Title: Need for Data Integration and Statistical Modeling for Various Purposes

Date/Time: September 25, 2019 1:00 PM to 2:30 PM

Speaker: Prof. Partha Lahiri, Joint Program in Survey Methodology (JPSM) and Department of Mathematics, University of Maryland, College Park, USA

Abstract: We present a few illustrative examples to explain potential needs for integrating information contained in multiple databases. For each example, we argue the usefulness of statistical modeling and the associated statistical methodology. We borrow our first example from Hansen and Hurwitz (1953), who showed how a simple regression model can be used to integrate information contained in two surveys in reducing various nonsampling errors. Our second example concerns an extension of this simple regression model to combine information from different sample surveys, administrative, and census data in improving direct estimation at granular levels. As for our third example, we consider a situation when the variables to be used in the intended data analysis are stored in multiple databases that do not have error-free unique identifiers. We will explain how models can be used to probabilistically link multiple databases and to provide valid inferences from such linked data that contains potential mismatch errors. Our fourth example concerns monitoring socio-economic and health statistics at shorter interval (e.g., every year) in order to evaluate the effectiveness of planning. Lack of available annual detailed data is the main impediment to undertake such an exercise. This needs massive resources in terms of cost and time to conduct annual survey to produce reliable estimates and it is difficult, especially for developing and underdeveloped countries, to set aside such massive resources for this purpose, given other prioritized vital socio-economic needs for their people. We will discuss how modeling can help to integrate multiple data sources to provide a viable solution to this problem. Our final example is related to traffic projection in real time. Traditional time series methods are not suitable for this purpose because of the considerable time they take for building and fitting models for traffic projections. We will discuss how historical big data on traffic patterns can be used to first build and fit a reasonable working model for short time traffic projection and then use the fitted model and current time data as an input for fast real time traffic projection. We will end the talk with a few concluding remarks.

Location: Bureau of Labor Statistics Janet Norwood Conference Center, Rooms 9

To be placed on the seminar attendance list at the Bureau of Labor Statistics, you need to e-mail your name, affiliation, and seminar name to wss_seminar@bls.gov (underscore after 'wss') by noon at least two days in advance of the seminar. Please bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station. Parking in the area of BLS is available at Union Station. For parking information see www.unionstationdc.com/parking. No validation is available from BLS for reduced parking rates.

WebEx:

https://dol.webex.com/dol/j.php?MTID=mfe4ca6f234cf67981502cc964f0ebb95

For audio: Call-in toll-free number (Verizon): 1-866-865-9536 (US), Attendee access code: 744 124 3