
Do Rothstein’s “Future Teacher Effects” Imply Value-Added Models (VAMS) Are Biased or Miss-Specified?

October 10 2013

**Presentation at Michigan State University,
Institute for Education Sciences Conference**

By Duncan Chaplin

**MATHEMATICA
Policy Research**

Co-Author and Paper

Dan Goldhaber, Center for Education Data and Research, University of Washington

**Goldhaber, Dan & Duncan Chaplin (2012),
“Assessing the ‘Rothstein Falsification Test.’ Does it Really Show Teacher Value-added Models are Biased?”
CEDR Working Paper, 2012-1.3,
Mathematica Working Paper, February 2012,
Calder Working Paper 71, 2012.**

Presented at APPAM, NBER, BLS, and Urban Institute

Main Contributions

- **Clarifying Rothstein Falsification Test based on Future Teacher effects**
- **Simulate Magnitude of Future Teacher Effects in Value Added Models**
- **Test to see Whether Future Teacher effects suggest Bias or Miss-specification**

Introduction

- **Rothstein's Falsification Test for VAMs**
 - “Future teacher effects” on students' past achievement in real data are large and jointly significant (QJE 2010)
 - This appears to suggest a need to adjust more for *dynamic* ability tracking
- **We find large and jointly significant future teacher effects with VAMS that are correctly specified and have no bias**

Policy Context

- **Growing interest in using student growth as a component of teacher evaluations**
 - Race to the Top, DCPS, Chicago Teachers, etc.
- **Falsification test used to call VAMs into question**
 - Rothstein and Darling-Hammond (2012) say that future teacher effects show that , “... VAMs wrongly attribute to teachers other influences on student performance that are present when the teachers have no contact with the students ...”
- **Falsification test now commonly used by researchers**
 - Harris and Sass (2010) – Districts for Research
 - Koedel and Betts (2009) – Revised VAM
 - Briggs and Domingue (2011) – Los Angeles VAM

Rothstein Points Made to Congress

- ***“Logically, 5th grade teachers can’t influence 3rd grade test scores.***
- ***So a VA model that identifies teachers’ true effects should show zero effects of 5th grade teachers on 3rd grade test scores.***
- ***But all extant VA models show large effects!***
- ***VA rewards or penalizes teachers for the kids they teach, not just for how well they do it.***
- ***Not fair to ELL teachers / special needs teachers / other specialists.”***
 - Rothstein in Congressional Testimony September 2011,
 - Made similar points in February 2012 paper in Phi Delta Kappan

Rothstein Test: Bias and Miss-Specification

- **Does the test show that VAMs are biased?**
 - No
 - He notes, negatively correlated errors are needed for bias
 - We find that the falsification test rejects without negatively correlated errors
- **Does the test show VAMs are Miss-Specified?**
 - No
 - He claims test shows that students are not randomly assigned conditional on the covariates used in the VAM
 - We find that the test rejects when we randomly assign students conditional on covariates in VAM

Start with Two-Teacher Example

- $A_5 = \lambda_v * A_4 + \beta_1 * \tau_{1_5} + e_5$
- A_5 = Achievement in grade 5
- τ_{1_5} = Has teacher 1 (teacher 0 omitted)
- β_1 = Effect of teacher 1 vs. teacher 0
- e_5 = Other factors
- λ_v = Previous learning effect
- Tracking based on prior achievement

VAM vs Falsification Test

VAM Equation: $A_5 = \lambda_v * A_4 + \beta_1 * \tau_{15} + e_5$ (EQ1)

Falsification Test: $A_3 = \lambda_f * A_4 + R1 * \tau_{15} + w_3$ (EQ2)

- RHS Same (g4 achievement and g5 teacher)
- Outcomes differ (G3 achievement vs G5)
- R1 is “future” teacher effect from Rothstein Test
- $R1=0$, but $R1_{hat} \neq 0$ if $cov(\tau_{15}, w_3 | A_4) \neq 0$
- $R1_{hat} \neq 0$ appears to “falsify” VAM, but it may not

Test is linear, but tracking is inherently non-linear

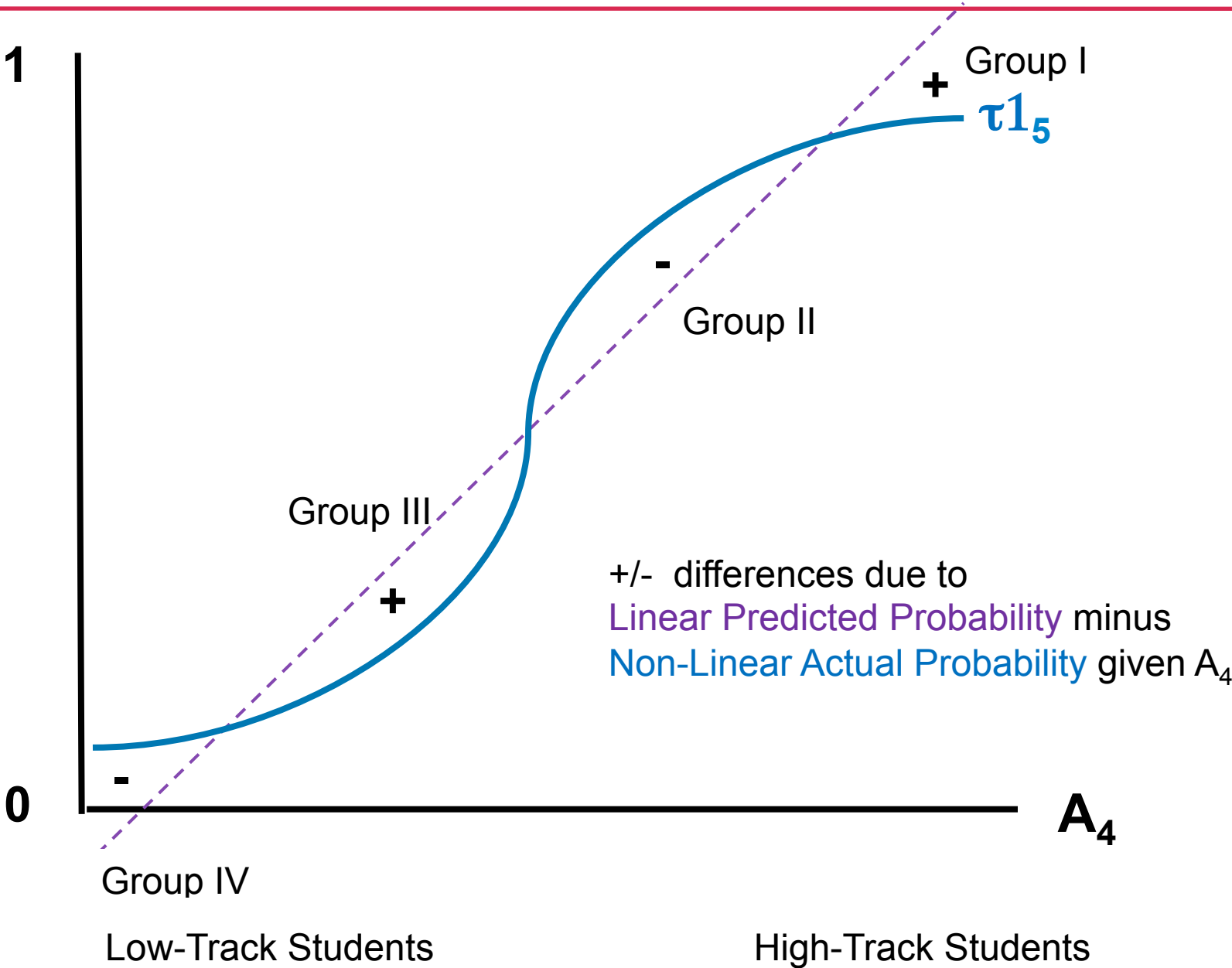
Non-Linearities and Rothstein Test

- $R1\hat{h}at$ = Estimated future teacher effect
- $E(R1\hat{h}at) = \text{cov}(\tau1^{r_5}, A^{r_3}) / \text{var}(\tau1^{r_5})$
- where $\tau1^{r_5}$ and A^{r_3} are residuals equal to the **actual** minus **predicted** values given A_4
- Correlated residuals imply $R1\hat{h}at$ non-zero

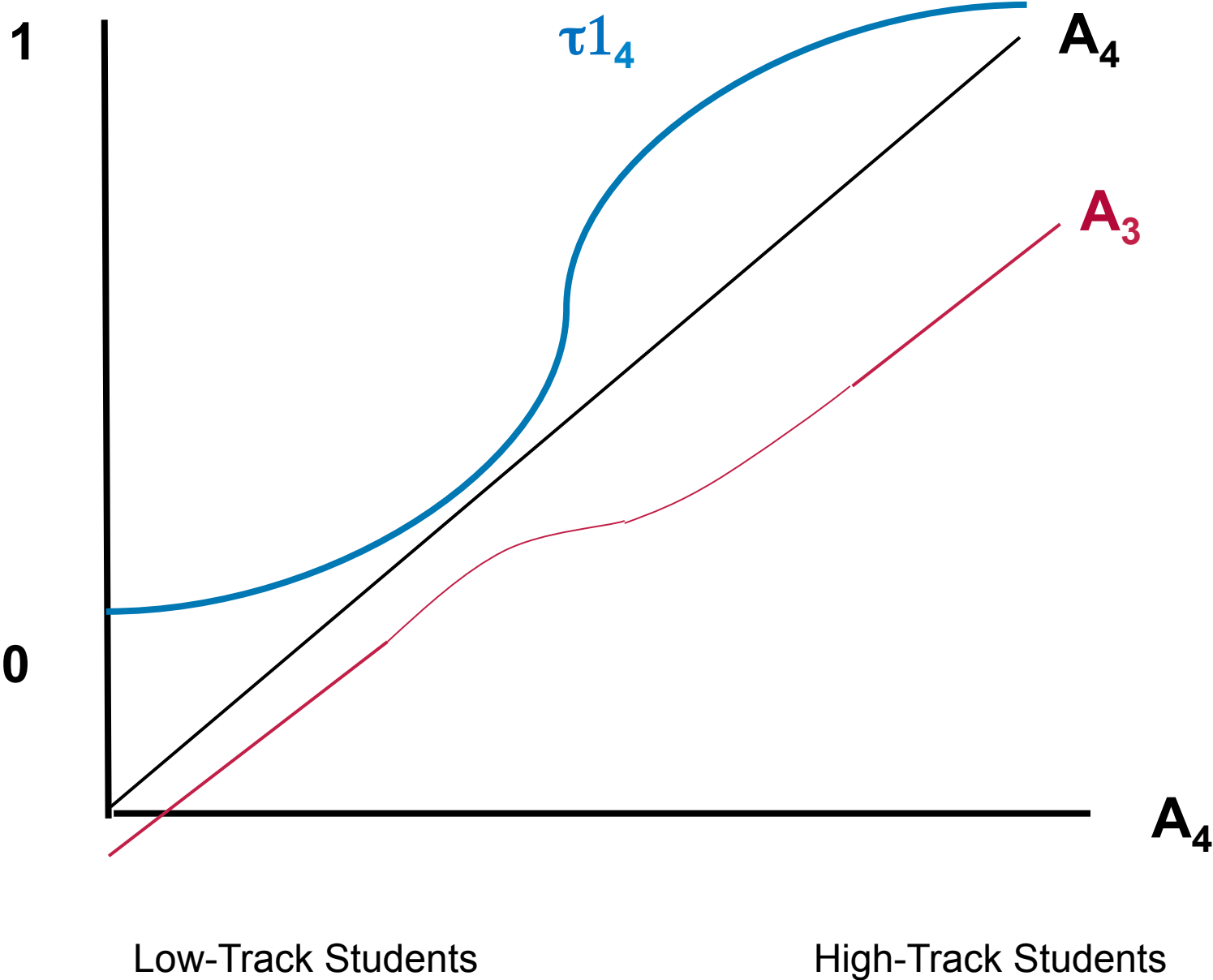
Graphing Residuals

- $E(\hat{R}_1) = \text{cov}(\tau_5^r, A_3^r) / \text{var}(\tau_5^r)$
- **Grade 5 Teacher:** $\tau_5^r = \tau_5 - \gamma \tau^* A_4$
- **Grade 3 Achievement:** $A_3^r = A_3 - \gamma A^* A_4$
- **High track students (with higher scores in grade 4) get the better teacher (with $\tau_5 = 1$)**

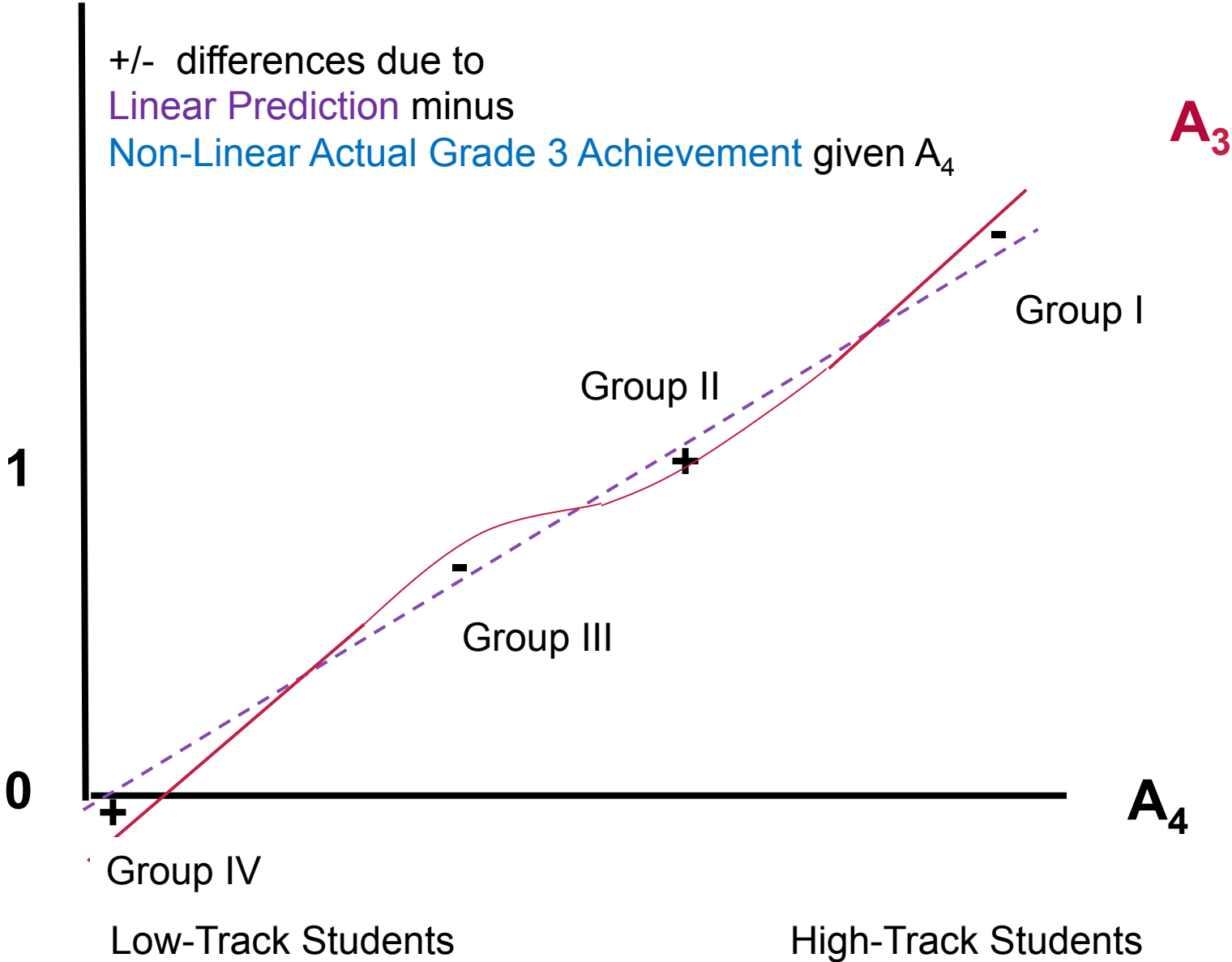
Grade 5 Teacher Residuals



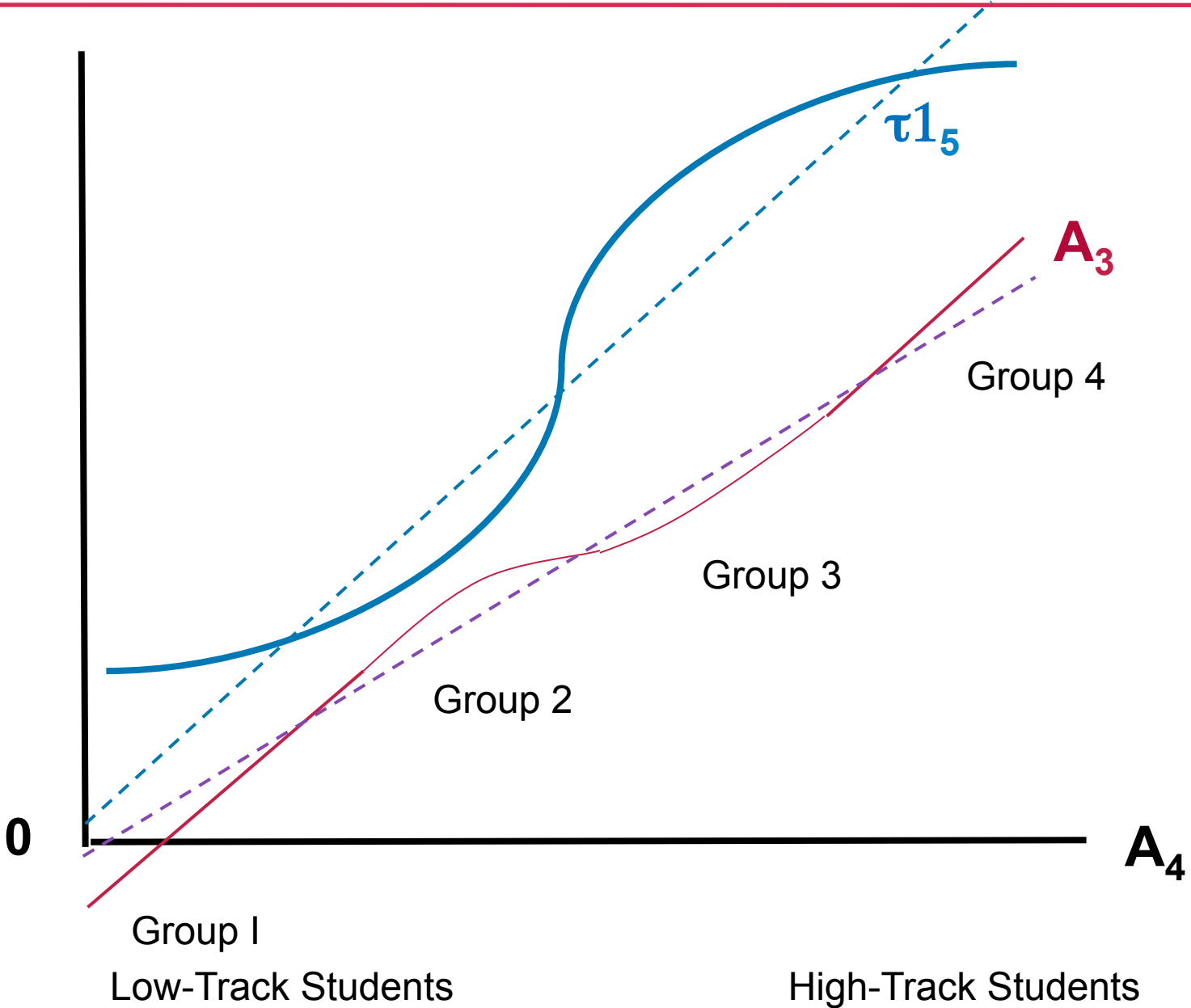
Grade 3 Achievement Residuals



Grade 3 Achievement Residuals (Continued)



Residuals: G3 Achievement and G5 Teacher



Does these Non-Linearities Matter?

Simulation Parameters

- **Parameters similar to Rothstein's**
- **Large Ns**
 - 600 teacher effects
 - 200 schools,
 - 3 teachers per school (4th omitted),
 - 20 students per teacher
 - 16,000 students
- **Other Parameters Standard**
 - Errors jointly normal, Std Dev = 0.4
 - Teacher effects Std Dev = 0.10
 - λ between 0.91 and 0.95 to keep Std Dev Achievement close to 1 across years

Estimation Methods for Miss-Specification

- **Conduct Rothstein falsification test with**
 - random assignment,
 - miss-specified model, and
 - correctly specified model
- **Calculate the std dev of the future teacher effect estimates**
- **Calculate the std dev of the current teacher effect estimates**
- **Errors are uncorrelated, estimates are unbiased**

Simulation Results re: Miss-Specification

Topic	Subtopic	Random Assignment	Miss-Specified VAM	Correctly Specified VAM
Grade 5 Teachers Tracked on				
	Random Error	X	X	X
	Grade 3 Scores		X	
	Grade 4 Scores		X	X
	Correlated Errors			
Results				
	Future Teacher Effects	No	Yes	Yes
	Bias	No	No	No
	Std Dev Future Teacher Effect Estimates	0.132	0.331	0.149
	Std Dev Current Teacher Effect Estimates	0.139	0.132	0.137

Estimation Methods for Bias

- **Test for bias by comparing true teacher effects to estimated teacher effects**
- **Calculate Correlation of True Teacher Effects and Estimated Teacher Effects, adjusted for estimation error**
- **Run with and without correlated errors**

Simulation Results re: Bias

Topic	Sub topic	Random Assignment	Miss-Specified VAM	Correctly Specified VAM	Errors Not Correlated
<i>G5 Students Tracked on:</i>					
	Grade 3 scores		X		
	Grade 4 scores		X	X	X
	Omitted variable				
	Correlated Errors	X	X	X	
<hr/>					
<i>Results:</i>					
	Future Teacher Effects	No	Yes	Yes	Yes
	Bias	No	Yes	Yes	No
	Correlation with truth	0.97	0.77	0.99	1.00

Note: Correlations with truth adjusted for estimation error.

Robustness Checks

- **With and without measurement error in test scores**
- **With previous grade teachers added to falsification test (a version Rothstein estimated)**
- **Switched lagged and double-lagged achievement**
- **With non-linear lagged achievement polynomials in the falsification test regression**
- **Same results**
 - Rothstein test continues to reject even when VAM is correctly specified (i.e. students are randomly assigned conditional on lagged achievement) and when there is no bias.

Conclusions

- **Rothstein has influenced both policy makers and researchers**
 - Appears to suggest VAM misspecification
- **We find Rothstein Future Teacher test rejects**
 - Without bias
 - Without misspecification
- **Future Teacher test may not provide useful guidance on using VAMs for either policy or research purposes**
- **Other evidence suggests small bias**
 - Our theoretical results regarding negatively correlated errors
 - Kane and Staiger (2008), Rothstein (2010), and Chetty et al (2012)

For More Information

- **Duncan Chaplin**
 - DChaplin@mathematica-mpr.com
- **Dan Goldhaber**
 - DGoldhab@u.washington.edu
- **Our paper is available at**
 - www.mathematica-MPR.com
 - www.cedr.us and
 - www.caldercenter.org