NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

08-13

Spatial Fay-Herriot Models for Small Area Estimation with Functional Covariates

Aaron T. Porter, Scott H. Holan, Christopher K. Wikle and Noel Cressie

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Spatial Fay-Herriot Models for Small Area Estimation with Functional Covariates

Aaron T. Porter[1], Scott H. Holan[2], Christopher K. Wikle[3], Noel Cressie[2, 4]

Abstract

The Fay-Herriot (FH) model is widely used in small area estimation and uses auxiliary information to reduce estimation variance at undersampled locations. We extend the type of covariate information used in the FH model to include functional covariates, such as social-media search loads, or remote-sensing images (e.g., in crop-yield surveys). The inclusion of these functional covariates is facilitated through a two-stage dimension reduction approach that includes a Karhunen-Loéve expansion followed by stochastic search variable selection. Additionally, the importance of modeling spatial autocorrelation has recently been recognized in the FH model; our model utilizes the conditional autoregressive class of spatial models in addition to functional covariates. We demonstrate the effectiveness of our approach through simulation and through the analysis of American Community Survey data. We use Google Trends search curves as functional covariates to analyze changes in rates of household Spanish speaking in the eastern half of the United States.

Keywords: American Community Survey; Bayesian hierarchical modeling; Google Trends; Spatial statistics; Stochastic search variable selection.

[1](to whom correspondence should be addressed) Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65211, porterat@missouri.edu

[2]Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65211-6100

[3]Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65211

[4]National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia.

1 Introduction

The Fay-Herriot (FH) model (Fay and Herriot, 1979) is one of the primary tools used in small area estimation (SAE) (e.g., Jiang et al., 2011, Roy, 2007, You and Zhou, 2011, among others). Model-based estimates are widely used in SAE as they represent a way to borrow strength across locations and thereby reduce the variance of the small area estimate (Rao, 2003). These models utilize scalar auxiliary information to obtain an "indirect" estimate of the small-area variable of interest, rather than a direct survey estimate. Critically, the FH model is structured in such a way as to guarantee model-based variance reduction in the variable of interest relative to that of the direct survey estimate (Rao, 2003, Chapter 4).

As government budgets remain flat or decline, auxiliary information that is relatively inexpensive and readily available, but still representative of the population under consideration, is of substantial interest. Functional covariates based on internet sources, social media, or other sources (e.g., remotely sensed image data) may augment or replace scalar auxiliary information for a wide variety of surveys. The advantage of these types of covariates is that they are often readily available and provide significant information related to a diverse set of demographic and other survey outcomes. For instance, Twitter tweets or Google searches can be associated with a precise location and searched for specific terms or hashtags. Alternatively, dimension-reduced representations of satellite imagery could be used as auxiliary information in modeling outcomes from agricultural surveys.

Not surprisingly, many federal agencies (including Bureau of Labor Statistics among others) have now realized the potential importance of harnessing these massive, readily available data sources. Methodologies relying on "web-scraping" for the collection of data and use of retail scanner and social-media data have emerged as avenues of particular interest (e.g., see the article by Michael W. Horrigan in *Amstat News*, January 2013, pp. 25-27). Consequently, it is extremely important that sound and effective statistical methodology be

developed to accommodate this abundantly rich class of "Big Data" resources.

Functional data analysis (FDA) methodology allows for the use of curves, images, and other "objects" as either independent or dependent variables in a statistical framework (e.g., Ramsay and Silverman, 2005, 2006). The use of FDA in a (generalized) linear statistical modeling framework is well developed, with a substantial amount of research occurring over the last decade. For example, James (2002) and Müller and Stadtmüller (2005) develop generalized linear models with functional covariates. In addition, Yao et al. (2005) consider such models in a longitudinal framework and Goldsmith et al. (2011) develop penalization methods for regression-model selection with functional covariates. From a Bayesian perspective, Baladandayuthapani et al. (2008) work with spatially correlated functional data and Crainiceanu et al. (2009) develop multilevel functional regression models.

Survey sampling followed by SAE is commonly implemented by official-statistics agencies, but in this article we propose a shift from the usual FH model in two ways. We propose a spatial FH model that uses functional and/or image covariates as auxiliary information. Examples of such covariates include Google Trends curves, Twitter hashtag counts, or remotely sensed satellite imagery. The use of social media and other internet-based predictors is a developing field (see, e.g., Signorini et al., 2011). However, statistical modeling of such data in the context of SAE using functional covariates in spatial FH models remains undeveloped.

Our model proceeds from a Bayesian perspective and, thus, it allows a natural quantification of uncertainty through the posterior distributions. Further, we demonstrate the importance of accounting for spatial correlation often present in SAE. The Bayesian paradigm provides a natural hierarchical framework for incorporating latent spatial random effects. In particular, we propose a FH model that utilizes conditional autoregressive (CAR) random effects. Finally, we use functional covariates that are curves generated from Google Trends (Google, 2012), in a statistical model of state-level American Community Survey (ACS) data (`http://www.census.gov/acs`).

The ACS is an on-going survey performed by the United States Census Bureau that provides single-year and multiyear estimates for a large number of demographic variables. Publicly available data (known as "multiyear estimates") provide one-year estimates for areas with large populations, and three- and five-year period estimates for smaller areas, such as census tracts. The public-use microdata samples (PUMS) are also available for a diverse set of variables and can be used to model smaller geographies, known as public use microdata areas (PUMAs) (see `http://www.census.gov/acs/www/data_documentation/public_use_microdata_sample/` for comprehensive details). The methodology we present here could also be used to fit statistical models to PUMS.

Typically, one would perform SAE on smaller geographies than states, such as at the county or census-tract level. Our reason for analyzing data using each state as a unit is that currently the Google Trends data is available at the state level (although one can also obtain search data for the ten largest cities in any state). For any particular problem, social scientists may find Twitter data and other social-media data at smaller geographies than the state level, to which our SAE methodology could be applied.

The structure of this paper is as follows. We first introduce the motivating data in Section 2. We provide the methodological details of our approach in Section 3, and we demonstrate the variance-reduction properties of our model via simulation in Section 4. An analysis using these techniques in the context of ACS data on changes in household Spanish-speaking is given in Section 5. We close with a discussion in Section 6.

2 Motivating Data: The American Community Survey

The rate of change of Spanish-speaking persons in the home in different areas of the country may provide insight into immigration patterns as well as provide a marker for socio-economic factors. The standard errors of the ACS estimates for language variables tend to be larger

than most other variables in the survey, and this is even true at larger geographies, such as the state level. To improve estimates, we incorporate Google Trends data (Google, 2012) as auxiliary information in a framework that uses a spatial FH model with functional covariates. Google Trends provide state-level weekly time series indicating scaled search loads in various categories (e.g., see Figure 1).

By considering Google Trends searches that contain commonly used Spanish words, we are able to develop a proxy measure for household Spanish-speaking. It is reasonable to assume that individuals who speak Spanish at home are more likely to perform internet searches in Spanish. The ubiquitous presence of Google, as well as many social-media services, make these searches a readily available source of data.

When determining which Google Trends data should be used as a proxy for the pattern of household Spanish speaking, our approach was to analyze the Google searches of relatively common Spanish words. Several candidate words were selected, and we found relatively high search volume for the words "y," "el," and "yo," which mean "and," "the," and "I" in English, respectively. These words rarely appeared in searches in other languages. We base our simulation study (Section 4) and application (Section 5) on these search results.

Google Trends data present several issues that must be addressed prior to analysis. The first issue is related to the way that Google Trends data are defined.[5] Although they can be scaled and normalized to a fixed time point by state, the raw data cannot be directly accessed (Google, 2012). This means that the values of the Google Trends data cannot be compared between states, and only within state comparisons are valid. To remedy this problem, we fix the time frame of $2008 - 2009$ as our period of interest. We standardize each curve to have

[5]The Google Trends data used in this article were downloaded prior to October 2012. Subsequently, Google changed the normalization applied to the data and, therefore, the Google Trends data, as presented here, are no longer available for download; however, they are available upon request from the corresponding author. Nevertheless, all of the results presented in this article are equally applicable to the currently available Google Trends data.

a within-curve mean of zero and a within-curve standard deviation of one. This results in curves with the same scale from state to state, which facilitates extraction of curve features, rather than spurious differences in magnitude.

Because we have considered search loads from $2008 - 2009$, we need to perform some standardization of the outcome. The outcome we consider for each state is defined as

$$\frac{\%\text{ households speaking Spanish in 2009} - \%\text{ households speaking Spanish in 2008}}{\%\text{ households speaking Spanish in 2008}}. \quad (1)$$

The western and eastern halves of the country may behave differently with regard to rates-of-change of Spanish-speaking; so, for illustration, we restrict our analysis to 20 states and the District of Columbia in the eastern half of the United States. This yields 21 locations of interest, many of which have traditionally had a low number of native Spanish speakers. As a consequence, relatively large changes may appear, but the margins of error (MOE) for the ACS estimates of Spanish speaking tend to be larger in the eastern half of the country. Considering small areas in the eastern half gives the FH model the potential to provide a great deal of improvement when compared to the public-use ACS estimate.

Iowa, Mississippi, Arkansas, Virginia, West Virginia, Delaware, Rhode Island, Vermont, New Hampshire, and Maine are excluded from our analysis. There were two reasons that a state was excluded from consideration. The first is that the search load for more than 20% of the weeks under consideration did not meet the threshold that Google Trends uses to indicate search loads. When the threshold was not met, Google Trends reports the value to be zero. Removing states with 20% or more zeroes helped to mitigate Google Trends' censoring of the data. The second reason a state was eliminated was because after January 1, 2010, Google Trends redefined, and presumably improved, their algorithm for tagging searches to a location (Google, 2012). Certain states, such as Virginia, exhibited markedly different behavior after that date, which casts doubt on the accuracy of the search loads during the period 2008-2009 that we considerd. Thus, we excluded these states from our

analysis. The final count of the small areas is $n = 21$, and they are listed in Table 1.

The approach presented here is certainly not unique to estimating rates of household Spanish-speaking. Internet searches or social-media sources contain high-dimensional data that, in principle, could be used in many applications of SAE, thus increasing the auxiliary information that could be used to improve survey-based estimates.

3 Functional Covariates in the Fay-Herriot Model

The model we propose can be viewed as an extension of the traditional FH model. Specifically, we propose including functional covariates as a source of auxiliary information, along with a random effect that captures spatial correlation. To model the spatial correlation, we use a CAR structure.

For $i = 1, \ldots, n$, the traditional FH model is given by

$$Y_i = \theta_i + \epsilon_i, \tag{2}$$

$$\theta_i = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta}_x + u_i, \tag{3}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$ and $u_i \sim N(0, \sigma_u^2)$, with all error terms, $\{\epsilon_i\}$ and $\{u_i\}$, independent. Here, θ_i is the superpopulation mean of the parameter of interest for small area i, Y_i is a design-unbiased estimate of θ_i, and the variance of ϵ_i, σ_i^2, is estimated based on the survey design and assumed known. The auxiliary information at small area i is a q-dimensional vector of scalar covariates denoted by $\mathbf{x}_i$, with associated parameters $\boldsymbol{\beta}_x$ and the intercept is given by β_0.

There is an alternate representation of (3): If we let $[A|B]$ represent the conditional distribution of a generic random quantity A given the generic random quantity B, then (3) can be written as:

$$[Y_i|\theta_i, \sigma_i^2] = (2\pi\sigma_i^2)^{-1/2} \exp\left\{-\frac{1}{2}(Y_i - \theta_i)/\sigma_i^2\right\}.$$

We call the distribution $[Y_i|\theta_i, \sigma_i^2]$ the "data model" following the hierarchical modeling terminology in Cressie and Wikle (2011) in order to clarify that the data responses are specified conditionally on the superpopulation mean and sampling error.

3.1 Dimension-Reduced Functional Covariates

Let $z_{ij}(t)$ denote the j-th functional covariate defined over time domain $\mathcal{T}$ and associated with the i-th small area. Note that one could also include spatially correlated functional covariates (e.g., Baladandayuthapani et al., 2008) or image covariates (e.g., Holan et al., 2010, 2012) in this framework. However, for illustration, we focus here on temporal functional covariates.

An extension of model (3), that includes J functional covariates, can be written as

$$\theta_i \;=\; \beta_0 + \sum_{j=1}^{J} \int_{\mathcal{T}} \beta_j(t) z_{ij}(t)\, dt + \mathbf{x}_i' \boldsymbol{\beta}_x + u_i; \;\; i = 1, \ldots, n, \tag{4}$$

where $\{\beta_j(t) : t \in \mathcal{T}\}$ is a square-integrable functional parameter associated with the j-th functional covariate. Now, for each j, assume that $\{\phi_{jk}(t) : k = 1, \ldots, \infty\}$ forms a complete orthonormal basis in $\mathcal{T}$. Then, we have the unique representation,

$$z_{ij}(t) = \sum_{k=1}^{\infty} \xi_{ij}(k) \phi_{jk}(t), \tag{5}$$

where $\{\xi_{ij}(k) : k = 1, 2, \ldots\}$ are expansion coefficients of $\xi_{ij}(\cdot)$, the j-th functional covariate associated with the i-th small area. We also have the unique representation,

$$\beta_j(t) = \sum_{k=1}^{\infty} b_j(k) \phi_{jk}(t), \tag{6}$$

where $\{b_j(k) : k = 1, 2, \ldots\}$ are the expansion coefficients of $\beta_j(\cdot)$, the j-th square-integrable functional parameter. From the orthonormality property of the basis functions and upon substitution of (5) and (6), (4) can be alternatively expressed as:

$$\theta_i \;=\; \beta_0 + \sum_{j=1}^{J} \sum_{k=1}^{\infty} b_j(k) \xi_{ij}(k) + \mathbf{x}_i' \boldsymbol{\beta}_x + u_i. \tag{7}$$

In principle, any complete orthonormal basis set could be used to represent the functional covariates. In our analysis, we utilize a Karhunen-Loéve (K-L) expansion; see Jolliffe (2002, Chapter 12), Cressie and Wikle (2011, Chapters 4, 5), and the references therein. Following Cressie and Wikle (2011, Chapter 5), assume $\{z_{ij}(\cdot)\}$ are stochastic processes with $E(z_{ij}(t)) = 0$, and for $t, t' \in \mathcal{T}$, define the temporal covariance function for the j-th functional covariate as $C_{0,j}(t,t') = E(Z_{ij}(t)Z_{ij}(t'))$, which is assumed to be invariant across small areas (see Cressie and Wikle, 2011, p. 267, for an analogous definition of a spatial covariance function that is invariant in time). Thus, the subscript "0" serves to remind us that this is effectively a spatio-temporal covariance function for "lag 0" in space and is invariant over all spatial small areas. Then, assuming this covariance is continuous and square-integrable, we can write

$$C_{0,j}(t,t') = \sum_{k=1}^{\infty} \lambda_{jk}\psi_{jk}(t)\psi_{jk}(t'),$$

where $\lambda_{j1} \geq \lambda_{j2} \geq \cdots$ are the eigenvalues and $\{\psi_{jk}(\cdot) : k = 1, 2, \ldots\}$ are the orthonormal eigenfunctions that solve the Fredholm integral equation (e.g., Papoulis, 1965, p. 457-461),

$$\int_{\mathcal{T}} C_{0,j}(t,t')\psi_{jk}(t')dt' = \lambda_{jk}\psi_{jk}(t); \;\; k = 1, 2, \ldots, t \in \mathcal{T}. \tag{8}$$

Because the eigenfunctions, $\{\psi_{jk}(\cdot) : k = 1, 2, \ldots\}$, form a complete orthonormal basis, $z_{ij}(t)$ can be written as,

$$z_{ij}(t) = \sum_{k=1}^{\infty} \xi_{ij}(k)\psi_{jk}(t), \tag{9}$$

where $\{\xi_{ij}(k) : k = 1, 2, \ldots\}$ are uncorrelated, mean-zero, variance $\{\lambda_{jk} : k = 1, 2, \ldots\}$ random variables, respectively. Thus, one can see that the K-L temporal basis functions $\{\psi_{jk}(t)\}$ in (9) play the role of the general temporal basis functions $\{\phi_{jk}(t)\}$ in (5).

In practice, for T discrete times $\{t_1, t_2, \ldots, t_T\}$, the empirical temporal basis functions, $\widetilde{\boldsymbol{\psi}}_{jk} \equiv (\widetilde{\psi}_{jk}(t_1), ..., \widetilde{\psi}_{jk}(t_T))'$, are obtained from a numerical solution of (8). For cases where

the discrete times are equally spaced, this is equivalent to solving the spectral decomposition of the empirical temporal covariance matrix (e.g., Cressie and Wikle, 2011, Chapter 5): $\widehat{\mathbf{C}}_{0,j} = \widetilde{\mathbf{\Psi}}_j \widetilde{\mathbf{\Lambda}}_j \widetilde{\mathbf{\Psi}}_j'$, where $\widetilde{\mathbf{\Psi}}_j \equiv \{\widetilde{\boldsymbol{\psi}}_{j1}, \ldots, \widetilde{\boldsymbol{\psi}}_{jT}\}$, $\widetilde{\mathbf{\Lambda}}_j \equiv \text{diag}(\widetilde{\lambda}_{j1}, \ldots, \widetilde{\lambda}_{jT})$, and $\widehat{\mathbf{C}}_{0,j} \equiv (n-1)^{-1} \sum_{i=1}^{n} (\mathbf{z}_{ij} - \widehat{\boldsymbol{\mu}}_j)(\mathbf{z}_{ij} - \widehat{\boldsymbol{\mu}}_j)'$, for $\widehat{\boldsymbol{\mu}}_{ij} \equiv n^{-1} \sum_{i=1}^{n} \mathbf{z}_{ij}$ and $\mathbf{z}_{ij} \equiv (z_{ij}(t_1), \ldots, z_{ij}(t_T))'$. Note, in some applications, one may consider $\widehat{\boldsymbol{\mu}}_j \equiv \widehat{\mu}_{\cdot j}\mathbf{1}$, where $\widehat{\mu}_{\cdot j}$ is the grand mean, $\widehat{\mu}_{\cdot j} \equiv (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} z_{ij}(t)$. A comprehensive discussion of issues associated with the calculation of empirical basis functions in the discrete K-L framework can be found in Cressie and Wikle (2011, Chapter 5).

In practice, the summation in (7) must be truncated, such that

$$\theta_i = \beta_0 + \sum_{j=1}^{J} \sum_{k=1}^{K_j} b_j(k)\xi_{ij}(k) + \mathbf{x}_i'\boldsymbol{\beta}_x + u_i, \tag{10}$$

where $K_j < T$. Then equations (2) and (10) together represent a FH model that includes both scalar and functional covariates. Typically, K_j is chosen such that some predetermined percentage (e.g., 95%) of variation in the function is retained. That is, K_j is the smallest integer $K \leq T$ such that $\sum_{k=1}^{K} \widetilde{\lambda}_{jk} / \sum_{k=1}^{T} \widetilde{\lambda}_{jk} \geq 0.95$. However, in our framework, this only represents an initial phase of dimension-reduction. Subsequent dimension-reduction proceeds by stochastic search variable selection (SSVS) (George and McCulloch, 1993, 1997).

Our Bayesian SSVS requires prior distributions for the components of $\mathbf{b}_j \equiv (b_j(1), \ldots, b_j(K_j))'$, $j = 1, \ldots, J$, and of $\boldsymbol{\beta}_x$ in (10). In general, when interest resides in a substantial number of submodels, as is the case in the examples we consider, SSVS algorithms provide an effective means of model selection (e.g., see George, 2000, for a comprehensive overview). For example, recall that $\boldsymbol{\beta}_x = (\beta_1, \beta_2, \ldots, \beta_p)'$ consists of p (potential) covariates. Consider the prior distribution,

$$\beta_\ell | \gamma_\ell \sim \gamma_\ell N(0, c_\ell \tau_\ell^2) + (1 - \gamma_\ell) N(0, \tau_\ell^2); \quad \ell = 1, \ldots, p, \tag{11}$$

where conditional independence of $\{\beta_\ell\}$ is assumed, and γ_ℓ are specified at the next level of the hierarchy to have independent Bernoulli(π_ℓ) distributions, with parameter $0 < \pi_\ell < 1$,

for $\ell = 1, \ldots, p$. In this context, π_ℓ represents the prior probability that β_ℓ should be included in the model, and $\gamma_\ell = 1$ indicates that the ℓ-th variable ($\ell = 1, \ldots, p$) is included in the model. Now, typically, c_ℓ, τ_ℓ, and π_ℓ are taken as fixed hyperparameters; George and McCulloch (1993, 1997) present several alternatives for their specification. Specifically, they recommend taking τ_ℓ to be small so that when $\gamma_\ell = 0$ it is sensible to specify an effective prior for β_ℓ that is close to zero. Additionally, in general, it is advantageous to take c_ℓ to be large (greater than 1) so that if $\gamma_\ell = 1$, then the prior favors a non-zero β_ℓ. For $j = 1, \ldots, J$, selection of the elements of $\mathbf{b}_j$ proceeds in an identical manner to selection of the elements of $\boldsymbol{\beta}_x$, with prior mutual independence between $\{\mathbf{b}_j\}$ and $\boldsymbol{\beta}_x$ assumed. For further discussion surrounding SSVS as it relates to functional data modeling, see Holan et al. (2010, 2012) and the references therein.

3.2 Spatial Random Effects

In extensions of the basic FH model, the vast majority of papers in the literature assume independent Gaussian latent random effects for $\mathbf{u} = (u_1, \ldots, u_n)'$. Instead, the model we propose assumes spatially correlated random effects based on the CAR model. Other spatial models, such as SAR models and geostatistical models have been used (see, e.g., Sengupta and Cressie, 2013, for a review and comparison of these). CAR modeling for estimation in small areas dates back to Besag et al. (1991); see also Leroux et al. (1999) and MacNab (2003), who utilize such a model to estimate rates for non-rare diseases in small areas. The CAR model has also been employed in the FH structure (e.g., Cressie, 1990, Gomez-Rubio et al., 2010, You and Zhou, 2011). In addition, Torabi (2011) has implemented the intrinsic CAR (ICAR) model to account for the spatial effects in a spatio-temporal hierarchical Bayesian FH model. We utilize the same ICAR structure here, now in the presence of functional covariates.

The use of an ICAR prior allows the latent spatial characteristics of the data to be modeled directly, which facilitates the borrowing of strength across spatial units. The ICAR formulation is due to Besag et al. (1991). In this setting, define

$$u_i|\{u_{j\neq i}\} \sim N\left(\sum_{i\sim j}\frac{u_j}{w_{i+}}, \frac{\sigma_u^2}{w_{i+}}\right), \tag{12}$$

where the notation "$i \sim j$" denotes that small areas i and j are neighbors (e.g., they share a border), and the term w_{i+} indicates the number of neighbors associated with small area i. The ICAR model defined by (12) yields an Intrinsic Gaussian Markov Random Field (IGMRF) (Rue and Held, 2005), which corresponds to an improper prior distribution in the hierarchical model we propose. The precision matrix of this IGMRF has the form

$$\mathbf{\Sigma}_u^{-1} = \sigma_u^{-2}(\mathbf{D}_w - \mathbf{W}),$$

where $\mathbf{D}_w$ is a diagonal matrix with element (i, i) equal to w_{i+}, the number of neighbors of small area i. Further, the (i, j)-th element of $\mathbf{W}$ equals one if small areas i and j are neighbors, and zero otherwise. The diagonal of $\mathbf{W}$ is set to zero since small area i is not a neighbor of itself.

The improper prior is due to a linear dependency in the columns of $(\mathbf{D}_w - \mathbf{W})$, which can be seen by post-multiplying this matrix by a vector of ones and noting that it yields a vector of zeroes. Despite its impropriety, the ICAR prior distribution is often used, as it yields a proper posterior distribution for many commonly used data models, such as the Gaussian, Poisson, and Binomial distributions. The ICAR prior implies a smoother spatial process than can be obtained from a CAR prior, which facilitates more borrowing of strength between spatial units. A "sum-to-zero" constraint, $\sum_{i=1}^n u_i = 0$, is needed allow the intercept term in the model to be estimable; if not enforced, the intercept and spatial latent effects are linearly dependent. Fast algorithms for sampling $\mathbf{u}$ subject to $\sum_{i=1}^n u_i = 0$ can be found in Rue and Held (2005), which we will need in our analysis..

In conjunction with a Gaussian data model, the ICAR prior yields a proper Gaussian posterior distribution for $\{u_i : i = 1, \ldots, n\}$. This makes the ICAR (and CAR models in general) convenient for modeling the spatial dependency in the FH framework, where Gaussian data models are typically assumed. In a hierarchical modeling framework, of which the FH model is a special case, the posterior distribution is often simulated using a Gibbs sampler, made up of a sequence of Gibbs steps. When an ICAR or CAR prior is used with a non-Gaussian data model, some of the Gibbs steps are more likely to involve Metropolis sampling or numerical approximations.

4 Simulation Study

The simulation study we consider is designed to evaluate the performance of our model (2), (10), (11), and (12) using simulated data that is calibrated to behave like our motivating ACS example. In particular, we consider the effect of using functional covariate information and spatial correlation, both within the FH framework. We assess performance in terms of reduction of variance of the small area estimates.

Using the expansion coefficients from (10), based on the detrended time series (see Step 2, Appendix A), we generated 250 data sets according to the algorithm found in Appendix A. For each data set we performed a FH-CAR-SSVS analysis and our MCMC algorithm consisted of 100,000 iterations with the first 2,000 discarded for burn-in. As mentioned, the FH model yields a guaranteed decrease in estimation variance, and the long chains are required to achieve sufficiently low Monte Carlo error that these variance improvements can be verified. In this setting, all of the full conditional distributions are of standard forms and straightforward to derive. As such, Gibbs sampling was used for all model parameters. The full conditional distributions of the parameters can be found in Appendix B. The model used

for simulating the data $\{\widehat{Y}_i\}$ is

$$\begin{aligned}\widehat{Y}_i &= \widehat{\theta}_i + \epsilon_i \\ \widehat{\theta}_i &= \beta_0 + \sum_{k=1}^{K} b(k)\widehat{\xi}_i(k) + u_i,\end{aligned}$$

where $\widehat{\xi}_i(k)$ is derived from $\widehat{z}_i(t) - \overline{z}$, with $\widehat{z}_i(t)$ being time series simulated to behave like the Google Trends curves for the search term "el" according to the algorithm in Appendix A (i.e., see Step 5 of Appendix A). That is, the data are simulated using just one curve for each small area. Finally, for this simulation, $\{u_i\}$ where assumed to follow the ICAR structure specified in (12) with parameters detailed in Step 8 of Appendix A.

The estimated model for each of the 250 data sets consisted of (2), with the following model for θ_i

$$\theta_i = \beta_0 + \sum_{k=1}^{13} b(k)\xi_i(k) + u_i.$$

In this case, $\{u_i : i = 1, \ldots, n\}$ follows the ICAR model given in (12), with $\sigma_u^2 \sim IG(.001, .001)$ and a "sum-to-zero" constraint imposed on the elements of $\mathbf{u}$. Finally, we assumed $\beta_0 \sim N(0, \sigma_\beta^2)$, with $\sigma_\beta^2 \sim IG(0.001, 0.001)$.

Our primary interest is the reduction in the variance associated the estimate of the survey quantity of interest. To evaluate the variance reduction, we compute

$$\frac{\sigma_i^2 - \text{var}(\widehat{\theta}_i)}{\sigma_i^2} \times 100\%, \tag{13}$$

for $i = 1, \ldots, n$. For each of the 250 simulated data sets, three analyses were performed. The first analysis was performed using the Spatial FH model with functional covariates described in (10) (henceforth called the "SFFH" model). The second model was a FH model with functional covariates and independent Gaussian spatial effects, as is typically done in the FH framework (henceforth called the "FFH" model). The third model includes an ICAR prior on the latent spatial effects, but ignores the functional predictors and contains no

auxiliary information (henceforth called the "Spatial Only" model). In general, one would also include a nonspatial nonfunctional FH model that utilizes only scalar covariates, but we choose not to since our simulated data does not include scalar covariates. Estimates of the mean of the variance reductions can be found in Table 1.

As illustrated in Table 1, the spatial only model does not perform as well as the FH models containing auxiliary information (i.e., the FFH and SFFH models). This is not surprising, because the auxiliary information is key to the reduction of variance in the FH model. However, note that spatial autocorrelation is important in this simulation study. There was greater variance reduction in 18 of 21 locations when using the SFFH model with functional covariates as compared to the FFH model. Two locations that did not show improvement were the District of Columbia and Maryland. The District of Columbia only borders Maryland (recall we removed Virginia from consideration), and Maryland has only one other neighbor (Pennsylvania). The relationship between Washington D.C. and Maryland is the likely explanation for these two locations performing worse in the case where a spatial structure is included in the model. Minnesota is the third location, and Minnesota has only a single neighbor, which leads to poor spatial fitting (e.g., large mean squared error of the latent spatial effect). We conclude that, while the spatial structure does not guarantee a reduction in variance at every location in the FH framework, an ICAR model is an effective way to achieve a reduction in the average estimation variance when spatial autocorrelation is present.

Additionally, our SSVS algorithm selects several eigenfunctions of the search term "el" to be of interest. The primary eigenfunctions of interest are of higher order: the sixth, ninth, and eleventh principal components are selected in 61%, 62%, and 65%, respectively, of the 250 models fitted. Additionally, the fourth, tenth, twelfth, and thirteenth principal components are selected in 55% to 60% of the 250 models fitted. This indicates that high-frequency components in the models are important. The first three principle components

are selected in fewer than 50% of the 250 models fitted, indicating that these low-frequency components are selected less often relative to the high-frequency components and less often than the prior would suggest. This provides further evidence towards the importance of functional covariates and in particular their high-frequency components.

5 Google Trends Data to Improve ACS Estimates

Recall that we utilize a prior distribution for SSVS that consists of a mixture of normals to distill the important features of the functional covariates of the searches for "y," "el," and "yo" (Section 3.1). When employing the SSVS procedure, it is typically advantageous to ensure that all of the covariates are on the same scale. Otherwise, certain components may be selected based solely on their relative magnitude. Therefore, in addition to the standardization discussed in Section 2, in our model, all 39 principle components under consideration (i.e., the eigenvectors $\{\xi_{jk} : k = 1, \ldots, K_j,\ j = 1, \ldots, 3\}$ in Section 3.1) were standardized by subtracting the mean of the component and dividing by the standard deviation of the component. This yielded functional principle components with a mean of zero and a standard deviation of one.

The model we use here differs from the simulation study in that we utilize all three search terms "y," "el," and "yo" as our functional information (see Figure 1). The model used then becomes

$$\begin{aligned} Y_i &= \theta_i + \epsilon_i \\ \theta_i &= \beta_0 + \sum_{j=1}^{3} \sum_{k=1}^{K_j} b_j(k)\xi_{ij}(k) + \mathbf{x}_i'\boldsymbol{\beta}_x + u_i. \end{aligned}$$

For our purposes, π_ℓ in (11), the SSVS portion of the model, was fixed at 0.5 for all ℓ, as this yields equal contributions to the likelihood whether a variable is included or not, and it can be considered non-informative in this sense. The terms c_ℓ and τ_ℓ were considered

equal for all components, yielding two hyperparameters, c and τ, which were chosen via a sensitivity analysis. Specifically, we allowed τ^2 to take values 10^{-3}, 10^{-4},and 10^{-5}, and c to take values 10 and 100. A factorial (sensitivity) experiment was performed in order to select the values of c and τ for our analysis. In this experiment, we chose the values of c and τ that yielded the lowest mean posterior variance in a leave-one-out cross-validation scheme. For each combination of c and τ, one small area at a time was removed for 40,000 iterations, and the chain allowed to burn in for 2,000 iterations. The remaining 38,000 iterations for this area was then used for inference. The MCMC algorithm was run separately for each small area left out. The leave-one-out cross-validation scheme is designed to protect against model overfitting. Our factorial design selected $\tau^2 = 10^{-5}$ and $c = 10$ as producing the lowest mean posterior variance when averaged over all $\{\theta_i\}$.

For $c = 10$ and $\tau^2 = 10^{-5}$, the MCMC algorithm was run for 100,000 iterations with the first 2,000 being discarded for burn in. The functional principle components selected by the SSVS algorithm in over 50 percent of models were the ninth principal component of "y" (77% of models), the first principal component of "y" (55%), the tenth principal component of "y" (55%), the seventh principal component of "el" (70%), the fifth principal component of "el" (57%), and the tenth principal component of "yo" (59%). All other principal components were selected with frequency smaller than the prior π=0.5 (ranging from 36% to 48%), suggesting that the data selects against these other components. It is worth noting that four of the six components selected in over fifty percent of the models are high-frequency terms (i.e., the fifth principle component or higher). It would appear that high-frequency features of the Google Trends data are the primary predictors of the rates of change of household Spanish-speaking. These components may be detecting shocks in the search load, indicating large search volumes resulting from instances when some of the Spanish speaking community within a state searches for news or other stories of interest.

The variance reductions provided by this model can be found in Table 2, and the im-

provements are illustrated in Figure 2. Additionally, we report the results of an analysis using the SSVS with independent random spatial effects and the time series of "y," "el," and "yo" (labeled as the "FFH model") as well as a model that only utilizes an intercept and spatial random effects with an ICAR correlation structure (called the "Spatial Only" model). Of note is the advantage of using the spatial structure in the FH model. The SFFH outperforms both the FFH model and the Spatial Only model in 15 of 21 small areas, while it performs worse than the other two models in only 3 of 21 small areas. These three small areas are the District of Columbia, Maryland, and Connecticut. The issue of poorer estimation in the District of Columbia and Maryland was previously discussed in the simulation study. Based on our simulation study, the poorer estimation of Connecticut is likely due to the particular data set rather than a systematic issue. This conclusion is derived from the fact that in our simulation, for particular data realizations, states other than DC and Minnesota were estimated more poorly using the SFFH model than the other two models (FFH and Spatial Only). In short, when using a SFFH model, these states are estimated better on average across all 250 realizations, though this is not guaranteed for any particular realization. These results argue strongly for the use of spatially correlated latent random effects and further demonstrate the utility of functional covariates in the FH framework.

6 Discussion

FH models have a celebrated history, owing to their versatility in small area estimation. To increase the usefulness of this class of models, we have extended them to include functional covariates along with spatial correlation. Importantly, we have demonstrated that functional covariates can be effectively utilized to improve estimation in the public-use ACS data. Further, we have emphasized the importance of the spatial relationships between small areas in our model, and we have illustrated the importance of a spatial prior in the FH structure.

The fully Bayesian procedure incorporating the dimension-reducing SSVS provides an automated method for feature selection and selection among different candidate models. The model selection is tuned to minimize the variances of $\{\theta_i\colon\ i = 1, \ldots, n\}$. However, it would also be possible to consider other model properties, when selecting SSVS hyperparameters.

The issue of spatial autocorrelation has been addressed systematically, and we have illustrated, via model-based simulation and through our motivating ACS data, that priors inducing spatial autocorrelation yield greater reduction in estimation variance than non-spatial priors. We have also found that the reduction in variance is not guaranteed to be greater for every location. Even so, these results argue strongly for spatial priors to be used in the FH framework.

Due to data limitations, we have applied our approach using Google Trends data that are available at the state level, but not for smaller areas. Twitter data are another source of functional covariates, and they are available at finer spatial resolutions. However, the drawback of using Twitter data is that they are not as readily available. Finally, our model is also generally applicable with image data, such as remotely sensed scenes of land-use/land-cover, which may result in a key use of this technique in agricultural surveys.

Acknowledgments

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.

Appendix A: The Simulation Algorithm

The following algorithm was used to generate the functional covariates and the data for the simulation study presented in Section 4.

Step 1: Consider the Google Trends time series for the search term "el" at location i. Denote this quantity by $\mathbf{z}_i = (z_i(t_1), ..., z_i(t_T))'$. Let $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_n]$ be a $T \times n$ matrix containing the Google trends time series associated with the search term "el."

Step 2: Subtract the time-dependent mean of the matrix $\mathbf{Z}$, namely $\bar{\mathbf{z}} \equiv n^{-1}(\sum_{i=1}^{n} \mathbf{z}_i)$, from each column of $\mathbf{Z}$ to obtain $\mathbf{Z}^*$, a matrix of detrended time series.

Step 3: Consider the $T \times T$ empirical covariance matrix $\mathbf{S}^* \equiv \mathbf{Z}^*\mathbf{Z}^{*'}/(n-1)$. Let $\mathbf{S}^* = \mathbf{\Phi}^*\mathbf{\Lambda}^*\mathbf{\Phi}^{*'}$ be the usual spectral decomposition of $\mathbf{S}^*$. Here, $\mathbf{\Phi}^*$ represents the discretized eigenfunctions for the functional covariate "el."

Step 4: Project the detrended time series onto these eigenfunctions: $\mathbf{A} = \mathbf{\Phi}^{*'}\mathbf{Z}^*$. Let the scale of the eigenfunctions be denoted by $\boldsymbol{\tau} \equiv \text{diag}(\mathbf{A}\mathbf{A}'/n)$.

Step 5: Generate a new set of functional curves, $\widehat{\mathbf{z}}_i = \bar{\mathbf{z}} + \mathbf{\Phi}^*\boldsymbol{\psi}_i$, where $\boldsymbol{\psi}_i \sim MVN(\mathbf{0}, \boldsymbol{\tau})$, and define $\widehat{\mathbf{Z}} \equiv [\widehat{\mathbf{z}}_1, \ldots, \widehat{\mathbf{z}}_n]$.

Step 6: Next we simulate a set of responses. First, obtain the weighted least squares estimates $\widehat{\mathbf{b}}^*$ from $\mathbf{Y} = \mathbf{\Phi}^*\mathbf{b}^* + \boldsymbol{\epsilon}$, where $\mathbf{b}^*$ denotes the vector of coefficients associated with the Google Trends functional time series for the search term "el," $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \text{diag}(\sigma_1^2, \ldots, \sigma_n^2))$, and $\mathbf{Y}$ denotes the n-dimensional vector of observed small-area responses from the ACS, namely (1).

Step 7: For the new functional covariate $\widehat{\mathbf{Z}}$, perform Steps 1 – 3 in order to obtain simulated discretized eigenfunctions, $\widehat{\mathbf{\Phi}}^*$, for the simulated functional covariate.

Step 8: Generate simulated responses $\widehat{\mathbf{Y}}$ according to $\widehat{\mathbf{Y}} = \widehat{\mathbf{\Phi}}^* \widehat{\mathbf{b}}^* + \mathbf{u} + \boldsymbol{\epsilon}$, where $\mathbf{u} \sim ICAR(\sigma_u^2)$, $\sigma_u^2 \sim IG(21/2, 0.004)$, and $\boldsymbol{\epsilon} \sim MVN\left(\mathbf{0}, \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)\right)$. The generating distribution for σ_u^2 was chosen to yield variances similar to those estimated from the ACS data analyzed in Section 6.

Appendix B: Full Conditional Distributions

Here we provide the forms of the full conditional distributions for the SFFH model utilized in Section 5. We define $\mathbf{\Upsilon}$ as a diagonal matrix with $\mathbf{\Upsilon}_{\ell\ell} = c\tau^2\gamma_\ell + \tau^2(1-\gamma_\ell)$ and $\mathbf{\Sigma}_\epsilon$ to be diagonal matrix with $\mathbf{\Sigma}_{\epsilon,ii} = \sigma_i^2$. The term $\mathbf{\Xi} = [\boldsymbol{\xi}_1(1), \ldots, \boldsymbol{\xi}_J(K_J)]$ denotes a matrix with columns $\boldsymbol{\xi}_j(k) = (\xi_{1j}(k), \ldots, \xi_{nj}(k))'$. The scalar n represents the number of locations under consideration. For our analysis, the value is 21 and we let $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_J')'$ denote the concatenated vector of $\{\mathbf{b}_j\}$. The scalar $K = \sum_j K_j$ represents the dimension of $\mathbf{b}$ and, for our analysis, this value equals 39. The scalars a_1 and a_2 denote the shape and scale parameters in the $IG(a_1, a_2)$ prior for σ_u^2 and σ_β^2. For our analysis we set $a_1 = a_2 = 0.001$. Under this notation, the full conditional distributions have the following forms.

1. $\mathbf{b} \sim MVN(\boldsymbol{\mu}_b, \mathbf{\Sigma}_b)$, where $\mathbf{\Sigma}_b = (\mathbf{\Xi}'\mathbf{\Sigma}_\epsilon^{-1}\mathbf{\Xi} + \mathbf{\Upsilon}^{-1})^{-1}$ and $\boldsymbol{\mu}_b = \mathbf{\Sigma}_b\mathbf{\Xi}'\mathbf{\Sigma}_\epsilon^{-1}(\mathbf{y} - \mathbf{1}\beta_0 - \mathbf{u})$.

2. $\mathbf{u} \sim MVN(\boldsymbol{\mu}_u, \mathbf{\Omega}_u) I_{\{\sum_{i=1}^n u_i = 0\}}$, where $\mathbf{\Omega}_u = (\mathbf{\Sigma}_\epsilon^{-1} + \sigma_u^{-1}\{\mathbf{D}_w - \mathbf{W}\})^{-1}$, $\boldsymbol{\mu}_u = \mathbf{\Omega}_u\mathbf{\Sigma}_\epsilon^{-1}(\mathbf{y} - \mathbf{1}\beta_0 - \mathbf{\Xi}\mathbf{b})$, and $I_{\{\cdot\}}$ denotes the indicator function.

3. For $j = 1, \ldots, J$ and $\ell = 1, \ldots, p_j$,

$$\gamma_{j\ell} \sim Bern\left(\frac{f(b_{j\ell}|\gamma_{j\ell} = 1)}{f(b_{j\ell}|\gamma_{j\ell} = 1) + f(b_{j\ell}|\gamma_{j\ell} = 0)}\right),$$

   where $f(\cdot)$ is the pdf of the normal prior associated with $b_{j\ell}$, and $Bern(p)$ denotes a Bernoulli distribution with probability p. For model identifiability, we require the number of selected $\{b_{j\ell}\}$ be less than the number of locations (i.e., $\sum_j \sum_\ell \gamma_{j\ell} \leq 20$).

In cases where the SSVS prior selected this number greater than 20, the set $\{\gamma_{j\ell}\}$ was re-sampled. Note, this occurred infrequently (i.e., in less than 10 percent of the samples).

4. $\sigma_u^2 \sim IG(a_1 + n/2, a_2 + \mathbf{u}'(\mathbf{D}_w - \mathbf{W})\mathbf{u}/2)$.

5. $\beta_0 \sim N(\mu_{\beta_0}, \widetilde{\sigma}^2_{\beta_0})$, where $\widetilde{\sigma}^2_{\beta_0} = (\mathbf{1}'\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{1} + {\sigma^2_{\beta_0}}^{-1})^{-1}$ and $\mu_{\beta_0} = \widetilde{\sigma}^2_{\beta_0}\mathbf{1}'\boldsymbol{\Sigma}_\epsilon^{-1}(\mathbf{y} - \boldsymbol{\Xi}\mathbf{b} - \mathbf{u})$.

6. $\sigma^2_{\beta_0} \sim IG(a_1 + 1/2, a_2 + \beta_0^2/2)$.

Finally, the inclusion of scalar covariates is straightforward. That is, sampling $\boldsymbol{\beta}_x$ in (4) using an SSVS prior would proceed in a similar manner to sampling the functional covariates (see Holan et al., 2012, for an example).

References

Baladandayuthapani, V., Mallick, B., Young Hong, M., Lupton, J., Turner, N., and Carroll, R. (2008). "Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis." *Biometrics*, 64, 64–73.

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration with two applications in spatial statistics (with discussion)." *Annals of the Institute of Statistical Mathematics*, 43, 1–59.

Crainiceanu, C., Staicu, A., and Di, C. (2009). "Generalized multilevel functional regression." *Journal of the American Statistical Association*, 104, 1550–1561.

Cressie, N. (1990). "Small-area prediction of undercount using the general linear model." In *Proceedings of the 1990 Symposium on the Measurement and Improvement of Data Quality*, 93–105. Statistics Canada, Ottawa, Canada.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, NJ: John Wiley and Sons.

Fay, R. and Herriot, R. (1979). "Estimates of income for small places: an application of James-Stein procedures to census data." *Journal of the American Statistical Association*, 74, 269–277.

George, E. (2000). "The variable selection problem." *Journal of the American Statistical Association*, 95, 1304–1308.

George, E. and McCulloch, R. (1993). "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, 88, 881–889.

— (1997). "Approaches for Bayesian variable selection." *Statistica Sinica*, 7, 339–374.

Goldsmith, J., Bobb, J., Crainiceanu, C., Caffo, B., and Reich, D. (2011). "Penalized functional regression." *Journal of Computational and Graphical Statistics*, 20, 830–851.

Gomez-Rubio, V., Best, N., Richardson, S., Li, G., and Clarke, P. (2010). "Bayesian Statistics Small Area Estimation." Tech. rep., Imperial College London. (http://eprints.ncrm.ac.uk/1686/). Unpublished.

Google (2012). "Google Trends." (http://www.google.com/trends/).

Holan, S., Wikle, C., Sullivan-Beckers, L., and Cocroft, R. (2010). "Modeling complex phenotypes: generalized linear models using spectrogram predictors of animal communication signals." *Biometrics*, 66, 914–924.

Holan, S., Yang, W., Matteson, D., and Wikle, C. (2012). "An approach for identifying and predicting economic recessions in real-time using time–frequency functional models." *Applied Stochastic Models in Business and Industry*, 28, 485–499.

James, G. (2002). "Generalized linear models with functional predictors." *Journal of the Royal Statistical Society: Series B*, 64, 411–432.

Jiang, J., Nguyen, T., and Rao, J. (2011). "Best predictive small area estimation." *Journal of the American Statistical Association*, 106, 732–745.

Jolliffe, I. (2002). *Principal Component Analysis*. New York, NY: Springer.

Leroux, B., Lei, X., and Breslow, N. (1999). "Estimation of disease rates in small areas: A new mixed model for spatial dependence." In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, vol. 116, 135–178. New York, NY: Springer.

MacNab, Y. (2003). "Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease." *Statistics in Medicine*, 22, 1761–1773.

Müller, H. and Stadtmüller, U. (2005). "Generalized functional linear models." *Annals of Statistics*, 33, 774–805.

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill.

Ramsay, J. and Silverman, B. (2005). *Applied Functional Data Analysis*. New York, NY: Springer-Verlag.

— (2006). *Functional Data Analysis*. New York, NY: Springer-Verlag.

Rao, J. (2003). *Small Area Estimation*. Hoboken, NJ: Wiley-Interscience.

Roy, A. (2007). "Empirical and Hierarchical Bayesian Methods with Applications to Small Area Estimation." Ph.D. thesis, University of Florida, Department of Statistics.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman & Hall/CRC.

Sengupta, A. and Cressie, N. (2013). "Empirical hierarchical modelling for count data using the Spatial Random Effects model." *Spatial Economic Analysis*. Forthcoming.

Signorini, A., Segre, A., and Polgreen, P. (2011). "The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic." *PLoS One*, 6, 5, e19467. doi:10.1371/journal.pone.0019467.

Torabi, M. (2011). "Hierarchical Bayes estimation of spatial statistics for rates." *Journal of Statistical Planning and Inference*, 142, 358–365.

Yao, F., Müller, H., and Wang, J. (2005). "Functional linear regression analysis for longitudinal data." *Annals of Statistics*, 33, 2873–2903.

You, Y. and Zhou, Q. (2011). "Hierarchical Bayes small area estimation under a spatial model with application to health survey data." *Survey Methodology*, 37, 25–36.

State	SFFH	FFH	Spatial Only
Alabama	**52.8**	44.0	14.1
Connecticut	**14.8**	13.7	3.3
District of Columbia	65.7	**75.1**	38.9
Florida	**1.4**	**1.4**	0.3
Georgia	**11.8**	7.6	1.8
Illinois	**5.3**	3.2	0.7
Indiana	**31.9**	23.4	6.3
Kentucky	**64.5**	51.2	18.6
Maryland	19.7	**21.1**	6.0
Massachusetts	**12.3**	11.3	3.0
Michigan	**20.6**	13.9	4.0
Minnesota	2.41	**29.6**	9.4
Missouri	**39.0**	31.2	10.2
New Jersey	**5.8**	5.2	1.2
New York	**2.8**	1.9	0.4
North Carolina	**11.2**	8.6	1.9
Ohio	**33.1**	23.9	6.6
Pennsylvania	**21.6**	14.9	3.7
South Carolina	**34.2**	30.9	8.9
Tennessee	**45.4**	32.6	9.8
Wisconsin	**28.1**	22.7	6.1

Table 1: Mean relative percentage decrease in variance estimates for the 21 small areas based on 250 simulated data sets for the spatial FH model with functional covariates (SFFH), the standard FH model with functional covariates (FFH), and a FH model using only spatial random effects (Spatial Only). Bolded values indicate the greatest variance reduction. Variance reduction is computed by Equation (13)

State	SFFH	FFH	Spatial Only	σ^2
Alabama	**34.3**	30.8	26.7	2.014e-3
Connecticut	1.0	9.3	**9.6**	3.472e-4
District of Columbia	60.8	67.2	**68.5**	7.268e-3
Florida	1.8	**2.0**	0.1	3.000e-5
Georgia	**11.8**	7.2	5.9	1.819e-4
Illinois	**4.6**	3.1	2.5	7.204e-5
Indiana	**24.5**	19.4	18.6	6.883e-4
Kentucky	**44.7**	39.6	43.7	2.487e-3
Maryland	17.0	**18.5**	17.0	6.732e-4
Massachusetts	9.7	**9.8**	9.0	3.159e-4
Michigan	**18.0**	10.8	0.2	5.239e-4
Minnesota	19.4	**22.2**	26.2	1.086e-3
Missouri	**32.2**	27.9	24.5	1.224e-3
New Jersey	**5.8**	4.9	3.6	1.184e-4
New York	**2.9**	1.7	1.4	4.159e-5
North Carolina	**9.6**	8.2	5.3	2.030e-4
Ohio	**31.2**	23.9	19.1	7.320e-4
Pennsylvania	**19.6**	15.6	11.1	3.901e-4
South Carolina	**24.7**	21.4	21.9	1.093e-3
Tennessee	**38.9**	29.8	24.3	1.160e-3
Wisconsin	**19.1**	17.0	18.0	6.638e-4

Table 2: Relative percentage decrease in variance estimates for the 21 small areas for the analysis of the ACS data for the spatial FH model with functional covariates (SFFH), the standard FH model with functional covariates (FFH), and a FH model using only spatial random effects (Spatial Only). Bolded values indicate the greatest variance reduction. Variance reduction is computed by Equation (13). The column with the heading σ^2 gives the known sampling variance of the relative change in percent household Spanish-speaking for each state.

Figure 1: Functional curves for the Google Trends search loads of "el," "yo," and "y" (see Section 2). To avoid clutter, we show only the first five time series, in alphabetical order (i.e., Alabama, Connecticut, District of Columbia, Florida, and Georgia), for each search term.

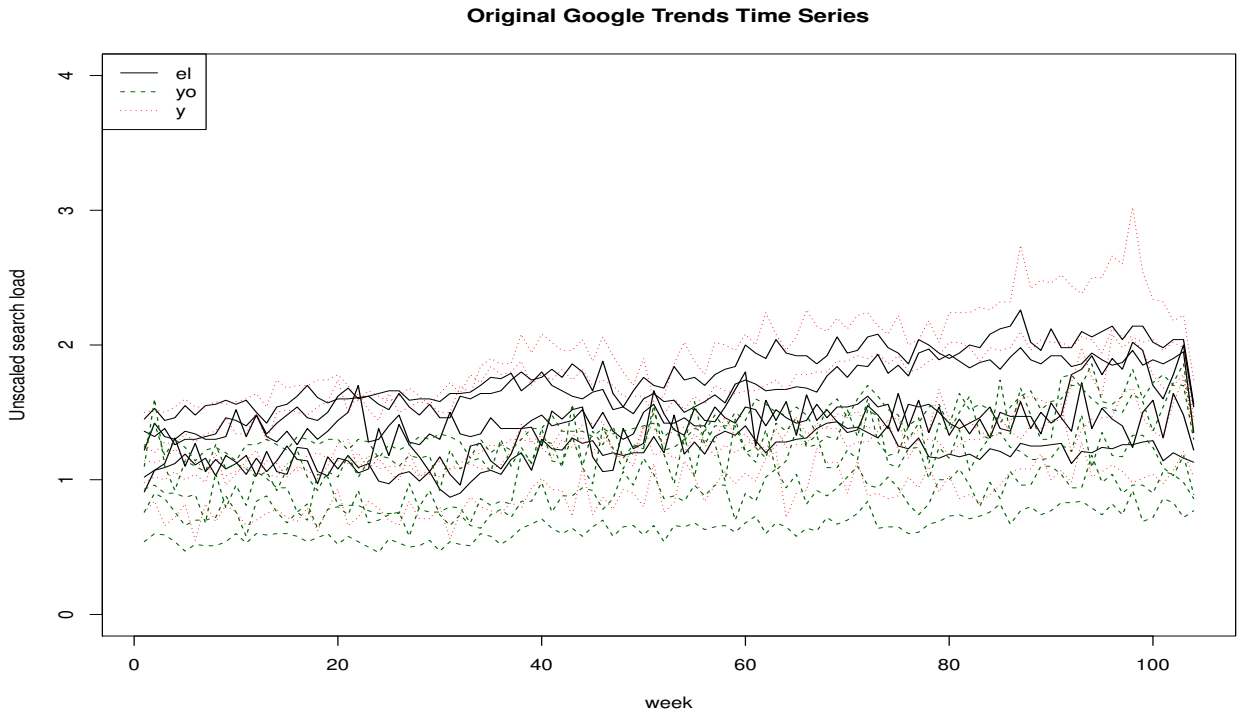

Figure 2: Relative percentage decreases in functional Fay-Herriot model variance versus ACS variance for the SFFH (upper left), the FFH model (upper right), and the Spatial Only model (lower left).