A Model Selection Study for Spatio-Temporal Change of Support

Andrew M. Raim[1*], Scott H. Holan[2,3], Jonathan R. Bradley[4], and Christopher K. Wikle[2]

[1]Center for Statistical Research and Methodology, U.S. Census Bureau
[2]Department of Statistics, University of Missouri
[3]Office of the Associate Director for Research and Methodology, U.S. Census Bureau
[4]Department of Statistics, Florida State University

Abstract

Spatio-temporal change of support (STCOS) methods are designed for statistical inference and prediction on spatial and/or time domains which differ from the domains on which the data were observed. Bradley, Wikle, and Holan (2015) introduced a parsimonious class of Bayesian hierarchical spatio-temporal models for STCOS for Gaussian data through a motivating application involving the American Community Survey (ACS), an ongoing survey administered by the U.S. Census Bureau that measures key socio-demographic variables for various populations in the United States. Importantly, their methodology provides ACS data users a principled approach to estimating variables of interest, along with associated measures of uncertainty, on customized geographies and/or time periods. In this work, we revisit use of the STCOS methodology to capture median household income at the county level. We use the Deviance Information Criterion (DIC) to study variations in the prior and in basis function specification, and to select a reasonable final model. This work makes use of an R package which is currently under development, whose aim is to make STCOS methodology broadly accessible to federal statistical agencies such as the Census Bureau, the ACS data-user community, and to the general R-user community.

Key Words: American Community Survey; Basis Functions; Bayesian; Change of Support; Moran's I Propagator; Spatio-Temporal.

1. Introduction

The American Community Survey (ACS) is an ongoing survey administered by the U.S. Census Bureau to measure key socioeconomic and demographic variables for the U.S. population. ACS data are available to the public through the American FactFinder (http://factfinder.census.gov) website from year 2005 through the present. Estimates have historically been released for 1-year, 3-year, or 5-year periods; 3-year estimates were discontinued after 2013. At their finest geography, ACS data are released at the census block-group level; however, estimates for an area are suppressed unless the area meets certain criteria. For example, an area must have a population of at least 65,000 for 1-year estimates to be released, but there is no population requirement for 5-year estimates (U.S. Census Bureau, 2016).

The Census Bureau releases yearly ACS estimates on a variety of geographies including states, counties, census tracts, and school districts. Because statistical agencies like the Census Bureau have direct access to the confidential survey microdata, releases for new geographies or period lengths can be prepared as needed. However, data users may be interested in custom geographies or nonstandard time periods, which are not provided by the agency. Spatio-temporal

*U.S. Census Bureau, Washington, DC, 20233, U.S.A. Email: andrew.raim@census.gov.

This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program.

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the NSF or the U.S. Census Bureau.

change of support (STCOS) methodology enables data users to compute model-based estimates for custom geographies and time periods using available ACS releases. The STCOS methodology developed in Bradley et al. (2015) makes use of spatio-temporal dependencies in direct survey estimates (and also incorporates associated direct variance estimates) through a Bayesian hierarchical model and provides estimates, predictions, and appropriate measures of uncertainty. STCOS methodology is not limited for use in applications involving ACS data, but was developed specifically with ACS in mind. See Gotway and Young (2002) Bradley et al. (2015), and the references therein for a review of the change of support literature.

The STCOS problem can be motivated graphically as follows. We will take median household income as our variable of interest throughout this article. Estimates for median household income are available at the county level for all publicly available 1-year and 5-year ACS releases, and 3-year releases prior to 2013. Figure 1 shows both direct estimates and associated variance estimates for the continental U.S.[1] We notice that 1-year and 3-year estimates have been suppressed for many counties. Each of the geographies containing direct estimates, which will be used to train the STCOS model, are referred to as *source supports*. Suppose there is interest in producing model-based estimates of median household income on the geography of congressional districts.[2] Geographies on which we want to produce estimates and predictions are referred to as *target supports*. Congressional districts as defined in the year 2015 are plotted in Figure 2; there are 433 total districts in the continental U.S. We emphasize that counties and congressional districts do not necessarily align, and the crux of the STCOS problem is to "translate" between the county-level observations and the congressional districts. The final type of support is the *fine-level support*, which for this example will be taken as the complete set of counties shown in Figure 2. The STCOS methodology works by translating each of the source supports to the fine-level support during training. Once the model is trained, estimates and predictions on target supports are obtained by translating from the fine-level support.

The STCOS model has already been proposed and fully specified by Bradley et al. (2015), but requires significant expertise (and time) to implement, which many potential users will not have. This article is part of an initiative to produce a user-friendly and efficient R package for the STCOS model. The software will take source and fine-level supports as input, fit the Bayesian hierarchical model via Gibbs sampling, and produce estimates and predictions on target supports of interest. The package is nearing completion, and we use this article as an opportunity to present a large-scale data analysis with a prototype version of the software. Using all available county-level ACS source supports for median household income, we carry out a model selection to choose from several prior covariance structures as well as some options for basis functions. A model is selected using the Deviance Information Criterion (Spiegelhalter et al., 2002), and some preliminary model-based results are shown on county and congressional district geographies.

The rest of the paper will proceed as follows. Section 2 describes how the public ACS data were obtained from the American FactFinder website for use in our study. Section 3 reviews the STCOS model and the Markov-Chain Monte Carlo (MCMC) algorithm (a Gibbs sampler). Section 4 documents the model selection study. Results based on the selected model are shown in Section 5. Finally, Section 6 concludes the paper.

[1] Alaska, Hawaii, and other U.S. territories have been excluded from our study to facilitate display of graphical results.

[2] Note that the Census Bureau releases ACS estimates on congressional districts. This geography serves as an illustrative example of change-of-support methodology in which model-based estimates can be compared to direct estimates.

Table 1: List of ACS datasets obtained from the American FactFinder website. "CD" is an abbreviation for congressional districts.

2015 ACS 5-year	2015 ACS 1-year	2014 ACS 5-year	2014 ACS 1-year
2013 ACS 5-year	2013 ACS 3-year	2013 ACS 1-year	2012 ACS 5-year
2012 ACS 3-year	2012 ACS 1-year	2011 ACS 5-year	2011 ACS 3-year
2011 ACS 1-year	2010 ACS 5-year	2010 ACS 3-year	2010 ACS 1-year
2009 ACS 5-year	2009 ACS 3-year	2009 ACS 1-year	2008 ACS 3-year
2008 ACS 1-year	2007 ACS 3-year	2007 ACS 1-year	2006 ACS
2005 ACS	2015 CD 1-year	2015 CD 5-year	

2. Preparation of ACS Data

The data used in this report were obtained from the American FactFinder website (http://factfinder.census.gov). Data were obtained using the "Advanced Search" option, entering "S1901" as the table name and selecting "County — 050" as the geography level. Comma separated value (CSV) files containing estimates and margins of error for household-level median income (in U.S. dollars) then become available for download. Through related menus in the Advanced Search section, we also obtained shapefiles corresponding to each of these datasets. The margin of error (MOE) represents half the width of the confidence interval $\bar{y} \pm z_{\alpha/2} \cdot \text{SE}$ with $\alpha = 0.1$; i.e., $\text{MOE} = 1.645 \cdot \text{SE}$. Because STCOS methodology requires a variance estimate, we make the transformation $\text{VAR} = \text{MOE}^2/1.645^2$. Congressional district data are obtained similarly as county-level data, except that "Congressional District — 500" is selected as the geography level. The 114th Congress contains years 2014 and 2015. Table 1 lists datasets from the American FactFinder website where are used in the remainder of the paper.

Shapefiles for the source, target, and fine-level supports must be based on the same geographic projection so that they are compatible for use in the model. We have ensured that source and target shapefiles match the projection used in the fine-level shapefile. We have selected the fine-level shapefile to be the one corresponding to the 2015 5-year estimates.

3. Review of the STCOS Model

Let $\mathcal{T} = \{T_L, \ldots, T_U\}$ denote times for which direct estimates are available and $\mathcal{L}$ denote the set of all possible time periods. For ACS data, $\mathcal{T}$ consists of the years 2005 through 2015[3] and $\mathcal{L} = \{1, 3, 5\}$ correspond to 1-year, 3-year, and 5-year period releases. Data may not have been released for all $(t, \ell) \in \mathcal{T} \times \mathcal{L}$; for example, 3-year estimates were discontinued after 2013. Let $(\mathcal{T} \times \mathcal{L})^*$ denote the subset of $\mathcal{T} \times \mathcal{L}$ that does correspond to a release. For each $(t, \ell) \in (\mathcal{T} \times \mathcal{L})^*$, the associated source support $D_{t\ell}$ is a collection of areas. For each area $A \in D_{t\ell}$, $Z_t^{(\ell)}(A)$ and $\sigma_{t\ell}^2(A)$ are the direct survey estimate and associated variance for the survey variable of interest. We will let $D_B = \{B_1, \ldots, B_{n_B}\}$ denote the fine level support.

The STCOS model is a Bayesian hierarchical model which can be described at a high level in three parts:

1. The data model is

$$Z_t^{(\ell)}(A) = Y_t^{(\ell)}(A) + \varepsilon_t^{(\ell)}(A), \quad \varepsilon_t^{(\ell)}(A) \overset{\text{ind}}{\sim} \text{N}(0, \sigma_{t\ell}^2(A)),$$

for $A \in D_{t\ell}$ and $(t, \ell) \in (\mathcal{T} \times \mathcal{L})^*$.

[3] 2016 ACS estimates became available during preparation of this article and are therefore not included in the study.

2. The process model is

$$Y_t^{(\ell)}(A) = \boldsymbol{h}(A)^\top \boldsymbol{\mu}_B + \boldsymbol{\psi}_t^{(\ell)}(A)^\top \boldsymbol{\eta} + \xi_t^{(\ell)}(A),$$
$$\boldsymbol{\eta} \sim \mathrm{N}(\boldsymbol{0}, \sigma_K^2 \boldsymbol{K}), \quad \xi_t^{(\ell)}(A) \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\xi^2),$$

for $A \in D_{t\ell}$ and $(t, \ell) \in (\mathcal{T} \times \mathcal{L})^*$.

3. The parameter model is

$$\boldsymbol{\mu}_B \sim \mathrm{N}(\boldsymbol{0}, \sigma_\mu^2 \boldsymbol{I}), \quad \sigma_\mu^2 \sim \mathrm{IG}(a_\mu, b_\mu), \quad \sigma_K^2 \sim \mathrm{IG}(a_K, b_K), \quad \sigma_\xi^2 \sim \mathrm{IG}(a_\xi, b_\xi).$$

Here we assume that direct estimates $Z_t^{(\ell)}(A)$ are a noisy observation of a latent process $Y_t^{(\ell)}(A)$. The variance of the noise $\varepsilon_t^{(\ell)}(A)$ is assumed to be $\sigma_{t\ell}^2(A)$, the variance of the direct estimate. The mean of the latent process $Y_t^{(\ell)}(A)$ consists of a coarse spatial trend $\boldsymbol{h}(A)^\top \boldsymbol{\mu}_B$ and a spatio-temporal random process $\boldsymbol{\psi}_t^{(\ell)}(A)^\top \boldsymbol{\eta}$. Conjugate priors are assumed for the coefficients and variance parameters from the previous two stages. The matrix $\boldsymbol{K}$, which appears in the prior covariance of $\boldsymbol{\eta}$, is assumed to be known and computable from the fine-level support. More explanation on each of these terms is given below.

The latent process model is motivated by the following construction. Define a continuous-space discrete-time process on $\boldsymbol{u} \in \bigcup_{i=1}^{n_B} B_i$, $t \in \mathcal{T}$,

$$Y(\boldsymbol{u}; t) = \delta(\boldsymbol{u}) + \sum_{j=1}^{\infty} \psi_j(\boldsymbol{u}; t) \cdot \eta_j,$$

where $\delta(\boldsymbol{u})$ is a large-scale spatial trend process and $\{\psi_j(\boldsymbol{u}, t)\}_{j=1}^{\infty}$ is a prespecified set of spatio-temporal basis functions. Integrating $Y(\boldsymbol{u}; t)$ uniformly over $u \in A$ and an ℓ-year period,

$$\begin{aligned} Y_t^{(\ell)}(A) = & \underbrace{\frac{1}{|A|} \int_A \delta(\boldsymbol{u})\, d\boldsymbol{u}}_{\text{large-scale spatial trend}} + \underbrace{\frac{1}{\ell |A|} \sum_{k=t-\ell+1}^{t} \sum_{j=1}^{r} \int_A \psi_j(\boldsymbol{u}; k) \cdot \eta_j \, d\boldsymbol{u}}_{\text{spatio-temporal random process}} \\ & + \underbrace{\frac{1}{\ell |A|} \sum_{k=t-\ell+1}^{t} \sum_{j=r+1}^{\infty} \int_A \psi_j(\boldsymbol{u}; k) \cdot \eta_j \, d\boldsymbol{u}}_{\text{remainder}} \\ = & \mu(A) + \boldsymbol{\psi}_t^{(\ell)}(A)^\top \boldsymbol{\eta} + \xi_t^{(\ell)}(A). \end{aligned}$$

For the remainder, we assume that $\xi_t^{(\ell)}(A) \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\xi^2)$. We make use of local bisquare basis functions for the small-scale spatio-temporal trend, which are of the form

$$\psi_j(\boldsymbol{u}, t) = \left[1 - \frac{\|\boldsymbol{u} - \boldsymbol{c}_j\|^2}{w_s^2} - \frac{|t - g_t|^2}{w_t^2}\right]^2 \cdot I(\|\boldsymbol{u} - \boldsymbol{c}_j\| \le w_s) \cdot I(|t - g_t| \le w_t).$$

These functions require specification of the number and location of spatial knot points $\boldsymbol{c}_j$, $j = 1, \ldots, r_{\text{space}}$, the number and location of temporal knot points g_t, $t = 1, \ldots, r_{\text{time}}$, the spatial radius w_s, and the temporal radius w_t. Some possibilities are investigated in Section 4.

Once the basis function is fully specified, we must compute it at the area level. For area A and an ℓ-year period, we take a Monte Carlo approximation

$$\psi_{jt}^{(\ell)}(A) \approx \frac{1}{\ell Q} \sum_{k=t-\ell+1}^{t} \sum_{q=1}^{Q} \psi_j(\boldsymbol{u}_q, k),$$

using a uniform random sample $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_Q$ on A.

Next, for the large-scale spatial trend process, we make the simplifying assumption that

$$\delta(u) = \sum_{i=1}^{n_B} \mu_i I(u \in A \cap B_i),$$

for a given area A. Then $\delta(u)$ takes on a constant value on each overlap $A \cap B_i$ for $B_i \in D_B$. Now, integrating over $u \in A$,

$$\mu(A) = \frac{1}{|A|} \sum_{i=1}^{n_B} \int_{A \cap B_i} \delta(u) du = \frac{1}{|A|} \sum_{i=1}^{n_B} \mu_i \int_{A \cap B_i} du = \sum_{i=1}^{n_B} \mu_i \frac{|A \cap B_i|}{|A|} = \boldsymbol{h}(A)^\top \boldsymbol{\mu}_B.$$

where we have defined

$$\boldsymbol{h}(A) = (|A \cap B_1|/|A|, \ldots, |A \cap B_{n_B}|/|A|)^\top.$$

The vector $\boldsymbol{h}(A)$ is computed from the source and fine-level supports and is therefore a known quantity in the analysis. For example, if the geographies are specified through shapefiles, modern libraries such as the `R` package `sf` (Pebesma, 2017) can readily compute overlaps between areas. The coefficient

$$\boldsymbol{\mu}_B = (\mu_1, \ldots, \mu_{n_B})^\top$$

is unknown, and must be estimated in the analysis. Note that $\boldsymbol{\mu}_B$ represents the change-of-support coefficient between the fine-level support and all other supports, and is the primary quantity of interest in the model. To simplify the remaining presentation, we may now write the model in vector form. Denoting $\mathcal{I} = ((A, t, \ell) : A \in \mathcal{D}_{t\ell}, (t, \ell) \in (\mathcal{T} \times \mathcal{L})^*)$ as the list of triples (A, t, ℓ) from the source supports taken in a particular order, we may write

$$\begin{aligned}
\boldsymbol{Z} &= \text{vec}\left(Z_t^{(\ell)}(A) : (A, t, \ell) \in \mathcal{I}\right), & \boldsymbol{H} &= \text{rbind}\left(\boldsymbol{h}_t^{(\ell)}(A)^T : (A, t, \ell) \in \mathcal{I}\right), \\
\boldsymbol{S} &= \text{rbind}\left(\boldsymbol{\psi}_t^{(\ell)}(A)^T : (A, t, \ell) \in \mathcal{I}\right), & \boldsymbol{\xi} &= \text{vec}\left(\xi_t^{(\ell)}(A) : (A, t, \ell) \in \mathcal{I}\right), \\
\boldsymbol{\varepsilon} &= \text{vec}\left(\varepsilon_t^{(\ell)}(A) : (A, t, \ell) \in \mathcal{I}\right), & \boldsymbol{V} &= \text{diag}\left(\sigma_{t\ell}^2(A) : (A, t, \ell) \in \mathcal{I}\right),
\end{aligned} \tag{3.1}$$

and $\boldsymbol{h}_t^{(\ell)}(A) \equiv \boldsymbol{h}(A)$. The notation $\text{vec}(\mathcal{S})$ is used to mean that a vector is constructed from the elements of $\mathcal{S}$, while $\text{diag}(\mathcal{S})$ represents a matrix whose diagonal elements consist of $\mathcal{S}$, and $\text{rbind}(\mathcal{S})$ represents a matrix with the elements of $\mathcal{S}$ as rows. The model can now be written

$$\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{\mu}_B + \boldsymbol{S}\boldsymbol{\eta} + \boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \text{N}(0, \boldsymbol{V}).$$

We require a matrix $\boldsymbol{K}$ to complete specification of the model. Suppose the fine-level support behaves according to the process

$$\begin{aligned}
\boldsymbol{Y}_t^* &= \boldsymbol{\mu}_B + \boldsymbol{\nu}_t, \\
\boldsymbol{\nu}_t &= \boldsymbol{M}\boldsymbol{\nu}_{t-1} + \boldsymbol{b}_t, \\
\boldsymbol{b}_t &\overset{\text{iid}}{\sim} \text{N}(\boldsymbol{0}, \sigma_K^2(\boldsymbol{I} - \boldsymbol{\mathcal{A}})^-),
\end{aligned} \tag{3.2}$$

for $t \in \mathcal{T}$, where $\boldsymbol{\mathcal{A}}$ is the adjacency matrix of D_B and $\boldsymbol{X}^-$ represents a generalized inverse of the matrix $\boldsymbol{X}$. That is, $\{\boldsymbol{Y}_t^*\}$ is a vector autoregressive (VAR) process in time and a conditional autoregressive (CAR) process in space. Let $\boldsymbol{\Sigma}_{y^*}$ denote the covariance matrix of $(\boldsymbol{Y}_t^* : t \in \mathcal{T})$

and $\boldsymbol{S}^*$ denote the basis function expansion on the fine-level geography. We obtain $\boldsymbol{K}$ as the minimizer of

$$\|\boldsymbol{\Sigma}_{y^*} - \boldsymbol{S}^* \boldsymbol{C} \boldsymbol{S}^{*\top}\|_{\mathrm{F}}, \quad \text{over } r \times r \text{ positive semidefinite matrices } \boldsymbol{C}, \tag{3.3}$$

using the Frobenius norm $\|\cdot\|_{\mathrm{F}}$. The solution to (3.3),

$$\boldsymbol{C}^* = (\boldsymbol{S}^{*\top}\boldsymbol{S}^*)^{-1}\boldsymbol{S}^{*\top}\boldsymbol{\Sigma}_{y^*}\boldsymbol{S}^*(\boldsymbol{S}^{*\top}\boldsymbol{S}^*)^{-1},$$

provides the best positive approximant to $\boldsymbol{\Sigma}_{y^*}$. Bradley et al. (2015) further discuss this approach within the context of the present model, and Higham (1988) discusses the positive approximant problem in a general setting. We may write $\boldsymbol{\Sigma}_{y^*} = \sigma_K^2 \tilde{\boldsymbol{\Sigma}}_{y^*}$ so that

$$\boldsymbol{C}^* = \sigma_K^2 \boldsymbol{K}, \quad \boldsymbol{K} = (\boldsymbol{S}^{*\top}\boldsymbol{S}^*)^{-1}\boldsymbol{S}^{*\top}\tilde{\boldsymbol{\Sigma}}_{y^*}\boldsymbol{S}^*(\boldsymbol{S}^{*\top}\boldsymbol{S}^*)^{-1} \tag{3.4}$$

and $\tilde{\boldsymbol{\Sigma}}_{y^*}$ and $\boldsymbol{K}$ are free of unknown parameters. We consider three possible structures for $\boldsymbol{K}$:

1. **Independence**: Take $\boldsymbol{K} = \boldsymbol{I}$ to assume no spatial or temporal covariance in $\boldsymbol{\eta}$.
2. **Spatial-only**: Let $\boldsymbol{\Sigma}_{y^*} = \sigma_K^2(\boldsymbol{I} - \boldsymbol{\mathcal{A}})^- \otimes \boldsymbol{I}_{|\mathcal{T}|}$, where $\otimes$ represents the Kronecker product, which assumes no temporal covariance, and take $\boldsymbol{K}$ as in (3.4).
3. **Random Walk**: Take $\boldsymbol{M} = \boldsymbol{I}$ so that the fine-level process defined in (3.2) is a vector random walk with nonstationary autocovariance function

$$\boldsymbol{\Gamma}(t,h) = \begin{cases} t\sigma_K^2(\boldsymbol{I} - \boldsymbol{\mathcal{A}})^- & \text{if } h \geq 0, \\ (t - |h|)\sigma_K^2(\boldsymbol{I} - \boldsymbol{\mathcal{A}})^- & \text{if } -t < h < 0. \end{cases}$$

This yields

$$\boldsymbol{\Sigma}_{y^*} = \begin{bmatrix} \boldsymbol{\Gamma}(1,1) & \boldsymbol{\Gamma}(1,2) & \cdots & \boldsymbol{\Gamma}(1,|\mathcal{T}|) \\ \boldsymbol{\Gamma}(2,1) & \boldsymbol{\Gamma}(2,2) & \cdots & \boldsymbol{\Gamma}(2,|\mathcal{T}|) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}(|\mathcal{T}|,1) & \boldsymbol{\Gamma}(|\mathcal{T}|,2) & \cdots & \boldsymbol{\Gamma}(|\mathcal{T}|,|\mathcal{T}|) \end{bmatrix}$$

as the covariance of $\{\boldsymbol{Y}_t^*\}$, and $\boldsymbol{K}$ is computed using (3.4).

We can derive a Gibbs sampler by considering the joint distribution

$$\begin{aligned} f(\boldsymbol{Z}, \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\mu}_B, \sigma_\mu^2, \sigma_K^2, \sigma_\xi^2) = {} & \mathrm{N}(\boldsymbol{Z} \mid \boldsymbol{H}\boldsymbol{\mu}_B + \boldsymbol{S}\boldsymbol{\eta} + \boldsymbol{\xi}, \boldsymbol{V}) \cdot \mathrm{N}(\boldsymbol{\xi} \mid \boldsymbol{0}, \sigma_\xi^2 \boldsymbol{I}) \cdot \mathrm{N}(\boldsymbol{\eta} \mid 0, \boldsymbol{K}) \\ & \times \mathrm{N}(\boldsymbol{\mu}_B \mid \boldsymbol{0}, \sigma_\mu^2 \boldsymbol{I}) \cdot \mathrm{IG}(\sigma_\mu^2 \mid a_\mu, b_\mu) \cdot \mathrm{IG}(\sigma_K^2 \mid a_K, b_K) \cdot \mathrm{IG}(\sigma_\xi^2 \mid a_\xi, b_\xi). \end{aligned}$$

Determining each of the full conditional distributions of μ_B, η, ξ, σ_μ^2, σ_K^2, and σ_ξ^2, we obtain the following Gibbs steps:

- $[\boldsymbol{\mu}_B \mid \bullet\,] \sim \mathrm{N}(\boldsymbol{\vartheta}_\mu, \boldsymbol{\Omega}_\mu^{-1})$,

$$\boldsymbol{\vartheta}_\mu = \boldsymbol{\Omega}_\mu^{-1}\boldsymbol{H}^\top\boldsymbol{V}^{-1}(\boldsymbol{Z} - \boldsymbol{S}\boldsymbol{\eta} - \boldsymbol{\xi}), \qquad \boldsymbol{\Omega}_\mu = \boldsymbol{H}^\top\boldsymbol{V}^{-1}\boldsymbol{H} + \sigma_\mu^{-2}\boldsymbol{I}.$$

- $[\boldsymbol{\eta} \mid \bullet\,] \sim \mathrm{N}(\boldsymbol{\vartheta}_\eta, \boldsymbol{\Omega}_\eta^{-1})$,

$$\boldsymbol{\vartheta}_\eta = \boldsymbol{\Omega}_\eta^{-1}\boldsymbol{S}^\top\boldsymbol{V}^{-1}(\boldsymbol{Z} - \boldsymbol{H}\boldsymbol{\mu}_B - \boldsymbol{\xi}), \quad \boldsymbol{\Omega}_\eta = \boldsymbol{S}^\top\boldsymbol{V}^{-1}\boldsymbol{S} + \sigma_K^{-2}\boldsymbol{K}^{-1}.$$

- $[\boldsymbol{\xi} \mid \bullet\,] \sim \mathrm{N}(\boldsymbol{\vartheta}_\xi, \boldsymbol{\Omega}_\xi^{-1})$,

$$\boldsymbol{\vartheta}_\xi = \boldsymbol{\Omega}_\xi\boldsymbol{V}^{-1}(\boldsymbol{Z} - \boldsymbol{H}\boldsymbol{\mu}_B - \boldsymbol{S}\boldsymbol{\eta}), \quad \boldsymbol{\Omega}_\xi^{-1} = \boldsymbol{V}^{-1} + \sigma_\xi^{-2}\boldsymbol{I}.$$

- $[\sigma_\mu^2 \mid \bullet\,] \sim \mathrm{IG}(\alpha_\mu, \beta_\mu)$, $\alpha_\mu = a_\mu + n_B/2$ and $\beta_\mu = b_\mu + \boldsymbol{\mu}_B^\top\boldsymbol{\mu}_B/2$.
- $[\sigma_K^2 \mid \bullet\,] \sim \mathrm{IG}(\alpha_K, \beta_K)$, $\alpha_K = a_K + r/2$ and $\beta_K = b_K + \boldsymbol{\eta}^\top\boldsymbol{K}^{-1}\boldsymbol{\eta}/2$.
- $[\sigma_\xi^2 \mid \bullet\,] \sim \mathrm{IG}(\alpha_\xi, \beta_\xi)$, $\alpha_\xi = a_\xi + N/2$ and $\beta_\xi = b_\xi + \boldsymbol{\xi}^\top\boldsymbol{\xi}/2$.

4. Model Selection

In order to apply the STCOS methodology, the analyst must make several modeling choices including the prior covariance structure for $\boldsymbol{K}$, the number of knot points used to define the basis functions, and the radius parameters in the basis functions. We carried out a model selection study using Deviance Information Criterion (DIC) to find appropriate choices. We label the three prior covariance structures discussed in Section 3, Independence, Spatial-only, and Spatial with Random-Walk, as IND, SP and RW respectively. Temporal knot points are fixed to be 2005, 2005.5, ..., 2014.5, 2015, so that there are 21 equally spaced points through $\mathcal{T}$ at half-year intervals. We take the temporal radius parameter $w_t = 1$ throughout the study. We select the spatial knot points via a space-filling design on D_B, which is implemented in the `R fields` package (Nychka et al., 2015). The spatial radius parameter w_s must be selected so that it is compatible with the projection used in the shapefiles of the supports; for example, $w_s = 1$ has a very different effect if shapefile distances are specified in meters than in kilometers. Therefore, we take a data-driven approach. Based on the selection of knot points $\{\boldsymbol{c}_{jt}\}$, where $\boldsymbol{c}_{jt} = (\boldsymbol{c}_j, g_t)$ for $j = 1, \ldots, r_{\text{space}}$ and $t = 1, \ldots, r_{\text{time}}$, we compute the distance matrix of $\{\boldsymbol{c}_{jt}\}$ and let $Q_{0.05}$ be the 0.05 quantile of the nonzero upper-triangular entries of the matrix. We then take $w_s = \tau_s \cdot Q_{0.05}$, where τ_s is a multiplier selected by the analyst. In our DIC study, we consider $\tau_s \in \{0.5, 1.0\}$ and $r_{\text{space}} \in \{250, 500\}$. Intuitively, a larger τ_s will yield more smoothing of spatial information in the basis, and more knot points will make use of finer geographical features. Figure 3 plots spatial knot points for the four combinations of τ_s and r_{space} used in this study. For each of these combinations, it can be seen that each point in the domain is within the radius of multiple spatial knot points.

Multicollinearity may be present in the columns of $\boldsymbol{S}$, which will become more pronounced as the number of knot points are increased. In our experience, this multicollinearity can cause severely slow convergence of the MCMC sampler. We therefore reduce the $n \times r$ matrix $\boldsymbol{S}$ using principal components analysis (PCA). Suppose $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\top$ is the eigendecomposition of $\boldsymbol{S}^\top\boldsymbol{S}$, and $\tilde{\boldsymbol{U}}$ contains the $\tilde{r}$ columns of $\boldsymbol{U}$ corresponding to the $\tilde{r} \leq r$ largest magnitude eigenvalues in $\boldsymbol{D}$. The transformation $T(\boldsymbol{S}) = \boldsymbol{S}\tilde{\boldsymbol{U}}^\top$ is then applied to obtain a reduced $\boldsymbol{S}$.[4] In our model selection study, we consider taking $\tilde{r}$ corresponding to 60%, 75%, and 90% of the variation in the eigenvalues.

To summarize, our exploratory model selection exercise considers four factors: the prior covariance structures IND, SP and RW, $\tau_s \in \{0.5, 1.0\}$, $r_{\text{space}} \in \{250, 500\}$, and eigenvalue proportions 60%, 75%, and 90%. For each combination of factors, we prepare the terms of the STCOS model and run the Gibbs sampler for 2000 iterations. We discard the first 500 iterations as a burn-in period and save every 10th remaining iteration. Thinning is useful to access convergence for this model because of the large storage required for $\boldsymbol{\mu}_B$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$. The maximum likelihood estimator (MLE) has been used as the initial value of the sampler in all cases. DIC is computed using the saved draws from MCMC sampling; smaller values of DIC indicate better fitting models. It is also important to check convergence of the Gibbs sampler; we have visually examined trace plots of the sampled chains (not shown) and detected no lack of convergence. Although only 2000 draws were used in each run (in the interest of completing the study in a reasonable amount of time) the resulting chains appeared adequate to compute DIC.

Table 2 presents our model selection results. Displayed DIC values are negative because density values evaluated at most observations were larger than 1. The table values indicate that the priors do not produce vastly different results, but the best model under IND achieves a slightly better DIC than the best model under the other options. It appears that $\tau_s = 0.5$ is

[4] Note that the same transformation $T(\boldsymbol{S})$ must also be applied to $\boldsymbol{S}^*$ in the computation of $\boldsymbol{K}$.

too small, and $\tau_s = 1.0$ provides a better fit in all cases. Likewise, taking more eigenvalues of $\boldsymbol{S}$ makes a large improvement to the fit. However, MCMC diagnostics suffered noticeably when using 90% of the eigenvalues and only 60% and 75% levels are considered further. Increasing the number of spatial knot points from 250 to 500 usually improved the fit. However, in some cases such as row 4 $(\text{IND}, 250, 1.0, 0.60)$ versus row 10 $(\text{IND}, 500, 1.0, 0.60)$, the fit became worse with more knot points; this may be due to too much reduction on the dimension of $\boldsymbol{S}$. After examining Table 2, we selected the model in row 11 $(\text{IND}, 500, 1.0, 0.75)$.

5. Results

Using the selected model from Section 4, we ran a longer MCMC with 10,000 iterations, discarding the first 1,000 as a burn-in period, saving every 10th remaining iteration, and taking the MLE as the initial value. Visual inspection of trace plots was used to assess mixing of the sampled chains, with no lack of convergence detected. We are primarily interested in draws of the mean

$$\mathrm{E}(Y_t^{(\ell)}(A) \mid \boldsymbol{\mu}_B, \boldsymbol{\eta}) = \boldsymbol{h}(A)^\top \boldsymbol{\mu}_B + \boldsymbol{\psi}_t^{(\ell)}(A)^\top \boldsymbol{\eta}, \tag{5.1}$$

where A is an area of interest (not necessarily in the source supports), or draws from the posterior predictive distribution

$$[Y_t^{(\ell)}(A) \mid \boldsymbol{\mu}_B, \boldsymbol{\eta}, \sigma_\xi^2] \sim \mathrm{N}\left(\boldsymbol{h}(A)^\top \boldsymbol{\mu}_B + \boldsymbol{\psi}_t^{(\ell)}(A)^\top \boldsymbol{\eta}, \sigma_\xi^2\right) \tag{5.2}$$

using draws from the posterior distribution based on the observed $\boldsymbol{Z}$ and $\boldsymbol{V}$. The posterior predictive distribution (5.2) generally produces larger variability than the posterior mean (5.1). We take the sample mean of MCMC draws from either distribution as a point estimate, and the sample standard deviation (SD) to measure variability in the respective distribution.

Figures 4 and 5 compare county-level direct estimates and SDs with model-based versions using (5.1); Figure 4 displays 1-year estimates while Figure 5 displays 5-year estimates. The model appears to be effective in capturing spatial patterns in both 1-year and 5-year estimates, in areas where direct estimates have been released. In most areas, the model-based SD is substantially smaller than the direct SD, which reflects the model's increased precision as it is able to borrow strength across the history of ACS releases. Figure 6 compares direct and model estimates on the geography of congressional districts. The results here do not match as well as they do for counties, but some general patterns in the direct estimates can be seen in the model-based estimates. For example, districts along the east coast are seen to have larger incomes than surrounding areas. Figures 7 and 8 present another comparison of the county-level 5-year estimates and congressional district 1-year estimates. Scatter plots in Figure 7 show that there is a high correlation between direct and model-based estimates for congressional districts, although not as ideal as the correlation at the county level. Figure 8 plots the direct estimates as a series in increasing order and shows the model-based estimates as a separate series. This presents another view of the increased variability in the congressional district estimates, and also shows that we have systematically underestimated larger direct estimates to an extent.

6. Conclusions

In this article, we have presented a small study using an R package which is being developed for STCOS methodology. The study led to several interesting conclusions. The choice of prior for $\boldsymbol{K}$—the covariance of the coefficient for the spatio-temporal random process—did not appear to have a major impact on model fit. It is usually desirable in a data analysis that results

should not change drastically with the choice of prior. The independence prior represents the simplest possibility, and avoids computation needed for other choices. We also found that a high dimensional use of the bisquare basis without adequate dimension reduction led to poor MCMC mixing. Taking the spatial radius term in the bisquare basis as a function of the geography—the 0.05 quantile of pairwise distances between knot points—appeared to be a reasonable choice, allowing each point in the geography to interact with multiple knot points. The model produced estimates on the source support geographies (counties) which were very close to the direct estimates. We have observed similarly good performance in estimating a source support when that particular support was left out of the training set (those results were not shown here). The model was also generally able to capture trends of direct estimates on the target support of congressional districts, but not as well as on the source supports. Further investigation of congressional district results is warranted as improvements may be possible.

We are currently in the process of completing the R package for public release. Details on the software and its use will be published in the near future.

References

Jonathan R. Bradley, Christopher K. Wikle, and Scott H. Holan. Spatio-temporal change of support with application to American Community Survey multi-year period estimates. *Stat*, 4(1):255–270, 2015.

Carol A. Gotway and Linda J. Young. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648, 2002.

Nicholas J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988.

Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain. *fields: Tools for spatial data*. University Corporation for Atmospheric Research, Boulder, CO, USA, 2015. URL www.image.ucar.edu/fields. R package version 9.0.

Edzer Pebesma. *sf: Simple Features for R*, 2017. URL https://CRAN.R-project.org/package=sf. R package version 0.5-1.

David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64 (4):583–639, 2002.

U.S. Census Bureau. American Community Survey data suppression, September 2016. URL https://www.census.gov/programs-surveys/acs/technical-documentation/data-suppression.html.

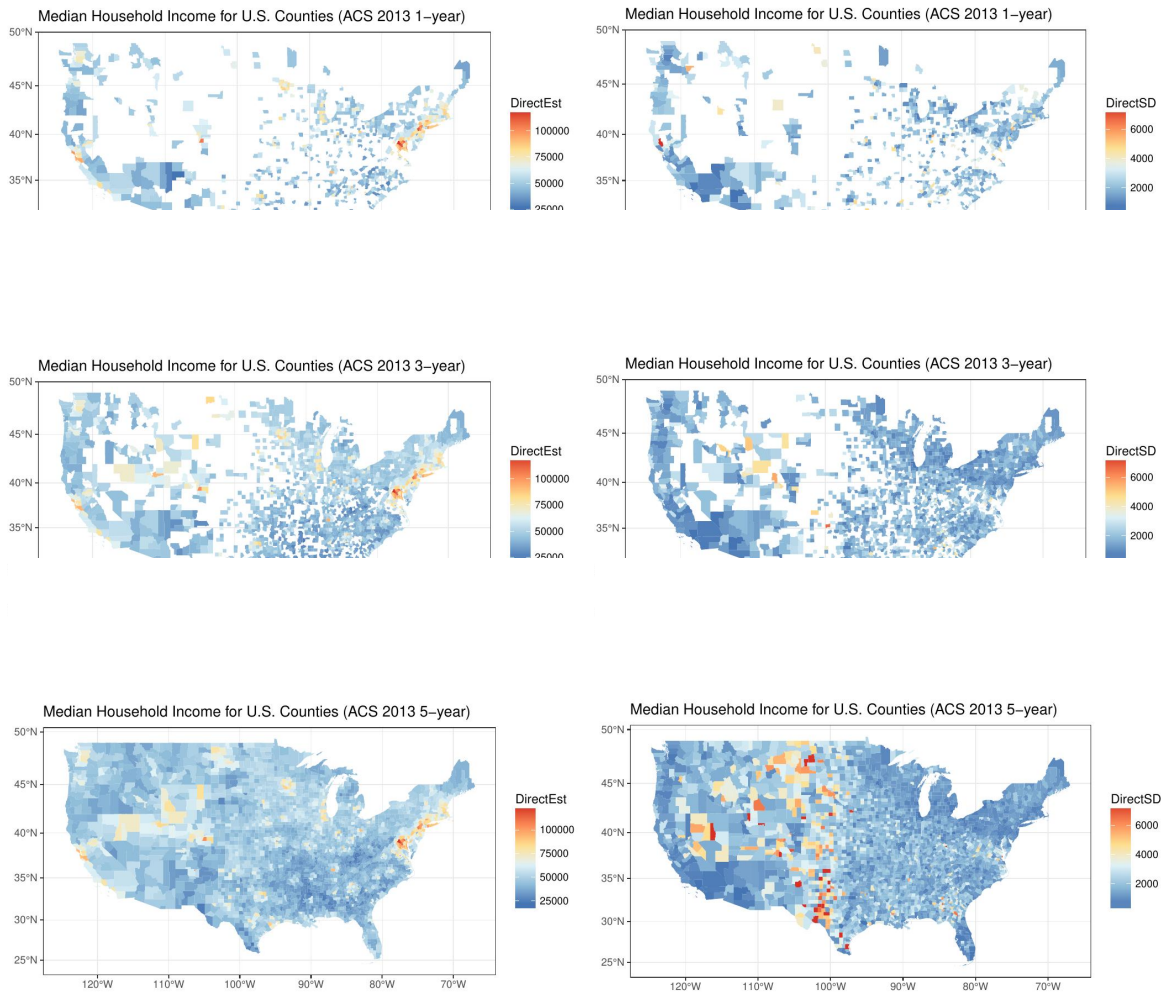


Figure 1: County-level ACS data for median household income in the year 2013. The left column shows direct estimates and the right column displays SDs. The first, second, and third rows correspond to 1-year, 3-year, and 5-year period estimates, respectively.

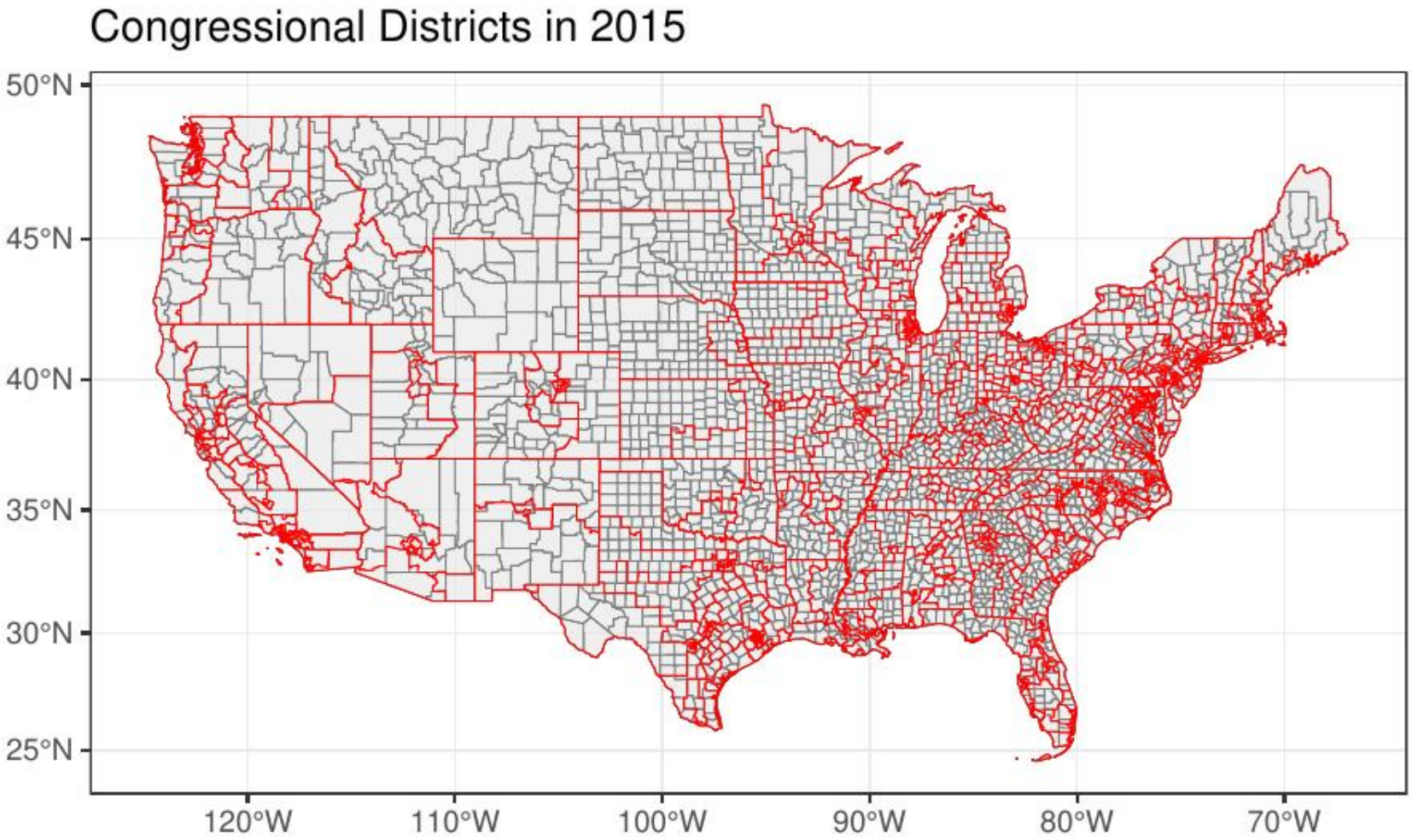
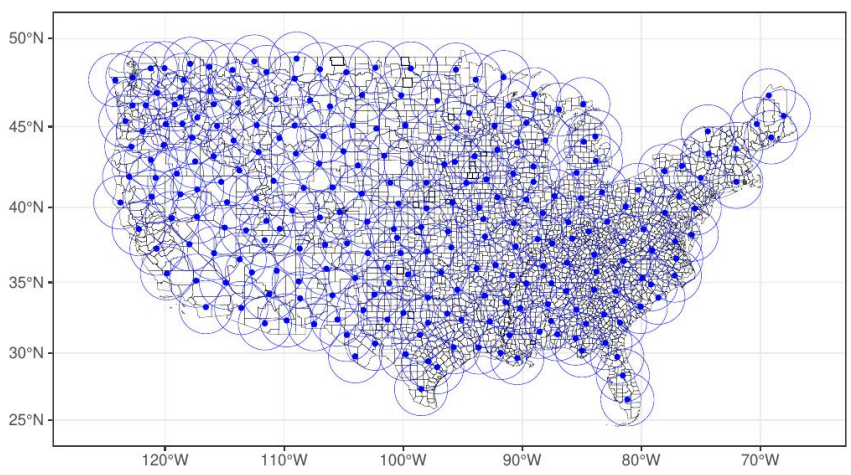
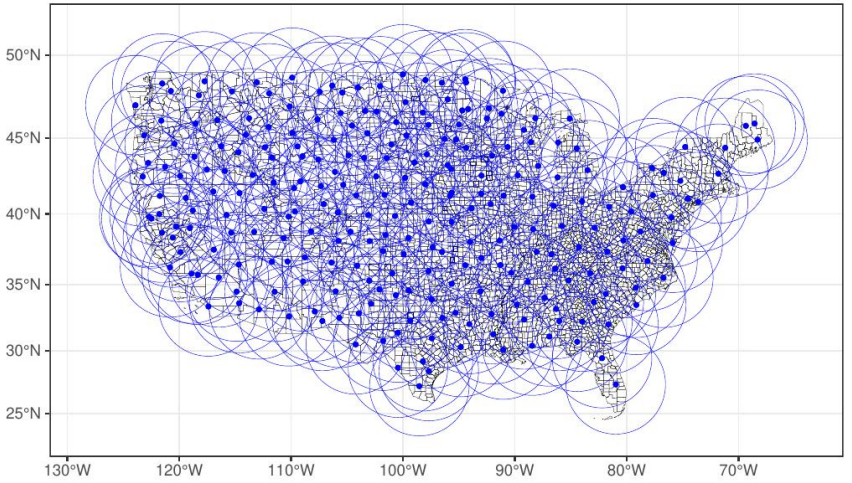


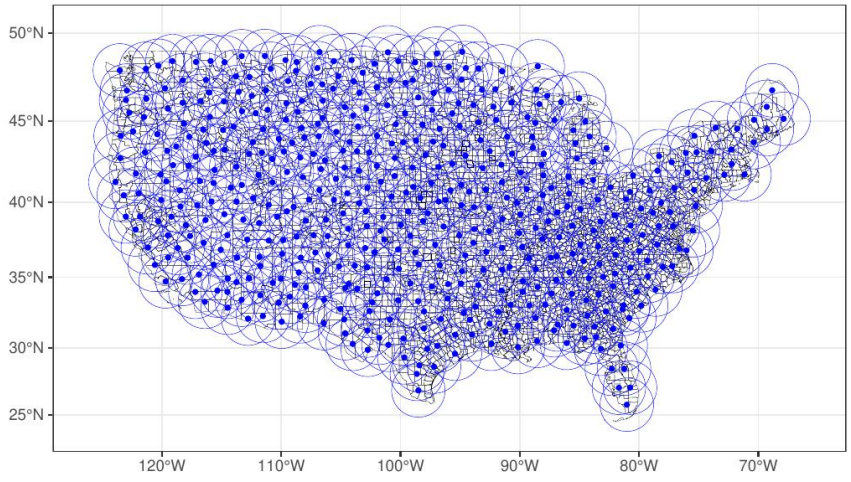
Figure 2: Congressional district geography overlaid on counties for the year 2015.



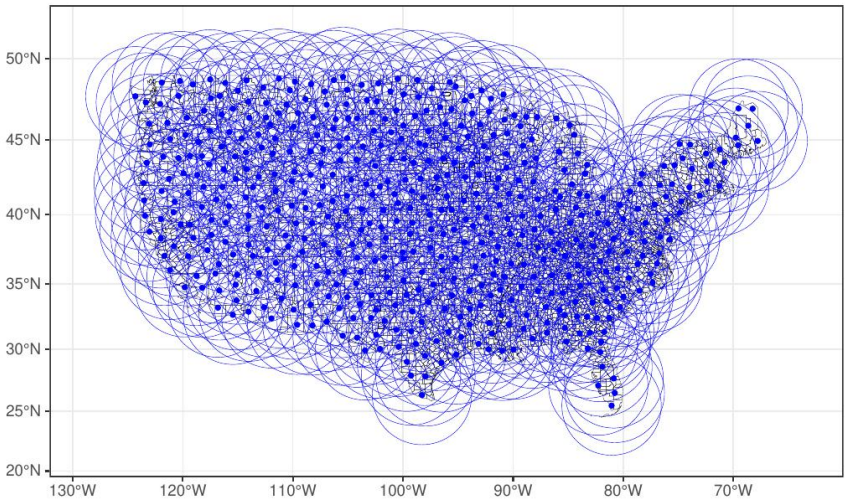
(a) 250 spatial knot points, $w_s = 0.5$



(b) 250 spatial knot points, $w_s = 1.0$



(c) 500 spatial knot points, $w_s = 0.5$



(d) 500 spatial knot points, $w_s = 1.0$

Figure 3: Space-filling designs for basis functions using 250 or 500 spatial knot points and radius $\tau_s \in \{0.5, 1.0\}$.

Table 2: Model selection by DIC.

	Prior	r_{space}	τ_s	Eigenvalue %	DIC
1	IND	250	0.5	60	-23213.99
2			0.5	75	-28011.33
3			0.5	90	-32298.14
4			1.0	60	-25167.15
5			1.0	75	-29415.06
6			1.0	90	-33951.35
7		500	0.5	60	-23747.65
8			0.5	75	-29870.56
9			0.5	90	-32989.99
10			1.0	60	-23535.65
11			1.0	75	-30653.63
12			1.0	90	-34533.95
13	SP	250	0.5	60	-23213.42
14			0.5	75	-28001.35
15			0.5	90	-32235.30
16			1.0	60	-25167.26
17			1.0	75	-29425.25
18			1.0	90	-33918.53
19		500	0.5	60	-23745.41
20			0.5	75	-29853.79
21			0.5	90	-32799.24
22			1.0	60	-23536.62
23			1.0	75	-30652.96
24			1.0	90	-34453.51
25	RW	250	0.5	60	-23208.09
26			0.5	75	-27994.73
27			0.5	90	-32306.21
28			1.0	60	-25165.68
29			1.0	75	-29407.43
30			1.0	90	-33903.14
31		500	0.5	60	-23738.00
32			0.5	75	-29845.44
33			0.5	90	-33104.12
34			1.0	60	-23533.74
35			1.0	75	-30652.96
36			1.0	90	-34465.37

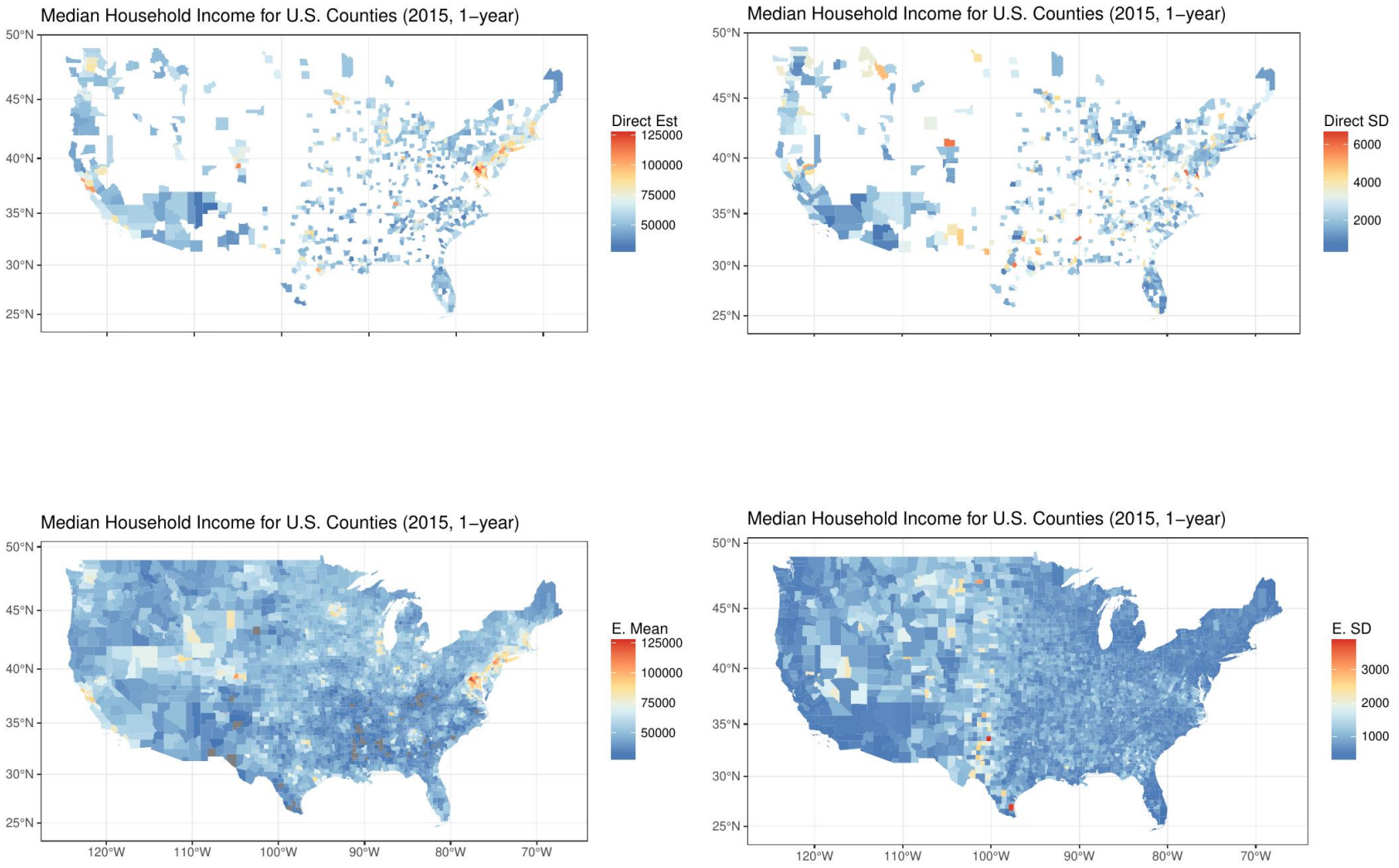


Figure 4: Direct and model-based estimates for 1-year 2015 counties. The top plots display direct estimates and SDs from ACS. The bottom plots show model-based estimates based on the posterior mean and SD of $\mathrm{E}(Y_t^{(\ell)}(A))$.

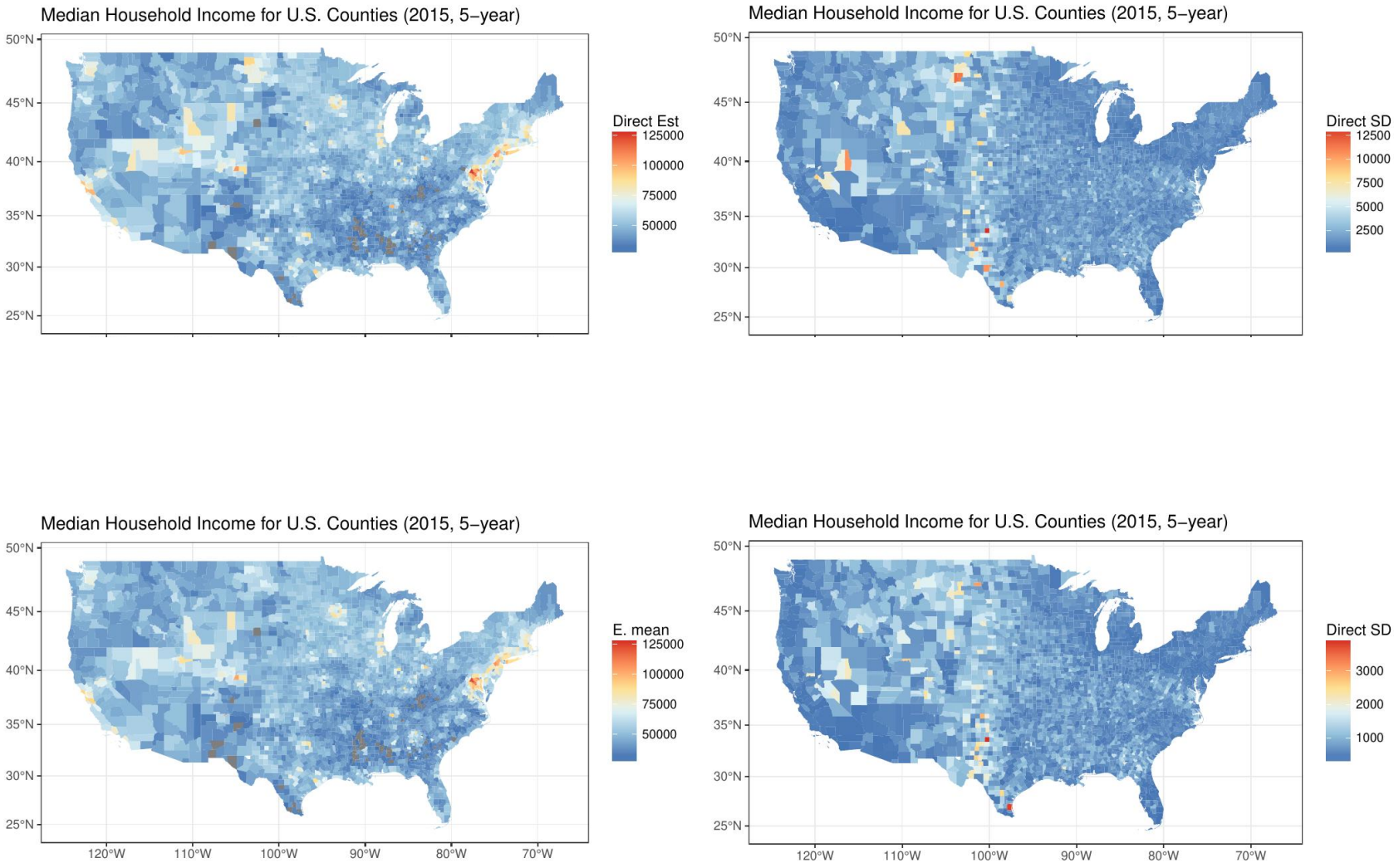


Figure 5: Direct and model-based estimates for 5-year 2015 counties. The top plots display direct estimates and SDs from ACS. The bottom plots show model-based estimates based on the posterior mean and SD of $\mathrm{E}(Y_t^{(\ell)}(A))$.

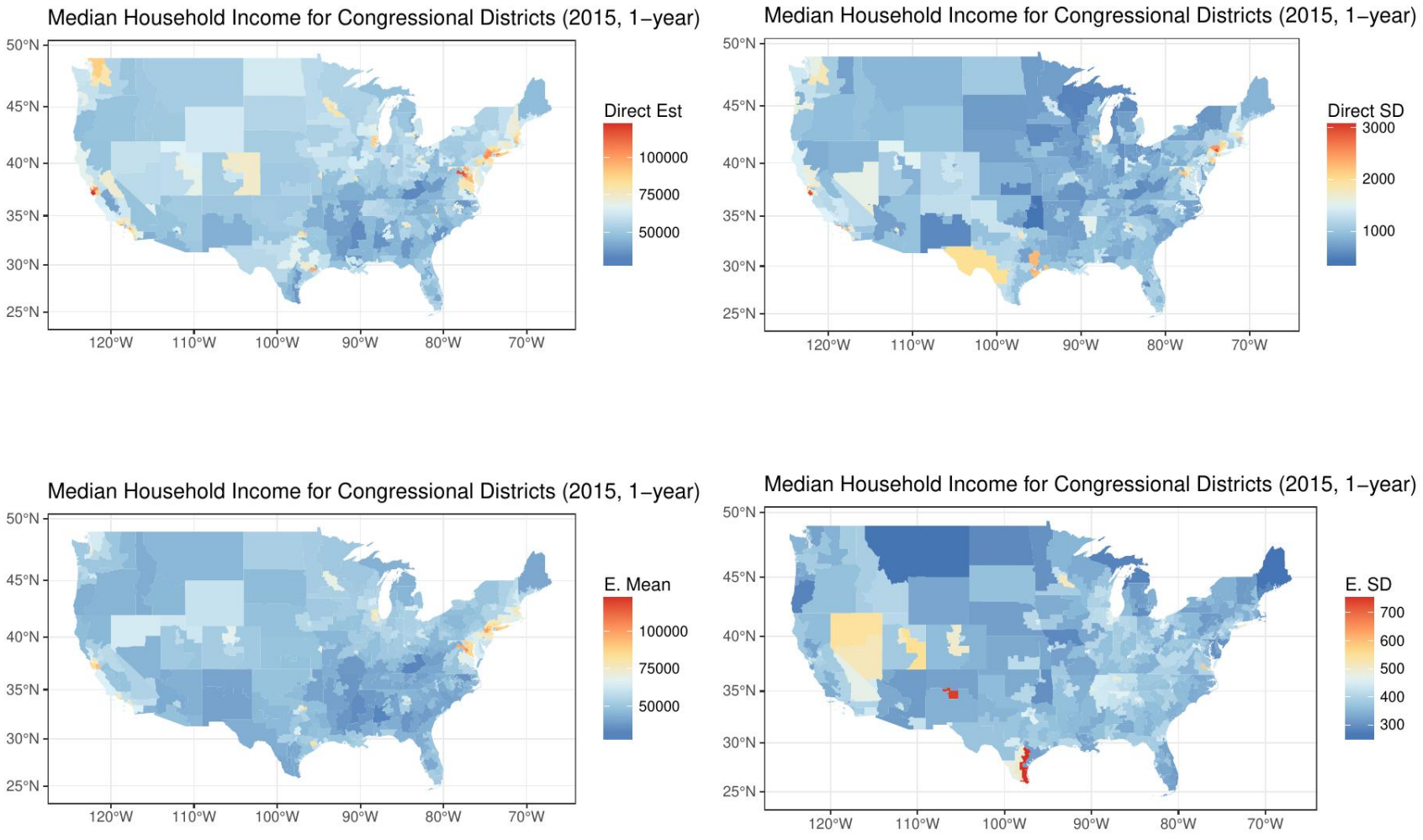
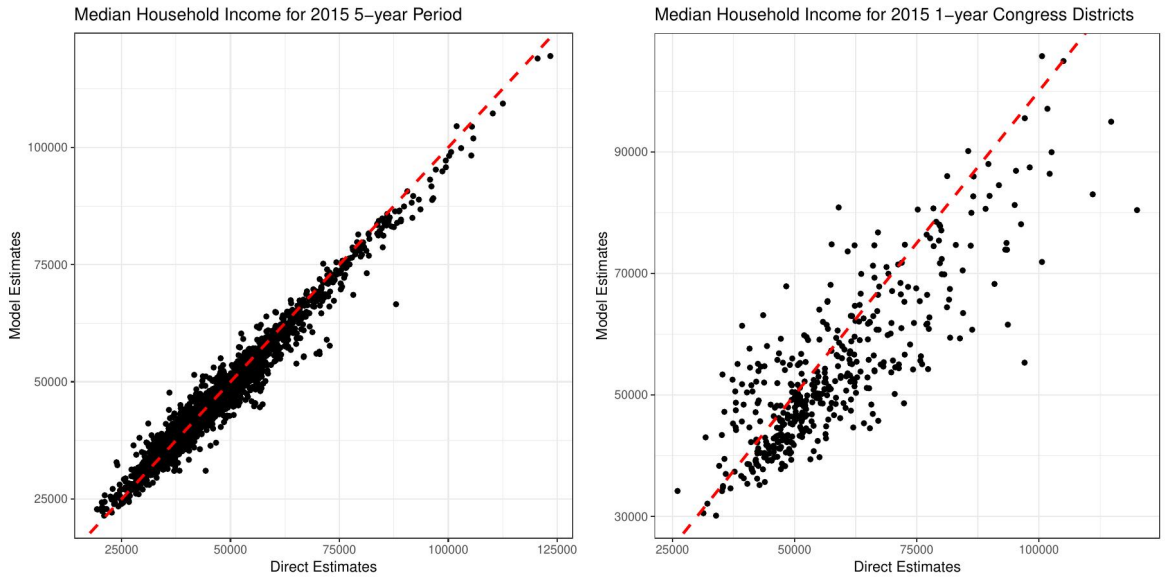
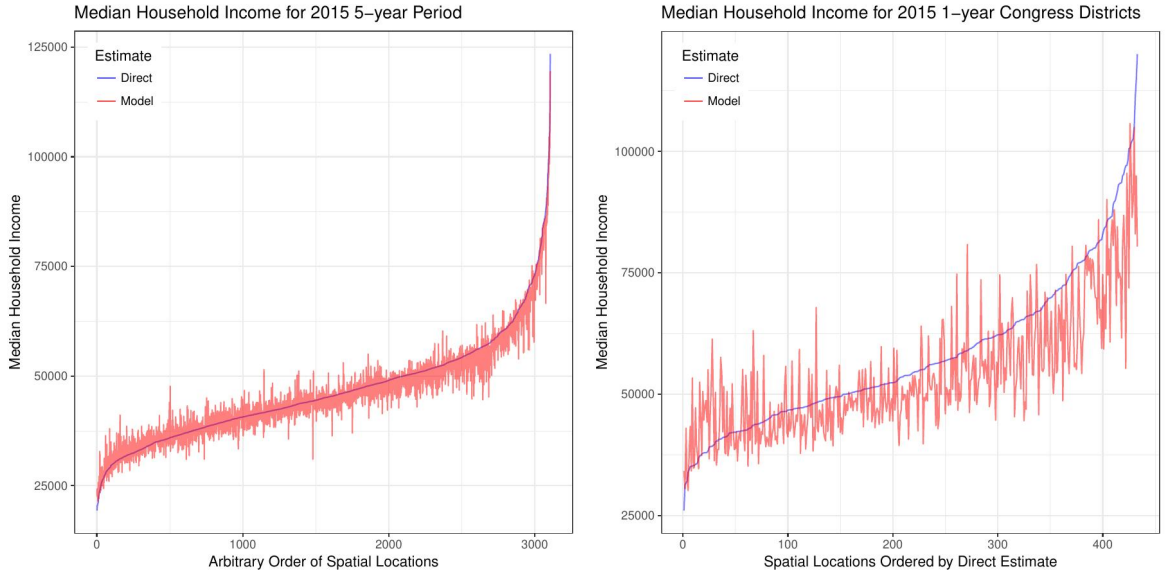


Figure 6: Direct and model-based estimates for year 2015 congressional districts. The top plots display direct estimates and SDs from ACS. The bottom plots show model-based estimates based on the posterior mean and SD of $\mathrm{E}(Y_t^{(\ell)}(A))$.



(a) Counties, 5-year period. (b) Congressional Districts, 1-year period.

Figure 7: Scatter plots of 2015 direct ACS estimates versus estimates based on the posterior mean of $\mathrm{E}(Y_t^{(\ell)}(A))$. Sample correlation between the two sets of estimates in (a) is 0.9814, while in (b) the correlation 0.8295.



(a) Counties, 5-year period. (b) Congressional Districts, 1-year period.

Figure 8: Plot of 2015 direct ACS estimates versus estimates based on the posterior mean of $\mathrm{E}(Y_t^{(\ell)}(A))$. Locations are ordered by direct estimate from smallest to largest.