On the Development and Extension of Bayesian Historical Borrowing

Sinan Yavuz

March 1, 2021

2021 OSG All Hands Meeting

Table of Contents

Introducing Our Team

David Kaplan, Ph.D.

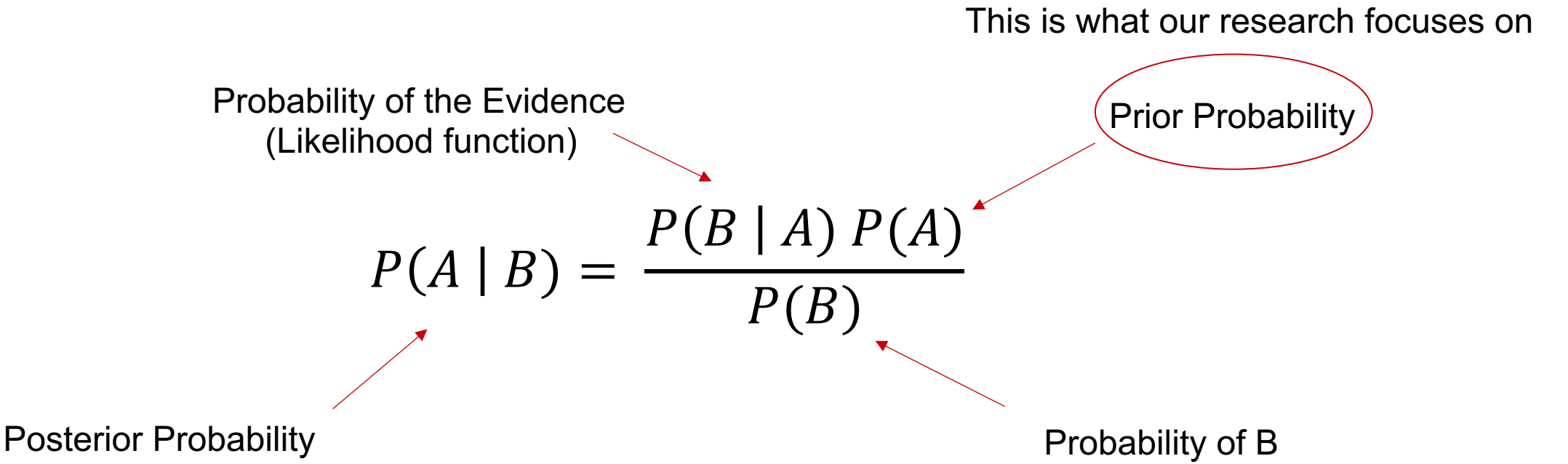
Cassie Chen, Ph.D.

Sinan Yavuz

Weicong Lyu

Motivation of the Study (1)

Bayes' theorem allows us to update probabilities after obtaining new data. Here is the generic equation for the two events A and B;



Motivation of the Study (2)

- Default option on many software is non-informative prior information.
- What if you have some prior knowledge (previous data set, research, or belief) which you want to add into your analyses?
  - The answer is "*mostly*" straightforward but requires a bit advanced knowledge and thinking.
- What if you have multiple previous data sets? Such as;
  - PISA (the Program for International Student Assessment)
  - ECLS (the Early Childhood Longitudinal Study)

Brief Information of PISA

- PISA is a triennial international survey started in 2000 and conducted by OECD.
- Goal is to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students.
- In 2018, 600,000 students, statistically representative of 32 million 15-year-old students in 79 countries and economies participated.
- We already have 7 cycles of datasets.
- Question is how to incorporate this previous knowledge properly.

Bayesian Dynamic Borrowing (1)

- Is a specific method under Bayesian historical borrowing.
- A method for systematically incorporating prior historical data.
- Prior strength depends on heterogeneity among historical data and current data
  - More homogeneous more borrowing and vice versa

Bayesian Dynamic Borrowing (2) – Some Math

- $\boldsymbol{\beta}$ is the vector of regression coefficients
- H is the number of historical cycles of data D^H $(h = 1, 2, \ldots, H)$
- $\boldsymbol{\beta^1}, \boldsymbol{\beta^2}, \ldots, \boldsymbol{\beta^H}$ parameters of interest in each historical data set
- D^0 current data set and $\boldsymbol{\beta^0}$ is its parameters of interest
- Priors for parameters of interest;

Bayesian Dynamic Borrowing (3) – More Math

- Priors for parameters of interest;

$$\boldsymbol{\beta}^0, \boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^{H-1}, \boldsymbol{\beta}^H \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\beta}^0 & 0 & \ldots & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{\beta}^1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{\Sigma}_{\beta}^{H-1} & 0 \\ 0 & 0 & \ldots & 0 & \boldsymbol{\Sigma}_{\beta}^{H} \end{bmatrix},$$

For simplicity, we assume

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \mathrm{diag}(\tau^2, \ldots, \tau^2).$$

- We also need priors for priors (hyper priors)

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} \sim N(\boldsymbol{\mu}, \mathbf{T})$$

$$\tau^2 \sim \mathrm{IG}(\delta, \lambda),$$

- We assume regression coefficients across cycles are generated from a population with common means and covariance matrices.

Bayesian Dynamic Borrowing (4) – Multilevel Model

Individual i's outcome score for the current data set

n x p predictor matrix for student i

p x 1 column vector of individual-level regression coefficients that vary across G schools

$$y_i^0 \sim N(\boldsymbol{X}_i^0 \boldsymbol{\beta}_{g[i]}^0, \sigma_y^{2[0]}) \qquad \text{for } i = 1,\ldots,n$$

G x Q matrix of group-level predictors

Q x 1 column vector of Q group-level regression coefficient

$$\boldsymbol{\beta}_g^0 \sim N(\boldsymbol{Z}_g^0 \boldsymbol{\Gamma}^0, \boldsymbol{\Sigma}_\beta^0) \qquad \text{for } g = 1,\ldots,G$$

$$\boldsymbol{\Gamma}^0 \sim N(\boldsymbol{\Gamma}_\gamma^0, \boldsymbol{\Sigma}_\gamma^0)$$

Bayesian Dynamic Borrowing (5) – Multilevel Model

We need priors for some other parameters as well!

$$y_i^0 \sim N(\boldsymbol{X}_i^0 \boldsymbol{\beta}_{g[i]}^0, \sigma_y^{2[0]})$$

$$\sigma_y^{[0]} \sim \text{Half-Cauchy}(\mu_y, \sigma_y),$$

Location

Scale

$$\boldsymbol{\beta}_g^0 \sim N(\boldsymbol{Z}_g^0 \boldsymbol{\Gamma}^0, \boldsymbol{\Sigma}_\beta^0)$$

$$\boldsymbol{\Sigma}_\beta^0 \sim \text{IW}(\mathbf{R}_\beta, \nu_\beta),$$

Positive definite scale matrix

Degrees of freedom

$$\boldsymbol{\Gamma}^0 \sim N(\boldsymbol{\Gamma}_\gamma^0, \boldsymbol{\Sigma}_\gamma^0)$$

$$\boldsymbol{\Sigma}_\gamma^0 \sim \text{IW}(\mathbf{R}_\gamma, \nu_\gamma),$$

Bayesian Dynamic Borrowing (6) – Multilevel Model with Historical Data

$$\boldsymbol{\beta}_g^0, \boldsymbol{\beta}_g^1, \ldots, \boldsymbol{\beta}_g^{H-1}, \boldsymbol{\beta}_g^H \sim N(\mathbf{B}_g, \mathbf{T}_{B_g}),$$

$$\mathbf{B}_g = (\mathbf{B}_g^0, \mathbf{B}_g^1, \ldots, \mathbf{B}_g^{H-1}, \mathbf{B}_g^H)$$

$$\mathbf{T}_{B_g}^{H+1} = \begin{bmatrix} \boldsymbol{\Sigma}_{B_g}^0 & & & & \\ & \boldsymbol{\Sigma}_{B_g}^1 & & & \\ & & \ddots & & \\ & & & \boldsymbol{\Sigma}_{B_g}^{H-1} & \\ & & & & \boldsymbol{\Sigma}_{B_g}^H \end{bmatrix},$$

$$\boldsymbol{\Sigma}_{B_g}^{H+1} \sim \mathrm{IW}(\mathbf{R}^{H+1}, \delta^{H+1})$$

Dynamic borrowing can be controlled with this priors for the first and second level

$$\mathbf{T}_\Gamma = \begin{bmatrix} \boldsymbol{\Sigma}_\Gamma^0 & & & & \\ & \boldsymbol{\Sigma}_\Gamma^1 & & & \\ & & \ddots & & \\ & & & \boldsymbol{\Sigma}_\Gamma^{H-1} & \\ & & & & \boldsymbol{\Sigma}_\Gamma^H \end{bmatrix}$$

$$\mathrm{IG}(\delta, \lambda).$$

$$\boldsymbol{\gamma}^0, \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^{H-1}, \boldsymbol{\gamma}^H \sim N(\boldsymbol{\Gamma}, \mathbf{T}_\Gamma)$$

The Power Prior

- Another approach of incorporating historical data.
- Not specifically dynamic insofar as the likelihood of the current data is not directly incorporated into the power prior.

$$p(\theta^0|D^h, a^h) \propto p(D^h|\theta^0)^{a^h} p(\theta^0|\omega^0),$$

scalar prior parameter, weights the historical data relative to the probability of the current data

hyperparameter for the initial prior

Empirical Example

- PISA data, the U.S. sample from 2003 to 2018, in total 31,823 students. Ranges from 4838 to 5611 per cycle.
- Student level variables;
  - Gender,
  - The highest educational level of either parent,
  - A summary index of all household and possession items,
  - Immigration background,
  - The language spoken at home
  - Math achievement score (PV1MATH)
- School level variables;
  - Teacher shortage in the school
  - Student-teacher ratio

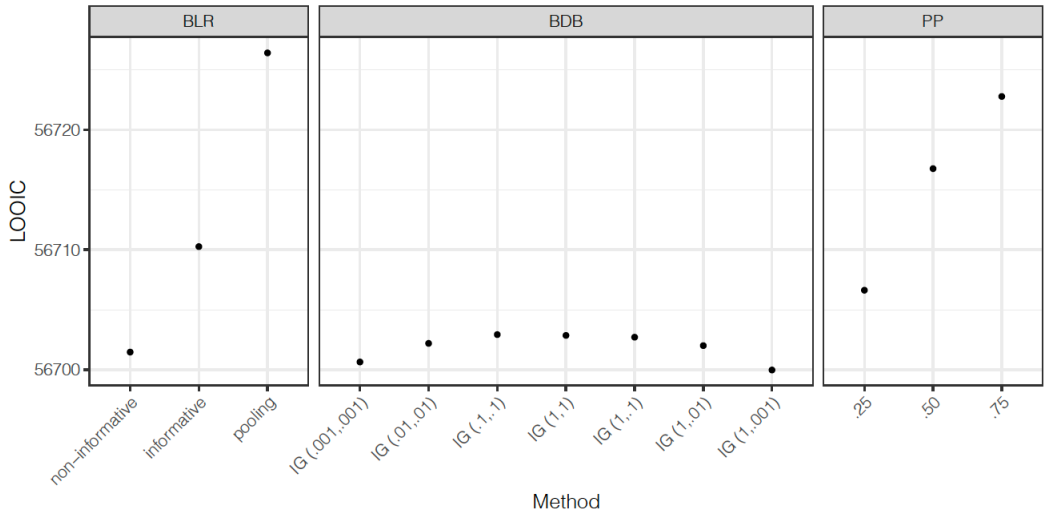
Empirical Example – Computational Problems and OSG (1)

- We mainly used R software and Rstan package.
- Problems of running gigantic Rstan study;
  - Work can’t be stopped and restarted
  - One chain can work on only one CPU.
  - The shortest replication (sets of all conditions) take approximately 2 hours to run.
  - Case study has a larger sample size and sometimes took up to two weeks.
- We started with 13 different prior conditions for single level case study and 20 different for the multilevel case study (ended with 12 for multilevel case).

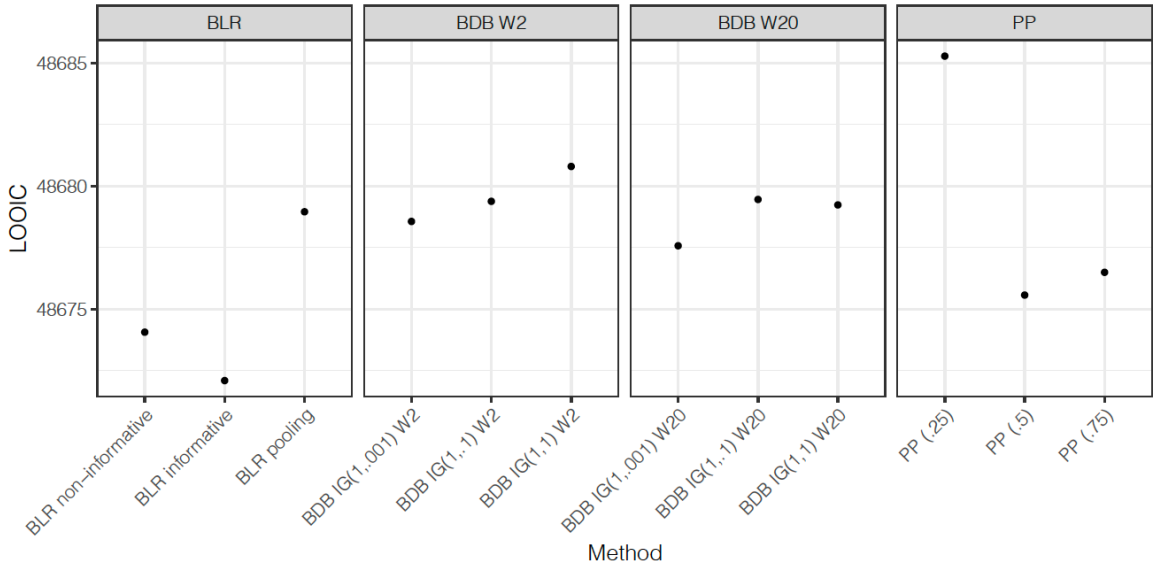
Empirical Example – Computational Problems and OSG (2)

- Each condition of single level case study took about 2 days and we ran it on our computers.
- But multilevel case study took weeks, we ran some part of it on CHTC system with a permission. But still eventually we had to drop some in between conditions.

Empirical Example – Results (3)

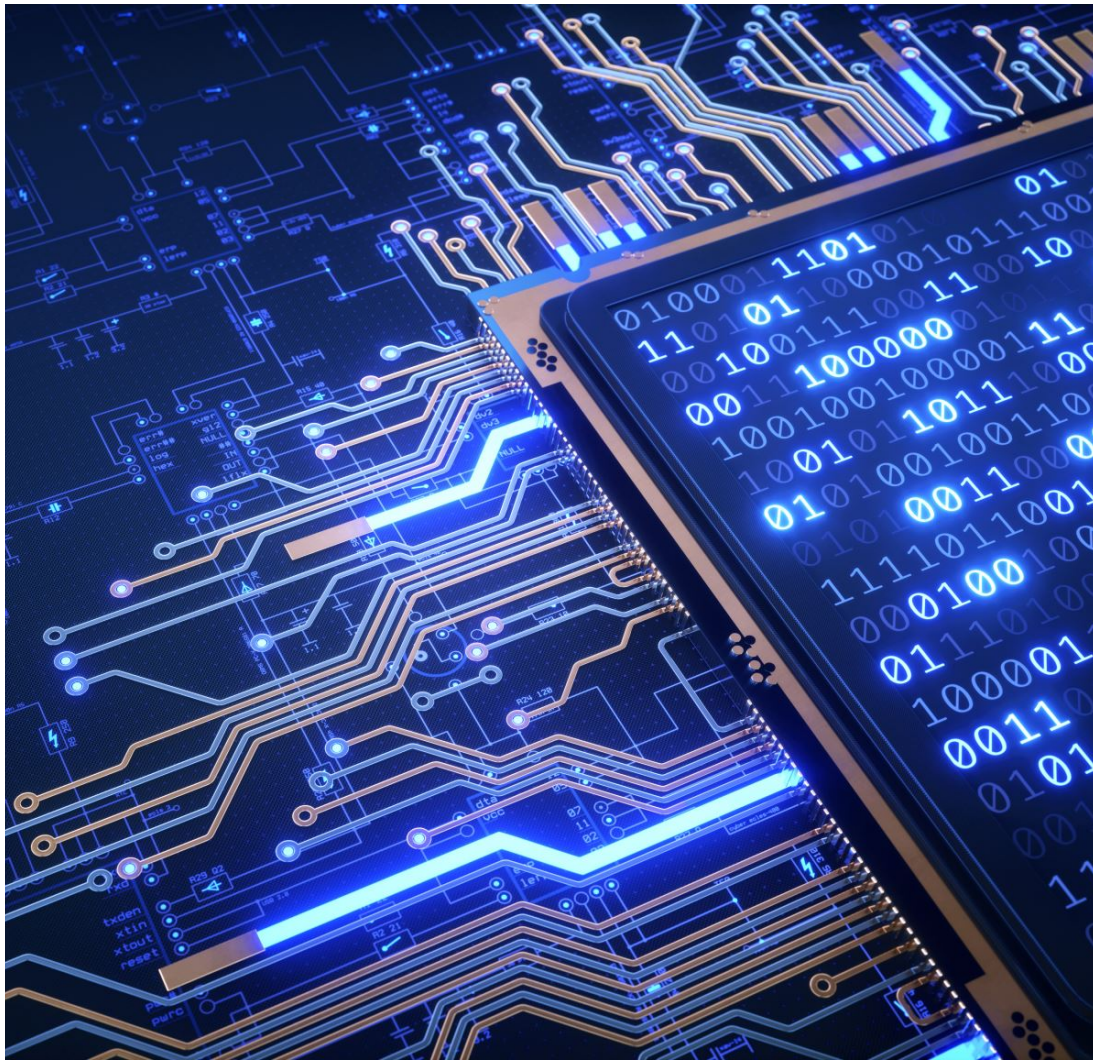


Single level



Multilevel

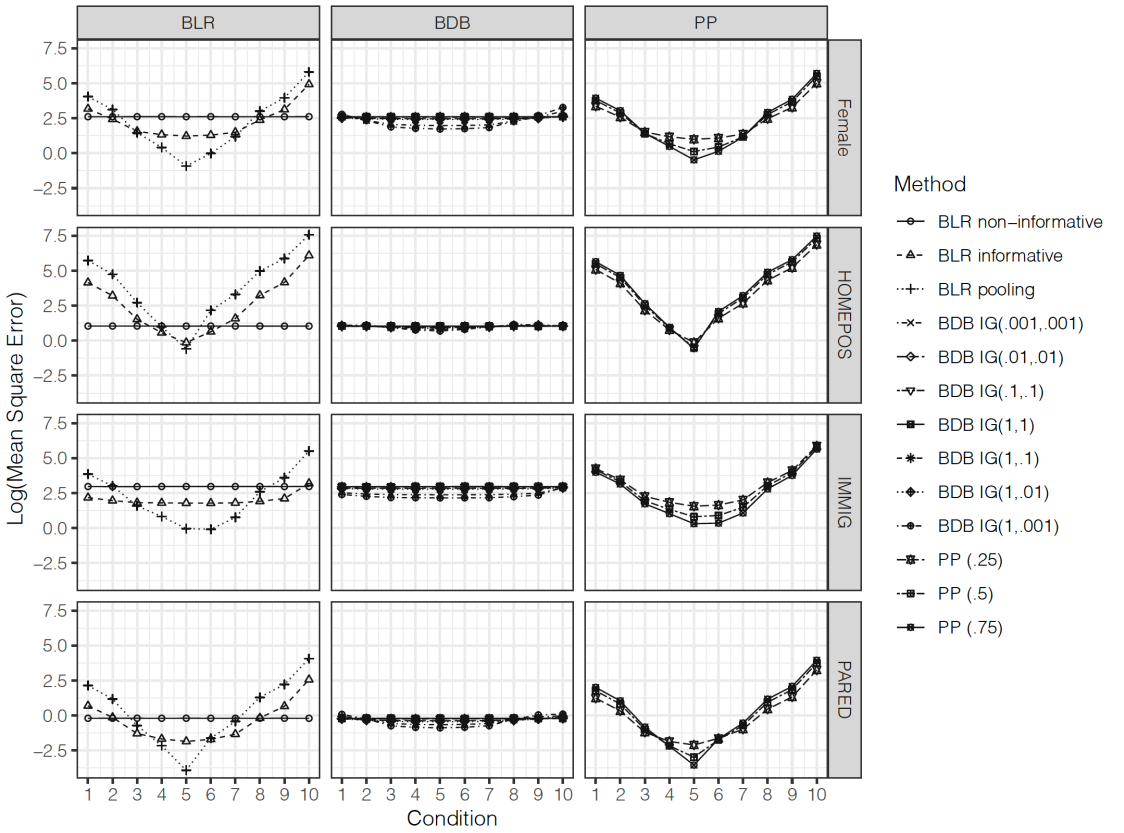
Simulation Study – Single Level

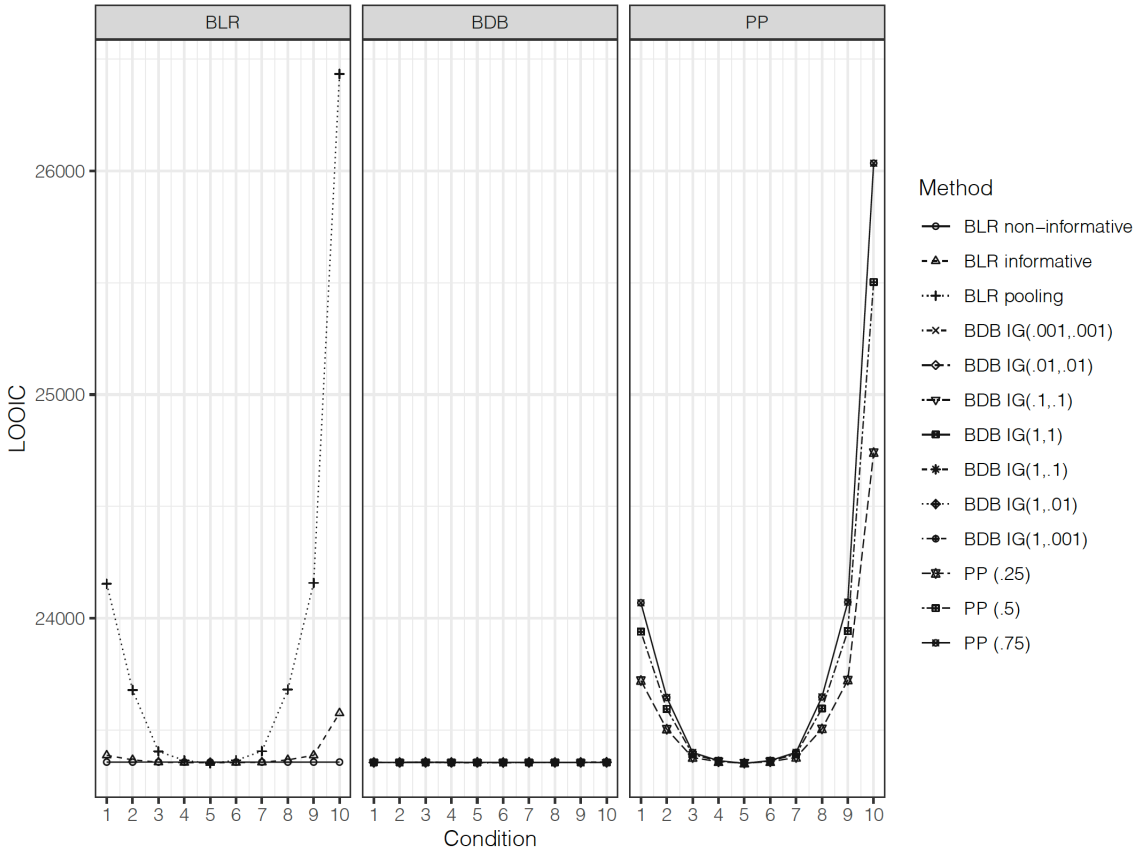


- We studied for both single and multilevel cases 500 replications in each.
- Sample sizes
  - Single level 100, 500 and 2000.
- 10 different heterogeneity conditions among historical cycles and the current cycle.
- 13 different prior conditions.
- 3 x 10 x 13 = 390 conditions
- 390 x 500 = 195,000 analyses
- Each take from 5 mins to 2 hours.
- Average is approximately 20 mins.
- 195,000 x 20mins ~ 7.4 years

Of course, in case everything goes smoothly, and we need to run only one time!

Simulation Study – Single Level Results



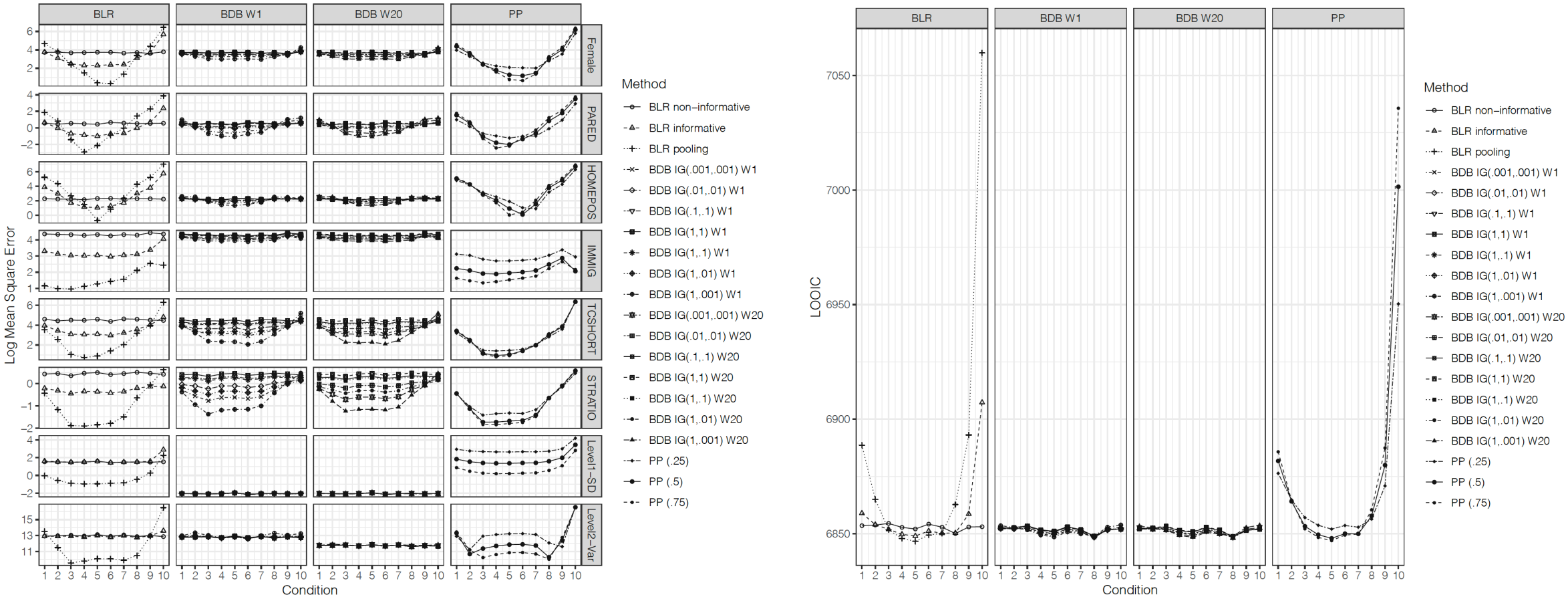


Simulation Study – Multilevel

- Sample size
  - 10 schools, 20 students
  - 10 schools, 40 students
  - 30 schools, 20 students
  - 30 schools, 40 students
- Similar design to single level, but with 20 prior distribution conditions
- 4 x 10 x 20 x 500 = 400,000
- This time it took a lot longer because of the hierarchical design.
- Approximately each analyses took 2 hours.
- Roughly, 400,000 X 2 ~ 91 years!!!

It still took us 3-4 months to finalize this part of the study on CHTC and OSG.

Simulation Study – Multilevel Results



Conclusions

- The goal is to develop a novel method for dynamically borrowing information.
- We extended BDB to handle data arising from multistage sampling designs.
- We compared our method to conventional Bayesian linear regression and power priors.
- Finally, BDB is a prudent choice for combining information across studies, particularly when the degree of heterogeneity is either unknown or known to be extreme relative to the current data.

- The research reported in this paper was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D190053 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. The authors are grateful to Merve Sarac for valuable research assistance.
- This research was performed using the computing resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

Thank you for your attention

Any question or comments

syavuz@wisc.edu