A Case for Transforming the Criterion of a Predictive Validity Study

Brian F. Patterson & Jennifer L. Kobrin

The College Board

In "Validity Research on College and Graduate School Admission Tests"

Annual Meeting of the American Educational Research Association

April 10th, 2011

Predictive Validity

- Key inference from studies of predictor-criterion relationship: the relationship is dependable in the specified setting (Messick, 1988, p. 36)
- Ordinary least squares (OLS) regression is commonly used for studies of predictive validity
- In a wide variety of settings this model is adequate and can provide high quality evidence to support the validity of a particular use of a test

CollegeBoard inspiring minds™

Criterion Choice in Higher Education

- The validation of undergraduate college admissions measures commonly considers first-year grade point average (FYGPA) as a criterion
- FYGPA is meaningful as we expect it to be related to:
  a) subsequent college performance;
  b) probability of being retained to the second year; and
  c) probability of graduating within some finite timeframe.

CollegeBoard inspiring minds™

Some Known Problems with FYGPA

- Students take different courses
  - Smits, Mellenbergh, & Vorst. (2002) proposed imputing individual course grades.
- Courses vary in grading practices
  - Stricker, et al. (1994) review adjustments to FYGPA for different grading standards across departments
- Courses vary in difficulty
  - Young (1990) proposes using IRT-based methods to adjust course grades

Stakeholders are Still Drawn to FYGPA

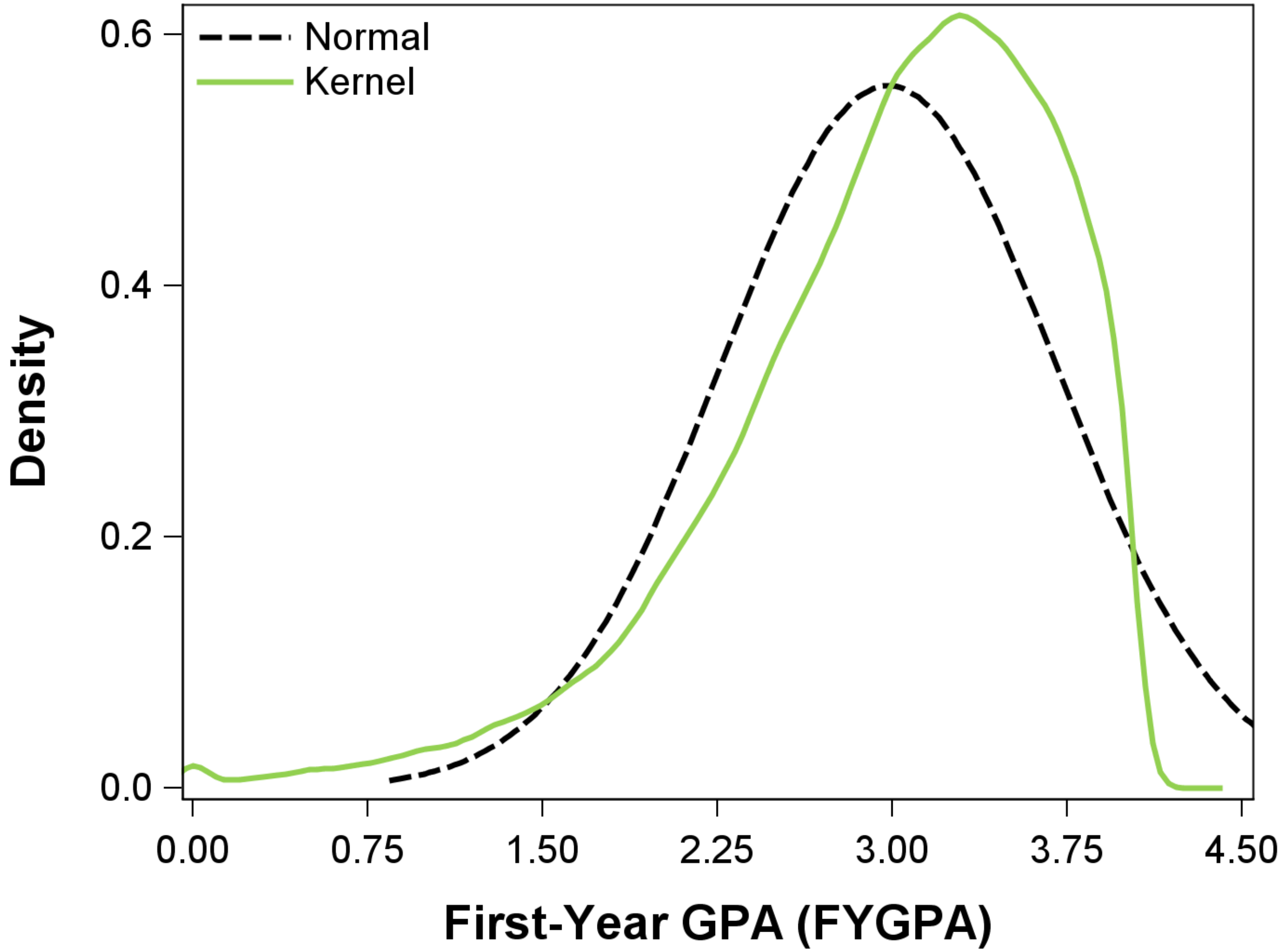
- Despite the avowed issues, stakeholders still feel that FYGPA is a good heuristic for early college performance
  - Admissions officers may consider predicted FYPGA in the admissions process
  - This prediction problem is presents certain problems
- Problem for OLS model: non-normal residuals
  - How can we retain the intuitive appeal of using FYGPA, but still overcome the non-normality of residuals?

CollegeBoard
inspiring minds™

Sample & Measures

- Sample
  - cohort of 173,963 entering 129 4-year institutions in the fall of 2008 as first-year students
  - institutions varied on size, admittance rate, control
- Measures
  - self-reported high school GPA from SAT Questionnaire
  - SAT critical reading, mathematics, writing
  - institution-supplied FYGPA

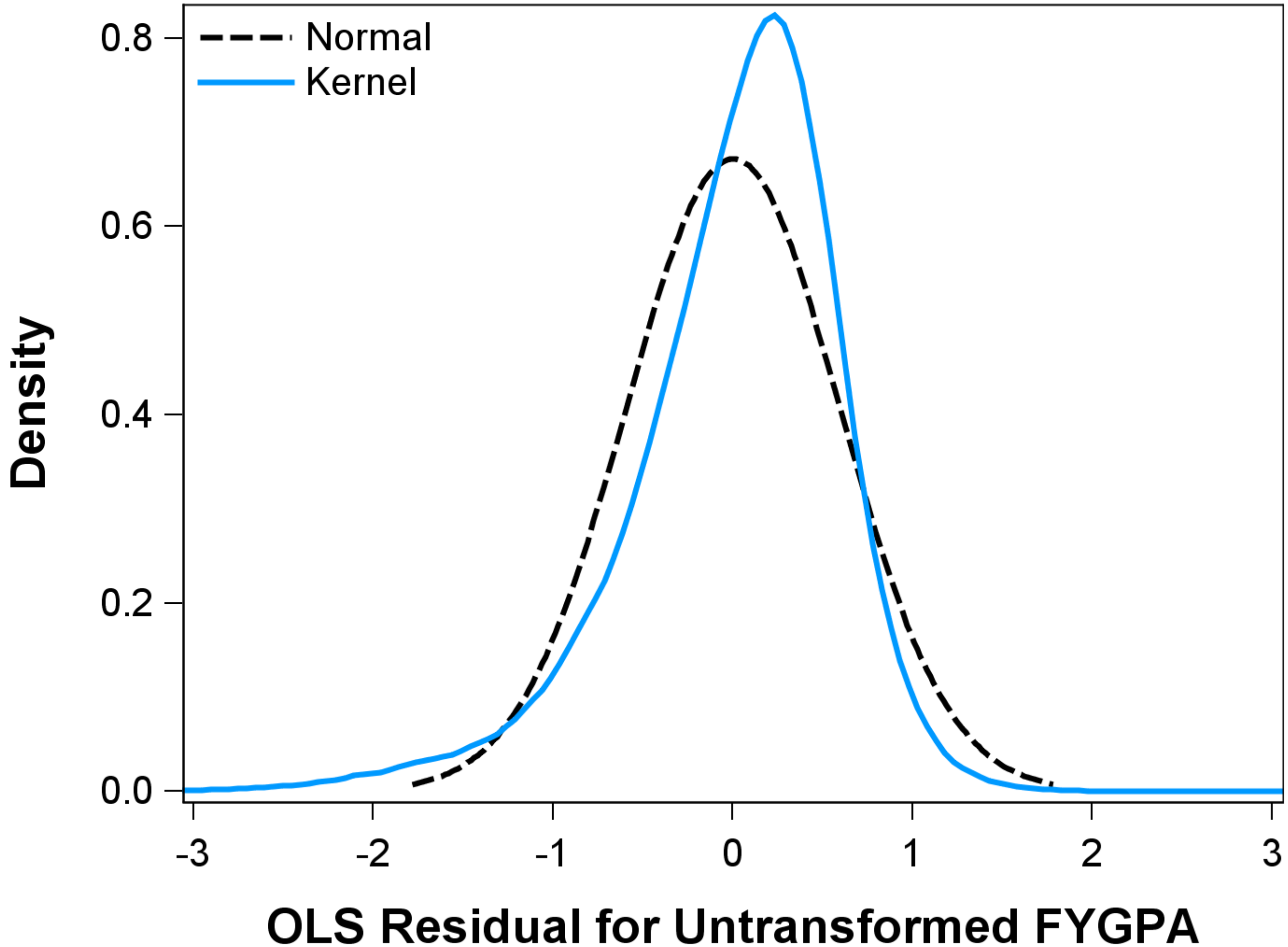
Criterion is Negatively Skewed



Note. Student-level density for 173,963 students across 129 institutions.

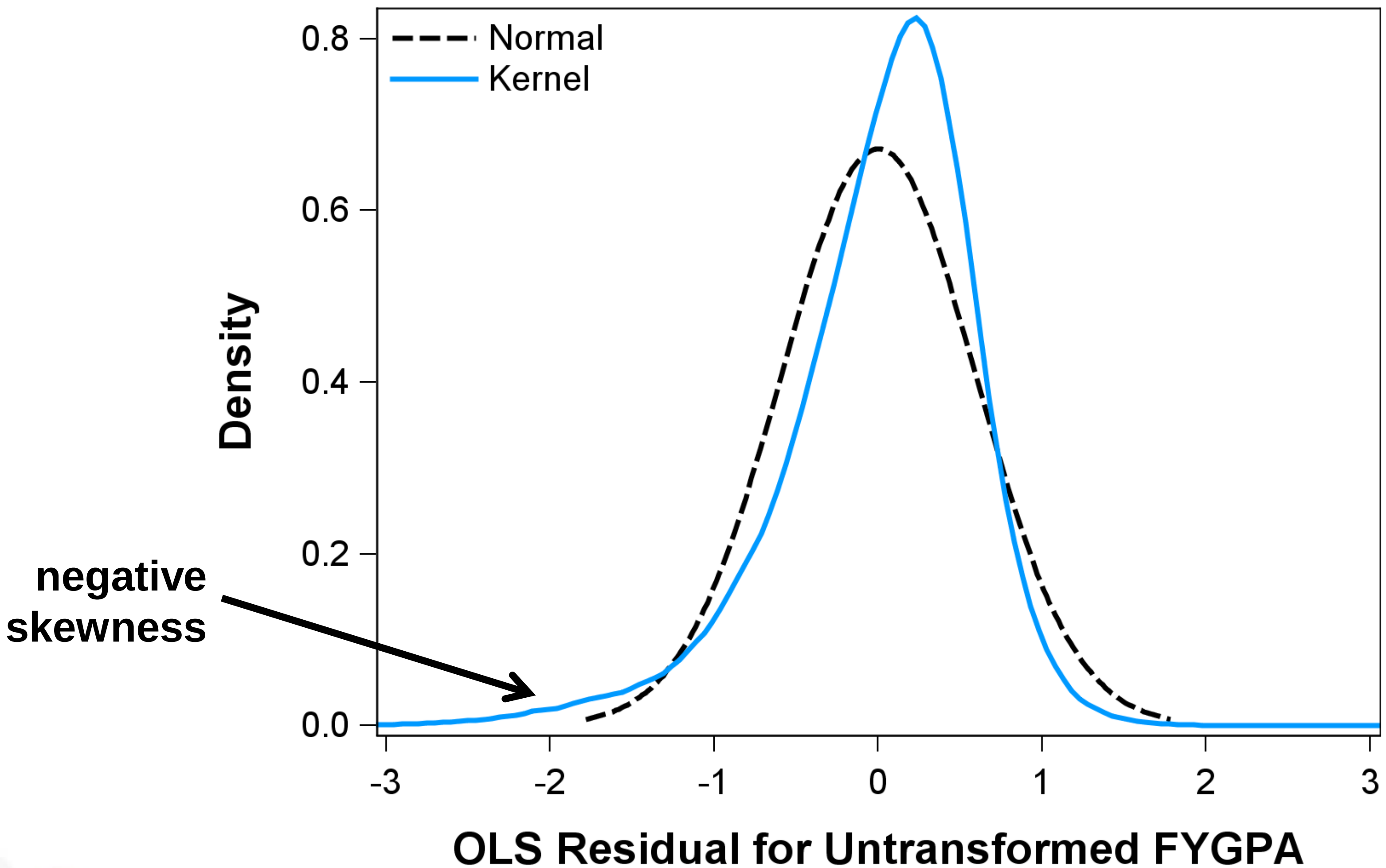
CollegeBoard inspiring minds™

Residuals: Negative Skew & Positive Kurt



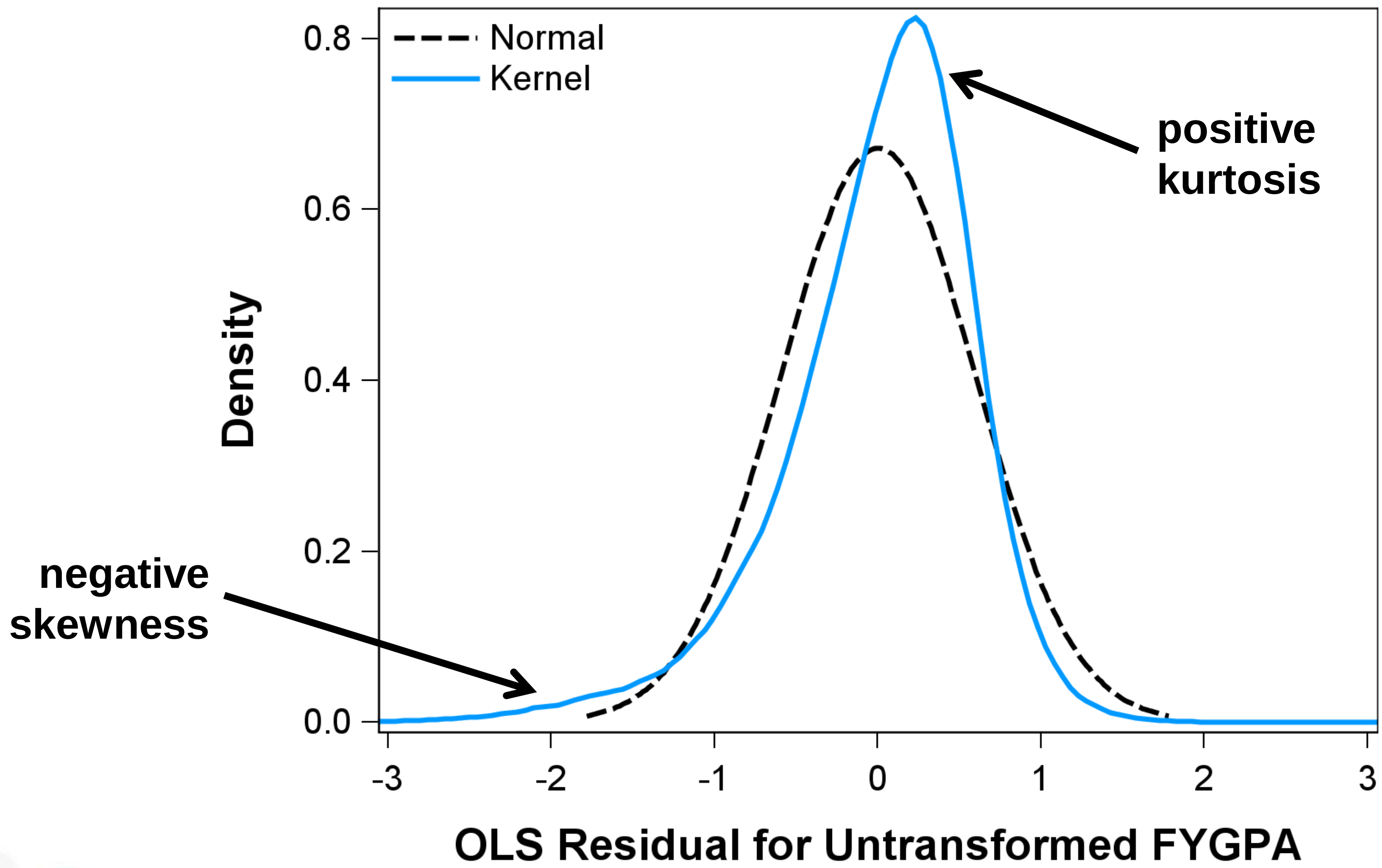
Note. Student-level density for 173,963 students across 129 institutions.

Residuals: Negative Skew & Positive Kurt



Note. Student-level density for 173,963 students across 129 institutions.

Residuals: Negative Skew & Positive Kurt



Note. Student-level density for 173,963 students across 129 institutions.

Understanding Residual Skewness

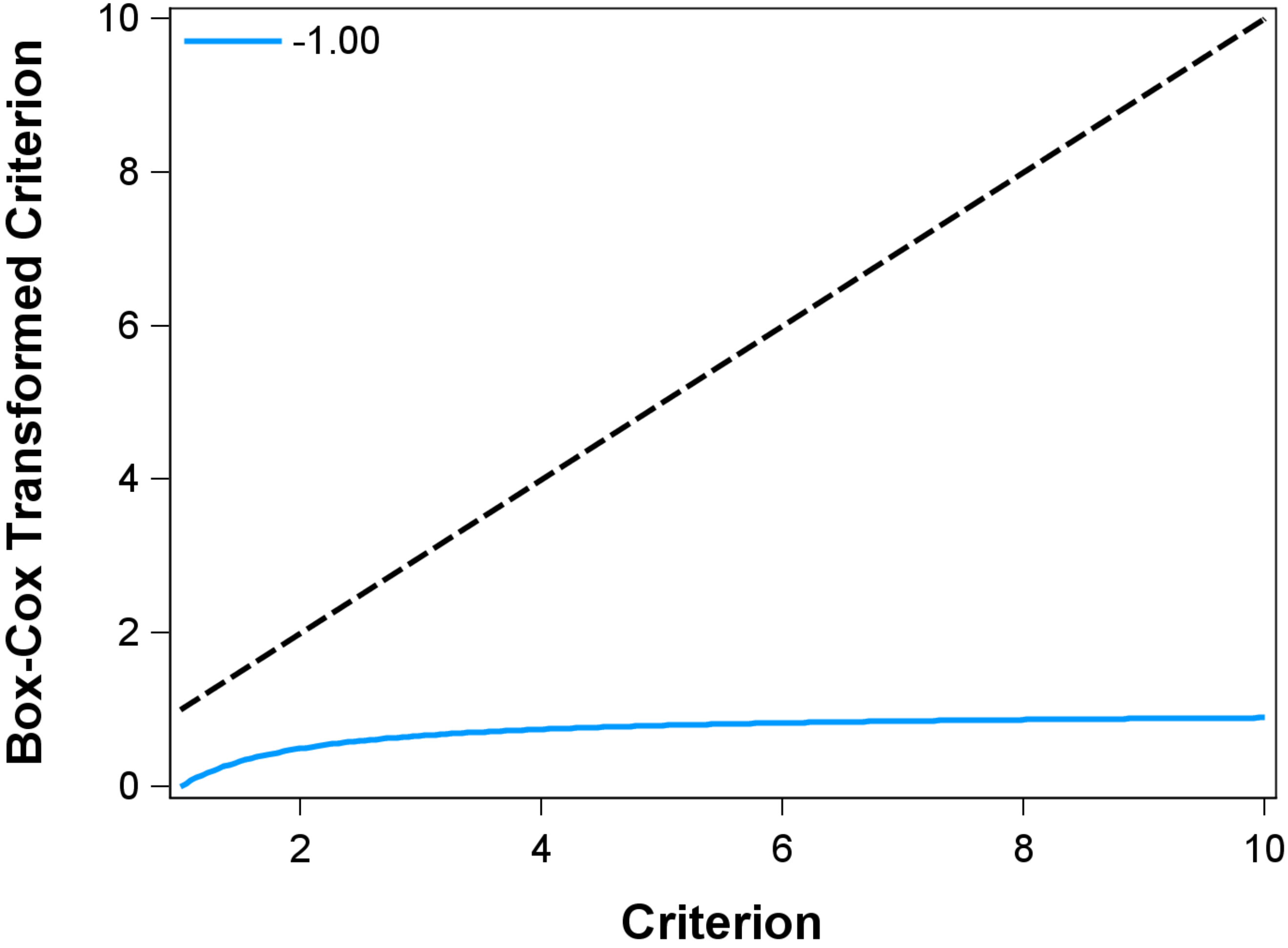
- Residuals exhibit negative skew because there are fewer, more extreme negative values and more, but less extreme positive values
  - In other words, a small number of students performed much worse than expected under the proposed model, while relatively more students slightly over-performed
  - Could additional—perhaps non-cognitive—measures account for that under-performance?
    - Colloquially, consider the awful roommate effect
  - Is there a ceiling effect, since the mean FYGPA is very close to the top of the scale?

Box-Cox (1964) Transformation

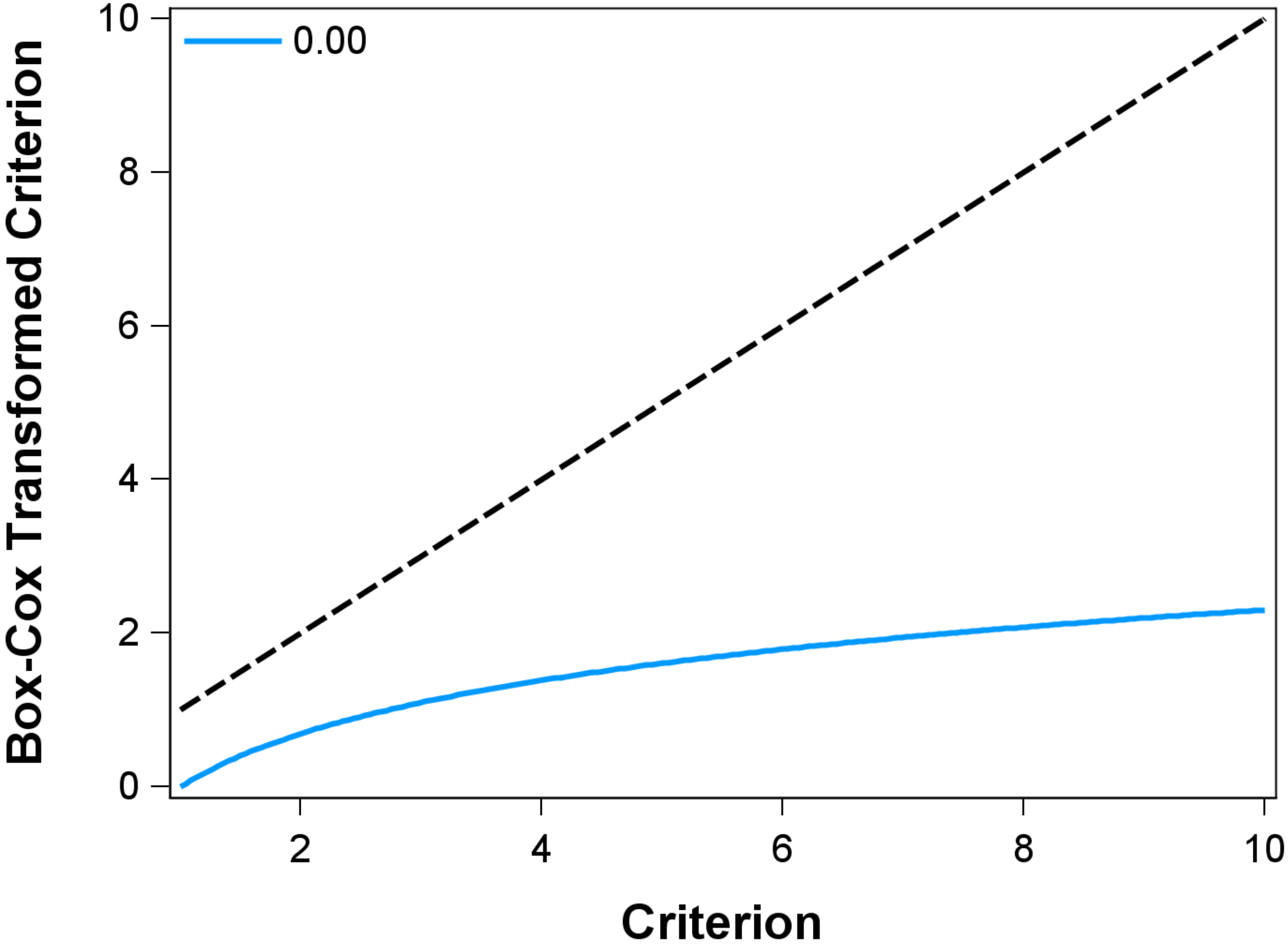
- Box & Cox's (1964) proposed transformation
  - $\mathbf{Z_{ij} = (Y_{ij}^{\lambda} - 1) / \lambda}$ for $\lambda \neq 0$
- Properties
  - monotonic (i.e., maintains ordering)
  - flexible functional form (e.g., useful for pos. / neg. skew)
  - when $\lambda = 1$, the transformed values only differ from original variable by a constant
  - may reduce root mean square error of prediction
  - λ is unknown and either must be estimated or based on prior research or subject-matter knowledge

CollegeBoard inspiring minds™

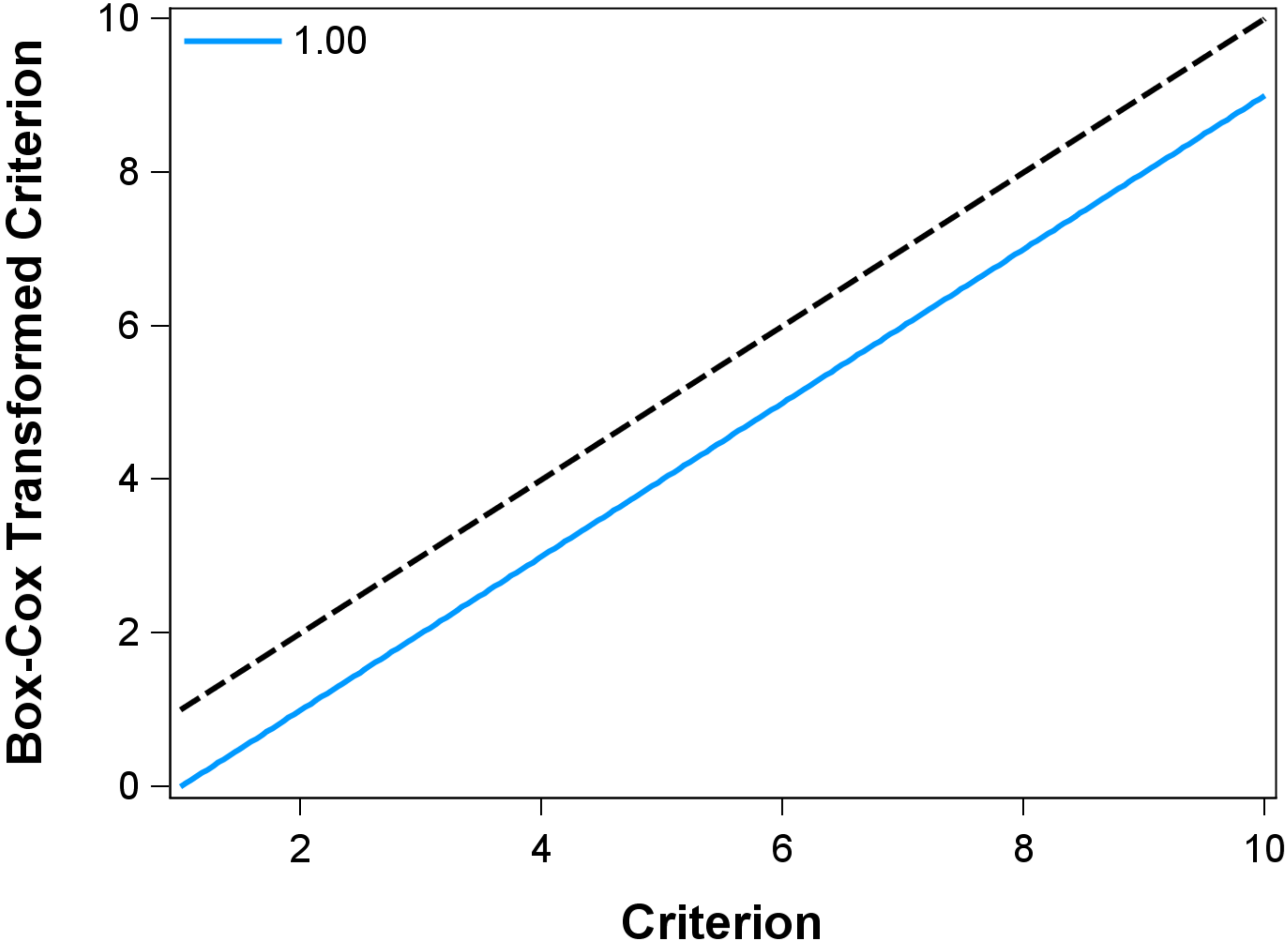
Examples of Box-Cox Transformation



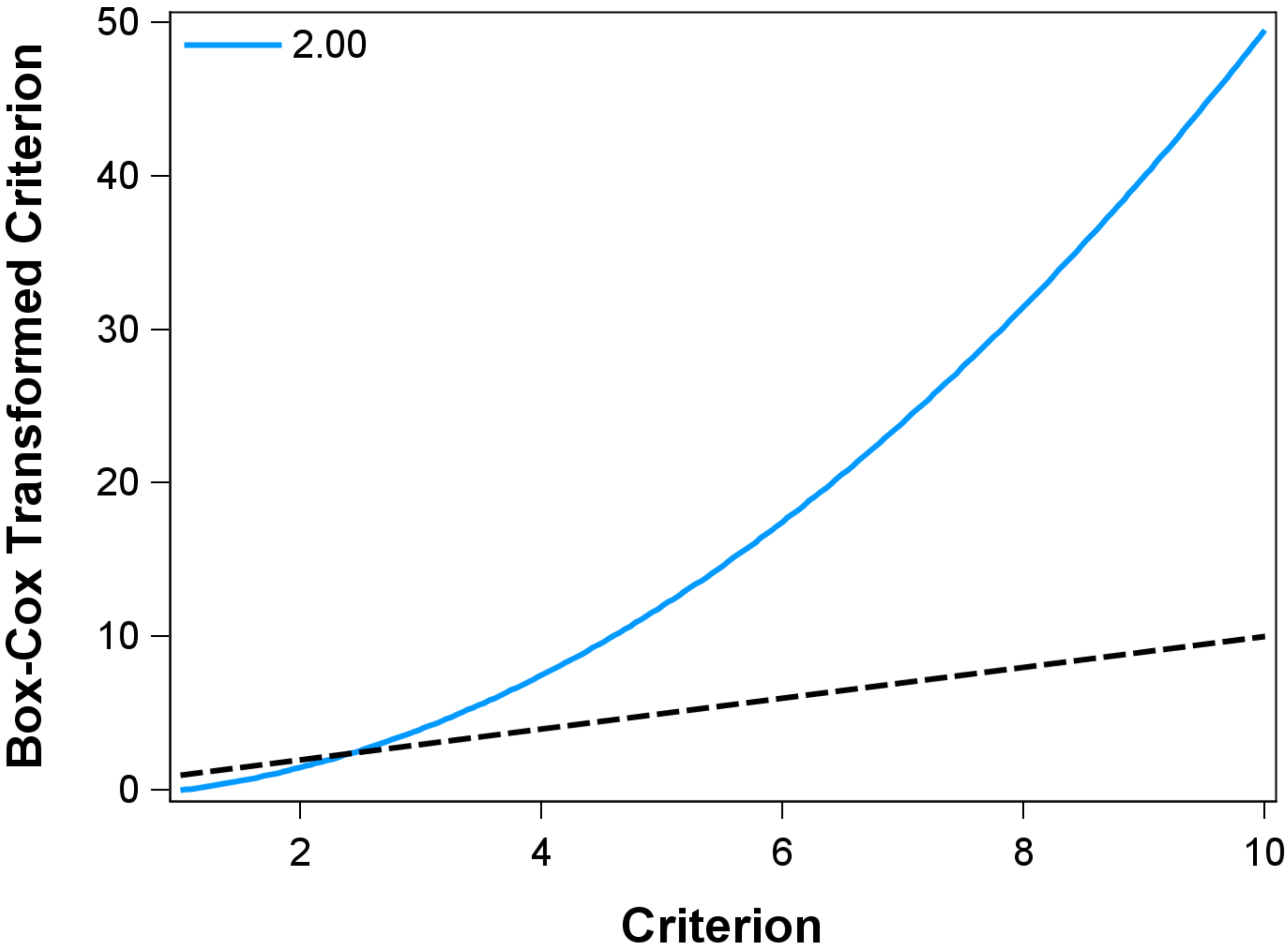
Examples of Box-Cox Transformation



Examples of Box-Cox Transformation



Examples of Box-Cox Transformation



Predictions on Original Scale

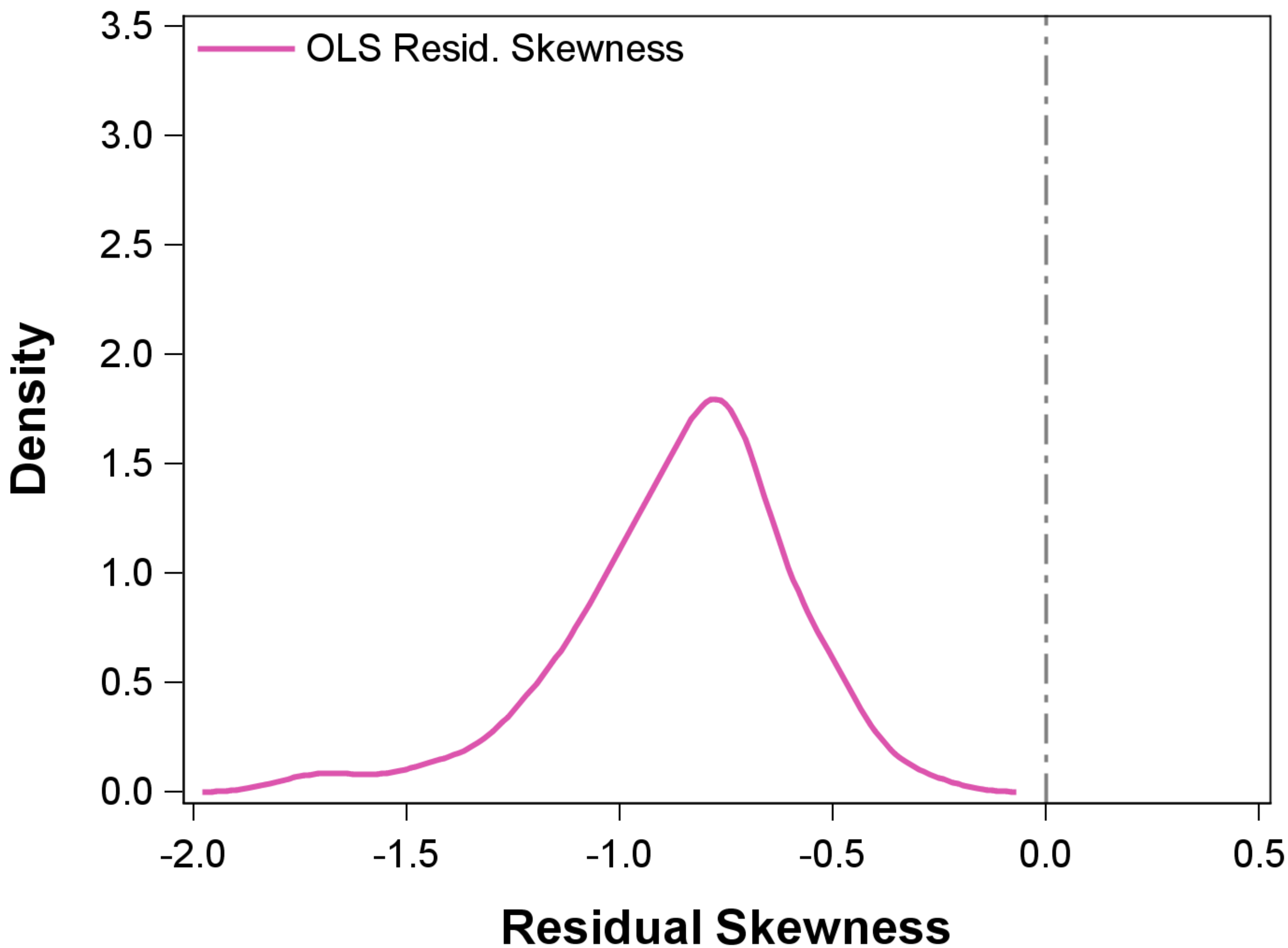
- One appeal of FYGPA is a typical 0 to 4 scale
  - The scale of the Box-Cox transformation is different
  - Predictions are on that unfamiliar scale
  - Solution: back-transform onto the FYGPA scale
  - $Y_{ij}^{Hat} = (\lambda^{Hat} \cdot Z_{ij}^{Hat} + 1) \wedge (1 / \lambda^{Hat})$
- Not all values may be back-transformed
  - If the predicted value is less than -1 and λ is not an integer, we cannot back transform
  - Consider restricting λ to integer values

Sample Estimates of Lambda

	λ
Mean	2.25
SD	0.68
Minimum	0.97
25th Pctile	1.76
Median	2.22
75th Pctile	2.68
Maximum	4.51

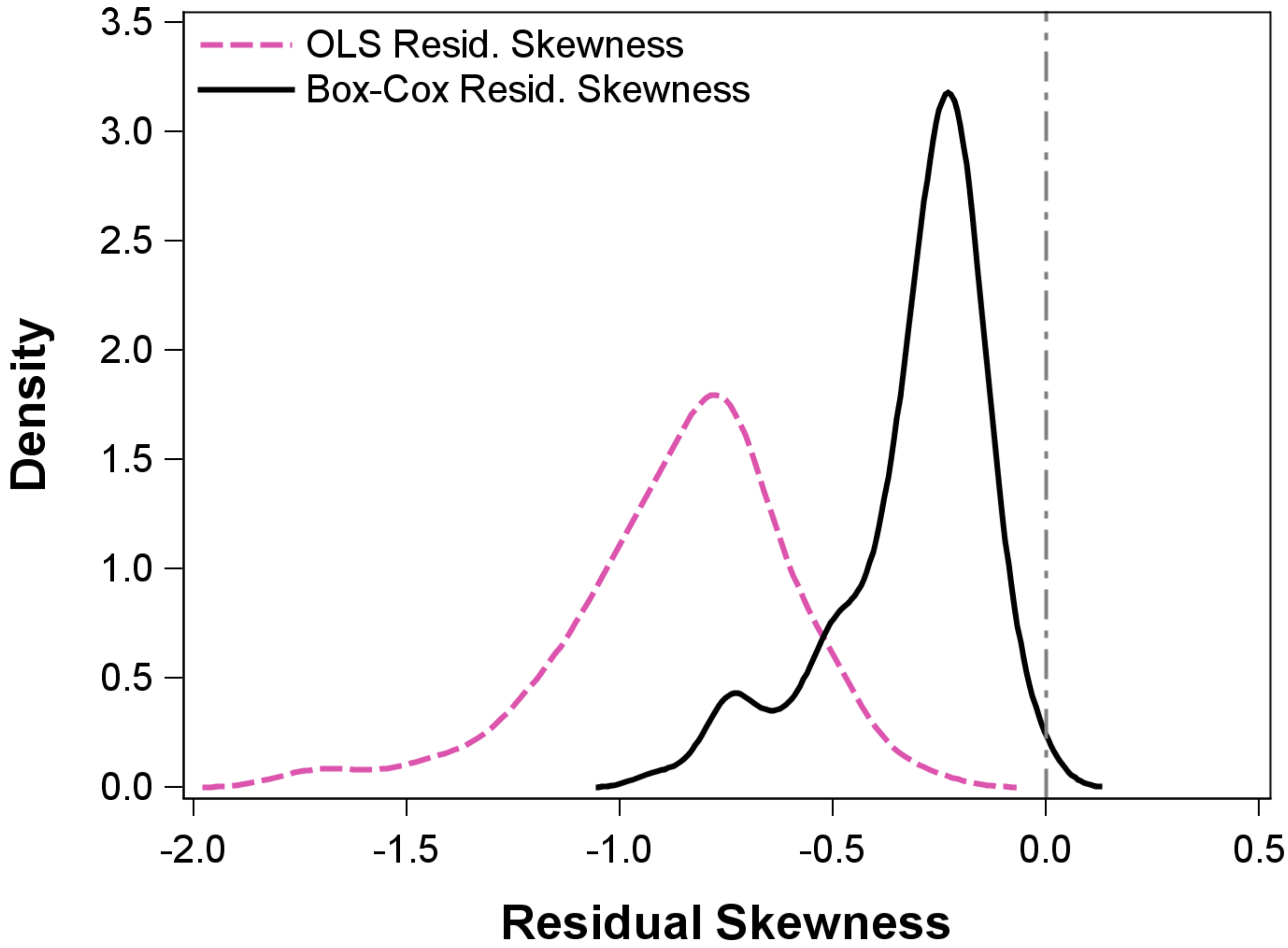
- Comments on estimates
  - simple mean λ-hat = 2.25
  - all λ-hat are positive
  - all λ-hat $\geq$ 1, when rounded
  - 50% of sample estimates range from 1.76 to 2.68
- Makes sense, with negative skewness
  - $\lambda > 1$ expands scale

Comparison of Residual Skewness



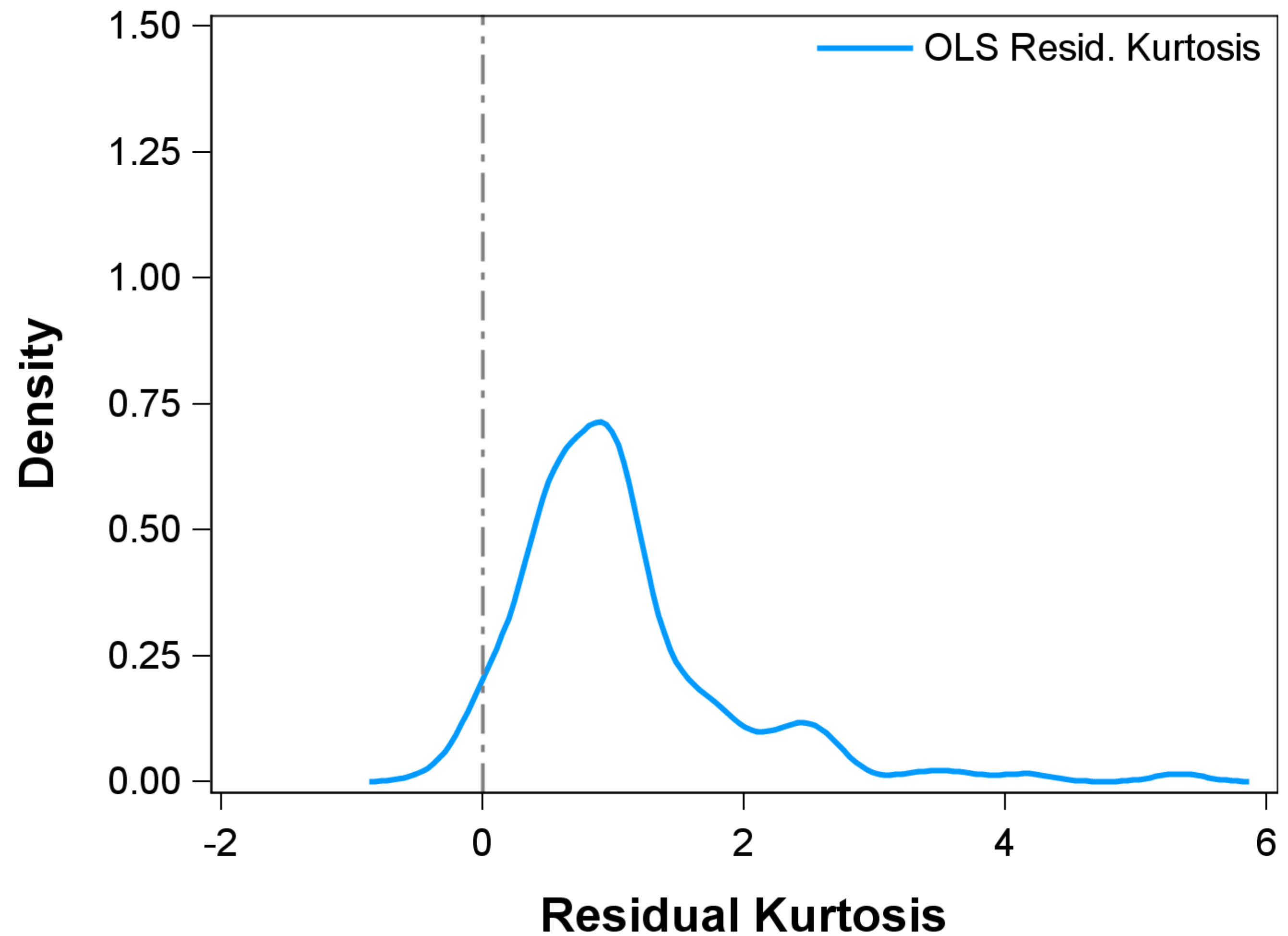
Note. Institution-level density for 129 institutions and 173,963 students.

Comparison of Residual Skewness



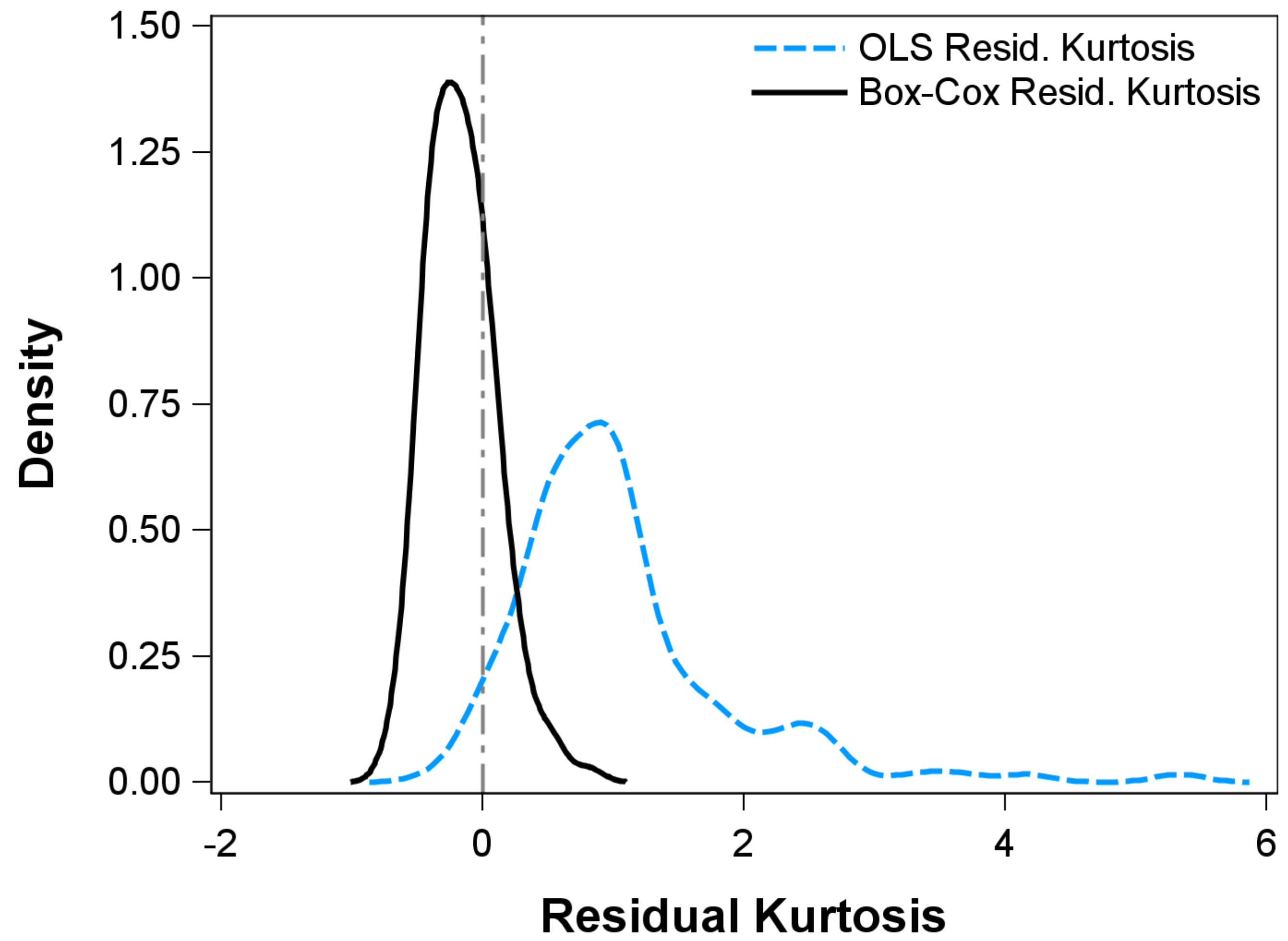
Note. Institution-level density for 129 institutions and 173,963 students.

Comparison of Residual Kurtosis



Note. Institution-level density for 129 institutions and 173,963 students.

Comparison of Residual Kurtosis



Note. Institution-level density for 129 institutions and 173,963 students.

Comparison of Root Mean Square Error

	RMSE	
Statistic	**OLS**	**BC**
Mean	0.58	0.27
SD	0.16	0.07
Minimum	0.31	0.10
25th Pctile	0.45	0.22
Median	0.55	0.26
75th Pctile	0.69	0.32
Maximum	0.98	0.46

- Reduces RMSE
  - In our sample, the RMSE for the Box-Cox model is about half of what it is for the OLS model
  - May use Box-Cox transformation even with normal residuals
  - More precise estimates of FYGPA are possible

CollegeBoard inspiring minds™

Summary & Limitations

- In this model, FYGPA residuals are not normal
- Using a Box-Cox transformation on FYGPA reduces non-normality
- Original scale may be recovered for prediction
- There are limitations of this approach for both prediction and inferences around criterion validity:
  - Original scale may not be recoverable for all predictions
  - The need to estimate λ increases Type I Error, unless properly adjusted

Questions, Comments, Suggestions

- Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board presentations do not necessarily represent official College Board position or policy.

- Please forward any questions, comments, and suggestions to:
  - bpatterson@collegeboard.org

CollegeBoard inspiring minds™

References

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological), 26*(2), 211-252.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33- 45). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Smits, N., Mellenbergh, G.J. and Vorst, H.C.M. (2002). Alternative Missing Data Techniques to Grade Point Average: Imputing Unavailable Grades. *Journal of Educational Measurement, 39*(3), 187-206.

Stricker, L.J., Rock, D.A., Burton, N.W., Muraki, E., and Jirele, T.J. (1994). Adjusting college grade point average criteria for variations in grading standards: a comparison of methods. *Journal of Applied Psychology, 79*(2), 178-183.

Young, J.W. (1990). Adjusting the cumulative GPA using item response theory. *Journal of Educational Measurement, 27*(2), 175-186.

CollegeBoard inspiring minds™