Spatial Modeling with Spatially Varying Coefficient Processes

Alan E. Gelfand, Hyon-Jung Kim, C. F. Sirmans, Sudipto Banerjee

Presenters: Halley Brantley and Chris Krut

September 28, 2015

Introduction

Want to model selling price of single family homes in Baton Rouge.

Covariates

- Age
- Square feet of liv
- Square feet of ot
- # of bathrooms

See Figure 1. Pg. 392

Linear Model

$$Y(\mathbf{s}) = X(\mathbf{s})\beta + W(\mathbf{s}) + \epsilon$$

$\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$
$X(\mathbf{s}) = (\mathbf{1}, Age, LivingArea, OtherArea, Bathrooms)$
$\epsilon \sim N(0, \tau^2)$ independent error term

Spatial Components

1. $\mathbf{s_i}$ is the location of observation i
2. Spatial random effect $W(\mathbf{s}) \sim N(0, \sigma^2 \mathbf{H})$
3. $(\mathbf{H})_{ij} = \rho(\mathbf{s_i} - \mathbf{s_j} | \phi)$
4. $cov(Y(s_i), Y(s_j) | \beta, \tau^2, \phi) = \sigma^2 \rho(s_i - s_j | \phi)$

Challenge: βs may not be constant across space

- Possibility 1: divide area into similar blocks.
  - Must assume coefficient is constant on specified area.
  - Arbitrary scale of resolution
  - Can't interpolate the surface
- Possibility 2: Assume β is a polynomial or spline function of lat/lon.
  - Arbitrary specification of polynomial could be too inflexible.
  - Splines are more flexible, but need to specify number and location of knots.

Gaussian Process

Definition

$Y(s)$ is a Gaussian process with mean function $\mu(s)$ and covariance function $cov(Y(s), Y(s')) = H(s, s')$ if for every subset of locations $s_1, ..., s_n$ the vector $\tilde{Y} = (Y(s_1), ..., Y(s_n))^T$

$$\tilde{Y} \sim MVN_n(\tilde{\mu}, \mathbf{H}) \quad (1)$$

where $\tilde{\mu} = (\mu(s_1), ..., \mu(s_n))^T$ and $\mathbf{H}$ is a matrix such that $\{\mathbf{H}\}_{ij} = H(s_i, s_j; \phi)$.

To be a valid covariance function $H(s, s' : \phi)$ must be positive semidefinite in the sense that it generates covariance matrices $\mathbf{H}$ which are positive semi definite $v^T \mathbf{H} v \geq 0$[1, Pg. 80].

Spatial Intercept Model: Process Model

- $Y(s) = \beta_0 + x(s)\beta_1 + \boldsymbol{\beta}_0(s) + \varepsilon(s)$
- $E(\varepsilon(s)) = 0$
- $cov(\varepsilon(s), \varepsilon(s')) = \tau^2 I(s = s')$
- $\beta_0(s) \sim$ Gaussian Process(0,$H(s, s'; \phi, \sigma^2)$)
- $H(s, s'; \phi, \sigma^2) = \sigma^2 \rho(s - s'; \phi)$.
- $\tilde{\beta}_0(s) = \beta_0 + \beta_0(s)$
- Assign Prior distributions to remaining parameters $(\beta_0, \beta_1, \sigma_0^2, \tau^2, \phi_0)$

Spatial Intercept Model: Discretized Model

- Observe $Y(s)$, $x(s)$, over set of locations $s_1, ..., s_n$.
- $Y(s_i) = \beta_0 + x(s_i)\beta_1 + \beta_0(s_i) + \varepsilon(s_i)$
- $\varepsilon(s_i) \sim N(0, \tau^2)$
- $(Y(s_i)|\beta_0, \beta_1, \beta_0(s_i)) \sim N(\beta_0 + x(s_i)\beta_1 + \beta_0(s_i), \tau^2)$
- $\boldsymbol{\beta_0} = (\beta_0(s_1), ..., \beta_0(s_n))$
- $\boldsymbol{\beta_0} \sim MVN(0, \sigma_0^2 \mathbb{H}_0(\phi_0))$
- Assign Prior distributions to remaining parameters $(\beta_0, \beta_1, \sigma_0^2, \tau^2, \phi_0)$

Spatial Slope Model: Process Model

- $Y(s) = \beta_0 + x(s)\beta_1 + x(s)\beta_1(s) + \varepsilon(s)$
- $E(\varepsilon(s)) = 0$
- $cov(\varepsilon(s), \varepsilon(s')) = \tau^2 I(s = s')$
- $\beta_0(s) \sim$ Gaussian Process(0,$H(s, s'; \phi, \sigma^2)$)
- $H(s, s'; \phi, \sigma^2) = \sigma^2 \rho(s - s'; \phi)$).
- $\tilde{\beta}_1(s) = \beta_1 + \beta_1(s)$
- Assign Prior distributions to remaining parameters $(\beta_0, \beta_1, \sigma_0^2, \tau^2, \phi_0)$
- The process $Y(s)$ ends up being heterogeneous and non-stationary.
  - $var(Y(s)|\beta_0, \beta_1, \tau^2, \sigma_1^2, \phi_1) = x^2(s)\sigma_1^2 + \tau^2$
  - $cov(Y(s), Y(s')|\beta_0, \beta_1, \tau^2, \sigma_1^2, \phi_1) = x(s)x(s')\sigma_1^2\rho_1(s - s'; \phi_1)$

Spatial Slope Model: Discretized Model

- Observe $Y(s)$, $x(s)$, over set of locations $s_1, ..., s_n$.
- $Y(s_i) = \beta_0 + x(s_i)\beta_1 + x(s_i)\beta_1(s_i) + \varepsilon(s)$
- $\varepsilon(s_i) \sim N(0, \tau^2)$
- $(Y(s_i)|\beta_0, \beta_1, \beta_1(s_i)) \sim N(\beta_0 + x(s_i)\beta_1 + x(s_i)\beta_1(s_i), \tau^2)$
- $\boldsymbol{\beta_1} = (\beta_1(s_1), ..., \beta_1(s_n))$
- $\boldsymbol{\beta_1} \sim MVN(0, \sigma_1^2 \mathbb{H}_1(\phi_1))$
- Assign Prior distributions to remaining parameters $(\beta_0, \beta_1, \sigma_0^2, \tau^2, \phi_0)$

Spatial Varying Coefficients Model

$$Y(s) = \beta_0 + x(s)\beta_1 + \beta_0(s) + x(s)\beta_1(s) + \varepsilon(s)$$

- $Y(s_i) = \beta_0 + x(s_i)\beta_1 + \beta_0(s_i) + x(s_i)\beta_1(s_i) + \varepsilon(s_i)$
- Vector of spatial random effects at observed locations: $\boldsymbol{\beta} = (\beta_0(s_1), \beta_1(s_1), ..., \beta_0(s_n), \beta_1(s_n))^T$
- $\boldsymbol{\beta} \sim MVN(0, \mathbb{H}(\phi) \otimes T)$
- Positive definite matrix: $T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$
- $cov(\beta_0(s_i), \beta_0(s_j)) = \rho(s_i - s_j : \phi)T_{11}$
- $cov(\beta_1(s_i), \beta_1(s_j)) = \rho(s_i - s_j : \phi)T_{22}$
- $cov(\beta_0(s_i), \beta_1(s_j)) = \rho(s_i - s_j : \phi)T_{12}$

Multiple Regression with Spatially Varying coefficients

$$Y(s) = \boldsymbol{X}^T(s)\tilde{\boldsymbol{\beta}}(s) + \varepsilon(s)$$

$$Y(s) = \boldsymbol{X}^T(s)\mu_\beta + \boldsymbol{X}^T(s)\boldsymbol{\beta}(s) + \varepsilon(s)$$

- $\boldsymbol{X}^T(s) = (1, x_1(s), ..., x_p(s))$
- $\tilde{\boldsymbol{\beta}}(s) = \mu_\beta + \boldsymbol{\beta}(s)$
- $\mu_\beta = (\beta_0, \beta_1, ..., \beta_p)$
- $\boldsymbol{\beta}(s) = (\boldsymbol{\beta}_0(s), \boldsymbol{\beta}_1(s), ..., \boldsymbol{\beta}_p(s))$

Multiple Regression with Spatially Varying coefficients

$$\boldsymbol{y} = X(1 \otimes \mu_\beta) + X\boldsymbol{\beta} + \varepsilon$$

- $\boldsymbol{y} = (Y(s_1), ..., Y(s_n))^T$
- $X = \begin{pmatrix} \boldsymbol{X}(s_1) & & & \\ & \boldsymbol{X}(s_2) & & \\ & & \ddots & \\ & & & \boldsymbol{X}(s) \end{pmatrix}$
- $\mu_\beta = (\beta_0, \beta_1, ..., \beta_p)^T$
- $\boldsymbol{\beta} = (\beta_0(s_1), ..., \beta_p(s_1), ..., \beta_0(s_n), ..., \beta_p(s_n))^T$
- $\boldsymbol{\beta} \sim MVN(0, \mathbb{H}(\phi) \otimes T)$
- T is a $p + 1 \times p + 1$ positive definite matrix.
- $cov(\boldsymbol{\beta}_i(s), \boldsymbol{\beta}_j(s')) = \rho(s - s' : \phi)T_{i+1j+1}$

Marginalized Likelihood

$$L(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y}) = \int f(\boldsymbol{y}|X, \mu_\beta, \boldsymbol{\beta}, T, \phi, \tau^2)\pi(\boldsymbol{\beta})d\boldsymbol{\beta}$$

$$\begin{aligned} L(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y}) = |(X(H(\phi) \times T)X^T + \tau^2 I)|^{-1/2} * \\ \exp(-\frac{1}{2}(\boldsymbol{y} - X(\mathbf{1} \otimes \mu_\beta))^T * \\ (X(H(\phi) \times T)X^T + \tau^2 I)^{-1} * \\ (\boldsymbol{y} - X(\mathbf{1} \otimes \mu_\beta))) \end{aligned}$$

Slice Sampling

- Want to sample from posterior $f(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y})$
- Introduce an auxiliary variable $U \sim unif(0, L(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y}))$ and use the fact

$$f(\mu_\beta, T, \tau^2, \phi, U|\boldsymbol{y}) = I(U < L(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y}))\pi(\mu_\beta, T, \tau^2, \phi)$$

- Sample U from $unif(0, L(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y}))$.
- Sample $\mu_\beta, T, \tau^2, \phi$ from $\pi(\mu_\beta, T, \tau^2, \phi)$ subject to $I(U < L(\mu_\beta, T, \tau^2, \phi|\boldsymbol{y}))$.

Sampling $\boldsymbol{\beta}(s)$

1. We can obtain the Marginal Posterior
   $f(\tilde{\boldsymbol{\beta}}|, y) = \int f(\tilde{\boldsymbol{\beta}}|\mu_\beta, T, \phi, \tau^2, y) * f(\mu_\beta, T, \phi, \tau^2|y)$
2. Full Conditional is of a known form

   $f(\tilde{\boldsymbol{\beta}}|\mu_\beta, T, \phi, \tau^2, y) = N(Bb, B)$

   $B = (X^T X/\tau^2 + H^{-1}(\phi) \otimes T^{-1})^{-1}$

   $b = X^T y/\tau^2 + (H^{-1}(\phi) \otimes T^{-1})(1 \otimes \mu_\beta)$
3. We can generate from $f(\tilde{\boldsymbol{\beta}}|, y)$ using the samples from $f(\mu_\beta, T, \phi, \tau^2|y)$ and the full conditional distribution in 2.

Sampling $\boldsymbol{\beta}(s_{n+1})$ for a new location s_{n+1}

1. We can obtain the Marginal Posterior
$f(\tilde{\boldsymbol{\beta}}(s_{n+1})|, y) =$
$\int f(\tilde{\boldsymbol{\beta}}(s_{n+1})|\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi, \tau^2) f(\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi, \tau^2|y)$
2. Conditional distribution has a known form
$f(\tilde{\boldsymbol{\beta}}(s_{n+1})|\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi, \tau^2) = N(\overline{\mu}, \overline{\sigma})$

$\overline{\mu} = \mu_\beta + (h_{new}^T(\phi) H^{-1}(\phi) \otimes I)(\tilde{\beta} - 1_{n\times 1} \otimes \mu_\beta)$
$\overline{\sigma} = (I - h_{new}^T(\phi) H^{-1}(\phi) h_{new}(\phi))T$
$h_{new}(\phi) = (\rho(s_{n+1} - s_1; \phi), ..., \rho(s_{n+1} - s_1; \phi))^T$
3. Similar to before we can use posterior samples to generate $\tilde{\boldsymbol{\beta}}(s_{n+1})$

Predict $y(s_{n+1})$ from predictive posterior

1. The Predictive Posterior is given by
$f(y(s_{n+1})|y) = \int f(y(s_{n+1})|\tilde{\boldsymbol{\beta}}(s_{n+1}), \tau^2) *$
$f(\tilde{\boldsymbol{\beta}}(s_{n+1})|\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi) * f(\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi, \tau^2|y)$

2. Can sample from $f(y(s_{n+1})|y)$ by iteratively generating from $f(\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi, \tau^2|y)$, $f(\tilde{\boldsymbol{\beta}}(s_{n+1})|\tilde{\boldsymbol{\beta}}, \mu_\beta, T, \phi)$, $f(y(s_{n+1})|\tilde{\boldsymbol{\beta}}(s_{n+1}), \tau^2)$.

Application to housing prices

Compared 4 different models:

1. Allow all βs to vary spatially (multivariate).
2. Allow 3 of the βs to vary spatially.
3. Allow 2 of the βs to vary spatially.
4. Allow all βs to vary spatially (independently).

Model comparison was based on the goodness-of-fit and penalty for model complexity.

$$D = \sum_{(\mathbf{s},t)} (Y_{obs}(\mathbf{s},t) - \mu(\mathbf{s},t))^2 + k \sum_{(\mathbf{s},t)} \sigma^2(\mathbf{s},t)$$

Spatial Results

See Table 3, Pg. 392

Spatial Results

See Figure 4, Pg. 393

Space-Time Regression with Varying Coefficients

$$
\begin{aligned}
Y(s) &= \mathbf{X}^T(s,t)\tilde{\boldsymbol{\beta}}(s,t) + \varepsilon(s,t) \\
\tilde{\boldsymbol{\beta}}(s,t) &= \mu_\beta + \beta(s,t)
\end{aligned}
$$

1. $\beta(s,t) = \beta(s)$
2. $\beta(s,t) = \beta(s) + \alpha(t)$
   - 2.a $\alpha_k(t) = I(t=1)\alpha_{k1} + ... + I(t=M)\alpha_{kM}$
   - 2.b $\alpha(t) = (\alpha_0(t), ..., \alpha_P(t))^T$
     $cov(\alpha(t), \alpha(t')) = \rho_{(2)}(t-t';\gamma)T$
3. $\beta(s,t) = \beta^t(s)$(independent processes nested in t)
   $(cov(\beta^t(s), \beta^t(s')) = \rho(s-s';\phi^{(t)})T^t$
4. Separable covariance
   $cov(\beta(s,t), \beta(s',t')) = \rho^1(s-s';\phi) \times \rho^2(t-t':\gamma)T$

Spatio-Temporal Results

Model prices over 4 years.

Model 1: $\beta(\mathbf{s},t) = \beta(\mathbf{s})$

Model 2: $\beta(\mathbf{s},t) = \beta(\mathbf{s}) + \alpha(t)$

- Model 2a: $\alpha(t)$ as 4 iid time dummies
- Model 2b: $\alpha(t)$ as multivariate temporal process

Model 3: $\beta(\mathbf{s},t) = \beta^{(t)}(\mathbf{s})$(independent process at each t)

- Model 3a: common μ_b across t.
- Model 3b: $\mu_b^{(t)}$

Model 4: Separable spatio-temporal model.

See Table 6 Pg. 395

See Table 7 Pg. 395

Generalized Linear Model

$$
\begin{aligned}
f(y(s_i)|\theta(s_i)) &= h(y(s_i)) \exp(\theta(s_i)y(s_i) - b(\theta(s_i))) \\
L(\tilde{\beta}|\boldsymbol{y}) &= \exp\left(\sum y(s_i)X^T(s_i)\tilde{\beta}(s_i) - b(X^T(s_i)\tilde{\beta}(s_i))\right) \\
\theta(s_i) &= X^T(s_i)\tilde{\beta}(s_i) \\
\tilde{\beta} &\sim N(1_{n\times 1} \otimes \mu_\beta, H(\phi) \otimes T)
\end{aligned}
$$

- Constructing a satisfactory sampler is quite challenging.
- More work needs to be done on model fitting side.

Carl Edward Rasmussen.
Gaussian processes for machine learning.
2006.