Fitting models to areal data

- I will start with a simple model, using species diversity data
- Strong spatial dependence, $\hat{I} = 0.79$
- what is the mean diversity? How precise is our estimate?
- Sampling discussion:
  - The 64 squares are a systematic sample: Only have 1 sample
  - If treat as a simple random sample,
  - usual se probably wrong
- So shift to model-based inference
  - If assume each square independent, $\hat{\mu} = 6.09$, se $\hat{\mu} = 0.36$
  - But we know that model is wrong
  - If we assume spatially correlated, what changes?

Spatial inference

- If recognize the spatial correlation between nearby squares, $\hat{\mu} = 5.69$, se $\hat{\mu} = 1.69$
  - Difference in estimates is not too big (7% change)
  - Difference in se's is huge (470% change)
- This data set has 64 obs
- I find it very helpful to compare sample sizes for equal se's, rather than directly compare se's.
  - So, under the correlated data model, how many obs. needed to get se $= 0.36$?
  - A: 1410! Huge number! Spatial correlation matters!
  - (BTW, also for geostatistical models)
- Computed as $64 * (1.69/0.36)^2$
  - Correlated model: Variance from 64 obs is $1.69^2 = 2.856$
  - Variance from 128 obs will be $2.856/2$.
  - Variance from N obs will be $2.856/(N/64)$. Want $= 0.36^2$. Solve for N

Accounting for spatial correlation

- How did I get estimates and se's "recognizing the spatial correlation"?
- Three common approaches
  - 1) Simultaneous Autoregressive (SAR) Model
  - 2) Conditional Autoregressive (CAR) Model
  - 3) geostatistical model
- Geostatistical model can be used for either point or areal data
- We will consider each in turn.
- First need to talk about:
  - Describing correlated data (multivariate normal distributions)
  - Regression using matrices
  - Regression with correlated data

Multivariate normal distributions

- Usual 401 setup: $Y_i \overset{iid}{\sim} N(\mu_i, \sigma^2)$
  - Mean, μ_i, for each observation may be constant, dependent on treatment, or unique to each observation (e.g., $\beta_0 + \beta_1 X_i$)
  - Variance same for each obs. (no subscript on σ^2)
  - Independent observations
- Now, need to describe correlations among pairs of observations
- $Y_i \sim N(\mu_i, \sigma^2)$ is not sufficient
- Towards that goal: collect all the observations in a vector $\boldsymbol{Y}$
- $\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $\boldsymbol{Y}' = [Y_1\ Y_2\ \cdots\ Y_n]$

Multivariate normal distributions

- Collect the means into a vector: $\boldsymbol{\mu}' = [\mu_1\ \mu_2\ \cdots\ \mu_n]$
- Variance now becomes the variance-covariance matrix
- Consider 3 random variables, X, Y and Z
- VC matrix is a 3 x 3 matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{XY} & \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{XZ} & \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}$$

- When n values of Y, VC matrix is n x n

Variances and covariances

- Diagonal values are variances:

$$\sigma_X^2 = \text{Var } X = \text{E } (X - \mu_X)^2$$

- Off diagonal values are covariances:

$$\sigma_{XY} = \text{Cov } X,\ Y = \text{E } (X - \mu_X)(Y - \mu_Y)$$

  - Cov X, Y = Cov Y, X, so VC matrix is symmetric
  - variance X is covariance of X with itself: $E(X - \mu_X)^2 = E(X - \mu_X)(X - \mu_X)$
  - correlation between X and Y is

$$Cor X, Y = \frac{\text{Cov } X, Y}{\sqrt{\text{Var } X\ \text{Var } Y}}$$

  - Cov = 0 means Cor = 0
  - In general, independent means Cov = 0.
  - When observations have Multivariate normal distribution, Cov = 0 means independent

Multivariate normal distribution

- So can write VC matrix for "401" observations (independent, constant variance)

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 \boldsymbol{I}$$

- $\boldsymbol{I}$ is the identity matrix.
  - matrix equivalent of the scalar 1

Least squares regression, again

- Model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- Can write down (but won't) equations to estimate β_0 and β_1
- Do not generalize easily to more parameters, e.g., $Z_i = \beta_0 + \beta_1 X_i + \beta_2 Y_i + \varepsilon_i$
- Can use matrices to simplify everything

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \boldsymbol{X\beta} = \begin{bmatrix} 1\,\beta_0 + X_1\,\beta_1 \\ 1\,\beta_0 + X_2\,\beta_1 \\ 1\,\beta_0 + X_3\,\beta_1 \\ \vdots \\ 1\,\beta_0 + X_n\,\beta_1 \end{bmatrix}$$

- Model: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$
  - Can have many columns in $\boldsymbol{X}$, or just 1

Least squares regression, again

- For any regression model with independent errors:

$$\hat{\beta}_{ols} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

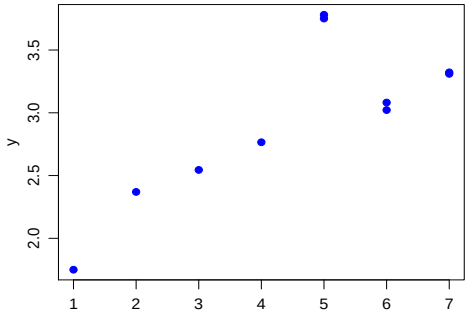
- $()^{-1}$ is the matrix inverse, equivalent of scalar reciprocal
- $\boldsymbol{X}\boldsymbol{X}^{-1} = \boldsymbol{I}$ and $\boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{I}$
- labeled OLS for ordinary least squares
  - will see another type of LS soon (for correlated obs)
- and Var $\hat{\beta}_{ols} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$

Least squares regression, again

- Example: constant mean, $Y_i = \mu + \varepsilon_i$
- $$\boldsymbol{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \boldsymbol{\beta} = [\mu]$$
- $$\boldsymbol{X}'\boldsymbol{X} = \mathbf{1}'\mathbf{1} = 1 + 1 + \cdots = n, \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = 1/n$$
- $$\boldsymbol{X}'\boldsymbol{Y} = \mathbf{1}'\boldsymbol{Y} = Y_1 + Y_2 + \cdots + Y_n = \Sigma Y$$
- so $\hat{\beta}_{ols} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} = \Sigma Y/n$ and Var $\hat{\mu} = \sigma^2/n$

Generalized Least Squares

- When errors are independent, each obs. is an additional piece of information
- Two positively correlated observations are less than 2 pieces of information



Generalized Least Squares

- Model:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \text{ Var } \varepsilon = \boldsymbol{\Sigma}$$

- Estimates:

$$\hat{\boldsymbol{\beta}}_{gls} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$$

- What happens when $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$?

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{gls} &= \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \\
&= \left(\boldsymbol{X}'(\sigma^2\boldsymbol{I})^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I})^{-1}\boldsymbol{Y} \\
&= \left(\boldsymbol{X}'(1/\sigma^2)\boldsymbol{X}\right)^{-1}\boldsymbol{X}'(1/\sigma^2)\boldsymbol{Y} \\
&= \sigma^2(1/\sigma^2)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} = \hat{\beta}_{ols}
\end{aligned}$$

- the GLS estimator simplifies to the usual OLS estimator
  - when observations are independent with constant variance

Generalized Least Squares

- Another way to think about GLS
- For most $\boldsymbol{\Sigma}$, can find a "square-root" matrix, so that $\boldsymbol{\Sigma} = \boldsymbol{C}'\boldsymbol{C}$
- Apply this idea to $\boldsymbol{\Sigma}^{-1}$ to get a square root matrix for the inverse, we will call this $\boldsymbol{B}$, so $\boldsymbol{\Sigma}^{-1} = \boldsymbol{B}'\boldsymbol{B}$
- Multiply all terms in the model by $\boldsymbol{B}$

$$\boldsymbol{BY} = \boldsymbol{BX}\beta + \boldsymbol{B}\varepsilon$$

- Look what happens to the errors:

$$\text{Var } \boldsymbol{B}\varepsilon = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}' = \boldsymbol{B}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{B}') = \boldsymbol{I}\boldsymbol{I} = \boldsymbol{I}$$

- Pre multiplying by square root of the inverse covariance matrix removes the correlation - transformed errors are now independent!

Generalized Least Squares

- So, if know the correlation matrix, can compute new $\boldsymbol{Y}^* = \boldsymbol{BY}$ and new $\boldsymbol{X}^* = \boldsymbol{BX}$
- and use OLS on $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$: $\boldsymbol{Y}^* = \boldsymbol{X}^*\beta + \varepsilon^*$
  - The regression coefficients from OLS on transformed variables are:
  - $\beta = \left[\boldsymbol{X}^{*'}\boldsymbol{X}^*\right]^{-1}(\boldsymbol{X}^{*'}\boldsymbol{Y}^*) = \left[(\boldsymbol{X}'\boldsymbol{B}')(\boldsymbol{BX})\right]^{-1}\left[(\boldsymbol{X}'\boldsymbol{B}')(\boldsymbol{BY})\right]$
  - $= (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y})$
  - which are the GLS estimates for the original model

Generalized Least Squares

- Take home: If you know $\boldsymbol{\Sigma}$, can compute the "best" estimates for any regression model
- using GLS
- A generalization (the Aitken model) says that all you need is the correlation part of $\boldsymbol{\Sigma}$
  - Write $\boldsymbol{\Sigma}$ as $\sigma^2\boldsymbol{C}$ where $\boldsymbol{C}$ is (now) a correlation matrix
  - only need to know $\boldsymbol{C}$, can estimate σ^2.
- Practical problem is that the correlation part almost always has to be estimated
- $\boldsymbol{\Sigma}$ (or $\boldsymbol{C}$) depends on unknown parameters.

SAR models

- General regression / ANOVA model
- $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$, if $\boldsymbol{X} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, then $\boldsymbol{Y} = \boldsymbol{\mu} + \varepsilon$
- model spatial correlation by allowing ε to depend on error values in neighboring regions

$$\varepsilon(s_i) = \Sigma_{j=1}^{N} b_{ij}\varepsilon_j + \nu(s_i)$$

- b_{ij} are the elements of the spatial dependence matrix expressing dependence among regions
- $\nu(s_i)$ is an independent random disturbance for each region. Usually assume $\nu(s_i) \overset{iid}{\sim} N(0, \sigma_\nu^2)$

SAR models

- What this model "means". Some examples:
- In all:
  - Focus on center observation (location s_5)
  - assume row standardized rook's neighbors, so $\{b_{ij}\}$ is

$$\begin{array}{ccc} 0 & 0.25 & 0 \\ 0.25 & \mathbf{0} & 0.25 \\ 0 & 0.25 & 0 \end{array}$$

- $\varepsilon(s_5) = 10$ Is this value large or small?
  - when b_{ij} is not zero, depends on neighbors
  - so look at $\nu(s_5)$ for $\varepsilon(s_5) = 10$
  - and different values for neighbors ε

- $$\begin{array}{ccc} 7 & \mathbf{8} & 10 \\ \mathbf{12} & 10 & \mathbf{8} \\ 11 & \mathbf{10} & 9 \end{array}$$
  - Only the bolded neighbors matter, $\Sigma b_{ij}\varepsilon(s_j) = 9.5$, $\nu(s_i) = 0.5$
  - large error but similar to neighbors, so independent contribution, $\nu(s_i)$ is small
- $$\begin{array}{ccc} 7 & \mathbf{16} & 10 \\ \mathbf{14} & 10 & \mathbf{12} \\ 11 & \mathbf{16} & 9 \end{array}$$
  - $\Sigma b_{ij}\varepsilon(s_j) = 14.5$, $\nu(s_i) = -4.5$
  - large error, but neighbors are larger, so independent contribution, $\nu(s_i)$ is negative

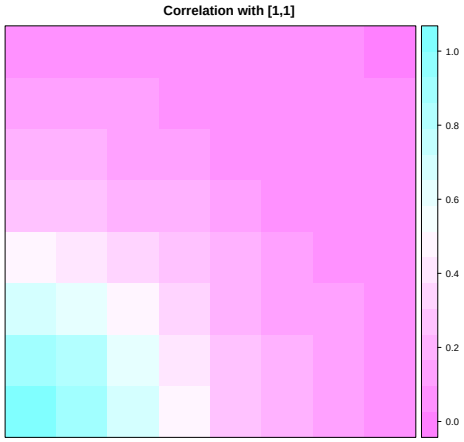
- $$\begin{array}{ccc} 1 & \mathbf{1} & 2 \\ \mathbf{3} & 10 & \mathbf{1} \\ 1 & \mathbf{2} & 3 \end{array}$$
  - $\Sigma b_{ij}\varepsilon(s_j) = 1.75$, $\nu(s_i) = 8.25$
  - much larger than neighbor errors, so independent contribution, $\nu(s_i)$ is large
- $$\begin{array}{ccc} -5 & \mathbf{-7} & -10 \\ \mathbf{-3} & 10 & \mathbf{-6} \\ -1 & \mathbf{2} & 0 \end{array}$$
  - $\Sigma b_{ij}\varepsilon(s_j) = -3.5$, $\nu(s_i) = 13.5$
  - neighbors suggest a negative error, so independent contribution, $\nu(s_i)$ is very large

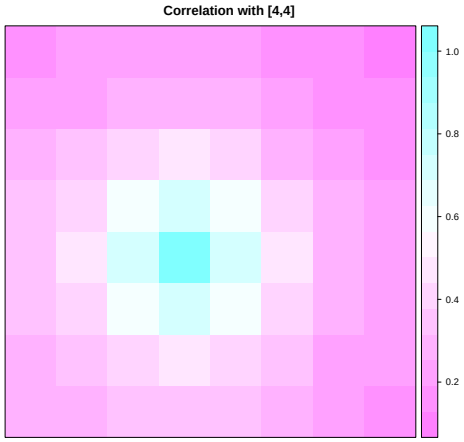
SAR models

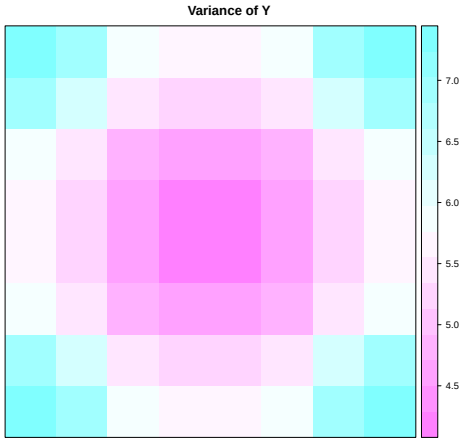
- When part of the variation in error values can be "explained" by neighbors, the "explained" part is removed under the SAR model.
- Both SAR and CAR models do this, but they define "explained" differently
- SAR model: applies to all locations simultaneously
- For one location: $\varepsilon(s_i) = \Sigma_{j=1}^{N} b_{ij}\varepsilon_j + \nu(s_i)$
- For all locations: $\boldsymbol{\varepsilon} = \boldsymbol{B}\boldsymbol{\varepsilon} + \boldsymbol{\nu}$, which means:
  - $\boldsymbol{\nu} = \boldsymbol{\varepsilon} - \boldsymbol{B}\boldsymbol{\varepsilon} = (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{\varepsilon}$
  - $\boldsymbol{\varepsilon} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\nu}$
- $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\nu}$
  - where $\boldsymbol{I}$ is the NxN identity matrix. $\boldsymbol{B}$ is the matrix of "connectivity" coefficients
  - and $\boldsymbol{\nu}$ is the vector of independent errors

SAR model in matrix terms

- If consider $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$, then the Var-Cov matrix of ε is $(\boldsymbol{I} - \boldsymbol{B})^{-1}\Sigma_\nu(\boldsymbol{I} - \boldsymbol{B}')^{-1}$
  - $\boldsymbol{B}$ can be arbitrary, but it has N(N-1) parameters. usually simplify using a model
  - usually write $\boldsymbol{B} = \rho W$, where ρ is a coefficient of spatial dependence and $\boldsymbol{W}$ is the spatial weight matrix
  - Note $\rho = 0 \Rightarrow$ locations are connected (elements of $\boldsymbol{W} > 0$), but not spatially dependent
- What does this model imply about VC matrix of the observations?
- Consider $\rho = 0.9$, $\boldsymbol{W}$ is rook's neighbors, row standardized, $\boldsymbol{\Sigma}_\nu = \sigma^2\boldsymbol{I}$
- Pictures on next few slides
  - Correlation declines with distance: good!
  - But Var Y not constant - largest in corners, smalest in middle of region

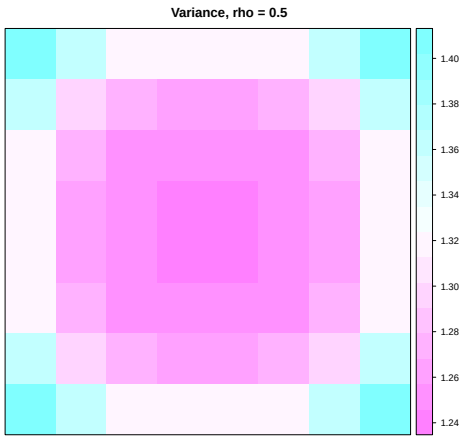


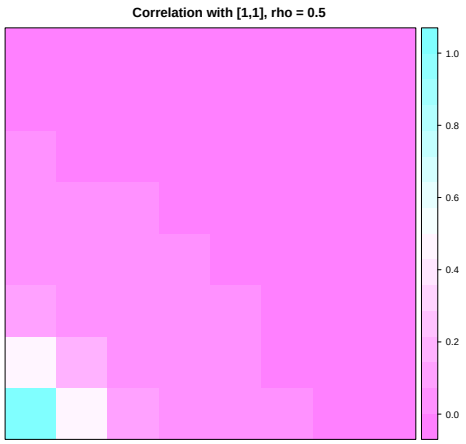


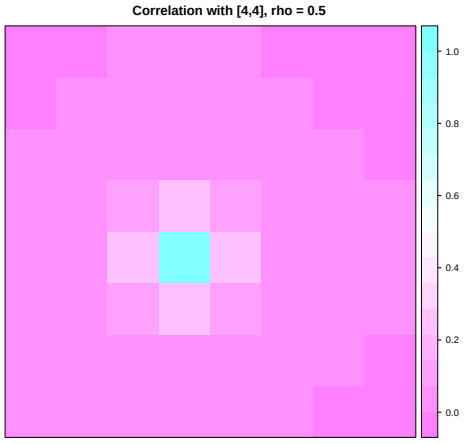


SAR models

- Notice modeling correlation between observations indirectly
  - $\boldsymbol{B}$ describes how one error value depends on the values of neighboring values
  - describing connections between observations, not correlation
  - Correlated errors is a consequence of those connections
  - As is unequal variances
- A geostatistical model directly describes correlations
- Preference is subject-matter dependent
  - Spatial econometrics, most areal data analysis, connections
  - Spatial statistics, geostatistical data, correlations
- Degree of non-constant variance depends on magnitude of ρ
- Similar plots for $\rho = 0.5$







Estimation for SAR models

- Reminder: model is $\boldsymbol{Y} = \boldsymbol{X}\beta + (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\nu}$,
- where $\boldsymbol{B} = \rho\boldsymbol{W}$
- $\boldsymbol{W}$ is the known spatial weight matrix
- IF ρ is known, then $\boldsymbol{B}$ is known, only need to estimate $\hat{\beta}$
- Easy:
  - 1) calculate $\Sigma_\varepsilon = (\boldsymbol{I} - \boldsymbol{B})^{-1}\Sigma_\nu(\boldsymbol{I} - \boldsymbol{B})^{-1}$ and use GLS, or
  - 2) Note that: $(\boldsymbol{I} - \boldsymbol{B})\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{B})(\boldsymbol{X}\beta) + \boldsymbol{\nu}$
    transform Y vector and X matrix, and you have an OLS problem.
- Usual situation: ρ is unknown, need to estimate
  - use maximum likelihood
  - general alternative to LS for any statistical problem
  - Have already seen the VC matrix for ε: $\boldsymbol{\Sigma}_\varepsilon = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Sigma}_\nu(\boldsymbol{I} - \boldsymbol{B})^{-1}$, where $\boldsymbol{B} = \rho\boldsymbol{W}$
  - mvNormal lnl is

$$-\frac{k}{2}\log 2\pi - \frac{1}{2}\log \mid \boldsymbol{\Sigma}_\varepsilon \mid -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\beta)'\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{y} - \boldsymbol{X}\beta)$$

Maximizing the mvN lnL

- Iterative algorithm.
- Key insight is that given $\boldsymbol{\Sigma}_\varepsilon$, mle of β is trivial (GLS)
- Assume $\rho = 0$, find OLS estimate of β
- Condition on β, use numerical maximization to find $\hat{\rho} \mid \beta$
- find GLS estimate of β for $\boldsymbol{\Sigma}_\varepsilon(\rho)$
- repeat last two steps until convergence.
- Traditional frequentist approach is to estimate β and Var β conditional on $\hat{\rho}$
- Bayesian approach incorporates uncertainty in $\hat{\rho}$ into uncertainty about β
- Not an issue for simple problems (e.g. $\boldsymbol{\Sigma}_\varepsilon = \sigma^2\boldsymbol{I}$) because in this case, $\hat{\beta}$ independent of σ^2
- Is in issue in these models because $\hat{\rho}$ and $\hat{\beta}$ are not independent.
- For sp diversity data, $\hat{\rho} = 0.914$.
  That is strong positive spatial dependence.

Useful things about likelihood

- test hypotheses using Likelihood Ratio Test (LRT)
- e.g. test $Ho: \rho = 0$
  - calculate lnL given $\rho = 0$ (need to maximize over β): lnL_0
  - calculate lnL at mle's of all parameters: lnL_A
  - $\text{lnL}_A \geq \text{lnL}_0$, because ρ probably not 0
  - but by how much? $\sim 0 \Rightarrow$ data consistent with $\rho = 0 \Rightarrow$ accept Ho
  - but how much is "too far" from 0?
  - General result: When Ho true, $-2(lnL_0 - lnl_A) \sim \chi^2_k$ where k is the difference in # parameters between the two models
  - This is an asymptotic result, but surprisingly effective in small samples
    - Here, $k = 1$ because testing hypothesis about 1 parameter
    - lnL_A = -112.137
      lnL_0 = -158.306
      $\Delta = -2(lnL_0 - lnL_A) = 92.34$
    - $\chi^2_{1,0.95} = 3.84$. Here $p << 0.0001$
  - LRT only works when one model is a simplification of another

Useful things about likelihood

- Model selection: e.g. compare different $\boldsymbol{W}$ matrices
  - AIC = -2 lnL + 2 k
  - k is # of parameters. spdep counts β, σ^2, and ρ
  - Can compare models with same number of parameters, or diff. # parameters
  - choose model with smallest AIC
  - Compare row-standardized (style='W') to binary (style='B') weights
  - Different models for connections between areas
  - $\Rightarrow$ Different variance and covariance relationships
    - Here, 3 parameters (μ, σ^2, and ρ)
    - Row std: AIC = 224.27 + 6 = 230.2742
    - Binary: AIC = 235.54 + 6 = 241.54
    - Choose model with Row std weights.

Reminders about interpretation of AIC

- Size of AIC irrelevant
  - doesn't matter whether best = 1000 or best = -250
  - Only comparisons among models fit to the same data
- Does not mean row std. is the "correct" model.
  - Only best among the set being considered.
  - model diagnostics still very important and useful
    - I don't know any that evaluate choice of weight matrix
- Smallest AIC is "best" model - Is anything almost as good?
  - Models with AIC within 2 units of the best are likely alternatives
  - Models with AIC > 10 units of the best are unlikely

Useful things about likelihood

- confidence interval for ρ by profile likelihood
  - Concept: Repeat LRT for many values of ρ
  - include inside 95% ci all values of ρ for which test has $p > 0.05$
  - here (0.82, 0.98)
- consequences of choice of ρ on inference about $\boldsymbol{\beta}$
  - $\rho = 0$ (independence): $\hat{\mu} = 6.09$, se $= 0.36$
  - $\rho = 0.914$: $\hat{\mu} = 5.69$, se $= 1.69$
- Two points:
  - just demonstrated non-independence of $\hat{\rho}$ and $\hat{\boldsymbol{\beta}}$ in a SAR model
  - se of $\hat{\mu}$ when account for correlation much larger
    - observations are positively correlated, so fewer "effective" observations

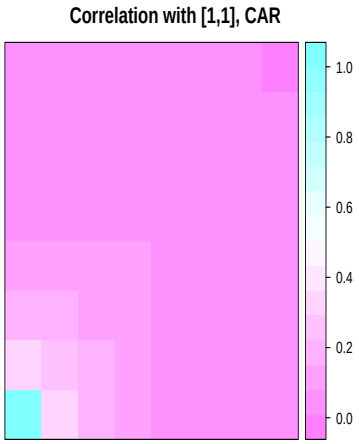
CAR models

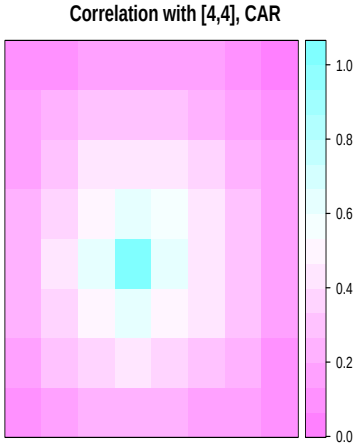
- SAR models all Y values simultaneously
- CAR models distrib of a each Y given the values of its neighbors

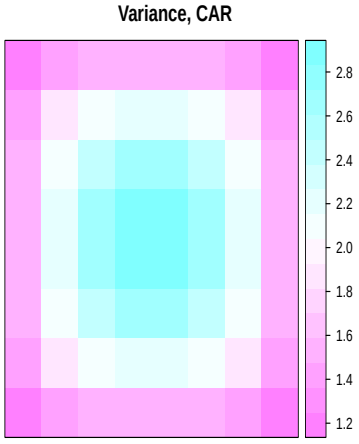
$$Y_i \mid Y_{-i} = \boldsymbol{X}_i\boldsymbol{\beta} + \Sigma_{j=1}^{N} c_{ij}(Y_j - \boldsymbol{X}_j\boldsymbol{\beta}) + \nu_i$$

- RHS same as SAR model
- The difference is that Y_{-i} now treated as fixed values when specifying distrib of Y_i
- Diff. VC matrix for obs.: $\boldsymbol{\Sigma}_\varepsilon = \text{Var } Y = (\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{\Sigma}_\nu$

- to be a valid VC matrix, requires some conditions:
  - ρ can not be too large
  - Definition depends on the weight matrix, $\boldsymbol{W}$
  - $\boldsymbol{C}$ must be symmetric, $C_{ij} = C_{ji}$
  - so my example now uses binary weights
- Practical difference: correlation pattern similar
  - Pictures on next slide
  - correl between pairs in middle higher among than between middle and edge
  - similar to SAR pattern, but details slightly different
- But pattern of variance quite different
  - biggest var in the middle of the area (more connections to other points)
- Fitting a CAR model gives: $\hat{\mu} = 5.10$, se $= 0.83$, $\hat{\rho} = 0.253$.
- 0.253 doesn't seem large, but close to maximum possible value







- Different variance and correlation pattern $\Rightarrow$ CAR and SAR are not the same models
- Which fits the data better? How will you answer this Q?

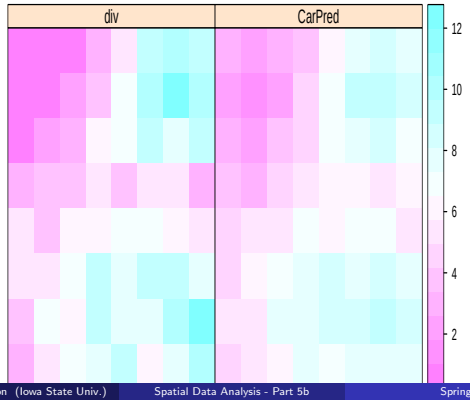
Model comparison

- I would use AIC

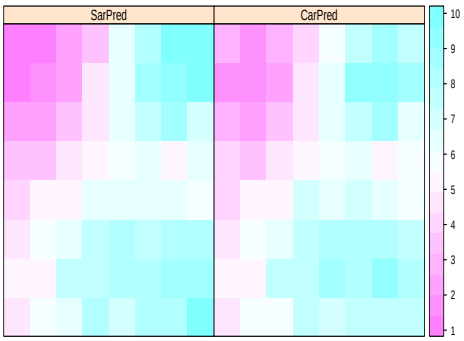
model	AIC
SAR	241.54
CAR	262.38
Indep	320.61

  - Clear dominance of SAR model
- Traditional interpretation
  - AIC w/i 2 of the top: worse model is reasonable competitor
  - AIC more than 10 from the top: worse model is very unlikely
- For Bayesian analysis, CAR models very, very popular
  - They are easy to use in an MCMC chain, because they are conditional distributions

Smoothing noisy areal data



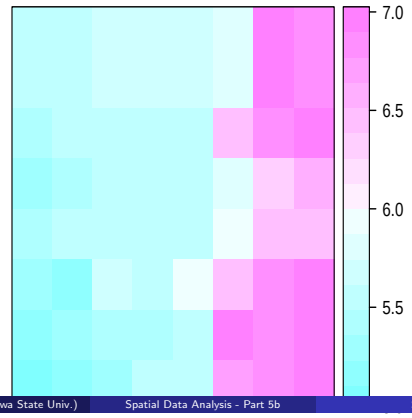
Comparison of smoothing using CAR and SAR



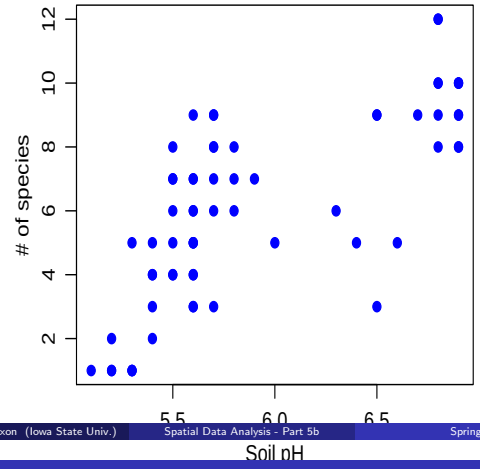
More complicated models

- What if you have X variables measured for each area?
- Easy to include in the model
- My example: soil pH measured in each area
- the small quadrats cross from igneous parent material (lower pH) to limestone parent material (higher pH)
- Residuals from OLS regression are spatially correlated
  - Moran's I: 0.62
  - less than I for observations = 0.79
- Pictures on next three slides

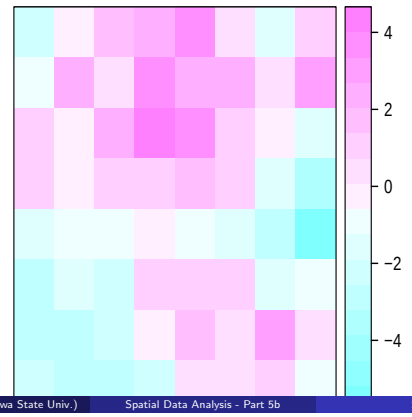
Soil pH



species vs. Soil pH



Residuals from linear regression

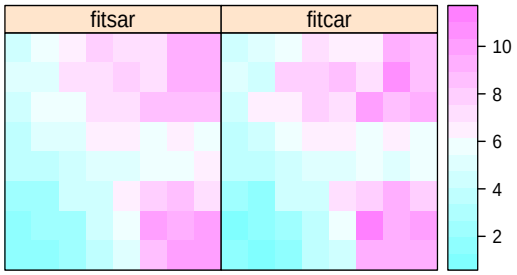


Results from pH models

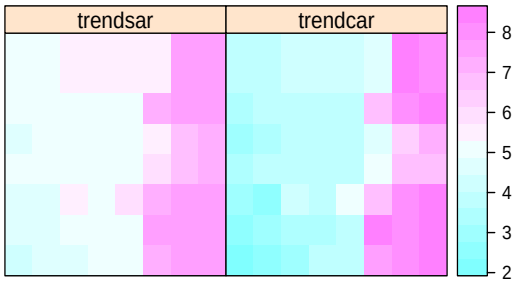
Model	$\hat{\beta}_{pH}$	se	$\hat{\rho}$	AIC
Indep.	3.56	0.43	–	275.06
SAR	1.93	0.66	0.86	225.77
CAR	3.28	0.52	0.26	233.78

- Pictures of the two fits on next slide

SAR and CAR fits



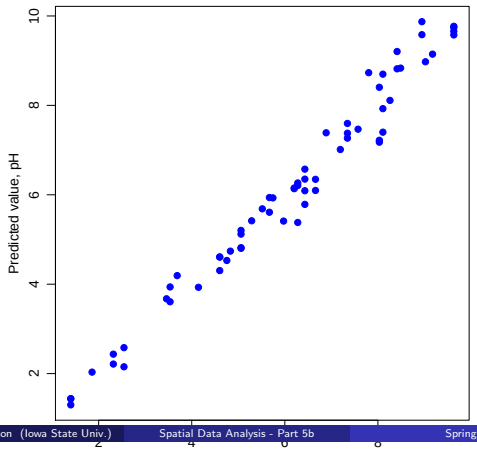
SAR and CAR fits: trend component = X beta



Why is SAR slope for pH much lower?

- When X variable is spatially correlated
  - X and the spatial correlation "fight" to predict Y
  - analogous to two correlated X variables: pH and "space"
- Commonly seen when using spatial correlated errors
  - "The effect you love disappears"
- For these data:
  - SAR fit: smaller β_{pH}, higher correlation
  - CAR fit: larger β_{pH}, smaller correlation

Comparison of intercept only to pH SAR models



Comparison of intercept only to pH SAR models