32,.48),
 tau.h = c(1,1,1,1,1) )
list(beta  = c(-.76,-.83,1.01,-.15,.24,-.78,-1.05,-.63,-.44,.32,-.48),
 tau.h = c(1,1,1,1,1) )

#DATA (Load the list and the matrix.)
list(N=46)
ID[] ideol[] aveelite[] pid[] demstr[] legideol[] policyli[]
1 -1.167923518 -0.984030422 1.195949091 1.752288983 -0.766087461 -1.434831088
3 -0.501636682 0.378122775 -0.608210211 -1.213293001 -0.505270107 -1.036507349
4 -0.515234166 -1.094632567 1.671191012 1.664661843 -0.025076815 -1.524453822
5 1.13008644 1.172712203 -0.115366721 0.003635298 1.607283184 1.492847748
6 0.803741764 0.497456732 -1.013045951 -1.133443471 -0.362298015 0.487080529
7 1.374844915 1.271671955 0.025445743 -0.017882776 1.315371264 1.194105063
8 0.314224944 -0.055554183 -0.484999336 -0.332355221 -0.360986519 1.114440292
9 -0.352062022 -0.911265829 -0.080163605 0.532768044 -0.453300584 -0.359357208
10 -0.433648224 -1.353674547 1.715194897 1.348115348 -0.669231366 -1.026549268
12 -1.820612611 -0.957835124 -1.84911976 -1.515839891 -2.1137742 -0.857261762
13 0.59977639 0.413049781 -0.12416749 -0.479091702 0.14938833 0.417373898
14 -0.297671308 0.03467391 -0.810628086 -1.083407906 -0.681067809 -1.18587881
15 0.137454927 1.169801553 -0.907436624 -0.679752706 0.767023047 0.447248173
16 -0.188889619 0.587684811 -1.479487145 -0.954559681 -0.022941616 0.248086348
17 0.178248028 -0.56490635 1.477573852 0.649431573 -0.286020907 -0.30956674
18 -1.154325775 -1.603984786 2.445659315 1.659995153 -0.889107118 -1.026549268
19 -0.025717344 0.817620913 -0.44979622 -0.530683036 0.522819828 -0.010824015
20 1.198074898 0.747766901 1.292757671 1.242599643 1.71946065 0.855529933
21 1.864361734 1.431754115 0.949527237 1.006680563 2.14434595 1.642219091
22 0.776546407 1.292046091 -0.476198567 -0.135323881 1.265915399 1.184146863
23 0.232638743 1.265850793 0.175058934 0.013486649 1.443596725 0.795781384
24 -1.480670321 -1.831010237 0.80871494 1.656884233 -0.96062398 -1.494579577
25 -0.134498905 -0.780289548 0.025445743 0.233332229 -0.624952334 -0.538602854
26 0.463799474 -0.218546872 -0.247378375 -0.660308836 -0.608889018 0.606577656
29 0.232638743 0.541115481 -1.541092567 -1.031816726 -0.120375778 -0.130321109
30 1.510821831 1.268761444 -0.212175259 -0.087880647 1.204605403 1.343476406
31 -0.202487363 -1.013136231 0.720707087 0.061189178 -0.883814961 -0.97675886
32 1.551614931 1.22510266 0.04304728 -0.654086687 0.839803458 2.120207363
33 -0.841578972 -1.406065004 1.011132742 1.259191734 -0.321283953 -0.946884556
34 -1.643842724 0.284984116 -1.444283987 -1.428212751 -0.773990285 -0.50872855
35 0.59977639 0.311179379 -0.273780681 -0.300207912 0.235546911 0.646409982
36 -1.73902641 -1.705855188 1.17834747 0.676912193 -1.2434212 -0.966800778
37 0.898925644 1.312420227 -0.009757373 0.190814978 1.471326841 1.393266814
38 0.531787932 0.797246846 -0.335386124 -0.568533597 0.456491081 1.014859357
```

```
39 1.687591847 0.651717661 0.456683779 1.011865533 0.890314788 0.686242368
40 -0.936762657 -1.088811405 0.36867601 1.335930343 0.058584088 -1.51449574
41 -1.303900434 -0.107944692 -0.484999336 -1.738018816 -1.564438348 -0.936926474
42 -0.284073564 -0.65804501 0.474285316 0.061448473 -0.560386038 -0.837345599
43 -1.181521262 -1.208145362 0.518289201 0.985940373 -0.127916404 -0.63818373
44 -1.834210355 -0.826858843 -2.192350026 -1.590504451 -1.894894092 -0.429063838
45 0.423006503 1.105768703 -1.109854489 -1.082370881 0.496026301 0.795781384
46 -0.460843451 -0.925818733 -0.37058924 0.661875563 -0.307844814 -0.827387458
47 1.170879541 0.637164756 0.025445743 -0.454981297 0.404776301 0.357625349
48 0.722155692 -0.585280417 1.107941322 1.007976729 0.141843308 0.128589258
49 0.545385675 1.178533365 -0.036159721 -0.065584997 1.410277576 1.23393739
50 -0.447245708 -0.512515824 -1.030647489 -1.278883786 -1.416806709 -0.687974197
END
```

We recommend the researcher open this script in OpenBUGS as a .txt file. There are three primary components in this script: the model, the initial values, and the data. The model, including all priors, is defined within a loop. The priors are listed in three lists (for three parallel chains we elect to run). The data are in two components: a list (consisting only of N=46 in this case) and a matrix that must end with the word END in all caps.

To get started in OpenBUGS, first select the 'Model' drop-down menu and select 'specification' to bring up the Specification Tool. Highlight the word "model" at the beginning of the model loop ($model\{$) and the click 'check model' on the Specification Tool. The banner at the bottom of the screen should display a message saying the model is syntactically correct. Next, load the data. First, highlight the word 'list' under the #DATA section of the script, and click 'load data' in the Specification Tool. In this model, the data list vector only contains the sample size, N. When fitting the CAR model, this list vector will also contain the adjacency matrix, including the weights and number of neighbors for each state. Next, highlight ID[] and click 'load data.' This loads the data when it is in matrix form, although it is possible to simply include all data in a list. After loading the data, enter the number of chains you wish to run, between 1 and 5. We run 3 chains in our analyses. After entering the number of chains, click 'compile,' and a "model compiled" message should appear at the bottom. Next, the user may choose between loading initial values of each parameter to be estimated, or having OpenBUGS generate initial values automatically. To load the initial values, click on the word *list* at the start of an initial value list, and then click 'load inits' to initialize the first chain. For the second chain, click on the word *list* for the second list of initial values and then click 'load inits' again. Repeat this process for the number of chains. Alternatively, clicking 'gen inits' once will have OpenBUGS generate all initial values immediately and automatically. After loading or generating initial values, "model initialized" should appear at the bottom of the screen.

Now select Sample Monitor Tool from the 'Inference' drop-down menu, and select 'update' from the 'Model' drop-down menu to bring up the Update Tool. The user should run some iterations of the model before setting monitors on any parameters. This is known as the burn-in period and is intended to give the model a head start toward converging. For this example, the user may simply run the default 1000 iterations by clicking 'update' in the Update Tool. (Usually, a user will want to run a longer burn-in.) A counter keeps track of the progress as iterations are run. Now that the burn-in period is complete, the user should set a monitor on every parameter of interest. To do this, simply type the name of every parameter defined in the model into the 'node' box in the Sample Monitor Tool, typing one at a time and clicking 'set' for each parameter. (It is not required that the user set a monitor on every parameter.) After all parameters have been entered into the Sample Monitor Tool, the user may either select one parameter from the 'node' drop-down menu or, as we recommend, enter an asterisk into the 'node' box to monitor all parameters simultaneously. The user may also select 'DIC' from the 'Inference' drop-down menu and click 'set' to obtain the model fit post-estimation. Also, the user may change the credible intervals to 90% (from the default 95%) by highlighting 5, median, and 95 in the 'percentiles' menu in the Sample Monitor Tool. With all parameters being monitored, enter 10,000 into the 'updates' window and click 'update' to run 10,000 iterations of the model. Keep in mind that if the user is running 3 chains, then this will run 10,000 iterations per chain for a total of 30,000 iterations.

After the iterations have completed, the user can now click 'stats' in the Sample Monitor Tool to get the estimates of posterior means and credible intervals for each parameter. Clicking 'DIC' in the DIC Tool gives the Deviance Information Criterion to assess model fit. Also in the Sample Monitor Tool are several tools for assessing model convergence, including history plots, autocorrelation plots, and modified Gelman-Rubin statistics ('brg diag'). Alternatively, we recommend the user assess convergence using the coda output. Click 'coda' in the Sample Monitor Tool. If the user is running 3 chains, 4 windows will appear: one coda index, and one window for each chain which contains a vector of estimated coefficient values. Save the index and each chain as .txt files, and see our Convergence Tests R script for the most reliable way to assess convergence. The user may also refer to our forest plots R script for an example of how to show the results using forest plots.

## 1.2   Using the `CAR.Proper` Command in R

Our recommendation is to use the CAR prior in OpenBUGS (or WinBUGS) as we just described, as it has the most coherent error structure and fits naturally with the Gibbs sampler estimation strategy. However, for users who prefer not to use OpenBUGS, we have written a program in R that allows researchers to estimate a "proper CAR" model in cases when the researcher wants to fit a *linear model* in a purely *cross-sectional* setting. This model is "proper" because, as Section 2 explains, the joint distribution of a CAR model is not a

proper distribution. Rather, the inverse variance-covariance matrix is singular (therefore meaning that the variance-covariance matrix itself cannot be solved for). With a ridge-style solution, we can force the inverse variance-covariance matrix to be nonsingular, giving us a proper distribution. The disadvantage of this is that it implies each spatial error term is a function of a *fraction* of the averages of its neighbors' errors, which is why OpenBUGS's solution is ideal. That said, users who are comfortable with this ridge fix can use our program.

Our R script entitled `CARproper.R`, available on our Dataverse page (http://dx.doi.org/10.7910/DVN/ADFBT7), creates the function `CAR.Proper`. This function estimates a Bayesian proper CAR model using a Metropolis-Hastings sampler. The function is structured as follows:

```
CAR.Proper(X, y, W, mcmc=1000, burnin=round(.1*mcmc),
    beta.mean=rep(0,times=dim(X)[2]),beta.var=diag(dim(X)[2])*100,
    sigma.shape=.001,sigma.rate=.001,delta.shape=.1,delta.rate=.1,
    ridge=.97, initials=c(rep(.1,dim(X)[2]),1,1,rep(.1,dim(X)[1])))
```

In this function, three arguments must always be specified by the user:

- `X` is a matrix of predictors,

- `y` is a vector of the dependent variable, and

- `W` is a spatial weights matrix.

Other arguments have default values:

- `mcmc` is the number of MCMC samples to draw (which ought to be raised above the 1,000 default),

- `burnin` is the length of the burn-in sample (defaulting to 10% of the `mcmc` sample if a number of iterations is not specified),

- `beta.mean` is a vector of the prior means of regression coefficients (which are assumed to have a joint multivariate normal prior) that defaults to 0 for each term,

- `beta.var` is the variance-covariance matrix for regression coefficients in the prior (again assumed to be multivariate normal) that defaults to independent coefficients with a variance of 100 each,

- `sigma.shape` and `sigma.rate` are the shape and rate parameters of an inverse gamma prior on the *heterogeneous* error variance that default to 0.001 each following the Best et al. (1999) recommendation,

6

- `delta.shape` and `delta.rate` are the shape and rate parameters of an inverse gamma prior on the *clustering* error variance that default to 0.1 each following the Best et al. (1999) recommendation,

- `ridge` is the propriety term that the spatial weight matrix is multiplied by to force the joint distribution of observations to be multivariate normal that defaults to 0.97 (and must be absolutely smaller than 1), and

- `initials` is the set of initial values for all parameters defaulting to 1 for the two variance terms and 0.1 for all other terms. This vector first lists values for regression coefficients, then each of the two variance terms, and finally all of the clustered error terms.

As an applied example of this function, the following lines of code estimate a proper CAR model for the last stage of the Erikson, Wright and McIver (1993) model, with policy liberalism as the dependent variable, and then summarizes the results:

```
ewm.5<-CAR.Proper(X=cbind(ewm$ideol,ewm$demstr,ewm$legideol),
    y=ewm$policyli, W=ewm.W, mcmc=100000,
    beta.mean=c(.44,-.32,.48),
    beta.var=diag(c(1/69.444,1/204.081,1/69.444)))
summary(ewm.5)
```

Here the term `X=cbind(ewm$ideol,ewm$demstr,ewm$legideol)` references a loaded data frame named `ewm` and binds the necessary predictor variables into a matrix. Because this is a standardized regression, a constant is omitted, but a constant could have easily been included if we had instead written: `X=cbind(1,ewm$ideol,ewm$demstr,ewm$legideol)`. Similarly, `y=ewm$policyli` calls our dependent variable from our data frame. The option `W=ewm.W` lists our spatial weights matrix, which we had to define earlier in the file as the matrix `ewm.W`. These three options would be enough to run the model, but we also lengthen our MCMC sample and change our priors. With `mcmc` we set the length of our sample to 100,000, with `beta.mean` and `beta.var` we set the prior means and variances of our coefficients to reflect Erikson, Wright and McIver's (1993) results.

For a fuller illustration of applying this model in practice, see the file `ewmProper.R` on our Dataverse page. After applying this structure to all five models in the *Statehouse Democracy* system of equations, our findings are presented in Figure A.1. Just as in the main text, the horizontal axis represents possible coefficient values, and the vertical axis names the input variable and which of the five dependent variables it affects. Points represent posterior means of parameters, while line segments represent 90% credible intervals. Black dashed lines with circle points represent independent errors, red solid lines with triangle points represent a

standard CAR model, and blue dotted lines with cross points represent the proper CAR model.

Figure A.1 shows that the means of most (but not all) coefficients tend to be closer to zero for the proper CAR model, and the credible intervals tend to be narrower. Outside of the figure, the variance of the clustered error terms is also squeezed downward by the proper CAR model, which likely is a culprit for why the proper CAR model's results behave as they do. So the implication of this figure is that the full model in OpenBUGS is best, but the proper CAR specification is an available option for those studying a linear model in a cross-sectional setting.

# 2  Estimation Approaches

The main text explains that, by using a Gibbs sampler, the full conditionals of a CAR are sufficient to estimate the Bayesian model. In fact, a Gibbs sampler is our usual approach because a CAR model is parameterized by defining the full conditionals. Therefore, our model is already structured for estimation by Gibbs sampling with no further analytical work. However, that is not our only option: If we apply Brook's Lemma (1964), we can obtain the joint likelihood of all observations, and hence the joint posterior of all observations. With a joint posterior we can estimate the model using a Metropolis-Hastings algorithm. While we use the Gibbs sampler for the real data examples, for the Monte Carlo experiments we use Metropolis-Hastings on the joint posterior for programming simplicity in replicate simulated samples.

For a linear model including both heterogeneous and CAR-structured random effects, Sun, Tsutakawa and Speckman (1999) define the joint posterior density. We can obtain the same estimates by sampling from the logarithm of the joint posterior density instead, which is:

$$-\frac{N}{2}\ln(\sigma_h^2) - \frac{Q}{2}\ln(\sigma_c^2) - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\phi})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\phi})}{2\sigma_h^2} - \tag{1}$$

$$\frac{\boldsymbol{\phi}'\mathbf{B}\boldsymbol{\phi}}{2\sigma_c^2} + \ln\{\mathcal{G}^{-1}(\sigma_h^2, 1, 1)\} + \ln\{\mathcal{G}^{-1}(\sigma_c^2, 1, 1)\}$$

Here, $N$ is the number of observations, $Q$ is the number of CAR random effects (equal to $N$ in a cross-sectional model), $\sigma_h^2$ is the variance of the heterogenous error term, $\sigma_c^2$ is the variance of the clustering error term, $\mathbf{y}$ is the vector of observed values of the dependent variable, $\mathbf{X}$ is a matrix of predictors for all observations, $\boldsymbol{\beta}$ is the vector of regression coefficients, $\boldsymbol{\phi}$ is a vector of spatial random effects for all observations, and $\mathcal{G}^{-1}$ represents an inverse gamma distribution. We generally use an inverse gamma prior on each of the variance terms, setting both the shape and scale parameters to 1 (though other distributions could apply). In this
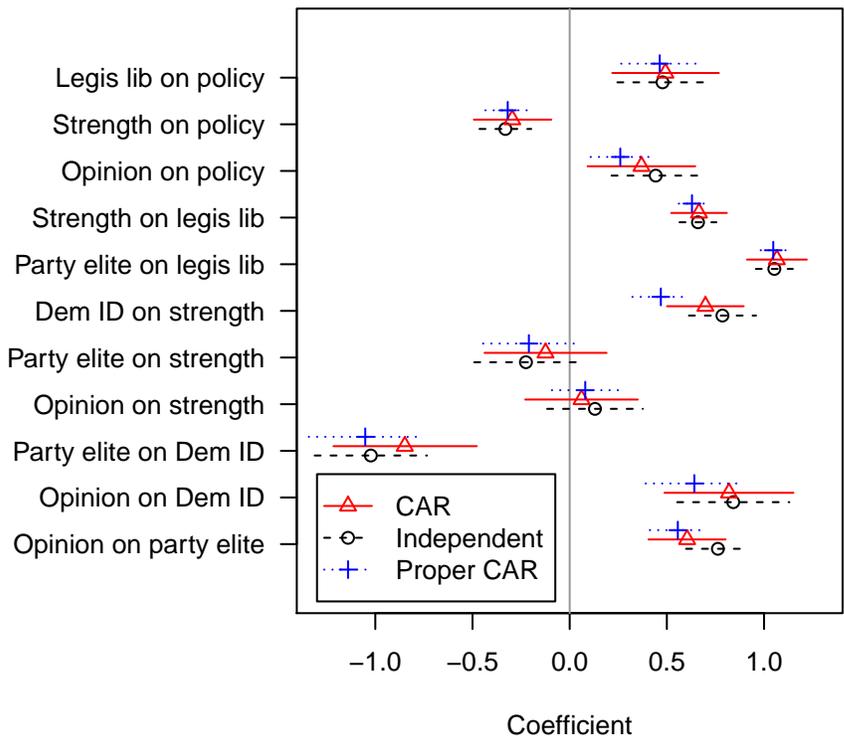
Figure A.1: Posterior means and 90% credible intervals of coefficients for the policy liberalism model using results from linear models, results from a CAR structure, and results from a proper CAR model.

model, we have assumed a flat prior on the coefficients, though we easily could include any of a variety of probability distributions for the coefficients as a prior.

As a final term from the log-posterior, $\mathbf{B}$ represents the correlation matrix among the various random effects, $\boldsymbol{\phi}$. We define $\mathbf{B} = \mathbf{D}_w - \rho\mathbf{W}$, where $\mathbf{D}_w$ is a diagonal matrix where each term on the diagonal is the sum of all weights for an observation, $\mathbf{W}$ is the spatial weight matrix (defined by neighbor contiguity to produce an average of neighbors), and $\rho$ is a ridge-style penalty that makes $\mathbf{B}$ nonsingular and thereby allows us to find $\mathbf{B}^{-1}$. In principle, setting $\rho = 1$ would be desirable because it would mean that the theoretical mean of each $\phi$ random effect would be the average of neighbors' random effects, but this leaves us with a nonsingular $\mathbf{B}$ matrix. In practice for the Monte Carlo experiment, we set $\rho = 0.97$ as the value closest to 1 that still allows computational matrix inversion in every simulated sample we try. In general, if we want to sample from this log-posterior density instead of the full conditionals, we have the option of using a Metropolis-Hastings algorithm as well—as we do in the Monte Carlo experiment.

# 3    Details on the Monte Carlo Experiment

In the main text, we report the results of a Monte Carlo experiment in which we contrast a CAR model with a simple Bayesian linear model when there is spatial correlation in the errors. When simulating data, we use the following data generating process:

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \theta_i + \phi_i \\
\theta_i &\sim \mathcal{N}(0, \sigma_h^2) \\
\boldsymbol{\phi} &\sim CAR(1/\sigma_c^2)
\end{aligned}
\tag{2}
$$

$\sigma_c^2$ is the variance of the spatial random effects (called $\boldsymbol{\phi}$) and $\sigma_h^2$ is the variance of the heterogeneous random effects (called $\boldsymbol{\theta}$). As the main text states, the first treatment variable in the Monte Carlo experiments is to manipulate the relative share of the error variance that is spatially correlated. The relative share of spatial clustering is:

$$
\psi = \frac{\sigma_c}{\sigma_c + \sigma_h}
\tag{3}
$$

For a given $\psi$ of interest, we would like $\sigma_c = \frac{\psi}{\sqrt{\psi^2 + (1-\psi)^2}}$ and $\sigma_h = \frac{1-\psi}{\sqrt{\psi^2 + (1-\psi)^2}}$. If these are the standard deviations of the respective disturbance terms, two results hold: (1) By substituting these values back into Equation 3, it is easily demonstrated that we recover the original $\psi$ itself meaning that these standard deviations do result in the proper share of clustered error variance that we seek. (2) With these parameterizations of the standard deviations, it always holds that $\sigma_c^2 + \sigma_h^2 = 1$ regardless of the value of $\psi$. The advantage

of this feature is that the total level of error variance is the same regardless of the level of spatial clustering. To do otherwise would vary the error variance while also varying the level of spatial clustering, violating the logic of an experiment.

With this in mind as our goal, simulating the heterogeneous error term is simple. We draw 100 simulated values from $\boldsymbol{\theta} \overset{i.i.d.}{\sim} \mathcal{N}(0, \frac{(1-\psi)^2}{\psi^2+(1-\psi)^2})$. Simulating the clustered error term ($\boldsymbol{\phi}$) involves two steps. First, we simulate an independent error term $\boldsymbol{\epsilon} \overset{i.i.d.}{\sim} \mathcal{N}(0, \frac{0.49\bar{m}\psi^2}{\psi^2+(1-\psi)^2})$. Here, $0.49\bar{m}$ is based on a constant that Bernardinelli, Clayton and Montomoli (1995) observed is necessary to equalize variances whenever disturbances are conditionally related through a neighbor matrix. $\bar{m}$ is the average number of neighbors that an observation has, and 0.49 is an approximate constant that also needs to be included.[1] The remainder of the variance term ($\frac{\psi^2}{\psi^2+(1-\psi)^2}$) is our target variance for $\boldsymbol{\phi}$, the clustered variance term. Intuitively, this term has to be weighted upward by $0.49\bar{m}$ in order to compensate for the induced error correlation process, which results in lowering the overall variance of the term.

To that end, the second step in simulating our clustered error term is to create the following matrix: $\boldsymbol{\Sigma}_\phi = (\mathbf{D}_W - \rho\mathbf{W})^{-1}$. This is the variance-covariance matrix of our $\boldsymbol{\phi}$ terms assuming our spatial neighbor structure. We create it by specifying that $\mathbf{W}$ is the neighbor weight matrix, $\mathbf{D}_W$ is a diagonal matrix in which each element is the sum of the elements on the corresponding row of $\mathbf{W}$, and $\rho$ is a ridge penalty that allows us to invert this difference.[2] We then conduct Cholesky decomposition to determine $\boldsymbol{\Sigma}_\phi = \mathbf{L}\mathbf{L}'$. This allows us to impose a correlation structure consistent with spatial clustering by defining $\boldsymbol{\phi} = \mathbf{L}\boldsymbol{\epsilon}$. The resulting $\boldsymbol{\phi}$ has the desired spatial correlation structure, and an approximate variance of $\frac{\psi^2}{\psi^2+(1-\psi)^2}$. Hence, we are able to manipulate the error structure across treatments as desired.

For our second experimental treatment variable, we also induce correlation in one of the predictor variables. For various levels of population $\psi$ in the errors, we compare the treatments with a correlated and an uncorrelated predictor. Specifically, in Equation 2 we induce correlation in $\mathbf{x_2}$. We did this using the same kind of Cholesky decomposition that was used to create a correlated $\boldsymbol{\phi}$ term.

For each treatment, we created 150 random samples and estimated the model on each sample. We estimated each of the two models (Bayesian standard linear and Bayesian CAR) using the Metropolis-Hastings algorithm with a burn-in of 10,000 iterations and a post burn-in sample of 100,000. We computed the posterior median from each sample as our point estimate of the parameter, computed the absolute difference between the median and the true parameter, and we average the absolute differences to get the mean absolute error

---

[1] In our example simulation, we use a map of 100 counties in North Carolina to project our data. There are 492 neighbor connections, so each county has 4.92 connections on average. For the interested reader, Bernardinelli, Clayton and Montomoli (1995) parameterize this weight in terms of standard deviation so their weighting constant is 0.7. Our parameterization is in terms of variance, so it is $0.7^2 = 0.49$.

[2] We set $\rho = 0.97$. Absent a ridge penalty $\mathbf{D_W} - \rho\mathbf{W}$ is singular so an inverse does not exist.

in each treatment. The results are reported in the main text.

# 4 Monte Carlo Experiment for a Dichotomous Outcome

In addition to the Monte Carlo analysis reported in the main article of how CAR models fared to standard regression with a linear functional form, we also conducted a Monte Carlo analysis for a binary outcome variable. In this experiment, we varied the level of spatial correlation in the latent utility. Then, with the simulated data we estimated both a standard Bayesian logistic regression and then a logistic regression in which the error was divided into a CAR-based spatial component and a heterogeneous component. When simulating the latent utility, we followed the exact process described in Appendix Section 3, and then created the dichotomous variable based on whether the utility was greater than zero. We still varied the true $\psi$ from 0.1 to 0.9 in increments of 0.1. Again, for each treatment we create 150 random samples. Each model still had a burn-in of 10,000 iterations and post burn-in sample of 100,000. For this model, data were simulated in R, but then the models were estimated using OpenBUGS through R's `R2OpenBUGS` package.

When estimating the model, we constrained the model so that the clustered standard deviation would be $\sigma_c = \frac{\pi}{\sqrt{3}} \frac{\psi}{\sqrt{\psi^2 + (1-\psi)^2}}$ while the heterogeneous standard deviation is $\sigma_h = \frac{\pi}{\sqrt{3}} \frac{1-\psi}{\sqrt{\psi^2 + (1-\psi)^2}}$. This guarantees that the standard deviations will satisfy Equation 3's specification of how the variance is distributed between the two parts, while also guaranteeing that the total variance fits our assumed total logistic variance of $\sigma_c^2 + \sigma_h^2 = \frac{\pi^2}{3}$. A logistic distribution's standard deviation can be manipulated to the specified values by treating the second ratio in each term as a scale parameter.

Figure A.2 presents our findings from this experiment. There were two predictors in the simulated model, called $x_1$ and $x_2$. Panel (a) shows the mean absolute error for the coefficient on $x_1$ and panel (b) shows this quantity for the coefficient on $x_2$. For each subfigure, the horizontal axis represents the true population share of unexplained variance in the utility that is attributable to spatial dependence, and the vertical axis represents the mean absolute error. The blue dashed spikes show the mean absolute error for a simple logistic regression, and the red solid spikes show the mean absolute error for a logistic regression with a CAR error structure.

The substantive results of Figure A.2 for logistic regression models resemble those of our linear model Monte Carlo experiment. At low levels of spatial variance (such as $\psi = 1$), the mean absolute error is quite similar for a traditional logistic regression and for a CAR logit model for both coefficients. Since there is the more complex parameterization with the CAR

12

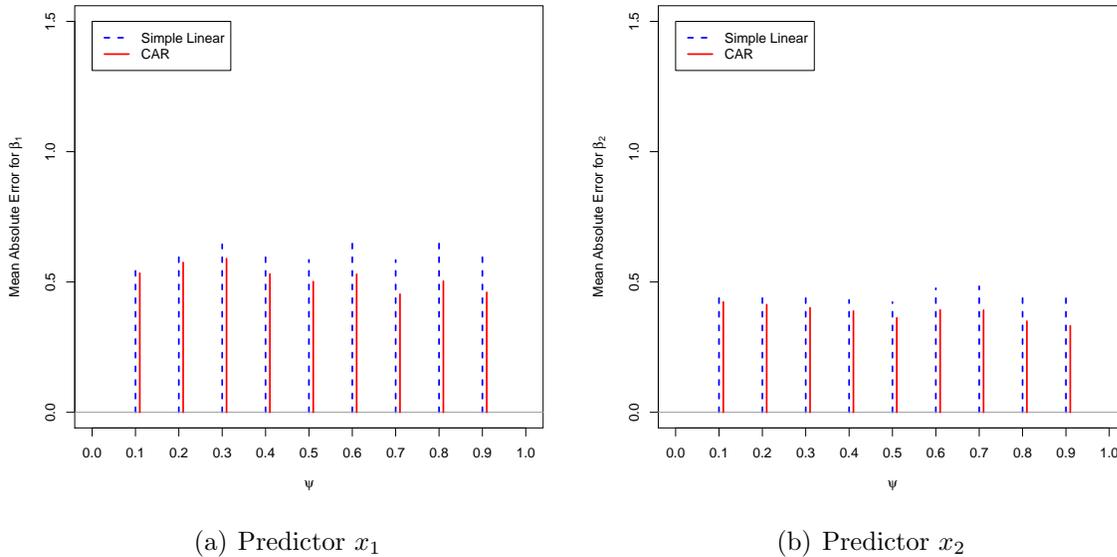(a) Predictor $x_1$            (b) Predictor $x_2$

Figure A.2: Mean absolute error of the coefficients of two predictors in a logistic regression model, contingent on the share of error variance with clustering.

model, it makes sense that there would be minimal pickup with spatial correlation is not much of a problem. However, as the share of the random utility that is spatially dependent rises, a gap starts to open up between the standard and CAR versions of the logit model, with CAR showing better values of the mean absolute error. This efficiency gain in the presence of greater spatial correlation is exactly what we would expect and exactly why the model is useful.[3]

# 5   Duration Data: Adoption of Lotteries

Another variety of data that is common in the study of state politics and policy is duration data, or the time until a policy is adopted or an action occurs. As a classic example, Berry and Berry (1990) report a model of the time until a state adopted an education lottery, using data from 1964 to 1986. As with all duration models there is right-censoring, meaning that some states never adopted a lottery in the time frame, but may have at a later date. Event

---

[3]We also conducted this experiment for a situation in which $x_2$ itself was spatially correlated. In this treatment, coefficients for $x_2$ showed a slightly wider efficiency gap in favor of the CAR model relative to the results presented in Figure A.2. The additional pickup was not as great as for the linear model, though. The Dataverse page contains code to replicate both versions of this experiment.

history models explicitly address this issue and produce unbiased results despite censoring.

Nevertheless, much more information is available about lottery adoption today. Figure A.3 shows the cumulative number of states with lotteries by year. To quantify this, by 1986, 27 states had adopted lotteries. A rapid diffusion among the states continued for the duration of the 1980s, and several more states continued to add programs at a slower rate over the next twenty years. Updating to 2013, 17 more states have created lotteries, bringing the total to 44. In other words, 74% of the right-censored observations from the original model are now observed completely. Therefore, an update of this model is a useful place to illustrate how to include spatial terms in a duration model.
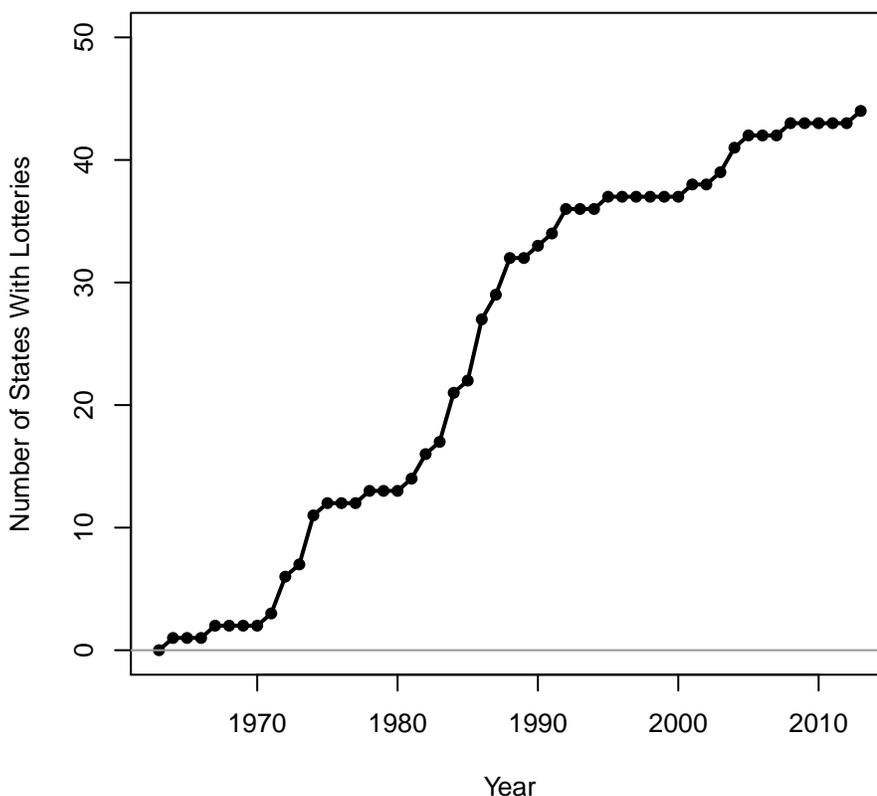


Figure A.3: Number of states with lotteries by year, 1963-2013.

Berry and Berry (1990) argue that policy diffusion is possible because voters and legislators in a state that neighbors a lottery state could see revenue flow across the border and thereby conclude that they needed their own lottery. In fact, Berry and Baybeck (2005) use the technology of Geographic Information Systems (GIS) to measure state residents' accessibility to other states' lotteries. With this information, they show that when a state's residents can easily play neighboring states' lotteries, the state is much more likely to adopt a lottery of its own to keep the lottery revenue in state. To capture this potential diffusion, Berry and Berry include the number of neighbors that have adopted lotteries prior to a given

year in the structural portion of the model, and they show significant evidence of diffusion. But does this result stand when we account for potential spatial error correlation?

In three regards these analyses differ from Berry and Berry (1990), although the two models we compare here are the same on all three counts. First, we update the data by 27 years to run from 1964-2013. Second, we include a quadratic trend over time. In this way, our analysis more closely resembles the analysis of lotteries by Mooney (2001) and Grossback, Nicholson-Crotty and Peterson (2004), each of which includes some form of time trend. To omit a time trend implies that the probability of adopting a lottery does not depend on how much time has elapsed, but that the baseline hazard of adopting a lottery is always the same. By including a parametric time trend, the model can allow that a state's probability of adopting a lottery may change over time, even if none of the other input variables changes.[4] Third, our models include individual frailty terms, meaning that some states may have an overall unmodeled tendency to adopt a lottery more quickly or more slowly, even holding the input variables constant. A frailty model includes a random effect for each state that is constant over time. Failure to account for unobserved heterogeneity in the states can lead to a higher Type I error rate than is reported (Box-Steffensmeier and Jones 2004; Hougaard 2000; Klein and Moeschberger 2003).

Equation 4 presents the full discrete time hazard model. $Y_{it}$ indicates whether a state has yet adopted a lottery in a given year, with states dropping-out after adoption:

$$
\begin{aligned}
Y_{it} &\sim \text{Bern}(\lambda_{it}) \\
\text{logit}(\lambda_{it}) &= \mathbf{x}'_{it}\beta + \phi_i \\
\beta_j &\sim \mathcal{N}(0, 100) \\
\phi_i &\sim CAR(3/\pi^2)
\end{aligned}
\tag{4}
$$

$\mathbf{x}_{it}$ is the vector of all covariates for an observation including an indicator for gubernatorial election years, an indicator for non-election or post-election years, lagged real per capita income, lagged state fiscal health, an indicator for unified party control, the proportion of a state that is fundamentalist Christian, the number of neighboring states having adopted a lottery, time, time squared, and a constant. $\lambda_{it}$ is the probability that state $i$ will adopt a lottery at time $t$ given that it has not adopted a lottery yet, $\boldsymbol{\beta}$ is a vector of population parameters, $\beta_j$ (where $j = 0, \ldots, 9$) refers to the individual elements of $\boldsymbol{\beta}$, and $\phi_i$ is the state-level spatial frailty term. Equation 4 also shows that the coefficients have vague normal priors. The model of Equation 4 follows the recommendation of Darmofal (2009) by forcing the frailty terms to be geographically correlated with one another by giving them a CAR prior.[5] The CAR prior's precision is $3/\pi^2$, so the variance of the random effects ($\pi^2/3$) will

---

[4]Time is included as an orthogonal quadratic polynomial, using the Cholesky decomposition method described by Bock (1975).

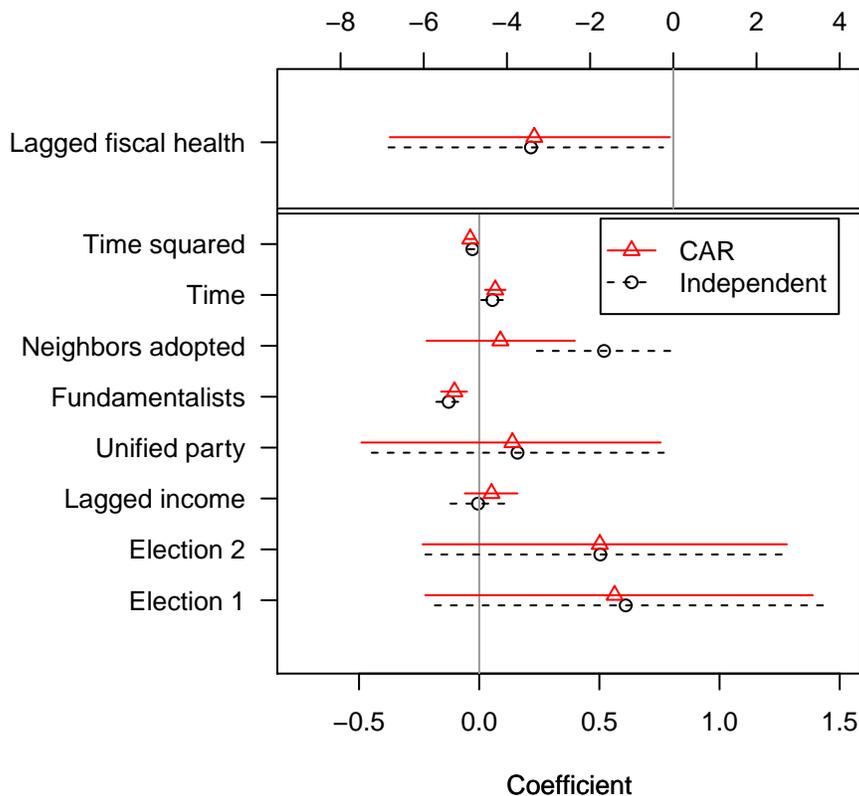[5]In assigning this prior, we match the specification of Berry and Berry (1990) by assuming the fol-

Figure A.4: Posterior means and 90% credible intervals of coefficients from the duration model of lottery adoption with independent frailties and with CAR frailties. The models were fit with a post burn-in MCMC sample of 600,000 in OpenBUGS 3.2.3.

match the assumed individual-level variance of a standard logistic distribution (also $\pi^2/3$). The comparison model has frailty terms that are independent of each other.[6]

Figure A.4 presents a forest plot of the results of the two duration models of time until lottery adoption. All coefficients plotted above the double bar are drawn on the scale of the top horizontal axis. Those plotted below the double bar are drawn on the scale of the bottom axis. Noteworthy right away is that the two election variables, lagged state income and whether a state was under unified party control, do not display robust effects in either model, whereas they did in the model reported in Berry and Berry (1990). The effect of income also drops-out in a non-frailty model with a time trend fit over the original data, suggesting that modeling the baseline hazard is critical. The other three effects may

lowing pairs of states are neighbors: Arkansas-Kentucky, Illinois-Michigan, New Jersey-Maryland, and Massachusetts-Maine.

[6]For the independent frailty term model, the precision of those frailties is 0.6732185. This is done by scaling $3/\pi^2$ with the constant $0.49\bar{m}$ where $\bar{m}$ is the average number of neighbors that a state has. This ensures that the variance of the frailty terms is approximately equal between the two rival models since spatial correlation tends to reduce overall variance.

have changed due to either the new data or the presence of frailties in general. Meanwhile, lagged fiscal health continues to have a negative and robust effect, which makes sense given that adopting a lottery can offer one solution to this problem. Further, states with a large percentage of fundamentalist Christians in are less likely to adopt a lottery quickly, thus lasting longer before they finally create a lottery (with three of the six states who still do not have a lottery scoring high on this variable: Alabama, Mississippi, and Utah).

Most coefficients do not shift much between the two models. However, one primary result does change: the model contains the temporally lagged number of neighboring states that have already adopted a lottery. Including the lagged number of neighbors with such a policy is an unbiased way to incorporate a spatial lag in an event history model (Kreitzer and Boehmke 2016). When we estimate this model with independent frailties, the coefficient for the number of neighbors with a lottery remains robustly positive, just as Berry and Berry (1990) found. Spatially-correlated frailties, alternatively, shift this coefficient's posterior distribution such that the effect is no longer robust. This implies that the diffusion process likely was picking up unmodeled structure in the error terms. In a model in which there is a horse race between the spatial lag term and our CAR-oriented spatial errors, the errors seem to win in this case. We should note that Grossback, Nicholson-Crotty and Peterson (2004) actually show that controlling for the ideological distance between adopter states and target states also leads the spatial lag's effect to lose statistical robustness, so a different result can emerge with an alternative theoretical specification. (We would add that a better theoretical specification is always the best choice rather than merely cleaning up errors.) Nevertheless, we still believe there is a worthwhile conclusion here that without accounting for spatial correlation in the errors, coefficients can be misleading, and a geographically correlated predictor like a diffusion variable can pick up this variance. Hence, accounting for this error structure is critical.

While all other coefficients' posterior means are similar regardless of whether the frailty terms spatially correlated, the CAR model's credible intervals tend to be a bit wider than the intervals with independent frailties (being wider for 6 of 9 coefficients). Hence, in two models that allow for unobserved heterogeneity, adding geographic information to the structure of the heterogeneity offers a more conservative estimate of model uncertainty, just as it did in the cross-sectional *Statehouse Democracy* model. The CAR model fits better with a DIC of 309 as opposed to a score of 323.2 for the model with independent frailties.

# 6    Panel Count Data: Clean Air Act Enforcement

To illustrate how to extend the hierarchical CAR methodology to panel data, this study replicates Monogan's (2013) analysis of state-level enforcement of the Clean Air Act, which itself built on Wood's (1992) study of enforcements from 1977-1985. As these prior studies

explain, clean air regulation requires activity from several levels of American government. The United States Environmental Protection Agency (EPA) and environmental agencies from each of the fifty states all play a regulatory role. This analysis asks how state-level factors shapes the number of actions states take to enforce clean air policy.

This example is relevant for geospatial data modeling because environmental policy at the state level has the potential to cluster among neighboring states due to a variety of unobserved features. For instance, the types of industry found in neighboring states are likely to be similar, due to similarities in available natural resources and similar labor markets. Even in a model that attempts to capture this—and the model estimated here does include manufacturing outputs as a predictor—it is difficult to discern what exact features of industry will be critical to shaping environmental policy, such as whether certain industries draw more regulation or whether labor relations call for greater policing. Additionally, on account of similar histories and views on the role of government (Elazar 1966), neighboring states are likely to have similar regulatory structures. Finally, it may be that certain regions have a greater commitment to cooperation on the environment, meaning that states within an area jointly enforce pollution regulations so that none of them experience excessive environmental consequences of air pollution. By contrast, in other areas states may be more inclined to free ride and let their pollution affect other downwind states. Since several of these concepts are difficult (if not impossible) to measure, the next best option is to account for the spatial correlation in the error term that these factors are likely to produce.

Figure A.5 plots the dependent variable, the number of enforcement actions taken from 2001-2009 by each of the states. The horizontal axis is the year of activity, and the vertical axis is the number of acts on a logged scale. Each line represents a state. The states vary widely in their average levels of enforcement, so we will include random effects by state in this model. (Whether or not the state random effects are spatially correlated is the comparison we will make.)

An important feature of these data is that the outcome variable is a count, so we will assume that the outcome variable has a Poisson distribution. With this alternative distribution choice, a Generalized Linear Mixed Model (GLMM) is the proper model specification. Also, Wood (1992) and Monogan (2013) both specified a dynamic functional form wherein there was temporal spillover in effects from one time period to the next. Our model thereby will resemble a specification by Brandt and Williams (2001) that allows dynamic effects with an event count as the dependent variable.

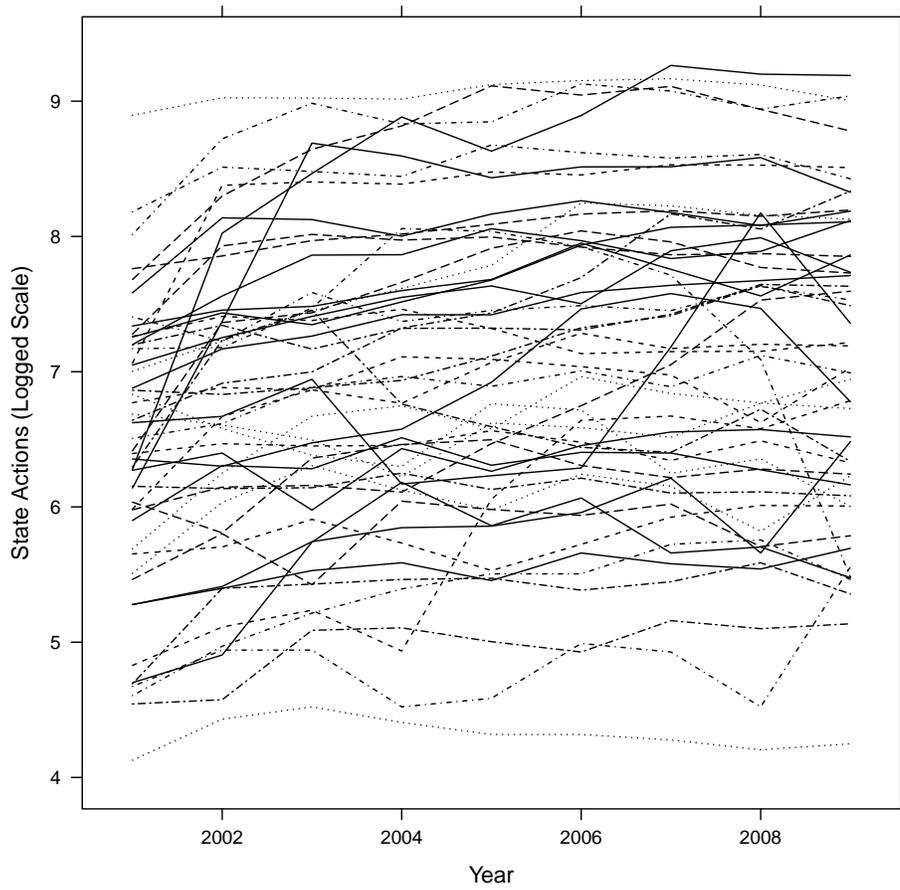To account for all of these features, Monogan (2013) estimates a state-space model for

Figure A.5: State-level Clean Air Act enforcement actions by state and year, 2001-2009.

panel data.[7] Equation 5 specifies the model as follows:

$$
\begin{aligned}
y_{it} &\sim \mathcal{P}(\lambda_{it}) \\
\lambda_{it} &= \rho y_{i,t-1} + \exp\{\mathbf{x}'_{it}\boldsymbol{\beta} + \phi_i\} \\
\beta_j &\sim \mathcal{N}(0, 100) \\
\rho &\sim \mathcal{B}(2, 5) \\
\phi_i &\sim CAR(1/\tau_c) \\
1/\tau_c &\sim \mathcal{G}(.1, .1)
\end{aligned}
\tag{5}
$$

In this equation, $y_{it}$ is the number of enforcement actions taken by state $i$ in year $t$, $\lambda_{it}$ is the Poisson parameter, $\rho$ is the first-order time-autoregressive parameter, $\boldsymbol{\beta}$ is a vector of population parameters, $\beta_j$ for $j = 0, \ldots, 8$ refers to the individual elements of $\boldsymbol{\beta}$, and $\phi_i$ is the state-level spatial random effect. $\mathbf{x}_{it}$ is the vector of all covariates for an observation including a lag of standardized federal enforcement actions, an indicator for unified Democratic control of state government, an indicator for unified Republican control of state government, the unemployment rate, manufacturing outputs, EPA grants, the number of lawsuits related to clean air originating in a state, and a constant. Equation 5 also shows that the coefficients have vague normal priors, the autoregressive parameter has a beta prior to constrain it to the interval $\rho \in (0, 1)$, the spatial random effects have a CAR prior, and the hyperprior on the precision of the random effects has a vague gamma distribution. Overall, this model assumes a Poisson distribution for the outcome, includes spatial random effects, and allows for temporal autoregression. Therefore, it should take care of all the issues facing these data: repeated observations of the same states, dynamic causality, a count process as the outcome variable, and geographic dependence.

Figure A.6 presents a forest plot of the results from the model in Equation 5 and contrasts them with the results of a similar model that does not force the random effects to be spatially clustered. Black circles with dashed lines represent the posterior means and credible intervals for the model with independent random effects, while red triangles with solid lines represent the model with CAR-structured random effects.

In Figure A.6, the results in the spatial versus non-spatial specifications of the model differ very little. The mean coefficients are approximately the same either way. Again, however, the credible intervals are consistently larger for the model with CAR-distributed random effects than for the model with independent random effects. As far as statistical inferences are concerned, placing a CAR prior on the random effects reduces the opportunity to commit a Type I error in this case as well. This pattern fits with much of what was reported in the models of the main text. Also, like the other models, the overall fit of the model with

---

[7]Additional details about the model are available in Blundell, Griffith and Windmeijer (2002). The technique is similar to the model Brandt and Williams (2001) use for a single time series.
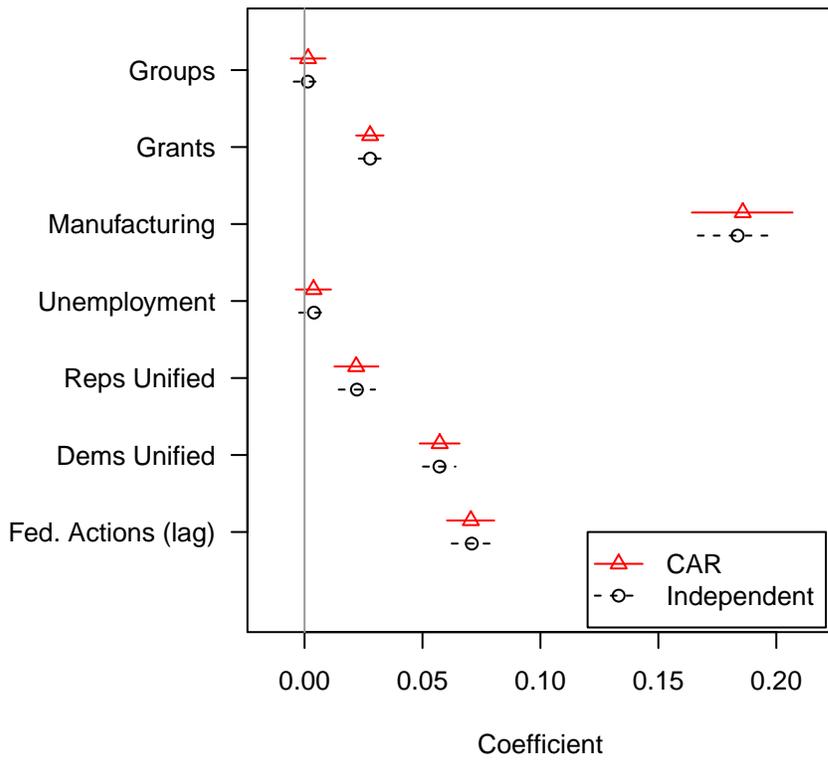
Figure A.6: Posterior means and 90% credible intervals of coefficients from the dynamic model of Clean Air Act enforcement from 2001-2009 with independent random effects and with CAR random effects. The models were fit with a post burn-in MCMC sample of 600,000 in OpenBUGS 3.2.3.

CAR-distributed state random effects is improved with a DIC score of 40280, as opposed to the model with independent random effects which has a DIC of 40350.

# 7  Statehouse Democracy, Ideological Distance

This paper deals with diffusion and clustering in terms of geographical space, or contiguity. However, diffusion may occur across non-geographical space as well. Boehmke and Pacheco (2016) discuss policy diffusion as a function of ideological similarity between states, similiarity of bureaucratic structure, and other pathways besides contiguity. Nicholson-Crotty and Carley (2016) find that policy diffusion occurs when policymakers observe successful policies in states with similar implementation strategies or regulatory environments. Volden (2016) demonstrates that abandoning policies diffuses across ideologically similar states when that policy is deemed a failure. Thus, the robust diffusion literature is growing even more as scholars uncover how policy diffusion occurs over a variety of non-geographic dimensions.

To demonstrate how scholars may examine diffusion or clustering in non-geographic space, we fit a CAR structure using an ideological distance matrix onto our cross-sectional example using the Erikson, Wright and McIver (1993) *Statehouse Democracy* data in section 4 of this paper. Using each state's measure of *legislative liberalism*, we obtain the weight of each state on another by taking the reciprocal of the absolute difference of the states' legislative ideology, or $\frac{1}{|z_i - z_j|}$ where $z$ is ideology and the terms $i$ and $j$ refer to distinct states. In this weights matrix, states whose legislatures are ideologically similar are closer neighbors. Note that while geographical weights matrices often contain binary values (1 if the states share a physical border and 0 otherwise), the ideological weights matrix contains continuous values, with higher values indicating more ideological similarity. (In both cases, we do row-standardize these weights.) Ideologically similar states could be more likely to have similar unmeasured traits than ideologically distant states; if they do, then their error terms will contain ideological autocorrelation. A CAR model can yield efficiency gains if there is strong ideological clustering.

Equation 6 represents the CAR structural equation model that follows the recursive structure from the *Statehouse Democracy* model (Erikson, Wright and McIver 1993: 129), which we also present in the main article. The $Y$ variables are, respectively, party elite liberalism, Democratic party identification, Democratic legislative strength, legislative liberalism, and

policy liberalism. $x$ is public opinion liberalism.

$$
\begin{aligned}
Y_{1i} &= \beta_1 x_i + \theta_{1i} + \phi_{1i} \\
Y_{2i} &= \beta_2 x_i + \beta_3 y_{1i} + \theta_{2i} + \phi_{2i} \\
Y_{3i} &= \beta_4 x_i + \beta_5 y_{1i} + \beta_6 y_{2i} + \theta_{3i} + \phi_{3i} \\
Y_{4i} &= \beta_7 y_{1i} + \beta_8 y_{3i} + \theta_{4i} + \phi_{4i} \\
Y_{5i} &= \beta_9 x_i + \beta_{10} y_{3i} + \beta_{11} y_{4i} + \theta_{5i} + \phi_{5i} \\
\beta_j &\sim \mathcal{N}(0, 1000) \\
\theta_{ki} &\sim \mathcal{N}(0, \sigma_k^2) \\
\phi_{ki} &\sim CAR(1/\lambda_k) \\
1/\sigma_k^2 &\sim \mathcal{G}(4.56420496, 2.1364) \\
\lambda_k &\sim \mathcal{G}(1.0, 1.0)
\end{aligned}
\tag{6}
$$

As Equation 6 shows, vague conjugate priors are assumed for the coefficients $(\beta_j)$. The $\theta$ terms refer to the independent disturbance in each equation, with each equation's disturbances having a homoscedastic variance. The $\phi$ terms refer to the ideological spatial clustering random effect in each equation, with a CAR prior on each. The variance for each of the random effects terms is freely estimated, and the hyperpriors for these parameters are based on the fair prior recommendation of Bernardinelli, Clayton and Montomoli (1995). In short, the model we estimate now is identical to what is reported in the paper, except for the spatial weights that structure the CAR errors are now driven by ideology rather than geography.

Figure A.7 is a forest plot that contrasts the results from the model in Equation 6 to those from a set of recursive Bayesian linear models. The left margin lists the input variable and the outcome it influences. (In other words, the top coefficient is the effect of legislative liberalism on policy.) The horizontal axis presents the values that a coefficient may take. The point characters are placed to reflect the mean of the posterior distribution of a coefficient. The line segments are placed to reflect the range of the 90% credible intervals, meaning there is a 0.9 posterior probability that the coefficient falls within that range. Solid blue lines with blue triangle characters present the results for a CAR model, while dashed black lines with circle characters present the results for a standard linear model.

As Figure A.7 shows, the mean coefficient estimates shifted only marginally, but the credible intervals show fairly large changes. All of the CAR credible intervals are narrower than the intervals from the model with independent errors, so we get substantial efficiency gains by accounting for ideological similarity in the error term. In fact, the narrower intervals yield robustness with respect to the effect of party elite liberalism on democratic legislative strength, which is now negative and robust while it was not robust in *Statehouse Democracy*. Further, the overall model fit is a better with the CAR model—the Deviance Information
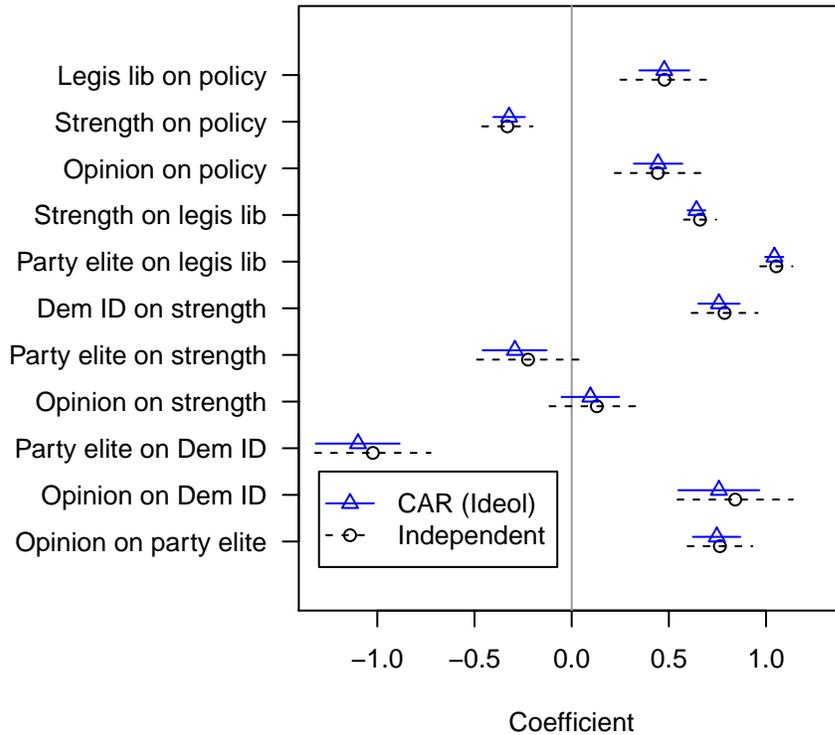
Figure A.7: Posterior means and 90% credible intervals of coefficients for the policy liberalism model using results from linear models and results from a CAR structure based on ideological distance. The models were fit with a post burn-in MCMC sample of 600,000 in OpenBUGS 3.2.3.

Criterion (DIC) for the CAR model is 288.3, which is a lower score and therefore a better penalized fit than the independent-error model's DIC score of 334.9.

It should be noted that all 5 estimates of $\hat{\psi}$ are lower than .5, meaning that while we assumed *a priori* that the share of variance explained by the heterogeneous and ideologically clustered error terms would be approximately equal, the estimates produced by the model all indicated less ideological clustering. The mean $\hat{\psi}$ values are 0.2619 for party elite liberalism, 0.385 for Democratic party identification, 0.4002 for Democratic legislative strength, 0.2334 for legislative liberalism, and 0.1814 for policy liberalism.

# 8 Appendix: Data and Analysis Notes

The sources of the data used in this article are listed as follows:

**Cross-Sectional Data: *Statehouse Democracy***

- Data from the replication of Erikson, Wright and McIver (1993) are from the authors.

**Multilevel Panel Data : Support for Welfare Policies**

- Data from the replication of Margalit (2013) are from the author.

**Duration Data: Adoption of Lotteries**

- Data from 1964-1986 in the duration analysis of lottery adoption are from Berry & Berry: http://mailer.fsu.edu/~wberry/garnet-wberry/a.html, accessed 20 February 2010.

- Lagged per capita income (remeasured for all of 1964-2013 for consistency): U.S. Department of Commerce, Bureau of Economic Analysis, http://bea.gov, accessed 7 May 2017. Inflation adjustment multiplier from Robert Sahr at Oregon State University, http://liberalarts.oregonstate.edu/spp/polisci/research/inflation-conversion-factors-convert-dollars-1774-estimated-2024-dollars-recent-year, accessed 7 May 2017.

- Lagged fiscal health: *Statistical Abstract of the United States*, 1964-2010. Print and electronic editions: http://www.census.gov, accessed 8 July 2010. For 2010-2013, updated with Census state government finances data (https://www.census.gov/govs/state/), accessed 7 May 2017.

- Partisan control of state government. 1987-2009: Klarner (2003), http://www.indstate.edu/polisci/klarnerpolitics.htm, accessed 3 March 2012. 2010-2013: http://www.ballotpedia.org, accessed 7 May 2017.

- Percent fundamentalist: Defined as the percentage of a state's population adhering to a fundamentalist Christian religion, as defined by (Smith 1987: Appendix 2). The Pearson correlation between this series and Berry & Berry's measure of religion in 1971 is 0.9998. Values from the 1971 study were used for years 1964-1978, values from the 1980 study were used for the years 1979-1989, values from the 1990 study were used for the years 1990-1998, values from the 2000 study were used for the years 1999-2009, and values from the 2010 study were used for the years 2010-2013. The data from each of the first four studies were highly correlated with each other, with the lowest correlation being 0.9710 between the 1971 and 2000 data. The data were downloaded from the Association of Religion Data Archives, TheARDA.com, and were collected

by the Association of Statisticians of American Religious Bodies and the following principal investigators:

- Johnson, D.W., P. Picard and Bernard Quinn. 1974. *Churches and Church Membership in the United States.* Glenmary Research Center: Washington, DC.
- Quinn, B., Herman Anderson, Martin Bradley, Paul Goetting and Peggy Shriver. 1982. *Churches and Church Membership in the United States.* Glenmary Research Center: Washington, DC.
- Bradley, M.B., Norman M. Green, Jr., Dale E. Jones, Mac Lynn and Lou McNeil. 1992. *Churches and Church Membership in the United States.* Glenmary Research Center: Washington, DC.
- Jones, Dale E., Sherry Doty, Clifford Grammich, James E. Horsch, Richard Houseal, Mac Lynn, John P. Marcum, Kenneth M. Sanchagrin and Richard H. Taylor. 2002. *Religious Congregations and Membership in the United States 2000: An Enumeration by Region, State and County Based on Data Reported for 149 Religious Bodies.* Glenmary Research Center: Nashville, TN.
- Clifford Grammich, Kirk Hadaway, Richard Houseal, Dale E. Jones, Alexei Krindatch, Richie Stanley, and Richard H. Taylor. 2012. *2010 U.S. Religion Census: Religious Congregations & Membership Study.* Association of Statisticians of American Religious Bodies. http://www.rcms2010.org/

**Panel Count Data: Clean Air Act Enforcement**

- State-level clean air enforcement actions by state and year: U.S. Environmental Protection Agency, Enforcement & Compliance History Online, http://www.epa-echo.gov/echo/idea_download.html, accessed 16 January 2011.

- Non-construction clean air grant allotments: U.S. Environmental Protection Agency, Grant Awards Database, http://yosemite.epa.gov/oarm/igms_egf.nsf/HomePage? ReadForm, accessed 12 April 2011.

- Manufacturing outputs by state, in chained 2005 dollars: Bureau of Economic Analysis, Regional Economic Accounts, http://www.bea.gov/regional/gsp/, accessed 25 May 2011.

- Unemployment by state: Bureau of Labor Statistics, http://data.bls.gov/, accessed 18 July 2010. 2009 data accessed 25 May 2011.

- Partisan control of state government: Klarner (2003), http://www.indstate.edu/ polisci/klarnerpolitics.htm, accessed 3 March 2012.

- Lawsuits under the Clean Air Act: Bureau of National Affairs (2001-2010).

# References

Bernardinelli, L., D. Clayton and C. Montomoli. 1995. "Bayesian Estimates of Disease Maps: How Important are Priors?" *Statistics in Medicine* 14:2411–2431.

Berry, Frances Stokes and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *The American Political Science Review* 84(2):395–415.

Berry, William D. and Brady Baybeck. 2005. "Using Geographic Information Systems to Study Interstate Competition." *American Political Science Review* 99(4):505–519.

Best, Nicola G., Lance A. Waller, Andrew Thomas, Erin M. Conlon and Richard A. Arnold. 1999. Bayesian Models for Spatially Correlated Disease and Exposure Data. In *Bayesian Statistics 6*, ed. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. New York: Oxford University Press.

Blundell, Richard, Rachel Griffith and Frank Windmeijer. 2002. "Individual Effects and Dynamics in Count Data Models." *Journal of Econometrics* 108:113–131.

Bock, R. Darrell. 1975. *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill.

Boehmke, Frederick J. and Julianna Pacheco. 2016. "Introduction to *SPPQ* Special Issue on Policy Diffusion." *State Politics and Policy Quarterly* 16(1):1–7.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. New York: Cambridge University Press.

Brandt, Patrick T. and John T. Williams. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." *Political Analysis* 9(2):164–184.

Brook, D. 1964. "On the Distinction between the Conditional Probability and the Joint Probability Approaches in the Specification of Nearest-Neighbour Systems." *Biometrika* 51:481–483.

Bureau of National Affairs. 2001-2010. *Environment Reporter—Cases*. Vol. 51-70 Washington, DC: The Bureau of National Affairs, Inc.

Darmofal, David. 2009. "Bayesian Spatial Survival Models for Political Event Processes." *American Journal of Political Science* 53:241–257.

Elazar, Daniel J. 1966. *American Federalism: A View from the States*. New York: Thomas Y. Crowell.

Erikson, Robert S., Gerald C. Wright and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States.* New York: Cambridge University Press.

Gelman, Andrew and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7:457–511.

Grossback, Lawrence J., Sean Nicholson-Crotty and David A. M. Peterson. 2004. "Ideology and Learning in Policy Diffusion." *American Politics Research* 32:521–545.

Hougaard, Philip. 2000. *Analysis of Multivariate Survival Data.* New York: Springer.

Klarner, Carl E. 2003. "Measurement of the Partisan Balance of State Government." *State Politics and Policy Quarterly* 3:309–319.

Klein, John P. and Melvin L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data.* 2nd ed. New York: Springer.

Kreitzer, Rebecca J. and Frederick J. Boehmke. 2016. "Modeling Heterogeneity in Pooled Event History Analysis." *State Politics and Policy Quarterly* 16(1):121–141.

Margalit, Yotam. 2013. "Explaining Social Policy Preferences: Evidence from the Great Recession." *American Political Science Review* 107(1):80–103.

Monogan, III, James E. 2013. Modeling Policy Adoption and Impact with Multilevel Methods. In *The SAGE Handbook of Multilevel Modeling*, ed. Marc A. Scott, Jeffrey S. Simonoff and Brian D. Marx. Thousand Oaks, CA: Sage.

Mooney, Christopher Z. 2001. "Modeling Regional Effects on State Policy Diffusion." *Political Research Quarterly* 54:103–124.

Nicholson-Crotty, Sean and Sanya Carley. 2016. "Effectiveness, Implementation, and Policy Diffusion: Or 'Can We Make That Work for Us?'." *State Politics and Policy Quarterly* 16(1):78–97.

Smith, Tom W. 1987. "Classifying Protestant Denominations." GSS Methodological Report No. 43.

Sun, Dongchu, Robert K. Tsutakawa and Paul L. Speckman. 1999. "Posterior Distributions of Hierarchical Models Using CAR(1) Distributions." *Biometrika* 86(2):341–350.

Volden, Craig. 2016. "Failures: Diffusion, Learning, and Policy Abandonment." *State Politics and Policy Quarterly* 16(1):44–77.

Wood, B. Dan. 1992. "Modeling Federal Implementation as a System: The Clean Air Case." *American Journal of Political Science* 36(1):40–67.