NSF–Census Research Network (NCRN)

Fall 2014 Meeting

September 11-12, 2014

ILR NYC Conference Center,

16 East 34th Street, 6th Floor
New York, NY 10016

(the most up-to-date meeting agenda can always be found at http://www.ncrn.info/event/ncrn-meeting-fall-2014)

This version last updated 2014-09-10.

Thursday, September 11, 2014

11:00-1:00 (by invitation only) PI-only meeting (lunch will be served) [Room D]

1:00-1:15 Opening remarks [TBD]

1:15-4:00 Research Session I: **Statistics of Non-standard Data** *(Organizer: Lars Vilhuber, Cornell, Chair: Connie Citro, National Academies)* [Room A/1]

- 1:15 **Jennifer Parker** (CDC) "Geographic linkages between NCHS population health surveys and air quality measures"
- 1:45 **Abe Dunn**, Eli Liebman, and Adam Shapiro (BEA), "Decomposing Medical-Care Expenditure Growth"
- 2:15 **Michael Spiller** (CDC) "Respondent-Driven Sampling Estimation and the National HIV Behavioral Surveillance System"
- 2:45 ***Discussion: Bill Eddy (CMU)***
- 3:15 Questions and discussion

4:00-4:30 Coffee break

4:30-5:30 Meeting PIs, Census Bureau, NSF [Room D]

7:00-9:00 Group dinner [registration required]

Friday, September 12, 2014

8:00-9:00 Coffee and breakfast

9:00-12:30 Research Session II: **Geo-spatial statistics** *(Organizers: Scott Holan, U of Missouri and Seth Spielman, University of Colorado-Boulder; Chair: Seth Spielman, University of Colorado-Boulder)* (each paper 25 minutes) [Room A/1]

- 9:00 Seth Spielman, "Designer Geographies: Tabulation zones that fit" (Colorado Node)
- 9:25 Daniel G. Brown, "Change in Impervious Surface Area in Southeastern Michigan Before and After the “Great Recession", (School of Natural Resources and Environment, University of Michigan, co-authored with Courtney R. Wilson)
- 9:50 Rebecca Steorts "Constrained Smoothed Bayesian Estimation"(CMU)
- 10:15 == Coffee break ==
- 10:25 Jonathan Bradley "Mixed Effects Modeling for Multivariate-Spatio-Temporal Areal Data" (University of Missouri) (Full paper)
- 10:50 Harrison Quick "Bayesian Marked Point Process Modeling for Generating Fully Synthetic Public Use Data with Point-Referenced Geography" (University of Missouri) (Full paper)
- 11:15 Discussant: Kevin Konty (Director of the Methodology Division of the New York City Department of Health)
- Questions

12:30- 1:30 Group lunch [Room A/1]

Conference ends.

Abstracts

Session I

Title: Decomposing Medical-Care Expenditure Growth

Author: **Abe Dunn**, Eli Liebman, and Adam Shapiro (BEA)

Abstract: Medical-care expenditures have been rising rapidly, accounting for almost one-fifth of GDP in 2009. In this study, we assess the sources of the rising medical-care expenditures in the commercial sector. We employ a novel framework for decomposing expenditure growth into four components at the disease level: service price growth, service utilization growth, treated disease prevalence growth, and demographic shift. The decomposition is conducted using MarketScan data, which is a convenience sample of medical claims from insurers and large employers that contains millions of enrollees. The data includes detailed information on the medical conditions of the enrollees and information about specific medical care procedures (e.g., X-ray exam of middle ear, radiation therapy, or nasal flu vaccine). The large and detailed data is necessary for accurately decomposing health care spending across years, since there are a large number of medical conditions and physicians and hospitals offer an enormous number of products and services. The decomposition shows that growth in prices and treated prevalence are the primary drivers of medical-care expenditure growth over the 2003 to 2007 period. There was no growth in service utilization at the aggregate level over this period. Price and utilization growth were especially large for the treatment of malignant neoplasms. For many conditions, treated prevalence has shifted towards preventive treatment and away from treatment for late-stage illnesses.

Title: Respondent-Driven Sampling Estimation and the National HIV Behavioral Surveillance System

Author: Michael Spiller (CDC)

Abstract: Respondent-driven sampling (RDS) is a link-tracing sampling method used to sample hidden or hard-to-reach populations. The CDC's National HIV Behavioral Surveillance System uses RDS in 20 large U.S. cities to sample persons who inject drugs and heterosexuals at increased risk for acquiring HIV, and hundreds of RDS studies have been conducted worldwide. RDS estimation was introduced in 1997 and has since undergone significant development. Because RDS is widely used for public health surveys, the literature on RDS estimation is quite active. This presentation discusses the statistical bases for and assumptions of RDS point and variance estimators.

Session II

Title: Designer Geographies: Tabulation zones that fit...

Author: Seth Spielman

Abstract: The US Census Bureau produces estimates at multiple geographic scales. Some scales are determined by jurisdictional/legal boundaries between states, counties, municipalities, and other geographic entities. Within these legally defined entities the census bureau also utilizes "fiat" boundaries which are drawn for the sole purpose of statistical tabulation. The design and maintenance of these fiat boundaries is exogenous to the surveys that they are used to tabulate. In this presentation we examine the idea of "designer" or "bespoke" tabulation geographies. Designer geographies balance the needs of data users for small area data against the strengths/limitations of a particular survey. The design of geographies is a complex problem, we demonstrate this idea by applying a region-building algorithm to the ACS public use data. The algorithm builds geographies subject to user-defined constraints (region size, shape, margin of error on input variables) and a user-defined objective function. We argue that "designer" geographies, built from existing tabulation units, might mitigate some of the problems of the ACS.

Title: Change in Impervious Surface Area in Southeastern Michigan Before and After the "Great Recession"

Authors: Courtney R. Wilson and Daniel G. Brown (School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109-1041)

Abstract: We examined the effect of "Great Recession," beginning in 2007, on the land-cover change across an urban to rural gradient in Southeastern Michigan. Using Landsat satellite imagery, we mapped changes in impervious surface area (ISA) during 2001 – 2005, a five-year period before the Great Recession, to those observed during 2007 – 2011, a five-year period during and after the recession. We used all available high-quality Landsat images taken during the growing season to estimate impervious surface areas in each year and calculated the rate of change over each five-year period. Factor analysis was used for data from the 2000 Census and the American Community Survey data from 2007-2011, to characterize tract-level variability in five factors. Spatial autoregressive models were used to evaluate the relationships between the rate of change in impervious surface area and the community-level factors over census tracts. The results did not supprot our hypothesis that, on average, impervious surface areas would increase at a faster rate, or decrease at a slower rate, before the recession when compared to the post-recession period. However, hypothesized regional differentiation was strong. The socioeconomic composition of individual communities was strongly associated with how the landscapes changed through time and space. Impervious surface areas increased in rural, exurban, and wealthy

communities at higher rates than urban and socially deprived communities. Additionally, differences in rates of change in impervious surface area showed that the recession served to level these differences somewhat, probably through a combination of decreased private-sector development in the suburbs and exurbs and increased public-sector in urban infrastructure. Overall, we demonstrate that the socioeconomic characteristics of communities have implications for both patterns of land cover and for the responses of land-cover-change processes to changing economic conditions. ***Key Words:*** Landsat, impervious surface area, economic recession

Title: Constrained Smoothed Bayesian Estimation

Author: Rebecca C. Steorts

Abstract**:** We develop methods of constrained Bayes estimation for small-area estimation. We deal with two kinds of constraints: those that require smoothness with respect to some form of similarity across areas, such as geographic proximity or a clustering by co-variates; and bench-marking constraints, requiring (weighted) means and variances of estimates to agree across levels of aggregation, or with external sources of information. We develop our tools for constrained estimation both geometrically, by projecting the unconstrained Bayes estimate into the feasible set, and decision-theoretically, by min-imizing the posterior risk. We show that our constrained estimators can be obtained as solutions to tractable convex optimization problems, and in some cases get closed-form solutions. Mean-squared errors of the constrained estimators are calculated via bootstrapping. Our techniques are free of distributional assumptions, and equally applicable whether the estimator is linear or non-linear, univariate or multivariate. We illustrated our methodology by applying it to data from the U.S. Census's Small Area Income and Poverty Estimates program.

Title: Bayesian Marked Point Process Modeling for Generating Fully Synthetic Public Use Data with Point-Referenced Geography

Author: Harrison Quick

Abstract: Many data stewards collect confidential data that include fine geography. When sharing these data with others, data stewards strive to disseminate data that are informative for a wide range of spatial and non-spatial analyses while simultaneously protecting the confidentiality of data subjects' identities and attributes. Typically, data stewards meet this challenge by coarsening the resolution of the released geography and, as needed, perturbing the confidential attributes. When done with high intensity, these redaction strategies can result in released data with poor analytic quality. We propose an alternative dissemination approach based on fully synthetic data. We generate data using marked point process models that maintain both the statistical properties and the spatial dependence structure of the confidential data. We illustrate the approach using data consisting of mortality records from Durham, North Carolina.

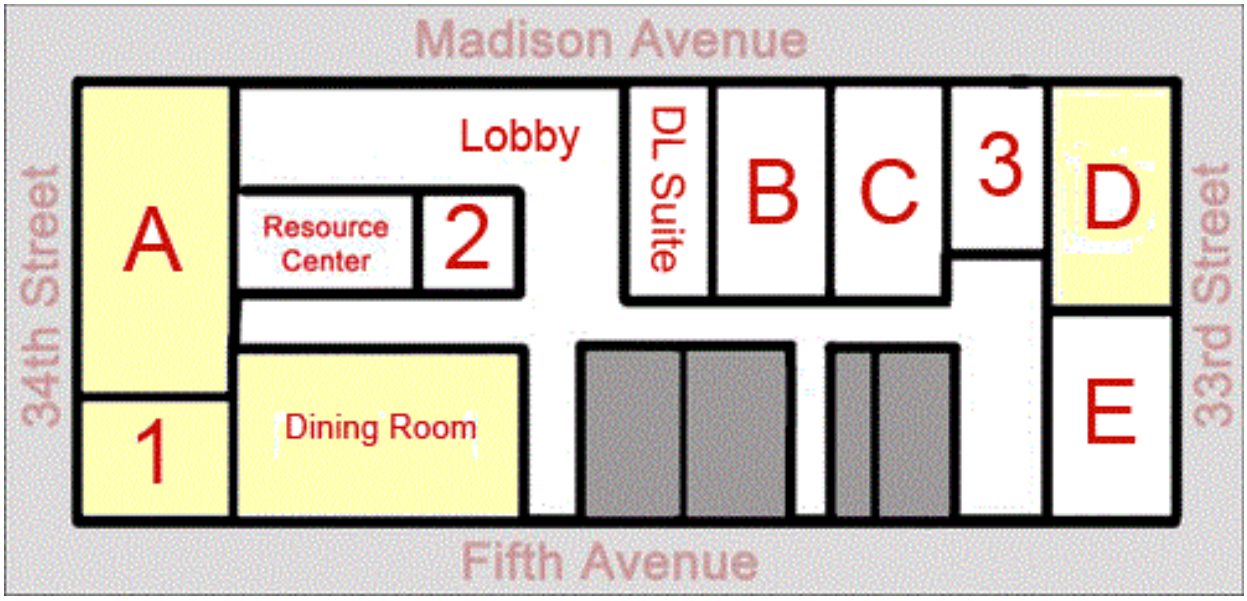
Title: Mixed Effects Modeling for Multivariate-Spatio-Temporal Areal Data

Author: Jonathan Bradley:

Abstract: There are often data sources available that estimate and report related variables of interest that are referenced over geographic regions and time; however, there are very few general statistical methods that one can readily use that incorporate multivariate-spatio-temporal dependencies. As such, we introduce the multivariate-spatio-temporal mixed effects model (MSTM) to analyze areal data with multivariate-spatio-temporal dependencies. We adopt a fully Bayesian approach for prediction and for uncertainty quantification. To demonstrate the proposed methodology, we consider examples obtained from the Bureau of Labor Statistics (BLS), and the US Census Bureau. In the first application, we estimate unemployment rates over counties of Missouri based on data from two different surveys: the American Community Survey (ACS; US Census Bureau) and the Local Area Unemployment Statistics (LAUS; BLS) program. This example demonstrates that one can use the MSTM to combine data from multiple surveys, which is a general topic of independent interest. Then, in the second application, we jointly analyze several earnings variables using a massive multivariate-spatio-temporal dataset obtained from the US Census Bureau, namely the Longitudinal Employment Household Dynamic (LEHD; US Census Bureau) survey. This joint analysis of 7,530,037 observations is meant to illustrate that the MSTM can be used to efficiently model massive multivariate-spatio-temporal datasets.

Layout of the rooms at the ILR NYC Conference Center

https://www.ilr.cornell.edu/ccnyc/services/



Map of the location

https://www.ilr.cornell.edu/CCNYC/directions/index.html