Conditionally Unbiased Best Linear Predictors for Score Augmentation

Xiang Liu, Matthew S. Johnson, and Sandip Sinharay

Educational Testing Service

Abstract

The best linear predictor (BLP) has been proposed to combine different types of information in estimating true scores. The BLP is biased for individual examinees when conditioned on the true score. In this paper, we propose a conditionally unbiased BLP. Additionally, a least square method is introduced for parameter estimation.

Conditionally Unbiased Best Linear Predictors for Score Augmentation

Introduction

With the current popular trend of transition from traditional paper-pencil assessments to digital formats, an increasingly common task is to combine different sources of information in assessing some latent construct. For instance, in the context of writing assessments, in addition to the final product essays, digital platforms are now able to capture information on keystrokes during students' interaction with writing tasks. Thus, features related to writing process are now becoming available (e.g. Guo et al., 2019; Zhang et al., 2019). Furthermore, additional product features such as those related to natural language processing (NLP) could also be obtained through NLP programs (e.g. the e-rater® program; Attalí and Bursteín, 2006; Burstein et al., 2004). Then a natural question is whether and how different sources of information can be combined to augment the rater scores in making inferences on examinees' writing proficiency.

There has been some previous research exploring related methods. For example, Zhang and Deane (2015) used linear regression to predict adjudicated essay scores from a large number of keystroke logging features and other product features. Sinharay et al. (2019) examined the same prediction problem utilizing machine learning methods such as the random forests and boosting. A common assumption of these methods is that essay score, the prediction target, is treated as fixed and known. However, in educational and psychological measurement, prediction targets are almost always latent and measured with error. To address this issue, the best linear predictors (BLP; e.g. Haberman et al., 2015; Yao et al., 2019) has been proposed to combine sources of information (e.g. scores from other sections) along with manifest variables such as the rater scores to predict some latent true score (e.g. the writing proficiency). It should be noted that, due to some historical reasons, in the field of animal breeding where the methods of BLP and BLUP were first popularized, random effects are said to be "predicted" as opposed to be "estimated" for fixed effects. However, in our field of psychometrics and educational/psychological measurement, such distinction is not customary. Thus both terms, prediction and estimation, are used synonymously and interchangeably in this paper.

The BLP minimizes the mean squared error (MSE) of prediction among all linear predictors. An important property of the BLP is that, when averaged over the population, the expected prediction is equal to the expected true score. In other words, the BLP is unbiased. However, it exhibits shrinkage towards the population mean. Consequently, the BLP is biased conditional on the true score level (Robinson, 1991). This may lead to some serious fairness concern as the BLP could favor certain groups of examinees over others.

In this paper, we introduce a conditionally unbiased best linear predictor (CUBLP) for estimating latent variable from multiple sources that minimizes the prediction error among all linear predictors that are unbiased at all true score levels. The proposed method is flexible and can be applied to a wide variety of problems.

Methods

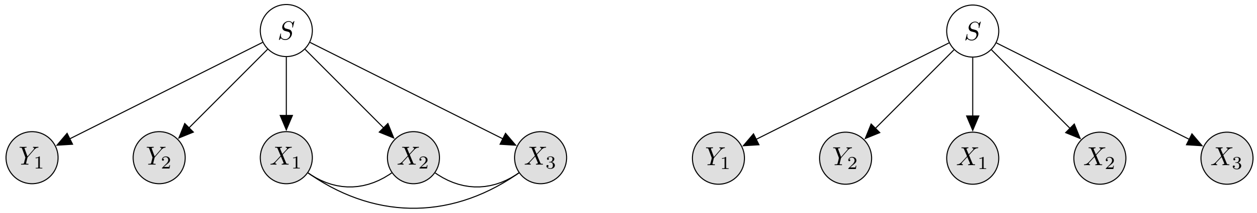
Let $\boldsymbol{W} = (\boldsymbol{Y}^\top, \boldsymbol{X}^\top)^\top$ denote observed random variables from a random examinee where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_J)^\top$ is the vector of variables measured with error (commonly referred to as manifest variables of some latent trait, e.g. essay scores from raters). and $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)^\top$ is the vector of variables measured without error (i.e. covariates that may correlate with the latent trait of interest, e.g. typing speed). For simplicity, we consider an unidimensional latent true score S and further assume that $\boldsymbol{W}$ and S are linearly related, i.e.

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix} = \boldsymbol{\alpha} + \boldsymbol{\lambda} S + \boldsymbol{\varepsilon}_w. \tag{1}$$

It is mathematically more convenient and simplifies discussions when $\boldsymbol{W}$ is centered, $E[\boldsymbol{W}] = 0$, which leads to

$$\boldsymbol{W} = \boldsymbol{\lambda} S + \boldsymbol{\varepsilon}_w. \tag{2}$$

For identifiability, the following constraints are imposed, $\mathbb{E}(S) = 0$ and $\mathbb{E}(\boldsymbol{\varepsilon}_w) = \mathbf{0}$. Furthermore, the variance-covariance matrix of the random effects, the latent variable S



(a) *Correlated residuals of X* (b) *Uncorrelated residuals of X*

Figure 1

Graphical representation of some example models

and the residuals $\boldsymbol{\varepsilon}_w$, is in the form

$$\mathrm{Cov}\begin{pmatrix} S \\ \boldsymbol{\varepsilon}_w \end{pmatrix} = \begin{bmatrix} \sigma_S^2 & 0 \\ 0 & \boldsymbol{\Sigma}_{\varepsilon_w} \end{bmatrix}. \tag{3}$$

In other words, the cross-covariance of S and $\boldsymbol{\varepsilon}_w$ is zero, i.e. $\mathrm{Cov}(S, \varepsilon_{w_i}) = 0, \forall i$. Notice that, similar to Haberman et al. (2015), we do not assume a specific structure for the variance-covariance matrix of the residuals, $\boldsymbol{\Sigma}_{\varepsilon_w}$. Thus, depending on the structure of $\boldsymbol{\Sigma}_{\varepsilon_w}$, Equation 2 describes a broad class of models (see Figure 1 for examples). For almost all real applications, we are interested in estimating the latent score S from the observed random variables with some estimator, $\hat{S} = g(\boldsymbol{W})$.

The best linear predictors

The BLP estimator approaches the estimation problem by finding a linear combination of the observed random variables, $\hat{S} = \boldsymbol{\gamma}_1^\top \boldsymbol{W}$, such that the mean squared error for prediction,

$$\mathrm{MSE} = E[(S - \boldsymbol{\gamma}_1^\top \boldsymbol{W})^2], \tag{4}$$

is minimized. The MSE can be expanded as

$$\text{MSE} = E[S^2 + (\boldsymbol{\gamma}_1^\top \boldsymbol{W})^2 - 2S\boldsymbol{\gamma}_1^\top \boldsymbol{W}] \tag{5}$$

$$= E[S^2] + E[(\boldsymbol{\gamma}_1^\top \boldsymbol{W})^2] - 2\boldsymbol{\gamma}_1^\top E[S\boldsymbol{W}] \tag{6}$$

$$= \sigma_S^2 + \text{Var}(\boldsymbol{\gamma}_1^\top \boldsymbol{W}) - 2\boldsymbol{\gamma}_1^\top E[\boldsymbol{\lambda} S^2 + S\boldsymbol{\varepsilon}_w] \tag{7}$$

$$= \sigma_S^2 + \boldsymbol{\gamma}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{W}} \boldsymbol{\gamma}_1 - 2\boldsymbol{\gamma}_1^\top \boldsymbol{\lambda}\sigma_S^2, \tag{8}$$

where $\Sigma_{\boldsymbol{W}}$ is the variance-covariance matrix of the observed random variables $\boldsymbol{W}$. From Equation 7 to Equation 8, we assume $E[\boldsymbol{\varepsilon}_w|S] = 0$, thus $E[S\boldsymbol{\varepsilon}_w] = E[SE(\boldsymbol{\varepsilon}_w|S)] = 0$. Solving the gradient function with respect to $\boldsymbol{\lambda}_1$,

$$\nabla\,\text{MSE} = \nabla(\boldsymbol{\gamma}_1^\top \boldsymbol{\Sigma}_{\boldsymbol{W}} \boldsymbol{\gamma}_1) - \nabla(2\boldsymbol{\gamma}_1^\top \boldsymbol{\lambda}\sigma_S^2) \tag{9}$$

$$= 2\Sigma_{\boldsymbol{W}} \boldsymbol{\gamma}_1 - 2\boldsymbol{\lambda}\sigma_S^2 = 0, \tag{10}$$

leads to the BLP coefficient $\boldsymbol{\gamma}_1 = \Sigma_{\boldsymbol{W}}^{-1}\boldsymbol{\lambda}\sigma_S^2$. Therefore, the BLP estimator of the latent score is $\hat{S} = (\Sigma_{\boldsymbol{W}}^{-1}\boldsymbol{\lambda}\sigma_S^2)^\top \boldsymbol{W}$. The model in Equation 2 and its associated assumptions imply that the variance-covariance matrix of $\boldsymbol{W}$ can be decomposed as

$$\Sigma_{\boldsymbol{W}} = \sigma_S^2 \boldsymbol{\lambda}\boldsymbol{\lambda}^\top + \boldsymbol{\Sigma}_{\varepsilon_w}. \tag{11}$$

By the Woodbury matrix identity (Fill & Fishkind, 1999), this decomposition of $\Sigma_{\boldsymbol{W}}$ leads to an alternative expression of the BLP coefficients,

$$\boldsymbol{\gamma}_1 = \frac{\Sigma_{\varepsilon_w}^{-1}\boldsymbol{\lambda}}{1/\sigma_S^2 + \boldsymbol{\lambda}^\top \Sigma_{\varepsilon_w}^{-1}\boldsymbol{\lambda}}. \tag{12}$$

The bias of the BLP in estimating S is

$$Bias(\hat{S}) = E[\hat{S} - S] \tag{13}$$

$$= E[(\sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^{-1})\boldsymbol{W} - \lambda^{-1}\boldsymbol{W} + \boldsymbol{\lambda}^{-1}\boldsymbol{\varepsilon}_w] \tag{14}$$

$$= \sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^{-1} E[W] - \boldsymbol{\lambda}^{-1} E[\boldsymbol{W}] + \boldsymbol{\lambda}^{-1} E[\boldsymbol{\varepsilon}_w] = 0. \tag{15}$$

In other words, when averaged over the entire population, the BLP is unbiased in estimating S.

The conditional bias of the BLP

We showed that the BLP is marginally unbiased. However, there is no guarantee that it has the same unbiasedness property for each individual. To assess the bias for estimating the latent true score for each individual, we need to evaluate the bias of the BLP conditional on latent true score levels, i.e.

$$\begin{aligned} E[\hat{S} - S|S] &= E[\sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^{-1} \boldsymbol{W}|S] - S \\ &= E[\sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^{-1} (\boldsymbol{\lambda} S + \boldsymbol{\varepsilon}_w)|S] - S \\ &= \sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^{-1} \boldsymbol{\lambda} S - S \\ &= S(\sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^{-1} \boldsymbol{\lambda} - 1). \end{aligned} \tag{16}$$

Equation 16 is illuminating that the conditional bias depends on two factors - the latent true score level S, and the ratio to which the total variance-covariance in the manifest variables can be explained by the latent true score. The conditional bias of the BLP is zero for all latent score levels $S = s, \forall s \in \mathbb{R}$ if and only if $\sigma_S^2 \boldsymbol{\lambda}^\top \Sigma_{\boldsymbol{W}}^\top \boldsymbol{\lambda} = 1$. That is, the latent true score S perfectly explains the variance-covariance of the observed random variables $\Sigma_{\boldsymbol{W}}$. But, for almost all realistic applications, this perfect relationship is rarely attainable. Thus, given a model, the estimation bias of the BLP is inversely proportional to the true score level for any individual. Figure 2 illustrates the conditional bias of the BLP across levels of S under low, medium, and high variance-covariance ratios. While the BLP is less biased in the medium level of S, it is positively biased for individuals with lower levels of the latent true score and negatively biased for those with higher levels of the latent true score. The demonstrated conditional bias of the BLP estimator may create threats to fairness and equity.

Conditionally unbiased BLP

To mitigate the potential threats to fairness and equity from conditional bias of BLP, we develop a conditionally unbiased BLP. Instead of finding the linear predictor, $\hat{S} = \boldsymbol{\gamma}_1^\top \boldsymbol{W}$, which simply minimizes the MSE for prediction as in Equation 4, the CUBLP

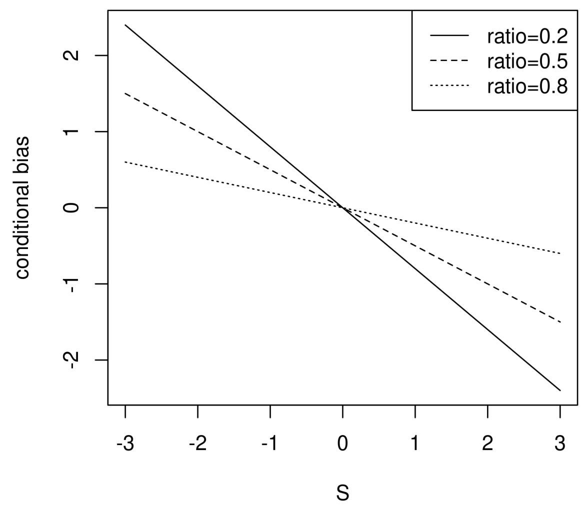


Figure 2

Conditional bias under different covariance ratios

is further subject to a constraint that it is conditionally unbiased, i.e.

$$E[\boldsymbol{\gamma}_1^\top \boldsymbol{W}|S] = \boldsymbol{\gamma}_1^\top E[\boldsymbol{W}|S] = \boldsymbol{\gamma}_1^\top(\boldsymbol{\lambda}S + E[\boldsymbol{\varepsilon}_w|S]) = S. \tag{17}$$

With the additional assumption that $E(\boldsymbol{\varepsilon}_w|S) = 0$, the constraint simplifies to

$$\boldsymbol{\gamma}_1^\top \boldsymbol{\lambda} = 1. \tag{18}$$

It follows that the CUBLP coefficients can be found by minimizing $\text{MSE} = E[(S - \boldsymbol{\gamma}_1^\top \boldsymbol{W})^2]$, subject to the constraint $\boldsymbol{\gamma}_1^\top \boldsymbol{\lambda} = 1$. This constrained optimization problem can be solved by using the method of Lagrange multipliers. The Lagrange function is defined as

$$\mathcal{L}(\boldsymbol{\lambda}_1, \delta) = -E[(S - \boldsymbol{\gamma}_1^\top \boldsymbol{W})^2] - \delta(\boldsymbol{\gamma}_1^\top \boldsymbol{\lambda} - 1), \tag{19}$$

where the Lagrange multiplier δ is a constant that ensures the constraint is met. The CUBLP coefficients $\boldsymbol{\lambda}_1$ is found by solving

$$\nabla_{\boldsymbol{\lambda}_1, \delta} \mathcal{L}(\boldsymbol{\lambda}_1^\top, \delta) = \mathbf{0}, \tag{20}$$

which is equivalent to solving the system of equations

$$\begin{cases} -2\boldsymbol{\Sigma_W}\boldsymbol{\gamma}_1 + 2\boldsymbol{\lambda}\sigma_S^2 = \delta\boldsymbol{\lambda} \\ \boldsymbol{\gamma}_1^\top\boldsymbol{\lambda} - 1 = 0 \end{cases}. \tag{21}$$

The solution leads to the CUBLP coefficients

$$\boldsymbol{\lambda}_1 = \frac{\boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1}\boldsymbol{\lambda}}{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1}\boldsymbol{\lambda}}. \tag{22}$$

Using the covariance matrix decomposition in Equation 11 and the Woodbury matrix identity, it is straightforward to verify that the CUBLP coefficients can be alternatively expressed in terms of the residual covariance matrix,

$$\boldsymbol{\lambda}_1 = \frac{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_w}\boldsymbol{\lambda}}{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_w}\boldsymbol{\lambda}}. \tag{23}$$

Parameter estimation

Notice that the derived BLP coefficients are expressed as a function of the population parameters $\boldsymbol{\lambda}$, and $\Sigma_{\boldsymbol{W}}$ or $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_w}$. In other words, they are treated as known. However, in most applications, these quantities would have to be estimated. We derive a least square estimator for the parameters in this section.

Let

$$\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda_Y} \\ \boldsymbol{\lambda_X} \end{pmatrix}$$

, where $\boldsymbol{\lambda_Y} = (\lambda_{Y_1}, \lambda_{Y_2}, \ldots, \lambda_{Y_K})'$ and $\boldsymbol{\lambda_X} = (\lambda_{X_1}, \lambda_{X_2}, \ldots, \lambda_{X_J})'$. Assume all observed variables are standardized and $E(S) = \mu_S = 0$ and $\sigma_S^2 = 1$. It leads to

$$\boldsymbol{\Sigma_W} = \boldsymbol{\lambda}\boldsymbol{\lambda}^\top + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}. \tag{24}$$

We further assume that the covariance of $\boldsymbol{Y}$ is fully explained by S. That is

$$\boldsymbol{\Sigma_Y} = \boldsymbol{\lambda_Y}\boldsymbol{\lambda_Y}^\top + \boldsymbol{\Psi_Y}, \tag{25}$$

where $\boldsymbol{\Psi_Y}$ is a diagonal matrix with residual variances of $\boldsymbol{Y}$ on the main diagonal. This is a common assumption for many popular measurement models. Now, we impose some additional restrictions on the residual covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. $\boldsymbol{Y}$ is assumed to be uncorrelated with $\boldsymbol{X}$ given S. Equivalently, $\mathrm{Cov}(\boldsymbol{\varepsilon}_{\boldsymbol{Y}}, \boldsymbol{\varepsilon}_{\boldsymbol{X}}) = \mathbf{0}$.

In many cases, it may be desirable or necessary to assume that $\lambda_{Y_j} = \lambda_Y$, $\forall j$. This equal discrimination assumption also simplifies the estimation of $\boldsymbol{\lambda}$. Consider a quadratic loss function,

$$L(\boldsymbol{\lambda}) = \sum_{k \neq k'} (\lambda_Y^2 - r_{Y_k Y_{k'}})^2 + \sum_k \sum_j (\lambda_Y \lambda_{X_j} - r_{Y_k X_j})^2, \tag{26}$$

where $r_{m,n}$ denotes the observed correlation between random variables m and n. Taking the gradient to minimize the loss function,

$$\nabla L = \begin{pmatrix} 4\lambda_Y \sum_{k\neq k'} (\lambda_Y^2 - r_{Y_k Y_{k'}}) + 2K\lambda_Y \sum_j \lambda_{X_j}^2 - 2\sum_j \lambda_{X_j} \sum_k r_{Y_k X_j} \\ \vdots \\ 2K\lambda_Y^2 \lambda_{X_j} - 2\lambda_Y \sum_k r_{Y_k X_j} \\ \vdots \end{pmatrix} = \mathbf{0}. \tag{27}$$

Solving Equation 27 leads to the unweighted least square estimator of the loadings,

$$\hat{\lambda_Y} = \sqrt{\frac{\sum_{k\neq k'} r_{Y_k Y_{k'}}}{K(K-1)}} \tag{28}$$

and

$$\hat{\lambda_{X_j}} = \frac{\sum_k r_{Y_k X_j}}{K\hat{\lambda_Y}}. \tag{29}$$

For unequal discrimination cases, the LS estimator can be obtained by iterating through

$$\hat{\lambda_{Y_k}} = \frac{\sum_{k'\neq k} \lambda_{Y_{k'}} r_{Y_k Y_{k'}} + \sum_j \lambda_{X_j} r_{Y_k, X_j}}{\sum_{k'\neq k} \lambda_{Y_{k'}}^2 + \sum_j \lambda_{X_j}^2}, \tag{30}$$

and

$$\hat{\lambda_{X_j}} = \frac{\sum_k \lambda_{Y_k} r_{Y_k X_j}}{\sum_k \lambda_{Y_k}^2} \tag{31}$$

until convergence.

Without any parametric distribution assumption on the data, the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{W}}$ is commonly estimated by the sample covariance matrix,

$$\hat{\boldsymbol{\Sigma}_{\boldsymbol{W}}} = \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{W}_i - \bar{\boldsymbol{W}})(\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top, \tag{32}$$

where N is the number of observations. If a multivariate Gaussian assumption is reasonable, the ML may be preferred, i.e.

$$\hat{\boldsymbol{\Sigma}_{\boldsymbol{W}}} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{W}_i - \bar{\boldsymbol{W}})(\boldsymbol{W}_i - \bar{\boldsymbol{W}})^\top. \tag{33}$$

A natural choice of an estimator for the residual covariance matrix is then

$$\hat{\boldsymbol{\Sigma}_{\varepsilon_w}} = \hat{\boldsymbol{\Sigma}_{\boldsymbol{W}}} - \hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^\top. \tag{34}$$

By substituting the population parameters with their estimates from the data, the BLP and CUBLP estimators become the empirical BLP and CUBLP estimators.

Simulations

Properties of the LS estimator

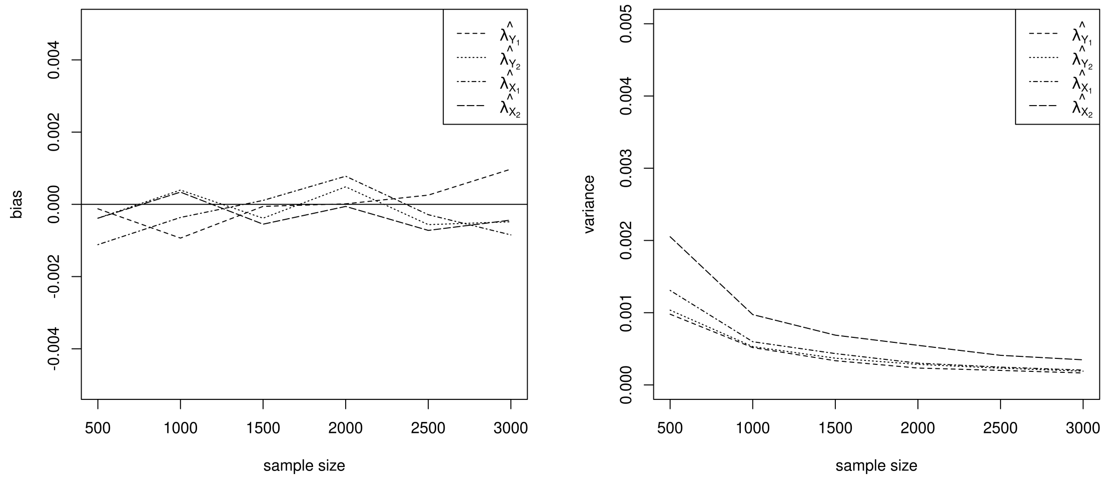
We do not formally investigate the theoretical properties of the LS estimator in this paper. However, a small simulation is provided here to offer some empirical evidence of their properties.

The loadings are fixed as $\boldsymbol{\lambda} = (0.8, 0.7, 0.6, 0.3)$. True scores S are randomly drawn from a standard normal distribution. Residuals $\boldsymbol{\varepsilon_W}$ are generated from a multivariate normal distribution, $N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon_W}})$ where the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon_W}} = \mathbb{I} - \boldsymbol{\lambda}\boldsymbol{\lambda}^\top$. The observations $\boldsymbol{W}$ are then computed according to the model in Equation 2. Using the generated data $\boldsymbol{W}$, the loadings are estimated using the least square method.

Figure 3 shows the bias and the variance of the LS estimator over 1000 replications for varying sample sizes. The estimator appears unbiased across sample sizes. Furthermore, the variance of the estimator decreases toward 0 as the sample size increases. While not a formal proof, this result suggests the LS estimator is likely a consistent estimator and has a good performance.

Comparison of the BLP and the CUBLP

We analytically showed how the BLP is biased conditional on the true score S and the unbiasedness of the CUBLP in previous sections. Here, we empirically verify the results



(a) *Bias* (b) *Variance*

Figure 3

Bias and variance of the LS estimator

using a simulation. The data are generated in the same fashion as described in the previous simulation. For the same data set $\boldsymbol{W}$, we computed both the BLP and the CUBLP. The results are in Figure 4. Consistent with the analytical results, there are not much noticeable difference between the two estimators if the true score is close to 0, and both are unbiased. However, as the true score moves away from 0 toward either extreme, the BLP starts to show biased predictions. For higher performers, the BLP underestimates on average. On the other hand, it overestimates on average for lower performers.

Real data analysis

To demonstrate the utilities of the proposed method, we analyze a real data set. As part of the speaking section of an English language proficiency test, students are asked to repeat what they have heard. The sentences vary in lengths and may situate in different scenarios such as presentations and campus tours. Each recorded response is scored by two different trained raters according to the rubric. The rating scale is from 0 to 5 with 5 being the highest possible. The two scores may agree or differ. In addition to the human raters,

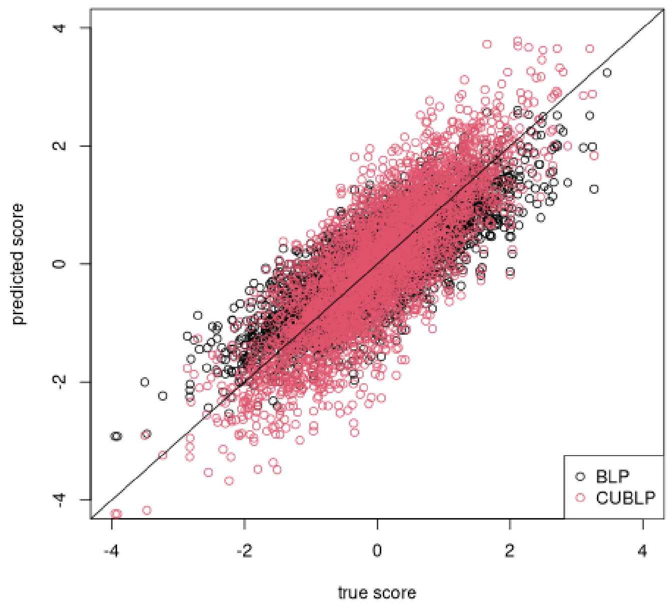


Figure 4

BLP vs. CUBLP in predicting true score.

the *SpeechRater*™system of Educational Testing Service (ETS), an automated scoring system for non-native speech, is used to evaluate the recorded response. The system provides features related to different dimensions of a speech. For the purpose of this demonstration, we computed composite scores on accuracy, pronunciation, fluency, and rhythm by aggregating the extracted features. These composite scores are standardized to have means of 0 and variances of 1.

There is a total number of 7690 observations. Since the raters are randomly assigned for each observation, we do not have any reason to believe the two human scores have a significant difference in discrimination. Therefore, we assume the loadings for the two human scores are equal. The estimated CUBLP coefficients, BLP coefficients, and the loadings are in Table 1. In this example, the estimated loadings on both human scores and accuracy dimension are quite large. As a result, a large amount of observed covariance can be explained well by the latent true score. In fact, plug in the estimators, the ratio

Table 1

Parameter estimates and predictor coefficients

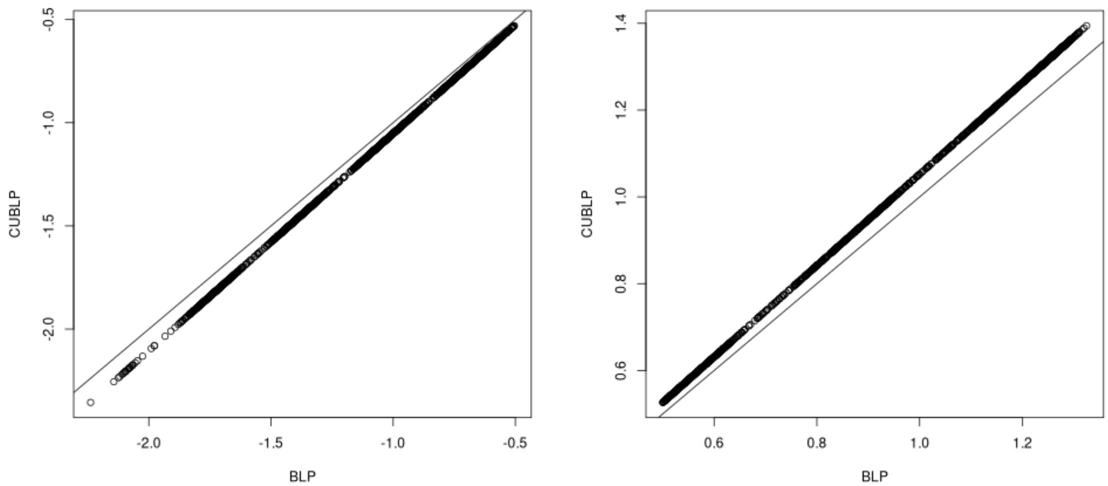
	CUBLP	BLP	$\hat{\boldsymbol{\lambda}}$
rating 1	0.456	0.433	0.944
rating 2	0.452	0.430	0.944
accuracy	0.144	0.137	0.839
pronunciation	0.009	0.009	0.333
fluency	0.017	0.016	0.400
rhythm	0.025	0.024	0.468

$\hat{\boldsymbol{\lambda}}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{W}}^{-1} \hat{\boldsymbol{\lambda}} = 0.95$. Therefore, the conditional bias of the BLP estimator would not be too large. This is reflected by the closeness of the estimated CUBLP coefficients an the BLP coefficients. But, compared to the CUBLP estimator, we still see the BLP underestimates for higher performers and overestimates for lower performers (see Figure 5).

Discussion

The improved technology enables the collection of more information related to the construct of interest from different sources. While it is tempting to combine these additional information with the traditional assessment data in making inferences on students' proficiency, researchers should be careful so that no systematic bias is introduced. The proposed CUBLP method addresses one specific source of bias - the algorithmic bias. The method does not generally rely on parametric distributional assumption. Consequently, it can be applied to a wide range of data and applications.

The current paper focuses on developing the CUBLP estimator. However, the



(a) *Lower performers*   (b) *Higher performers*

Figure 5

Comparison between the BLP and CUBLP

uncertainty of the estimator has not been investigated. Finding the variance or the asymptotic variance of the estimator should be an important topic of future research.

References

Attalí, Y., & Bursteín, J. (2006). Automated essay scoring with e-rater® V.2. https://doi.org/10.1002/j.2333-8504.2004.tb01972.x

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. *AI Magazine*, *25*(3), 27–36.

Fill, J. A., & Fishkind, D. E. (1999). The Moore-Penrose generalized inverse for sums of matrices. *SIAM Journal on Matrix Analysis and Applications*, *21*(2), 629–635. https://doi.org/10.1137/s0895479897329692

Guo, H., Zhang, M., Deane, P., & Bennett, R. E. (2019). Writing Process Differences in Subgroups Reflected in Keystroke Logs. *Journal of Educational and Behavioral Statistics*, *44*(5), 571–596. https://doi.org/10.3102/1076998619856590

Haberman, S. J., Yao, L., & Sinharay, S. (2015). Prediction of true test scores from observed item scores and ancillary data. *British Journal of Mathematical and Statistical Psychology*, *68*(2), 363–385. https://doi.org/10.1111/bmsp.12052

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, *6*(1), 15–32. https://doi.org/10.1214/ss/1177011926

Sinharay, S., Zhang, M., & Deane, P. (2019). Prediction of Essay Scores From Writing Process and Product Features Using Data Mining Methods. *Applied Measurement in Education*, *32*(2), 116–137. https://doi.org/10.1080/08957347.2019.1577245

Yao, L., Haberman, S. J., & Zhang, M. (2019). Penalized Best Linear Prediction of True Test Scores. *Psychometrika*, *84*(1), 186–211. https://doi.org/10.1007/s11336-018-9636-7

Zhang, M., Bennett, R. E., Deane, P., & van Rijn, P. W. (2019). Are There Gender Differences in How Students Write Their Essays? An Analysis of Writing Processes. *Educational Measurement: Issues and Practice*, *38*(2), 14–26. https://doi.org/10.1111/emip.12249

Zhang, M., & Deane, P. (2015). Process Features in Writing: Internal Structure and Incremental Value Over Product Features. *ETS Research Report Series*, *2015*(2), 1–12. https://doi.org/10.1002/ets2.12075