Supplementary Materials for

"Understanding Reliability from a Regression Perspective"

Yang Liu[1] Jolynn Pek[2] Alberto Maydeu-Olivares[3,4]

Contents

A Miscellaneous Theoretical Results **1**

A.1 Orthogonality in Regression 1

A.2 Equivalence to a Unit-Weight Simple Linear Regression 2

B Definitions of Reliability in Classical Test Theory **3**

C Connections to Kim (2012) **4**

C.1 Squared-Correlation and Parallel-Form Reliability 4

C.2 Reliability, PRMSE, and Squared Correlation 4

References **6**

[1]Department of Human Development and Quantitative Methodology, University of Maryland, College Park. Correspondence author. Email: yliu87@umd.edu

[2]Department of Psychology, the Ohio State University.

[3]Department of Psychology, University of South Carolina

[4]Faculty of Psychology, University of Barcelona

A Miscellaneous Theoretical Results

A.1 Orthogonality in Regression

The orthogonality between the regression function and its regression error is a general result that holds in all regression contexts. Consider regressing the outcome variable u onto (potentially multivariate) regressors $\mathbf{v}$. We have

$$u = \mathbb{E}(\underline{u}|\mathbf{v}) + \omega, \tag{S1}$$

in which the conditional expectation $\mathbb{E}(\underline{u}|\mathbf{v})$ is referred to as the *regression function*, and the *regression error* ω is the difference between the outcome and the regression function; $\omega = u - \mathbb{E}(\underline{u}|\mathbf{v})$. We adopt the same notational convention as the main document in which we underline a variable or vector to highlight its random nature. Next, we reiterate two fundamental properties about the regression function and its error error (see also Lord & Novick, 1968).

First, the regression error ω has mean zero, which is a direct consequence of the law of iterated expectation:

$$\mathbb{E}\underline{\omega} = \mathbb{E}\underline{u} - \mathbb{E}\left[\mathbb{E}(\underline{u}|\underline{\mathbf{v}})\right] = \mathbb{E}\underline{u} - \mathbb{E}\underline{u} = 0. \tag{S2}$$

Second, $\mathbb{E}(\underline{u}|\mathbf{v})$ is linearly independent of ω. To see this, we verify that the covariance between the two terms is zero:

$$\begin{aligned}
\mathrm{Cov}\left(\mathbb{E}(\underline{u}|\underline{\mathbf{v}}), \underline{\omega}\right) &= \mathrm{Cov}\left(\mathbb{E}(\underline{u}|\underline{\mathbf{v}}), \underline{u} - \mathbb{E}(\underline{u}|\underline{\mathbf{v}})\right) \\
&= \mathrm{Cov}\left(\mathbb{E}(\underline{u}|\underline{\mathbf{v}}), \underline{u}\right) - \mathrm{Var}\left(\mathbb{E}(\underline{u}|\underline{\mathbf{v}})\right) \\
&= \mathbb{E}\left[\underline{u}\mathbb{E}(\underline{u}|\underline{\mathbf{v}})\right] - \mathbb{E}\underline{u}\mathbb{E}\left[\underline{\mathbb{E}}(\underline{u}|\underline{\mathbf{v}})\right] - \mathbb{E}\left[\mathbb{E}(\underline{u}|\underline{\mathbf{v}})^2\right] + \left\{\mathbb{E}\left[\underline{\mathbb{E}}(\underline{u}|\underline{\mathbf{v}})\right]\right\}^2 \\
&= \mathbb{E}\left\{\mathbb{E}\left[\underline{u}\mathbb{E}(\underline{u}|\underline{\mathbf{v}})|\underline{\mathbf{v}}\right]\right\} - (\mathbb{E}\underline{u})^2 - \mathbb{E}\left[\mathbb{E}(\underline{u}|\underline{\mathbf{v}})^2\right] + (\mathbb{E}\underline{u})^2 \\
&= \mathbb{E}\left[\mathbb{E}(\underline{u}|\underline{\mathbf{v}})^2\right] - \mathbb{E}\left[\mathbb{E}(\underline{u}|\underline{\mathbf{v}})^2\right] \\
&= 0.
\end{aligned} \tag{S3}$$

The third equality in Equation S3 is a result of the definition of (population) covariance and variance. The fourth equality in Equation S3 follows from the law of iterated expectations, and the fifth equality is implied by $\mathbb{E}(\underline{u}|\mathbf{v})$ being a function of $\mathbf{v}$.

The results in Equations S2 and S3 apply to both the measurement and prediction decompositions (Equations 15 and 31 in the main document). In the measurement decomposition of an observed score x, the regression function and error are the true score $\tau_x = \mathbb{E}(\underline{x}|\boldsymbol{\eta})$ and the error score ε, respectively. In the prediction decomposition of a latent score ξ, the regression function and error are the expected *a posteriori* (EAP) score $\tilde{\xi} = \mathbb{E}(\underline{\xi}|\mathbf{y})$ and the resulting prediction error δ.

A.2 Equivalence to a Unit-Weight Simple Linear Regression

Using our general notation of regression (Equation S1), we next show that the potentially nonlinear regression of u onto $\mathbf{v}$ is equivalent to the simple linear regression of u onto $\mathbb{E}(\underline{u}|\mathbf{v})$ with intercept zero and slope one. Because $\mathbb{E}(\underline{u}|\mathbf{v})$ is a measurable function of $\mathbf{v}$, the σ-algebra generated by $\mathbb{E}(\underline{u}|\underline{\mathbf{v}})$ is smaller than that generated by $\underline{\mathbf{v}}$. By the law of iterated expectations, we have

$$\mathbb{E}\left[\underline{u} \mid \mathbb{E}(\underline{u}|\mathbf{v})\right] = \mathbb{E}\left[\mathbb{E}(\underline{u}|\mathbf{v}) \mid \mathbb{E}(\underline{u}|\mathbf{v})\right] = \mathbb{E}(\underline{u}|\mathbf{v}). \tag{S4}$$

Thus, when we regress the outcome u onto the original regression function $\mathbb{E}(\underline{u}|\mathbf{v})$, the resulting regression function is equivalent to the original regression function itself, leading to a simple regression with intercept zero and slope one. Equation S4 is applicable to both measurement and prediction decompositions.

B Definitions of Reliability in Classical Test Theory

In the main document, we reviewed two equivalent definitions of reliability within the classical test theory (CTT) framework (see the subsection "Definition of Reliability"). First, reliability quantifies the proportion of variance in the observed score $\underline{x}$ that can be explained by its true score $\tau_x = \mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}})$. Second, reliability also equals to the squared correlation between the observed score $\underline{x}$ and its true score τ_x, provided the true score variance is positive.

A third definition of reliability can be defined using the concept of a parallel score. Let $\dot{x}$ be another (often hypothetically) observable score produced by the same person with the latent variables (LVs) $\boldsymbol{\eta}$, satisfying the true score formula $\dot{x} = \tau_{\dot{x}} + \dot{\varepsilon}$, where $\tau_{\dot{x}} = \mathbb{E}(\underline{\dot{x}}|\boldsymbol{\eta})$ is the true score of $\dot{x}$. Then, $\dot{x}$ is *parallel* to x if

(a) $\mathbb{E}(\underline{\dot{x}}|\boldsymbol{\eta}) = \mathbb{E}(\underline{x}|\boldsymbol{\eta})$ for all $\boldsymbol{\eta}$ (i.e., the true scores are equal for everyone in the population),

(b) $\mathrm{Var}(\underline{\varepsilon}) = \mathrm{Var}(\underline{\dot{\varepsilon}})$ (i.e., error score variances are equal),

(c) $\mathrm{Cov}(\underline{\varepsilon}, \underline{\dot{\varepsilon}}) = 0$ (i.e., the error scores are uncorrelated; Lord & Novick, 1968, p. 47).

Then it can be shown that reliability equals to the correlation between two parallel test scores (Lord & Novick, 1968, p. 58):

$$\mathrm{Rel}(\underline{x}) = \mathrm{Corr}\big(\underline{x}, \underline{\dot{x}}\big). \tag{S5}$$

The parallel score $\dot{x}$, often regarded as a concept, might not be necessarily computed from manifest variables (MVs). However, a more intuitive approach to construct parallel scores using MVs, which requires stronger assumptions, was described by Lord (1983; see also Kim, 2012). Let $\dot{\mathbf{y}}$ denote an independent and identically distributed copy of the original MV vector $\mathbf{y}$ conditional on the LVs $\boldsymbol{\eta}$; we may think of $\dot{\mathbf{y}}$ as independent responses to an equivalent form of the same measurement instrument produced by the same person. Treating the original observed score x as a function of $\mathbf{y}$, Lord (1983) directly set $\dot{x} = x(\dot{\mathbf{y}})$, which satisfies all the three requirements of a parallel score. The independence requirement, however, is stronger than having uncorrelated error scores.

C Connections to Kim (2012)

C.1 Squared-Correlation and Parallel-Form Reliability

Kim (2012) defined two types of reliability coefficients for LV scores under unidimensional item response theory (IRT) models: (a) squared-correlation and (b) parallel-form reliability. First, parallel-form reliability is based on the construction of parallel test forms, which leads to parallel LV scores, and thus is equivalent to reliability (see Section B). Second, squared-correlation reliability is closely related to but not always the same as the proportional reduction in mean squared error (PRMSE). Let η be the unidimensional LV in the IRT model and x denote an observable estimate of η. Squared-correlation reliability in Kim (2012) is defined as $\mathrm{Corr}^2(\underline{\eta}, \underline{x})$. When $\underline{x}$ is the EAP score of η, we have $\mathrm{PRMSE}(\underline{\eta}) = \mathrm{Corr}^2(\underline{\eta}, \underline{x})$ (see the section "Prediction Decomposition: Optimal Prediction of a Latent Score" in the main document). Note, however, that when other types of LV scores (e.g., maximum likelihood or maximum *a posteriori* scores) are used as x to estimate η, the squared correlation $\mathrm{Corr}^2(\underline{\eta}, \underline{x})$ is no longer the PRMSE of η.

C.2 Reliability, PRMSE, and Squared Correlation

Another important result established by Kim (2012) is that parallel-form reliability is at least as large as the squared-correlation reliability for the same pair of observed score x and unidimensional LV η (i.e., Equation 32 in Kim, 2012). In what follows, we present a generalization of this result.

Consider a measurement model that specifies the joint distribution of MVs $\mathbf{y}$ and LVs $\boldsymbol{\eta}$. Here, $\boldsymbol{\eta}$ can be multidimensional, which is more general than the context considered by Kim (2012). Define a pair of observed score x (a function of $\mathbf{y}$) and latent score ξ (a function of $\boldsymbol{\eta}$). Note first that $\mathbb{E}\underline{x} = \mathbb{E}\left[\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}})\right]$ and $\mathbb{E}(\underline{x}\underline{\xi}) = \mathbb{E}\left[\mathbb{E}(\underline{x}\underline{\xi}|\underline{\boldsymbol{\eta}})\right] = \mathbb{E}\left[\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}})\underline{\xi}\right]$ by the law of iterated expectations, and hence

$$\mathrm{Cov}\big(\underline{x}, \underline{\xi}\big) = \mathrm{Cov}\big(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}}), \underline{\xi}\big). \tag{S6}$$

Then, the squared correlation between x and ξ satisfies the following inequality:

$$\begin{aligned}\mathrm{Corr}^2\left(\underline{x},\underline{\xi}\right) &= \frac{\mathrm{Cov}^2\left(\underline{x},\underline{\xi}\right)}{\mathrm{Var}(\underline{x})\mathrm{Var}(\underline{\xi})} = \frac{\mathrm{Cov}^2\left(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}}),\underline{\xi}\right)}{\mathrm{Var}(\underline{x})\mathrm{Var}(\underline{\xi})} = \frac{\mathrm{Var}\left(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}})\right)}{\mathrm{Var}(\underline{x})}\frac{\mathrm{Cov}^2\left(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}}),\underline{\xi}\right)}{\mathrm{Var}\left(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}})\right)\mathrm{Var}(\underline{\xi})} \\ &= \mathrm{Rel}(\underline{x})\mathrm{Corr}^2\left(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}}),\underline{\xi}\right) \le \mathrm{Rel}(\underline{x}).\end{aligned} \tag{S7}$$

In Equation S7, the second equality is due to Equation S6, the third equality is obtained by multiplying $\mathrm{Var}\left(\mathbb{E}(\underline{x}|\underline{\boldsymbol{\eta}})\right)$ to both the numerator and the denominator, and the fourth equality follows from the definition of reliability (Equation 15). Taking $\xi = \eta$ in a unidimensional IRT model reproduces Equation 32 of Kim (2012). More generally, if we take x to be the observed EAP score of ξ, $x = \tilde{\xi}$, then $\mathrm{Corr}^2\left(\underline{\tilde{\xi}},\underline{\xi}\right) = \mathrm{PRMSE}(\underline{\xi})$. Equation S7 implies that reliability of the observed EAP score $\tilde{\xi}$ is at least as large as PRMSE of ξ. Furthermore, the equality is attained if and only if the correlation between the true EAP score $\tau_{\tilde{\xi}} = \mathbb{E}\left(\underline{\tilde{\xi}}|\underline{\boldsymbol{\eta}}\right)$ and ξ equals to one.

Similarly, we have $\mathbb{E}\underline{\xi} = \mathbb{E}\left[\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right]$ and $\mathbb{E}(\underline{x}\underline{\xi}) = \mathbb{E}\left[\mathbb{E}(\underline{x}\underline{\xi}|\underline{\mathbf{y}})\right] = \mathbb{E}\left[\underline{x}\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right]$ by the law of iterated expectations. Therefore,

$$\mathrm{Cov}\left(\underline{x},\underline{\xi}\right) = \mathrm{Cov}\left(\underline{x},\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right), \tag{S8}$$

which is a representation symmetric to Equation S6 and allows us to derive an alternative bound for the squared correlation between x and ξ:

$$\begin{aligned}\mathrm{Corr}^2\left(\underline{x},\underline{\xi}\right) &= \frac{\mathrm{Cov}^2\left(\underline{x},\underline{\xi}\right)}{\mathrm{Var}(\underline{x})\mathrm{Var}(\underline{\xi})} = \frac{\mathrm{Cov}^2\left(\underline{x},\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right)}{\mathrm{Var}(\underline{x})\mathrm{Var}(\underline{\xi})} = \frac{\mathrm{Var}\left(\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right)}{\mathrm{Var}(\underline{\xi})}\frac{\mathrm{Cov}^2\left(\underline{x},\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right)}{\mathrm{Var}(\underline{x})\mathrm{Var}\left(\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right)} \\ &= \mathrm{PRMSE}(\underline{\xi})\mathrm{Corr}^2\left(\underline{x},\mathbb{E}(\underline{\xi}|\underline{\mathbf{y}})\right) \le \mathrm{PRMSE}(\underline{\xi}).\end{aligned} \tag{S9}$$

If we take ξ to be the true score of x, $\xi = \tau_x$, then $\mathrm{Corr}^2(\underline{x},\underline{\tau}_x) = \mathrm{Rel}(\underline{x})$. It follows from Equation S9 that PRMSE of the true score τ_x is at least as large as reliability of x, with the equality attained if and only if x and the EAP of its true score $\tilde{\tau}_x = \mathbb{E}(\underline{\tau}_x|\underline{\mathbf{y}})$, are perfectly correlated.

References

Kim, S. (2012). A note on the reliability coefficients for item response model-based ability estimates. *Psychometrika*, *77*(1), 153–162. doi: /10.1007/s11336-011-9238-0

Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, *48*(2), 233–245. doi: 10.1007/bf02294018

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.