Title: New developments in small area estimation using a Bayesian-Frequentist Integrated approach
Speaker: Avi Singh, American Institutes for Research, Rockville, MD

Chair: Jonathan Lisic
Date and Time: Monday, February 22, 12:30 – 2:00 p.m.
Location: Bureau of Labor Statistics Conference Center #3

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 2 days in advance of the seminar, or call 202-691-7524 and leave a message. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station.

Sponsor: WSS Methodology Section

POC e-mail: asingh@air.org

WebEx event address for attendees: Go to
https://dol.webex.com/mw0401lsp13/mywebex/default.do?service=1&siteurl=dol&nomenu=true&main_url=%2Fmc0901lsp13%2Fe.do%3Fsiteurl%3Ddol%26AT%3DMI%26EventID%3D405132467%26UID%3D0%26Host%3DQUhTSwAAAALzWMuGH8MAGSZJMntEfvqAEchYI5yRuz8t9NL0hWEZcGutzSk6hrLemnGS0RszwKIDEFV9z32kdFM_--Mvk2iZ0%26FrameSet%3D2%26MTID%3Dm5c51a0d62958b3a6d9d8698a53fea0ec

For audio: Call-in toll-free number (Verizon): 1-866-747-9048 (US) Call-in number (Verizon): 1-517-233-2139 (US) Attendee access code: 938 454 2

Note: Particular computer configurations might not be compatible with WebEx.

Summary: Small area estimation (SAE) of finite population parameters such as domain totals under a model for sample survey data provides a natural ground for applications of the Bayesian approach because the parameters of interest are random and the domain sample sizes, as a rule, are not large enough to yield reliable frequentist interval estimates commonly based on normality assumptions. Even for fixed second order parameters such as the common variance and correlation of random effects (which admit consistent estimates for large number of small areas), a Bayesian solution might be preferable because frequentist methods may give rise to inadmissible or unreasonable point estimates. Assuming that the model is valid, the traditional Bayesian framework is in principle simple and attractive due to its prescriptive nature but it does not support various desirable practical requirements that can be built in a frequentist framework. These requirements are driven by user concerns for the validity of modeling assumptions, robustness of the estimates to departures from them, and for the key assumptions being amenable to frequentist-type easily interpretable model diagnostics under repeated sampling.

Since neither Bayesian nor frequentist methods are adequate by themselves for the SAE problem under consideration, it would be useful to develop a Bayesian-Frequentist Integrated (BFI) approach which starts with the specification of a frequentist model and the corresponding Bayesian model such that the two are equivalent when priors for the 'frequentist model' fixed parameters (first and second order) are not introduced. It is believed that a SAE system based on the BFI approach could generate user confidence by not being perceived as a black box. With this in mind, we extend the traditional Bayesian framework to include various frequentist-type features. We define a new type of posterior (to be termed as BFI-posterior) where the conditional posterior for random parameters given the 'frequentist model'

fixed parameters may not make full use of the available data while the marginal posterior for the 'frequentist model' fixed parameters does when treated as random in the Bayesian framework. The concept of BFI-posterior is somewhat akin to the concept of quasi-likelihood as an alternative to the maximum likelihood when the likelihood is not specified or is difficult to do so. However, BFI-posterior is not quasi as it does refer to a legitimate distribution although different from the traditional posterior.

Some of the key practical requirements needed for an SAE system (for balancing the trade-off between efficiency on the one hand and face-validity and weaker set of modeling assumptions for robustness on the other) that do not fit in a traditional Bayesian framework are:

(i) For SAEs at different levels of aggregation, there is need for a single underlying model for the sample total at the lowest level (or the building block) to render the exchangeability assumption more plausible and be able to derive models for sample totals at different higher levels without changing the original set of parameters. The 'frequentist model' fixed parameters are estimated using the lowest (BB-) level input data but the random effects are estimated using the input at the higher domain level under consideration in order to grant full say to the direct estimator and to reduce over-shrinkage.

(ii) For estimating parameters at the BB-level, it may not be possible to specify the variance-covariance structure of sampling errors and hence the likelihood without making additional modeling assumptions. Use of normal likelihood for sample totals is convenient in practice but approximation to normality becomes tenuous with the small sample size. Moreover, at the BB-level, the likely presence of zero direct estimates makes it even more problematic. Therefore, there is a need to group BBs to specify an approximate normal likelihood for the grouped total estimates as input to modeling.

(iii) Given 'frequentist model' fixed parameters, the posterior of the random effects obtained at higher levels using less informative or coarser data as input to the likelihood needs to be linked to the posterior of the same set of parameters obtained at the lower level using more informative or granular data as input so that SAEs at the lower level can be adjusted or benchmarked to sum to the corresponding SAE at the higher level. Here, the term 'linking' simply refers to the joint distribution which can be obtained empirically by linking the MCMC replicates. Thus, BFI can give rise to a benchmarked posterior to obtain benchmark-adjusted credibility intervals unlike the customary second step for benchmarking performed outside the Bayesian framework.

(iv) For data from repeated surveys where random effects in the model are connected over time, the posterior of random effects from the previous time needs to be linked (using MCMC replicates, for example) to the posterior at the current time for a given level of aggregated input data in order to obtain posterior of the change in small area parameters over time. This gives rise to the posterior of the change parameter without revising or updating previous time point posterior unlike the traditional approach which, although more efficient, updates by design using more information from the current time.

(v) For robustification as well as for borrowing additional strength, inclusion of new covariates, as in the frequentist modeling, is needed so that the estimating equations corresponding to new regression coefficients ensure synthetic estimates at the BB-level sum to corresponding reliable direct estimates at much higher levels. This type of exact built-in benchmarking in the frequentist framework becomes approximate in the Bayesian framework but inclusion of new covariates, nevertheless, is beneficial.

(vi)The BFI methodology needs to be amenable to familiar diagnostics using the frequentist version of the model. This is important for building user confidence because it is relatively easier for users, for example, to interpret behavior of estimated model residuals under sampling randomization given the parameters than the parameter randomization given the sample.

Potential applications of the BFI methodology to wellknown government programs (such as SAIPE, SAHIE, and LAUS) and to SAEs from surveys (such as NSDUH and NIS) will be discussed. The basic ideas underlying BFI can be found in a preliminary paper (Singh, 2013) in the FCSM Research Conference Proceedings.