Discussion of Papers in Mixed-Model Resampling Session

Eric V. Slud, U.S. Census Bureau, CSRM
Univ. of Maryland, Math. Dept.

Joint Statistical Meetings, Aug. 2016

Disclaimer

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are the author's and not necessarily the Census Bureau's.

3 Innovative Papers on Bootstrap in Mixed Models

• All 3 papers apply to Fay-Herriot Small-Area Model

• Chatterjee and Nguyen papers: goal is parametric bootstrap estimation of FH prediction parameters ($\hat{\beta}$ for Chatterjee, log(MSPE) for Nguyen) following model selection

Chatterjee paper further aims to extend consistency of bootstrap-based model selection to mixed models

• Nguyen and Pfefferman share goal of order $o(1/m)$ unbiased estimation, respectively of $\log(MSPE)$ and $MSPE$; estimand is the expectation of a function of MSPE under pre-defined standard Small-Area EBLUP prediction algorithm

Ansu Chatterjee paper

Preliminaries

(0) Underlying goal is generalization of Shao (1996) Bootstrap Model Selection idea to a linear mixed-effect (FH) model

(1) Question about paper's discussion of under- and over-fitting: why is underfitting harder to detect in small-area setting than in other (mixed-effect) regression settings ? With or without random effects, data analysts will look at residuals plots versus (fixed-effect) predictors and vs. other (non-included) covariates.

Goodness-of-fit tests in linear mixed-effect models:

Jiming Jiang (2001) *mixed model diagnostics* Ann. Stat. **29**

regarding residual distribution, and for goodness of fit geared to comparison of fixed-effect specifications, a chi-squared test:

Min Tang, Eric V. Slud, Ruth M. Pfeiffer (2014), Jour. Multivar. Anal.

(2) In Small-Area applications, often need some sort of variance prediction with external validity, like one based on cross-validation. (But this is not commonly done in small-area work.) Validation here is purely internal, via bootstrap.

Rescaling-Residual Bootstraps in Model Selection

Shao (1996) famously showed that m-out-of-n bootstrap of (x_i, y_i) tuples, with m growing but $o(n)$, results inconsistent bootstrap-based model selection (i.e., with probability approaching 1 (as $n \to \infty$) the correct (MSE-prediction-minimizing) model known to be within a finite-dimensional full model is selected.

Shao showed same result with bootstrap samples (of size n) of $\sqrt{n/m}$ scaled-up residuals.

WiSE bootstrap is a similar idea, but with a single set of random scale multipliers to provide enhanced-noise residuals for all candidate models s, and Chatterjee claims the same consistent-selection property.

His bootstrapped FH with scaled-up iid multipliers $\sqrt{\tau_n}\,U_b$ the same for all models s is

$$Y_{sb} = \widehat{\theta}_{sb} + (\sqrt{\tau_n}\,U_b)\,\hat{B}\,R_s\;, \qquad R_s = Y_s - \widehat{\beta}_s X_s\;, \qquad \hat{B} = \text{weight}$$

noting that $Y = \widehat{\theta}_s + \hat{B}\,R_s$.

Shao does not say the bootstrapped data or parameter estimators look like the original data, but does claim that simultaneous-contrast CIs for β following model-selection for the bootstrap are valid. Does the same thing happen for Ansu's bootstrap ?

Shao's setting includes nonlinear regression and autoregressive time series. Does Ansu's ?

Jiang-Lahiri-Nguyen paper

(I) Rigorous proof of properties of *Inference After Selection* is a very current and challenging topic, as in the POSI project of Larry Brown, Andreas Buja and others.

Papers of Leeb & Pötscher in '90's clarified that model selection &inferences err when regression coefficients are (truly) $O(1/\sqrt{n})$. For some predictive purposes, as in Jiang et al., this does not matter. More broadly, if $O(1/\sqrt{n})$ (or even larger small) coefficients could be excluded, then the model selection for properly specified finite-order model could be ignored!

Yet Jiang, Nguyen and Lahiri prove that for properly specified mixed models, in general setting including Fay-Herriot, the non-uniform consistency of selected models is not an obstacle to $o(1/m)$ unbiasedness of $\log(MSPE)$.

(II) Nguyen paper and talk formulate the problem of estimating $\log(MSPE)$, but MSPE is important primarily for Confidence Intervals in settings where prediction bias is negligible. The results for $\log(MSPE)$ could lead to results for $MSPE$ via exponentiation and Delta method if the variance of the $\log(MSPE)$ estimator could be estimated consistently. This is probably attainable via bootstrap.

Pfeffermann-Correa Paper

Like many of Danny's most interesting contributions, this one has many moving parts !

Setting: iid data $\{Z_i\}_{i=1}^n \sim f(z, \psi)$
with particular interest in bias-correction for parameter $\theta = h(\psi)$

Triply indexed computations:

$1 \le t \le T$ indexes parameter vectors in neighborhood of $\hat{\psi}$
$1 \le \ell \le L$ indexes a set of functions $q_\ell(\hat{\theta}_t, \theta_t^*, \hat{\psi}, \psi_t^*)$
$1 \le b \le B_1 \times B_2$ indexes Bootstrap replications

(In simulation: $T = 200, L = 9, B_1 = 75, B_2 = 50$)

Notes

(0) Motivating goal is to improve on a double-bootstrap bias-correction procedure of Hall and Maiti (2006) in Small-Area Estimation based on aggregate (cluster) models

(1) Alternative parameter vectors are more than just parameters for repeated Monte Carlo estimates – they enable replication for statistical estimation of parameters within q_ℓ.

(2) T parameter vectors ψ_t split into training & validation sets, with validation size $V \gg n^4$ in theory part

(3) Separate large simulation of size $V \cdot G$ used to 'evaluate' true parameters $\theta_t,\ t \in V$

Additional Comments & Questions for Pfeffermann

(A) The idea works very well in simulation, improving on previous double-bootstrap bias corrections. But the Hall & Maiti (2006) double bootstrap theory requires either large n or sufficiently large B_2, so may not show to best advantage here. The theory part of this paper is relatively silent on how large B must be.

(B) This paper's empirical bootstrap bias-correction method puts looped (L, T) replications on top of the double bootstrap, so wouldn't it have been a fairer comparison to apply the double bootstrap using increased B_1, B_2 along with replicated ψ_t to balance the additional L, T computational effort ?

Some General Discussion

(1) *I question the suitability of the goal of $o(1/m)$-unbiasedness in practical work in light of likely model misspecification*.
In general, misspecifications of models by other parameters of order $O(1/\sqrt{m})$ would make this goal unattainable, yet that order of misspecification is the most one could test for with any asymptotic power !

But as in most Small Area literature, these papers are explicitly parametric without questioning correct full-model specification.

(2) There is need for extensions of these ideas to full-model size $p+1$ to grow with n.

References

Jiang, J. (2001) *Ann. Stat.* Mixed model diagnostics ...

Leeb, H. & Pötscher, B., various papers in the 2000's

Shao, J. (1996), *JASA* Bootstrap Model Selection.

Tang, M., Slud, E. and Pfeiffer, R. (2014) *JMVA* Goodness of fit tests for Linear Mixed Models.

Thank you !

Eric.V.Slud@census.gov