Bayesian Dynamic Borrowing of Historical Information With Applications to the Analysis of Large-Scale Assessments

David Kaplan

Department of Educational Psychology

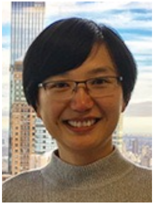
THE UNIVERSITY of WISCONSIN MADISON

ISBA 2022, Montreal, Canada

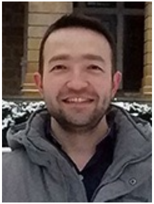
My Co-Authors

College Board

Department of Educational Psychology
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN–MADISON



Cassie Chen, Ph.D



Sinan Yavuz

Weicong Lyu

- The elicitation of substantive prior information is a difficult problem for subject-area researchers using Bayesian statistical methods.

- Researchers will often rely on software default settings that presume non-informative or weakly informative prior distributions for model parameters.

- One substantive area in which a wealth of historical information exists that can be leveraged to elicit prior distributions are *international large-scale educational assessments* (e.g. PISA)

- The current cycle of PISA involves a two-stage stratified sampling design that assesses 15 year-olds in schools for 79 countries/economies. The same design was used every three years since 2000.

Introduction

- The purpose of this paper is to develop and demonstrate the use of *Bayesian historical borrowing* – originally arising in the clinical trials literature (Viele, et al., 2014) – for single-level and multilevel models.
- Bayesian historical borrowing is a method for systematically incorporating prior historical data into current analyses.
- We focus on *Bayesian dynamic borrowing*, for which prior borrowing strength depends on heterogeneity among historical data and current data.
- The more homogeneous the historical data and current data are, the more borrowing there will be.

Review of Methods for Historical Borrowing

- **Power Priors**
  - Traditional power priors are inherently static insofar as the current data are not directly incorporated into the power prior itself.
- **Integrative Data Analysis**
  - AKA pooling. A static approach.
- **Bayesian Synthesis**
  - Bayesian updating. AGDP and AUDP
- **Bayesian Dynamic Borrowing (BDB)**
  - Viele (2014) proposed the idea of BDB for clinical trials which incorporates heterogeneity between the historical data and the current data into the specification of the prior.
  - Strong borrowing occurs when the historical data and the current data agree with each other and weak borrowing occurs when there are large discrepancies between the historical and current data.

Our Contribution

- Our paper contributes to the literature on historical borrowing by

  1. Extending BDB via hierarchical models to multilevel data structures with covariates for large-scale educational assessments,
  2. Comparing BDB to other methods of historical borrowing, including no borrowing, complete pooling and conventional power priors, all within two large simulation studies and two substantively meaningful case studies, and
  3. Focusing not only on typical measures of bias and mean squared error, but also on predictive criteria using loo cross-validation measures.

- Part 1 of this talk appears in *Psychometrika* (2022).

Single-Level BDB

- We specify a joint prior distribution over the historical and current data sets with the degree of heterogeneity across the data sets controlled by the variance of the joint distribution.
- Let H denote the number of available historical cycles of data D^H $(h = 1, 2, \ldots, H)$; for example, past cycles of PISA, and $\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots, \boldsymbol{\beta}^H$ be the parameters of interest in each historical data set.
- Let $\boldsymbol{\beta}^0$ be the parameters of interest in the current data set, D^0. Note that $\boldsymbol{\beta}$ can be a scalar or vector.

Single-Level BDB

- Dynamic borrowing with a Bayesian hierarchical structure can be written as:

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

with

$$\boldsymbol{\mu}_\beta \sim N(\boldsymbol{\mu}, \mathbf{T}) \tag{2}$$

and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is an $(H+1) \times (H+1)$ covariance matrix, where for this paper

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \text{diag}(\tau^2 \ldots \tau^2) \tag{3}$$

and

$$\tau^2 \sim \text{IG}(\delta, \lambda), \tag{4}$$

- Other priors for τ^2 can be implemented.

Single-Level BDB

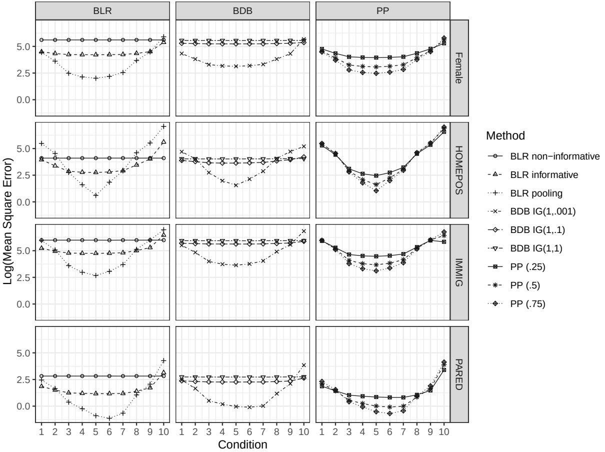
- The key parameter for dynamic borrowing is the variance of the joint prior distribution, τ^2.
- When historical data is consistent with the new data, the posterior distribution will be weighted toward a small τ^2 and there will be more borrowing from the historical information.
- When historical data and the new data differ greatly, the posterior distribution will be weighted toward a large τ^2 and there will be less borrowing.
- A more general covariance structure on Σ_β can be specified if desired.

Single-Level Simulation Study Design

- We simulate PISA-like data according to the case study linear regression model controlling the degrees of heterogeneity.
- For Conditions 1 to 4 and Conditions 6 to 9, the outcomes were generated using slopes that are 80%, 50%, 20%, or 10% below or above average historical slopes to reflect different degrees of heterogeneity.
- For Condition 5, the average of the slopes from the five historical cycles were used as the slopes for generating the outcome variable in the current data reflecting the completely *homogeneous* case.
- For Condition 10, coefficients with the opposite sign of the average historical slopes were used to explore the most extreme heterogeneous case.
- All reps (1000) and all iterations (10000) converged.

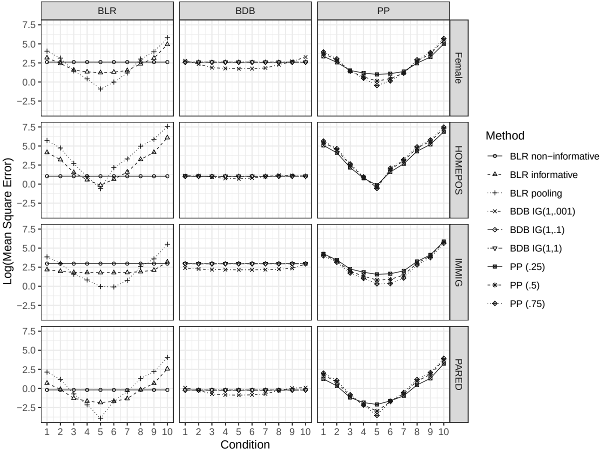
Single-Level Simulation Results

Figure 1: Log MSE (N=100)



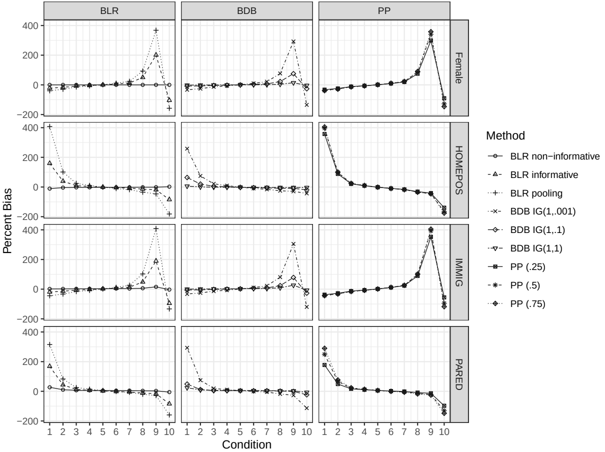
Single-Level Simulation Results

Figure 2: Log MSE (N=2000)



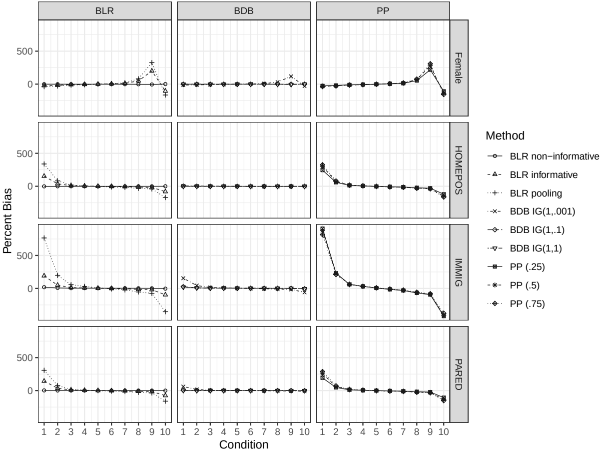
Single-Level Simulation Results

Figure 3: Percent Bias (N=100)



Single-Level Simulation Results

Figure 4: Percent Bias (N=2000)



Single-Level Simulation Results

Figure 5: LOOIC (N=100)

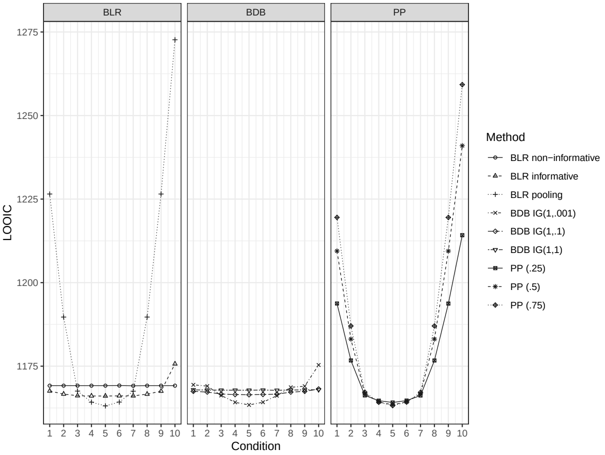
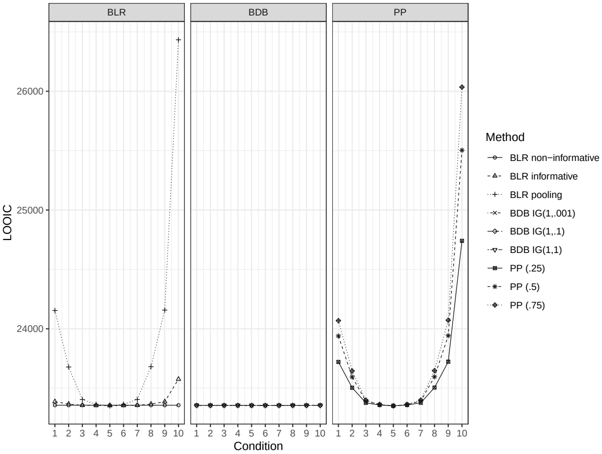


Figure 6: LOOIC (N=2000)



Multilevel BDB

- Applicable to the designs of international large-scale educational assessments.
- Let $\boldsymbol{\beta}_g^H$ represent the H historical student-level coefficients $(h = 1, 2, \ldots, H)$ and the current student-level coefficient, $\boldsymbol{\beta}_g^0$.
- The joint distribution of $\boldsymbol{\beta}_g^{H+1}$ is

$$\boldsymbol{\beta}_g^0, \boldsymbol{\beta}_g^1, \ldots, \boldsymbol{\beta}_g^{H-1}, \boldsymbol{\beta}_g^H \sim N(\mathbf{B}_g, \mathbf{T}_{B_g}), \tag{5}$$

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

Multilevel BDB

- The covariance matrix of the student-level coefficients, $\mathbf{T}_{B_g}$, is block diagonal,

$$\mathbf{T}_{B_g}^{H+1} = \begin{bmatrix} \mathbf{\Sigma}_{B_g}^{0} & & & & \\ & \mathbf{\Sigma}_{B_g}^{1} & & & \\ & & \ddots & & \\ & & & \mathbf{\Sigma}_{B_g}^{H-1} & \\ & & & & \mathbf{\Sigma}_{B_g}^{H} \end{bmatrix} \tag{6}$$

Multilevel BDB

- The joint distribution of school-level coefficients γ^{H+1} is assumed to be multivariate normal with mean $\mathbf{\Gamma}$ and covariance matrix $\mathbf{T}_{\Gamma}$ – viz.

$$\gamma^0, \gamma^1, \ldots, \gamma^{H-1}, \gamma^H \sim N(\mathbf{\Gamma}, \mathbf{T}_{\mathbf{\Gamma}}), \tag{7}$$

- The covariance matrix $\mathbf{T}_{\Gamma}$ under dynamic borrowing is

$$\mathbf{T}_{\Gamma} = \begin{bmatrix} \mathbf{\Sigma}_{\Gamma}^{0} & & & & \\ & \mathbf{\Sigma}_{\Gamma}^{1} & & & \\ & & \ddots & & \\ & & & \mathbf{\Sigma}_{\Gamma}^{H-1} & \\ & & & & \mathbf{\Sigma}_{\Gamma}^{H} \end{bmatrix} \tag{8}$$

Multilevel simulation study design

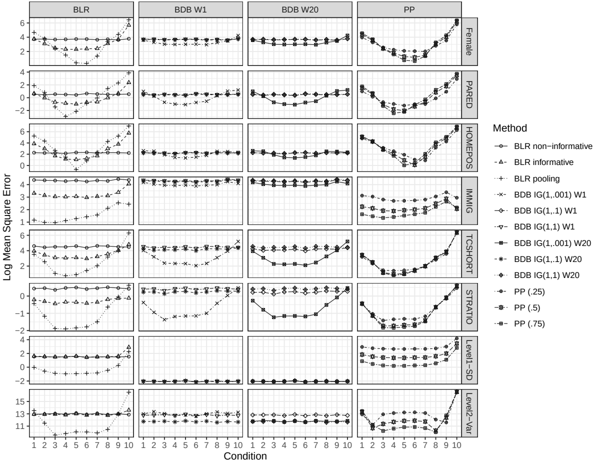
- We simulate PISA-like data controlling the degrees of heterogeneity.
- For Conditions 1 to 4 and Conditions 6 to 9, the outcomes were generated using slopes that are 80%, 50%, 20%, or 10% below or above average historical slopes to reflect different degrees of heterogeneity.
- For Condition 5, the average of the slopes from the five historical cycles were used as the slopes for generating the outcome variable in the current data reflecting the completely homogeneous case.
- For Condition 10, coefficients with the opposite sign of the average historical slopes were used to explore the most extreme heterogeneous case.
- We examine the case of 20 students nested in 30 schools.

Multilevel simulation study design

- We compare BDB to Bayesian linear regression with complete pooling (BLR) under complete pooling, non-informative priors, and informative priors.
- We also compare to power priors with different choices of the borrowing parameters.
- For BDB, we vary the informativeness of the covariance matrix of the random effects by manipulating the degree of informativeness in the IW$(\nu, \nu S)$ distribution where ν takes 1 (weak borrowing) or 20 (strong borrowing).
- $S = \sigma_S' \Omega \sigma_S$ is the baseline precision where $\sigma_S \sim$ Half-Cauchy$(0, 1)$ and $\Omega \sim$ LKJCorr(3) (Lewandowski, Kurowicka, & Joe, 2009).
- 30,000 iterations with 500 replications for each condition.

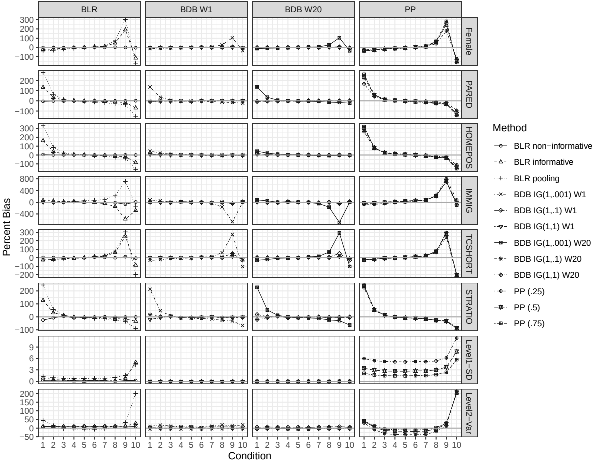
Multilevel Simulation Results

Figure 7: Log MSE 30/20



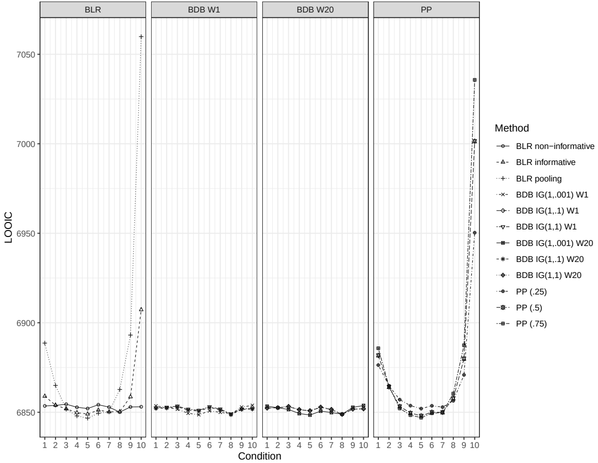
Multilevel-Level Simulation Results

Figure 8: Percent Bias 30/20



Multilevel Simulation Results

Figure 9: LOOIC 30/20



Multilevel Simulation Study Summary

- Most methods show smaller log MSE on coefficient estimates when there is more homogeneity between the current cycle and historical cycles.
- The non-informative prior condition is the most stable across different conditions while performing poorly in homogeneous conditions.
- BDB is stable and it offers good estimates under all conditions.
- Stronger borrowing generally leads to smaller MSE in BDB because the variance of the estimator is reduced, even at the price of possibly greater bias.
- BDB is much better than all other methods for level-1 standard deviation estimates.

Multilevel Simulation Study Summary

- BDB shows much smaller biases in estimates than other borrowing methods except condition 9 where even non-informative priors leads to large bias.
- Pooling and PP are seriously biased under all heterogeneous conditions.
- BDB is preferable to other borrowing methods in terms of LOOIC.
- BDB automatically borrows more historical information in homogeneous conditions and borrows less in heterogeneous conditions to achieve the balance between variance and bias.

Longitudinal Extensions

- We have applied Bayesian dynamic borrowing to longitudinal data collections.
- Our example is ECLS-K:1998 and ECLS-K:2011.
- Our interest is in the utility of borrowing with respect to forecasting in the growth curve modeling framework.
- We examined in-sample and pseudo-out-of-sample forecast scoring measures as our main outcomes.
- Under review at *Large-Scale Assessments in Education*

Longitudinal Extensions

- Let D^0 represent the current cycle ECLS-K:2011 and D^H $(h = 1 \ldots H)$ represent the historical cycles, where for our paper $H = 1$, ECLS-K:1998. The level-1 (within-student) model can be written as follows. Let

$$\boldsymbol{y}_{ig}^{0} = \boldsymbol{\Lambda}^{0}\, \boldsymbol{\eta}_{ig}^{0} + \boldsymbol{\epsilon}_{y(ig)}^{0} \tag{9}$$

and $\boldsymbol{\epsilon}_{y(ig)}^{0}$ is a $T \times 1$ vector of residuals, with diagonal covariance matrix

$$\boldsymbol{\Sigma}_{y_{ig}}^{0} = \begin{bmatrix} \sigma_{y_{1(ig)}}^{2} & & & & \\ & \sigma_{y_{2(ig)}}^{2} & & & \\ & & \ddots & & \\ & & & \sigma_{y_{(T-1)(ig)}}^{2} & \\ & & & & \sigma_{y_{T(ig)}}^{2} \end{bmatrix}. \tag{10}$$

- The level-2 (between-student) model allows the random growth parameters to be related to a set of time-invariant predictors.

$$\boldsymbol{\eta}^0_{ig} = \boldsymbol{\Gamma}^0_g \boldsymbol{x}^0_{ig} + \boldsymbol{\epsilon}^0_{\eta(g)}, \tag{11}$$

where $\boldsymbol{\epsilon}^0_{\eta(g)}$ is a $Q \times 1$ vector of residuals with covariance matrix

$$\boldsymbol{\Sigma}^0_{\eta_{ig}} = \begin{bmatrix} \sigma^2_{\eta^0_0} & 0 & 0 \\ \sigma_{\eta^0_1 \eta^0_0} & \sigma^2_{\eta^0_1} & 0 \\ \sigma_{\eta^0_2 \eta^0_0} & \sigma_{\eta^0_2 \eta^0_1} & \sigma^2_{\eta^0_2} \end{bmatrix}, \tag{12}$$

where it is usually assumed that $\boldsymbol{\Sigma}^0_{\eta_{ig}}$ is lower triangular allowing for the random growth parameters to be correlated conditional on the predictors.

Longitudinal Extensions

- The level-3 (between-school) model can be written as

$$\boldsymbol{\Gamma}_g^0 = \boldsymbol{\Pi}^0 \boldsymbol{w}_g + \boldsymbol{\epsilon}_{\Gamma_g}^0 \quad (13)$$

where $\boldsymbol{\epsilon}_{\Gamma_g}^0$ is a $Q \times 1$ vector of residuals with covariance matrix

$$\boldsymbol{\Sigma}_\Gamma^0 = \begin{bmatrix} \sigma^2_{\Gamma_{1g}} & & & & \\ & \sigma^2_{\Gamma_{2g}} & & & \\ & & \ddots & & \\ & & & \sigma^2_{\Gamma_{(M-1)2g}} & \\ & & & & \sigma^2_{\Gamma_{Mg}} \end{bmatrix}, \quad (14)$$

Longitudinal Extensions

- The full hierarchical Bayesian specification of the GCM model can be written as

$$\boldsymbol{y}_{ig}^{0} \sim N(\boldsymbol{\Lambda}^{0}\boldsymbol{\eta}_{ig}^{0}, \boldsymbol{\Sigma}_{y(ig)}^{0}), \tag{15}$$

$$\boldsymbol{\eta}_{ig} \sim N(\boldsymbol{\Gamma}_{g}^{0}\boldsymbol{x}_{ig}^{0}, \boldsymbol{\Sigma}_{\eta_{ig}}^{0}), \tag{16}$$

$$\boldsymbol{\Gamma}_{g}^{0} \sim N(\boldsymbol{\Pi}^{0}\,\boldsymbol{w}_{g}, \boldsymbol{\Sigma}_{\Gamma}^{0}), \tag{17}$$

$$\boldsymbol{\Pi}^{0} \sim N(\boldsymbol{\Omega}^{0}, \boldsymbol{\Sigma}_{\Pi}^{0}) \tag{18}$$

- Prior distributions on the residual covariance matrices are assumed to be Inverse-Wishart (IW). For the current data, we use

$$\sigma_{y_{ig}}^{0} \sim \text{Half-Cauchy}(\mu_y, \sigma_y), \tag{19}$$

$$\boldsymbol{\Sigma}_{\eta_{ig}}^{0} \sim \text{IW}(\mathbf{R}_{\eta}, \nu_{\eta}), \tag{20}$$

$$\boldsymbol{\Sigma}_{\Gamma}^{0} \sim \text{IW}(\mathbf{R}_{\Gamma}, \nu_{\Gamma}), \tag{21}$$

Longitudinal Extensions

- We adapted our cross-sectional multilevel modeling notation to the case of growth curve models.

- We borrow from historical cycles to estimate the growth parameters.

$$\boldsymbol{\eta}_{ig}^{0}, \boldsymbol{\eta}_{ig}^{1}, \ldots, \boldsymbol{\eta}_{ig}^{(H-1)}, \boldsymbol{\eta}_{ig}^{H} \sim N(\boldsymbol{\eta}_g, \mathbf{T}_{\eta_g}), \tag{22}$$

- The covariance matrix of the random growth parameters can be written as

$$\mathbf{T}_{\eta} = \begin{bmatrix} \boldsymbol{\Sigma}_{\eta_g^0} & & & & \\ & \boldsymbol{\Sigma}_{\eta_g^1} & & & \\ & & \ddots & & \\ & & & \boldsymbol{\Sigma}_{\eta_g^{(H-1)}} & \\ & & & & \boldsymbol{\Sigma}_{\eta_g^H} \end{bmatrix}. \tag{23}$$

Longitudinal Extensions

- Let $\boldsymbol{\Gamma}_g^H$ represent the H historical time-invariant regression coefficients $(h = 1, 2, \ldots, H)$ and let $\boldsymbol{\Gamma}_g^0$ represent the current time-invariant regression coefficients.
- The joint distribution of $\boldsymbol{\Gamma}_g^{H+1}$ is assumed to be multivariate normal with mean $\boldsymbol{\Gamma}_g$ and covariance matrix $\mathbf{T}_{\Gamma_g}$ – viz.

$$\boldsymbol{\Gamma}_g^0, \boldsymbol{\Gamma}_g^1, \ldots, \boldsymbol{\Gamma}_g^{(H-1)}, \boldsymbol{\Gamma}_g^H \sim N(\boldsymbol{\Gamma}_g, \mathbf{T}_{\Gamma_g}), \tag{24}$$

- $\mathbf{T}_{\Gamma_g}$, is block diagonal,

$$\mathbf{T}_{\Gamma_g} = \begin{bmatrix} \boldsymbol{\Sigma}_{\Gamma_g^0} & & & & \\ & \boldsymbol{\Sigma}_{\Gamma_g^1} & & & \\ & & \ddots & & \\ & & & \boldsymbol{\Sigma}_{\Gamma_g^{(H-1)}} & \\ & & & & \boldsymbol{\Sigma}_{\Gamma_g^H} \end{bmatrix}, \tag{25}$$

containing the variances and covariances of the regression coefficients within each historical data set.

Longitudinal Extensions

- Finally, let $\mathbf{\Pi}^H$ represent the H historical school-level regression coefficients and let $\mathbf{\Pi}^0$ represent the school-level regression coefficient for the current cycle.
- The joint distribution of $\mathbf{\Pi}^{H+1}$ is also assumed to be multivariate normal with mean $\mathbf{\Pi}$ and covariance matrix $\mathbf{T}_\Pi$ - namely

$$\mathbf{\Pi}^0, \mathbf{\Pi}^1, \ldots, \mathbf{\Pi}^{(H-1)}, \mathbf{\Pi}^H \sim N(\mathbf{\Pi}, \mathbf{T}_\Pi), \tag{26}$$

with

$$\mathbf{T}_\Pi = \begin{bmatrix} \mathbf{\Sigma}_{\Pi^0} & & & & \\ & \mathbf{\Sigma}_{\Pi^1} & & & \\ & & \ddots & & \\ & & & \mathbf{\Sigma}_{\Pi^{(H-1)}} & \\ & & & & \mathbf{\Sigma}_{\Pi^H} \end{bmatrix}. \tag{27}$$

Evaluating Predictions Under Historical Borrowing

- We evaluate historical borrowing for longitudinal data using two distinct approaches.

  1. In-sample simulations where the extant growth record is compared to the model-predicted growth record. We use Theil's measures.
  2. Pseudo out-of-sample performance where scoring rules are used to compare the predicted distribution of the outcome from the last wave of the current cycle to the known distribution of the outcome. We use KLD and LOOIC.

In-Sample Predictive Fit Measures

- Theil's Inequality Coefficient

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t^s - y_t^a)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t^a)^2}}, \tag{28}$$

- Theil's Bias Proportion

$$U^B = \frac{(\bar{y}^s - \bar{y}^a)^2}{\frac{1}{T}\sum_{i=1}^{T}(y_t^s - y_t^a)^2}, \tag{29}$$

- Theil's Variance Proportion

$$U^V = \frac{(\sigma^s - \sigma^a)^2}{\frac{1}{T}\sum_{i=1}^{T}(y_t^s - y_t^a)^2}, \tag{30}$$

Brief Results of the Longitudinal Case Study

- For predictive fit, different borrowing methods performed similarly for full samples and high poverty school samples. Theil's equality coefficients were nearly identical across all conditions.
- Pooling, AGDP, and power priors had relatively smaller RMSE between predicted and observed reading scores at the spring of 5^{th} grade and smaller KLD.
- No borrowing provided larger RMSE and KLD, but the smallest LOOICs.
- BDB with inverse-gamma prior (1, 0.001) had smaller LOOICs compared to pooling and power priors.
- AGDPs showed smaller RMSEs and KLDs compared to no borrowing and smaller LOOICs compared to pooling, power priors and BDB.

Brief Results of Longitudinal Simulation Study

- Most methods of historical borrowing perform similarly with respect to predictive performance and parameter recovery.
- Pooling and power priors performed relatively poorly across the conditions in this study, particularly when the current data and the historical data are heterogeneous.
- The findings regarding pooling under the homogeneous condition are surprising insofar as the homogeneous condition of our simulation study mimicked the desirable conditions for combining data from longitudinal studies as described in Hofer & Piccinin (2009)
- The current data and the historical data might still not be homogeneous enough.

Brief Results of Longitudinal Simulation Study

- Overall, the findings show that using aggregated data-dependent priors or simply using the current cycle of data with non-informative priors perform well.
- We only examined the realistic condition of one historical study. Kaplan, et al., (2022) found clear benefits of Bayesian dynamic borrowing over other methods of historical borrowing in the cross-sectional multilevel case study using PISA with five historical cycles, particularly with respect to prediction.
- Future research should expand the current study to consider many more cycles of longitudinal data.

Conclusions

- The overall pattern of results hold for varying student and school sample sizes
- Borrowing is not uniform across predictors. Sensitivity analyses are encouraged.
- Our results show that BDB is a prudent choice for combining information across studies, particularly with many historical cycles and when the degree of heterogeneity is either unknown or known to be extreme relative to the current data.
- A Shiny app to conduct Bayesian historical borrowing (BDB, power priors, pooling) can be found at `https://github.com/Bayesian-Methods-for-Education-Research/ShinyBHB`

Acknowledgements

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.

This research was performed using the compute 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.

The authors are grateful to Merve Sarac for valuable research assistance.

Thank You