Running head: CROSS-COUNTRY REGRESSION

Development and Application of Cross-Country Growth Regressions Using International Large-Scale Educational Assessments

David Kaplan
University of Wisconsin - Madison

Agnes Stancel-Piątak
Research and Analysis Unit
IEA Hamburg

Paper presented at the 2019 Annual Meeting of the Population Association of America.

DRAFT VERSION: PLEASE DO NOT CITE OR CIRCULATE

Abstract

This paper examines the utility of using international large-scale educational assessments for the purposes of forecasting country-level change over time in policy-relevant educational outcomes. We use the example of country-level change in math achievements for girls as an example. We adopt a fully Bayesian perspective by first estimating the change over time in girls' math achievement via Bayesian growth curve modeling with non-informative priors. Next, we regress the country estimated changes in girls math achievement on 15 explanatory variables. We account for model uncertainty through the use of Bayesian model averaging. Finally, we demonstrate the utility of our approach by forecasting the change over time in girls' math achievement for two countries. We close by arguing that our approach is useful for exploiting international large-scale educational assessments for purposes of prediction and forecasting, but we note that the collection of explanatory background variables are not currently designed for robust prediction. Future directions are discussed.

Development and Application of Cross-Country Growth Regressions Using International Large-Scale Educational Assessments

Introduction

Of critical importance to education policy is the monitoring of trends in educational outcomes over time. The United Nations Sustainable Development Goals identified Goal 4 as focusing on quality education for all. Goal 4.6 states

> “By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy."

If we wish to monitor progress toward these, and other, agreed-upon goals then it is necessary to develop optimally predictive models. It is a premise of this study that because international large-scale educational assessments (ILSAs) such as PISA, PIAAC, and TIMSS are longitudinal at the country level, they can be used to monitor trends in important educational outcomes. Inspired by Fernández, Ley, and Steele (2001b) in the economics domain, we apply a fully Bayesian cross-country growth growth modeling approach to TIMSS: Trends in International Mathematics and Science Study (Mullis, 2013). We demonstrate our approach to obtaining optimal prediction of growth by accounting for model uncertainty in the prediction of growth by employing Bayesian model averaging

The organization of this paper is as follows. In the next section we describe Bayesian growth curve modeling. This is followed by a discussion of Bayesian model averaging with attention paid to the elicitation of parameter and model priors. Next, we describe the data source and analysis steps for our example, focusing on change over time in girls mathematics achievement. This is followed by the results and then the discussion, where we outline the opportunities and challenges of using international large-scale educational data for prediction and forecasting.

Example: Change Over Time in Math Achievement for Girls

The data for this demonstration come the Trends in International Mathematics and Science Study TIMSS (Mullis, 2013). For countries with data back to 1995, TIMSS 2015 provides the sixth in a series of trend measures collected over 20 years. We restrict our attention to 5 waves of TIMSS starting with 1999 to provide more complete data. Our analysis sample consists of twenty-three countries that span the entire range of global GDP. Note that this is a very small sample size for our proposed demonstration. The outcome variable is **GirlsMathAchXX**, the country-level math achievement scores (first PV) for girls for the five waves (XX = 99, 03, 07, 11, 15) of TIMSS.

Bayesian Growth Curve Modeling

The first step in developing optimally predictive models of change over time in math achievement for girls is to obtain estimates of rate of change over all countries and for each country separately. To this end, we use the method of Bayesian growth curve modeling. We write the intra-country model as

$$y_{ti} = \pi_{0_i} + \pi_{1_i} a_{ti} + r_{ti} \tag{1}$$

where, for this paper y_{ti} is the average math achievement score for girls in country i at time t of the survey, π_{0_i} is the country-specific math achievement score at the beginning of the survey, π_{1_i} is the country-specific rate of change in girls' math achievement over the time interval a_{ti}, and r_{ti} is the residual term for country i at time t. Considerable flexibility is permitted in the specification of the time interval a_{ti}. For this study, the time intervals are specified to be 4 years apart corresponding to the cycle of the TIMSS assessment and is the same for each country. Thus, we can drop the individual country subscript i and write the time interval as a_t.

The inter-country model can be written generally in terms of a function of predictors of growth as

$$\pi_{qi} = \beta_{q0} + \sum_{k=1}^{Kq} \beta_{qk} x_{qi} + \epsilon_{qi}, \tag{2}$$

where π_{qi} are the q random coefficients (intercept π_{0_i} and rate of change π_{1_i}) that vary across the i countries, β_{q0} is the regression intercept, β_{qk} is the regression coefficient of the growth parameters on predictors x_{qi} for country i, and ϵ_{qi} is the regression disturbance terms. The model in Equation 2 is flexible enough to allow the growth parameters to be predicted by country level time-invariant covariates. For this study, we will concern ourselves with the prediction of the rate of change in country level math achievement for girls, π_{1_i}

Bayesian growth curve modeling requires priors on model parameters. For this paper, we use default non-informative priors on all model parameters. We use the R software program “blavaan" (Merkle & Rosseel, 2018), a latent variable modeling program that interfaces with JAGS (Plummer, 2016) to produce posterior distributions of the growth parameters

Bayesian Model Averaging

The Bayesian framework recognizes uncertainty in the choice of models used to predict and forecast growth. The uncertainty is due to not knowing whether the chosen model for predicting change in girls’ math achievement is the true data-generating model. Not accounting for model uncertainty can lead to “over confident inferences and decisions that are more risky that one thinks they are" (pg. 382 Hoeting, Madigan, Raftery, & Volinsky, 1999). A Bayesian approach to addressing the problem of model uncertainty is *Bayesian model averaging.*

Bayesian model averaging has had a long history of theoretical developments and practical applications. Early work by Leamer (1978) laid the foundation for Bayesian model averaging. Fundamental theoretical work on Bayesian model averaging was conducted in the mid-1990s by Madigan and his colleagues (e.g., Madigan & Raftery, 1994; Raftery, Madigan, & Hoeting, 1997; Hoeting et al., 1999). Additional theoretical work was conducted by Clyde (1999). Draper (1995) has discussed how model uncertainty can arise even in the context of experimental designs, and Kass and Raftery (1995) provide a review of Bayesian model averaging and the costs of ignoring model uncertainty. A more recent review of the general problem of model uncertainty can be found in Clyde and George (2004). Bayesian model averaging has been implemented in the R software programs “BMA" (Raftery, Hoeting, Volinsky, Painter, & Yeung, 2015) and “BMS" (Zeugner & Feldkircher, 2015).

In addition to theoretical developments, Bayesian model averaging has been applied to a wide variety content domains. A perusal of the extant literature shows Bayesian model averaging applied to economics (e.g., Fernández et al., 2001b), bioinformatics of gene express (e.g., Yeung, Bumbarner, & Raftery, 2005), weather forecasting (e.g., Sloughter, Gneiting, & Raftery, 2013), and causal inference within the propensity score framework (Kaplan & Chen, 2014; Wang, Parmigiani, & Dominici, 2012; Zigler & Dominici, 2014), to name just a few. An overview of Bayesian model averaging with applications to education policy research can be found in Kaplan and Lee (2018).

BMA Specification

To begin, let M_k, $k = 1, 2, \ldots, K$ be a set of competing models of growth. The posterior distribution of a predicted growth, Υ, given data y can be written as

$$p(\Upsilon|y) = \sum_{k=1}^{K} p(\Upsilon|M_k)p(M_k|y). \tag{3}$$

where $p(M_k|y)$ is the posterior probability of model M_k written as

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{l=1}^{K} p(y|M_l)p(M_l)}, \qquad l \neq k. \tag{4}$$

$p(y|M_k)$ is the integrated likelihood and $p(M_k)$ is the prior on the space of models. A key insight into BMA is that the quantity $p(M_k|y)$ is a measure of the probability that model M_k is the true data-generating model after having observed the data y, and of course, this quantity will likely be different for different growth models. Thus $p(M_k)$ expresses the uncertainty in model choice and are used as weights in the summation given in equation (3).

Default Priors

As with Bayesian growth curve modeling, BMA requires that priors be place not only on model parameters, but also on the space of models that could have possible generated the data. In the analysis we describe below, *unit information priors* are placed on model parameters. Unit information priors are weakly informative (data-based) priors that are diffused over the region of the likelihood where parameter values are considered mostly plausible, but not overly spread out. They can be considered priors for an individual with unbiased but weak prior information (Hoff, 2009). The unit information prior is equivalent to Zellner's g-prior (Zellner, 1986), where $g = 1/N$, and where N is the sample size.

Regarding the space of possible models, it is assumed that all models are equally likely to be the true model, and indeed, it is assumed that the true model is one of the models in the set – the so-called M-closed framework. Therefore, priors on the model space are are assumed equivalent for all models – namely, $1/M$, where M is the number of models. Other parameter and model priors can be specified.

Computational Issues

As pointed out by Hoeting et al. (1999), Bayesian model averaging is difficult to implement. In particular, they note that the number of terms in equation (3) can be quite large, the corresponding integrals are hard to compute, the specification of $p(M_k)$ may not be straightforward, and choosing the class of models to average over is also challenging.

The problem of reducing the overall number of models that one could incorporate in the summation of equation (3) has led to several interesting solutions. The solution used in this study is based on is based on the Metropolis-Hastings algorithm and is referred to as *Markov chain Monte Carlo Model composition* (MC^3).

Following Hoeting et al. (1999), the MC^3 algorithm proceeds as follows. First, let $\mathcal{M}$ represent the space of models of interest; in the case of our study, this would be the space of all possible combinations of explanatory variables used to predict change in girls' math achievement. The manner in which models are retained under MC^3 is as follows. First, for any given model currently explored by the Markov chain, say M_i, we can define a neighborhood for that model which includes one more variable and one less variable than the current model. So, for example, if our model has four predictors x_1, x_2, x_3 and x_4, and the Markov chain is currently examining the model with x_2 and x_3, then the neighborhood of this model would include $\{x_2\}$, $\{x_3\}$, $\{x_2, x_3, x_4\}$, and $\{x_1, x_2, x_3\}$. Now, a transition matrix is formed such that moving from the current model M_i to a new model M_j has probability zero if M_j is not in the neighborhood of M_i and has a constant probability if M_j is in the neighborhood of M_i. The model M_j is then accepted for model averaging with probability

$$\min\left\{1, \frac{pr(M_j|y)}{pr(M_i|y)}\right\}, \tag{5}$$

otherwise, the chain stays in model M_i. This form of MC^3 is also referred to as the *birth-death* sampler (Zeugner & Feldkircher, 2015)

Design

The steps in our approach are as follows.

1. Model the change in the girls math achievement and obtain estimates of change in each country using Bayesian growth curve modeling.

2. Validate the fit of the growth curve model.

3. Regress the country-level changes in girls math achievement on 15 country level predictors (averaged across cycles).

4. Use Bayesian model averaging to account for uncertainty in the prediction model.

5. Use BMA results to predict changes girls math achievement in two countries based on information from the remaining countries as well as the predictors of the two countries.

For this paper, we use the “BMS" software package (Zeugner & Feldkircher, 2015) in R. The *Markov Chain Monte Carlo Model Composition – M^3* (Madigan and Raftery, 1994) is used to reduce the space of possible models.Benchmark unit information priors based on Zellner’s g factor for the parameters based on Fernandez, Ley, & Steele (2001a). We also use the uniform prior on the space of models. Many other choices are possible.

It is important to note that this is a demonstration example only, and no presumption is made regarding the policy importance of results. Specifically, TIMSS (and other ILSAs) was not specifically designed for the purpose of probabilistic forecasting. The issue stems from the availability of relevant background explanatory variables developed for ILSAs. A thorough discussion of the design of background questionnaires for ILSAs is given in Kuger, Klieme, Jude, and Kaplan (2016), however what should be noted that these questionnaires were not developed for purposes of probabilistic forecasting. In addition to our outcome of interest, the 15 explanatory variables used in this study are give in Table 1.

Results

To begin, it may be of interest to provide a simple plot of country-level math achievement score for girls over the five waves of TIMSS to get a sense of the general trend. An inspection of Figure 1 shows that the change in mathematics achievement for girls is relatively flat. Figure 2 shows the empirical growth trajectories in girls math achievement for each country. Here we see that for most countries, the change over time in girls math achievement is also relatively flat, but some variability can be seen in countries such as #376.

The results of the Bayesian growth curve analysis are given in Table 2. Under the specification of non-informative priors (provided in the last column of Table 2), we find that the average 1999 math achievement score is approximately 500 with a small positive slope of .124 over the five waves of this study. The 95% posterior probability interval around the slope indicates that zero is a credible value for the slope, which is consistent with the visual inspection of Figure 1.

The fitted growth trajectory across countries and for each country separately can be see in in Figures 3 and 4. We see that the linear model provides a reasonably good fit to the empirical trajectory of girls' math achievement.

OLS Comparative Results

For comparison purposes, Table 3 provides the results of an ordinary least squares regression of the posterior slope values of each country on the predictors given in Table 1. Note that none of the predictors in this model reach conventional significance levels. The overall model R^2 is 0.70 which is statistical significant at the 0.05 level. However, it is important to note, that the OLS model does not account for model uncertainty. Rather, it is assumed that this OLS model is the true data generating model. Moreover, OLS can only provide a dichotomization of the evidence - namely whether a predictor is statistically significant or not.

BMA Results

Table 4 presents the BMA results for the predictors of change in girls math achievement. The interpretation of this table rests on recalling that BMA searches over a large space of possible models based on combinations of predictors, and weights each model by the posterior probability of each model. In this example, there were $2^{15} = 32,736$ possible models (not accounting for interactions). Under the algorithm, 2,011 models were visited. The cummulative posterior model probability across all of the models was 0.01, which is very small and indicates considerable model uncertainty. This degree of uncertainty is most probably due to the fact that these explanatory variables were not designed to predict and forecast change in math achievement for girls.

Continuing with the demonstration, the predictor Beh8avg (mean level of vandalism in the school) was found to have a posterior inclusion probability (PIP) of 0.70, meaning that this predictor appeared in 70% of the models explored. There is no specific rule-of-thumb regarding the importance of the variable on the basis of the PIP, and must be judged substantively. The averaged coefficient (post. mean) for Beh8Avg is -1.02 with a posterior standard deviation of 0.85 The column labeled (Cond.Pos.Sign) is the conditional probability that the coefficient is 0.00 given the model. Here we see that the conditional probability that the coefficient is positive is very small, indicating that the sign of the coefficient is very likely negative. Thus, countries reporting, on average, larger levels of vandalism are associated with decreasing change over time in girls' math achievement. The remaining predictors indicate very little importance as judged by the posterior inclusion probability – a result consistent with the OLS findings in Table 3. Again, it is important to note that this example is based on only 23 countries and the predictors used in this model were not developed to address the policy issue of change over time in girls' math achievement.

Cross-country specific results

It may be of interest to explore more deeply the impact of mean level of vandalism in the school on change in girls' math achievement. Figure 5 displays the posterior distributions of the regression coefficient relating mean reported amount of school vandalism (left) and mean difference in math self-concept between boys and girls (right) on the change in girls' math achievement. These posterior distributions are a result of adopting a Bayesian perspective to the problem. In particular, Figure 5 shows that Beh8avg is almost normal with a mean and median approximately equal. Moreover, we find that the value of zero lies outside the 95% posterior probability interval while we find that for the effect of country level average self-concept difference, zero lies in the 95% posterior probability interval. It is important to note that the information conveyed in diagrams such as these as well as the information in Table 4 allows one to address other intervals of interest. For example, with respect to the self-concept difference, it may be less important to know if zero is in the 95%

posterior interval, and instead calculate the probability that the effect is greater than some number of interest. Such a fine-grained analysis of the impact of these regressors on our math achievement outcome is not possible from a frequentist framework.

Country-specific forecasts

The primary goal of this paper is to examine whether ILSAs have utility in forecasting trends in important educational outcomes – in our example, forecasting trends in country-level math achievement for girls. To this end, our approach allows us to choose a particular country of interest and use the information provided by remaining countries as well as the explanatory data from the country of interest to forecast the trend. Figure 6 provides the forecasted change in girls' math achievement for country 840 (left) and country 764 (right). Based on the 95% posterior probabilities, we find fhat for country 840 a predicted change in girls' math achievement between approximately -1 and 1 is quite likely.

For country 764, another story emerges. The expected change in girls' math achievement is negative; however, we find that the actual values are quite far from the expected forecast. This suggests that either the forecasting model is inappropriate for country 764 or that this country is an outlier. To investigate this further, one could change the forecasting model settings. For example, one could choose a variety of combinations of priors on the model parameters as well as priors on the space of models itself (Fernández, Ley, & Steele, 2001a).

Conclusions

This paper demonstrates the potential of using TIMSS (and ILSAs generally) for Bayesian probabilistic forecasting, accounting for uncertainty in models and model parameters. A key advantage of BMA compared to the choice of a single model is that over long forecast periods, BMA is known to provide better predictive performance than choosing any single model. In addition, adopting a fully Bayesian approach to prediction and forecasting, unlike frequentist approaches, yields the full probability distribution of regression effects and the full probability distribution of forecasted values. Studying these full probability distributions provide richer information than frequentist approaches which tend to lead to a dichotomization of evidence. It is important to also note that in the case of outliner forecasts, alternative forecast settings can be examined and predictive accuracy can be compared. Such comparisons were outside the scope of this paper. Finally, we note again that this demonstration suffered from the fact that the background explanatory variables were never developed for the purposes of probabilistic forecasting. Future ILSA development should consider country-level BQ indicators for developing probabilistic educational forecasting models.

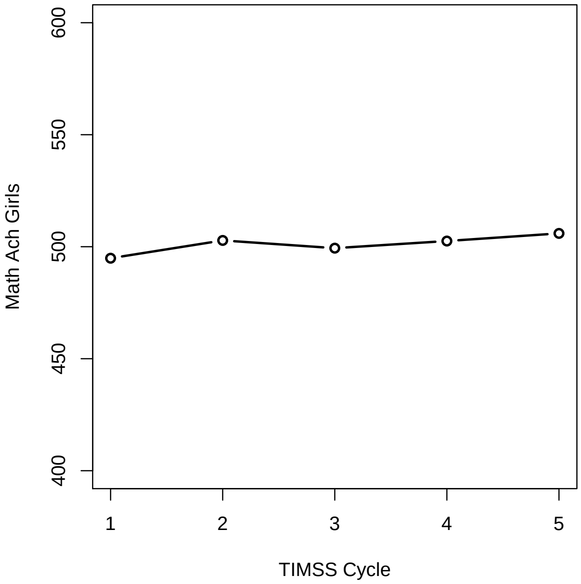

Figure 1. *Average girls math achievement across countries and cycles - first math plausible value.*

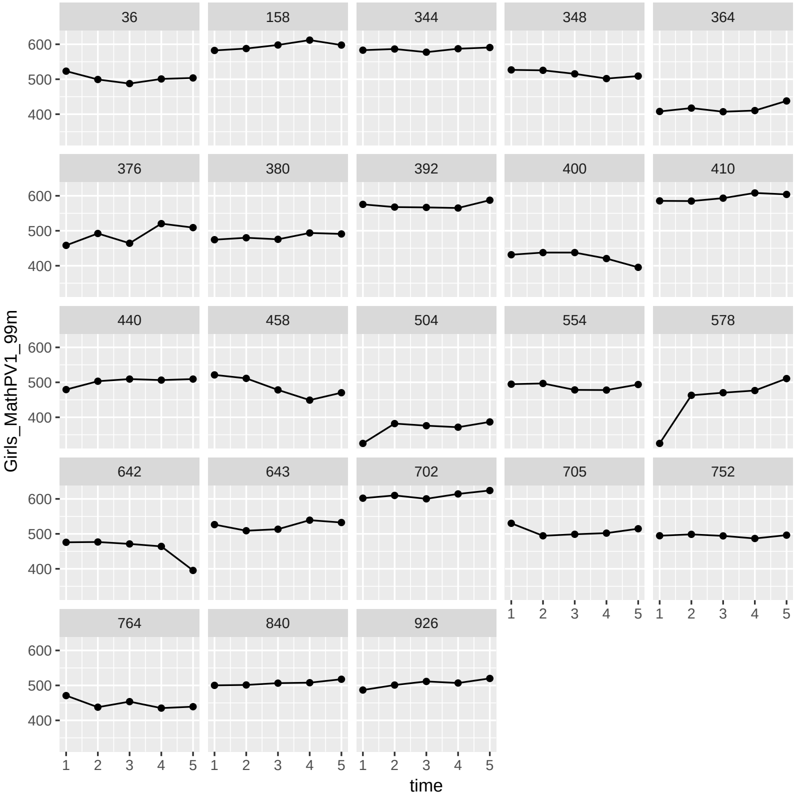

Figure 2. *Average girls math achievement for each country - first math plausible value*

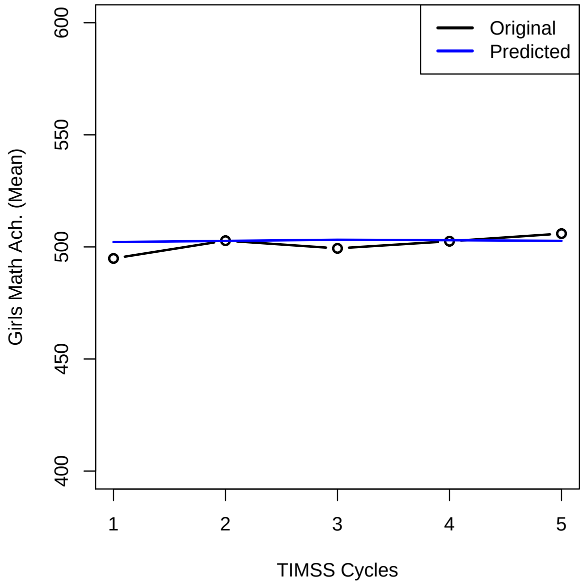

Figure 3. *Model predicted changes in girls math achievement across countries and cycles - first math plausible value.*

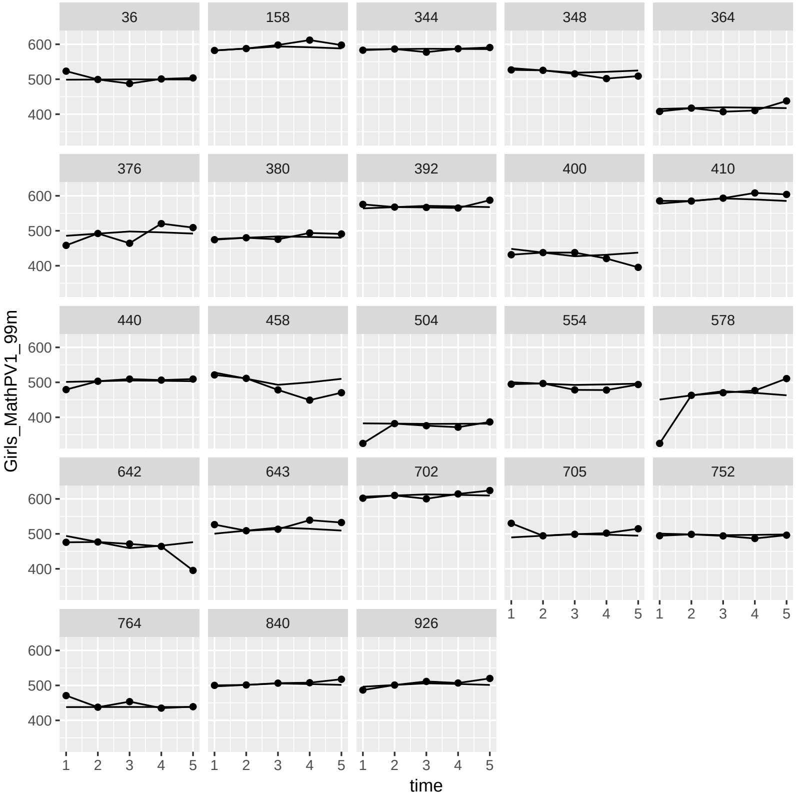

Figure 4. *Model predicted change in girls math achievement for each country – first math plausible value.*

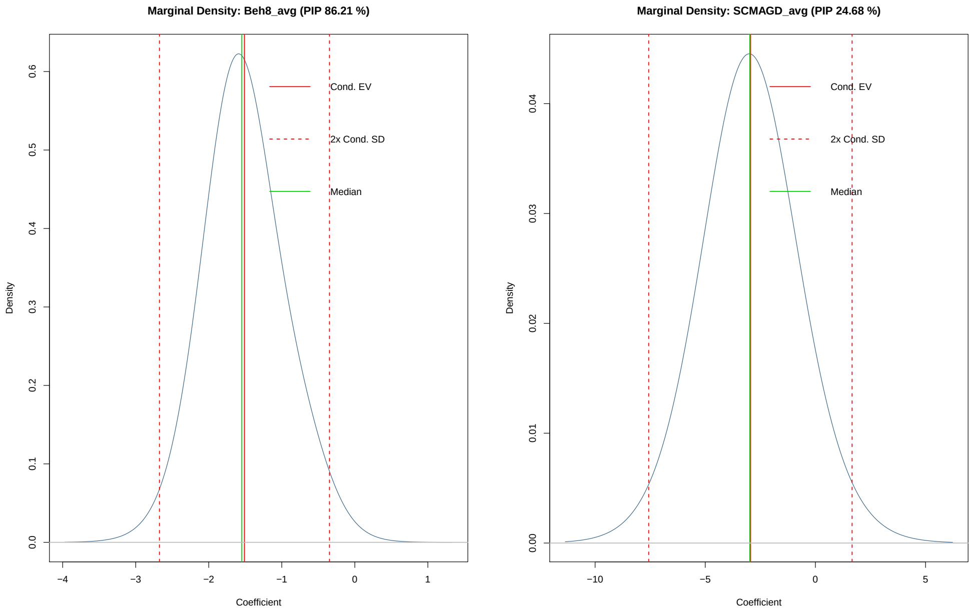

Figure 5. *Marginal density the regression of mean level of school vandalism (left) and math self-concept difference between boys and girls (right) the change in the girls math achievement.*

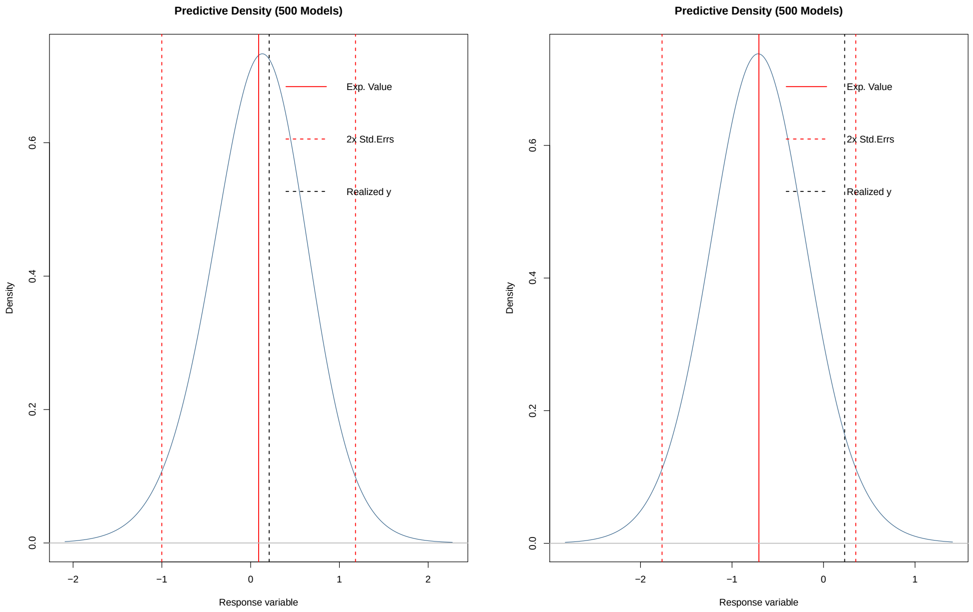

Figure 6. Predictive change in girls math achievement for country 840 (L) and country 764(R).

References

Clyde, M. A. (1999). Bayesian model averaging and model search strategies. In *Bayesian statistics 6* (pp. 157–185). Oxford: Oxford University Press.

Clyde, M. A., & George, E. I. (2004). Model uncertainty. *Statistical Science*, *19*, 81–94.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society (Series B)*, *57*, 55–98.

Fernández, C., Ley, E., & Steele, M. F. J. (2001a). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*, 381–427.

Fernández, C., Ley, E., & Steele, M. F. J. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, *16*, 563–576.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382–417.

Kaplan, D., & Chen, J. (2014). Bayesian model averaging for propensity score analysis. *Multivariate Behavioral Research*, *49*, 505–517.

Kaplan, D., & Lee, C. (2018). Optimizing prediction using Bayesian model averaging: Examples using large-scale educational assessments. *Evaluation Review*. doi: 10.1177/0193841X18761421

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kuger, S., Klieme, E., Jude, N., & Kaplan, D. (2016). *Assessing contexts of learning world-wide – Extended context assessment frameworks.* Dordrecht: Springer.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data.* New York: Wiley.

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainly in graphical models using Occam's window. *Journal of the American Statistical Association*, *89*, 1535–1546.

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(4), 1–30. doi: 10.18637/jss.v085.i04

Mullis, I. V. S. (2013). Foward. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks.* Boston, MA: TIMSS & PIRLS International Study Center Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Plummer, M. (2016). rjags: Bayesian graphical models using MCMC [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=rjags` (R package version 4-6)

Raftery, A. E., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. (2015, June 22). *Bayesian model averaging (BMA), version 3.12.* http://www2.research.att.com/ volinsky/bma.html.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*, 179–191.

Sloughter, J. M., Gneiting, T., & Raftery, A. E. (2013). Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Monthly Weather Review*, *141*, 2107–2119.

Wang, C., Parmigiani, G., & Dominici, F. (2012). Bayesian effect estimation accountng for adjustment uncertainty. *Biometrics*, *68*, 661–671.

Yeung, K. Y., Bumbarner, R. E., & Raftery, A. E. (2005). Bayesian model averaging: development of an improved multi-class, gene selection, and classification tool for microarray data. *Bioinformatics*, *21*, 2394–2402.

Zeugner, S., & Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS package for R. *Journal of Statistical Software*, *68*(4), 1–37. doi: 10.18637/jss.v068.i04

Zigler, C. M., & Dominici, F. (2014). Uncertainty in propensity score estimation: Bayesian methods for variable selection and model-averaged causal effects. *Journal of the American Statistical Association*, *109*, 95–107.

Author Note

Paper presented at the 2019 Annual Meeting of the Population Association of America. The authors would like to thank Dr. Justin Wild for valuable assistance with data preparation.

Table 1

Explanatory variables for BMA

Name	Definition
Migavg	Percent students born outside of country
Migparavg	Percent both parents born outside of country
Langavg	Percent students who do not speak lang. of tests at home
BOK1avg	Percent students with more than 100 books at home
SCSGDavg	Mean male - female self conf. in overall science ability
SCMAGDavg	Mean male - female self conf. in overall math ability
MathShort2avg	mean shortage of calculators
MathShort1avg	Mean shortage of computer software
MathShort3avg	mean shortage of library materials
Beh8avg	Mean level of school of vandalism
Beh10avg	Mean level of intimidation of students
Beh11avg	Mean level of injury of students
Beh12avg	Mean level of intimidation of teachers
Beh13avg	Mean level of injury to teachers
HDIavg	Human development index

Note: Predictors are averaged over time.

Table 2

Selected growth curve regression results

Parameter	Estimate	Post.SD	HPD.025	HPD.975	PSRF	Prior
intercept	500.135	3.068	494.226	506.233	1.000	dnorm(500,.1)
slope	0.124	0.473	-0.792	1.084	1.001	dnorm(0,1e-2)
Var(intercept)	2996.778	924.099	1484.445	4807.966	1.000	dwish(iden,3)
Var(slope)	4.441	1.856	1.229	8.135	1.002	dwish(iden,3)
Cov(int, slp)	4.074	26.302	-48.177	57.869	1.000	dwish(iden,3)

Note: Post.SD=Posterior standard deviation; HPD=Highest Posterior Density; PSRF=Potential Scale Reduction Factor; Prior = Prior distribution on model parameters.

Table 3

Comparative ordinary least squares results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.2479	7.6364	0.43	0.6834
Migavg	-5.6588	13.2035	-0.43	0.6811
Migparavg	5.9352	8.6090	0.69	0.5128
Langavg	-1.2739	1.6067	-0.79	0.4539
BOK1avg	1.2100	2.1677	0.56	0.5941
SCSGDavg	-2.0534	3.0639	-0.67	0.5242
SCMAGDavg	0.9237	4.3578	0.21	0.8382
MathShort1avg	-1.3632	1.5394	-0.89	0.4052
MathShort2avg	-1.1217	1.0129	-1.11	0.3047
MathShort3avg	2.3700	1.3881	1.71	0.1315
Beh8avg	-2.0428	1.2479	-1.64	0.1457
Beh10avg	-0.8360	1.7367	-0.48	0.6450
Beh11avg	2.3442	2.6054	0.90	0.3981
Beh12avg	0.7686	2.4958	0.31	0.7671
Beh13avg	0.1933	2.3494	0.08	0.9367
HDIavg	-3.9166	5.4068	-0.72	0.4923

Table 4

BMA results

	PIP	Post Mean	Post SD	Cond.Pos.Sign
Beh8avg	0.70	-1.02	0.85	0.00
Beh11avg	0.55	1.09	1.25	0.99
SCSGDavg	0.41	-1.15	1.81	0.00
MathShort3avg	0.40	0.38	0.67	0.96
MathShort2avg	0.39	-0.32	0.55	0.02
BOK1avg	0.31	0.36	0.91	0.98
Migavg	0.26	-0.35	2.93	0.42
SCMAGDavg	0.26	-0.65	1.71	0.04
Beh13avg	0.25	0.20	0.68	0.90
HDIavg	0.23	-0.41	1.68	0.19
Langavg	0.23	-0.24	0.65	0.02
MathShort1avg	0.22	0.01	0.41	0.54
Beh10avg	0.22	-0.03	0.31	0.27
Beh12avg	0.21	-0.06	0.47	0.22
Migparavg	0.17	0.29	1.80	0.71

Note: PIP=Posterior inclusion probability, Post Mean = BMA regression coefficient, Post SD = posterior standard deviation, Cond.Pos.Sign = posterior probability of a positive coef. conditional on inclusion. The posterior model probability of the top model was 0.011, indicating considerable model uncertainty.