Updating and Misspecification: Evidence from the Classroom

Marc-Antoine Chatelain[1] Paul Han[2] En Hua Hu[3] Xiner Xu[1]

[1]Toronto
[2]Competition Bureau
[3]Oxford

Motivation

- People use information to make decisions all the time:
  - Learning about restaurants on Google Maps' reviews
  - Reading an abstract before deciding whether to attend a seminar
  - Learning about candidates before voting

Motivation

- People use information to make decisions all the time:
  - Learning about restaurants on Google Maps' reviews
  - Reading an abstract before deciding whether to attend a seminar
  - Learning about candidates before voting
- Standard theory of belief updating is Bayesian
  - Widely documented violations
  - Failures of learning can be very costly

Motivation

- Why do people fail to learn from informative signals?
- What can we do to improve learning?

Motivation

- Standard explanations for failure of learning:
  - Behavioral channel: overconfidence, motivated beliefs, ego-utility
  - Cognitive channel: computational constraint, complexity of updating

Motivation

- Standard explanations for failure of learning:
  - Behavioral channel: overconfidence, motivated beliefs, ego-utility
  - Cognitive channel: computational constraint, complexity of updating
- Difficult to correct these *updating biases*

Motivation

- Standard explanations for failure of learning:
  - Behavioral channel: overconfidence, motivated beliefs, ego-utility
  - Cognitive channel: computational constraint, complexity of updating
- Difficult to correct these *updating biases*
- Bulk of the evidence is experimental

This paper

- We consider a perceptual channel: misspecification - failure to understand signal informativeness
    - Rich theory in misspecification and updating failures - empirics lacking
    - We study empirical relevance, estimate the impact and explore correctability
    - Environment: first-year students learning about their performance in a class

Conceptual Overview

Consider the following model:

$$g_{it} = \theta_i + \epsilon_{it}, \; \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- g_{it} - student grade
- θ_i - fixed ability, uncertain to the student
- ϵ_{it} - exogenous testing noise

Conceptual Overview

Consider the following simple conceptual model:

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it}, \ \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- Every period students see g_{it-1} and update on θ_i

Conceptual Overview

Consider the following simple conceptual model:

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- Every period students see g_{it-1} and update on θ_i
- Correct belief in σ^2 important - else even a Bayesian under/overreacts to past grades

Conceptual Overview

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it}, \; \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- To study misspecification and its impact on updating, we need:

Conceptual Overview

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- To study misspecification and its impact on updating, we need:
  1. Elicit belief about grades - belief evolution descriptives

Conceptual Overview

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- To study misspecification and its impact on updating, we need:
  1. Elicit belief about grades - belief evolution descriptives
  2. Estimate the actual testing noise σ^2 - non-Bayesian updating estimates

Conceptual Overview

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it}, \; \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- To study misspecification and its impact on updating, we need:
    1. Elicit belief about grades - belief evolution descriptives
    2. Estimate the actual testing noise σ^2 - non-Bayesian updating estimates
    3. Recover each student's subjective belief about σ^2 - effect of misspecification

Conceptual Overview

$$g_{it} = \theta_i + \eta_{it} + \epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- To study misspecification and its impact on updating, we need:

  1. Elicit belief about grades - belief evolution descriptives
  2. Estimate the actual testing noise σ^2 - non-Bayesian updating estimates
  3. Recover each student's subjective belief about σ^2 - effect of misspecification
  4. RCT to shock subjective belief about σ^2 - causal evidence

Literature and Contribution (Short)

- **Misspecification**: Berk 1966, Frick, Ijima & Ishii 2020, Bohren & Hauser 2021, Lanzani (forthcoming), Ba (WP)
- **Student beliefs**: Stinebrickner & Stinebrickner (2004, 2008), Zafar (2011), Zafar and Wiswall (2015)
- **Contributions**:
  1. Provide first empirical evidence of a fundamental theoretical channel, through a rich dataset and RCT
  2. Show that misspecification leads to failure of updating, and it can be corrected in an important economic setting

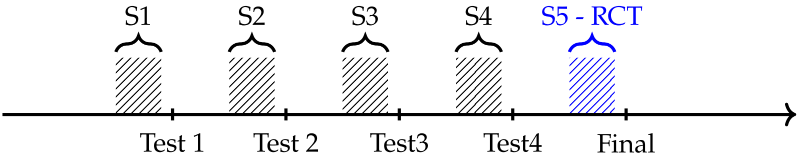
Data & Environment

Data

- We collect a unique dataset:
  - 1,500 students report their belief between each test (5 tests)
  - Elicit predictions for their grades incentive-compatibly
  - Elicit two measures of belief in testing noise
  - Not in this paper: track/elicit study hour, expected return to effort, willingness to study for grade, gender beliefs, teaching evaluation etc...

Data

- Surveys were conducted between each test after test grades were revealed



- Takeup incentivized with bonus grades: 88%/90% on 1st/5th survey

Variables

- This talk will focus on two variables:
    - Expected grade: $\hat{g}_{it}$
    - Belief about testing noise σ: measured two ways
- We elicit other measures of beliefs - incentivized whenever possible

Measures of Luck

You estimated that you would score **53** out of 100 in the final exam.

Making an accurate prediction is difficult because grades are determined by both **your MAT137 skills** and **luck.**

Let's call the difference between your actual grade and your predicted grade **"prediction error"**.

Prediction errors exist because either you were not very sure about your MAT137 skills when you made the prediction, or because you could not have possibly foreseen your luck (e.g. generous grading, poor sleep the night before,...).

How much (as a percent) do you think luck contributes to your prediction error?

Your answer last survey: 60

Error mostly caused by uncertainty about my skills — Error mostly caused by luck

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Percent prediction error caused by luck

Measures of Luck

- Prediction error: $(g_{it} - \hat{g}_{it})^2 = Var(\theta) + \sigma^2$
- Percentage of prediction explained by luck: $\frac{\sigma^2}{(g_{it} - \hat{g}_{it})^2}$

Measures of Luck

Luck can have a positive or negative impact on one's test scores.

How much <u>higher</u> (in points, out of 100) **do you think <u>you</u> would be able to score in the upcoming final exam, <u>if you were struck by good luck</u>?**

Your answer last survey: 5

Very little — A lot

0 5 10 15 20

Points

- This is the expected value of the noise, conditional on it being positive: $\mathbb{E}[\epsilon_{it}|\epsilon_{it}] \geq 0$.

Institutional Details

- Very difficult class: 64% of all students finish with a passing grade

Institutional Details

- Very difficult class: 64% of all students finish with a passing grade
- Correlation between test grades very high (0.8) - highly informative signals ($\sigma \approx 3.7$)

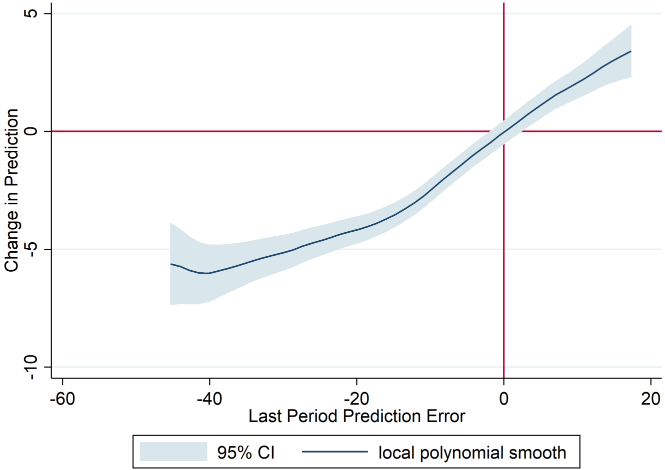
Institutional Details

- Very difficult class: 64% of all students finish with a passing grade
- Correlation between test grades very high (0.8) - highly informative signals ($\sigma \approx 3.7$)
- Instructors commit to no curving and do not release test average grade

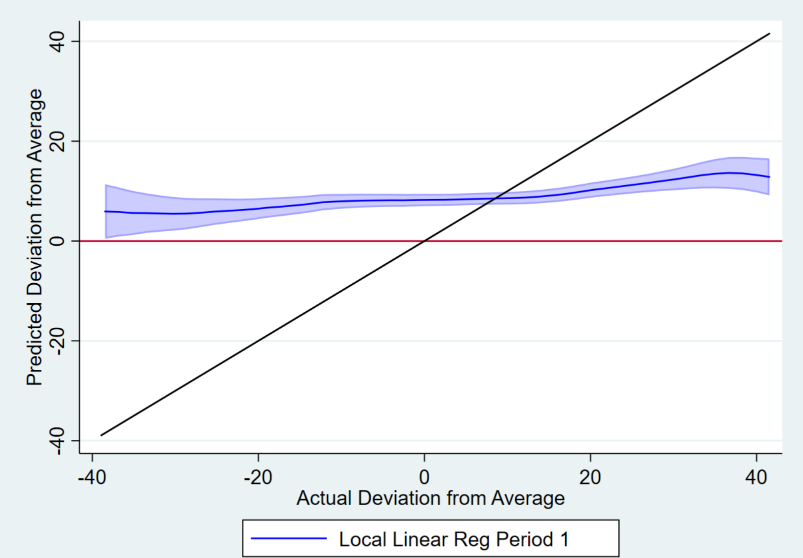
Descriptives

- Large prediction mistakes, no improvement over time:
  - Absolute prediction mistakes of 15 and 16 on the first and final tests

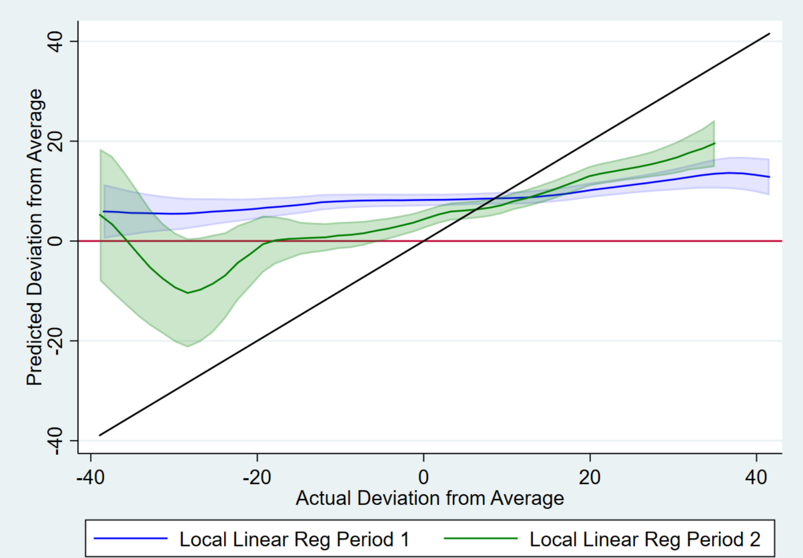
Descriptives - Prediction change condition on prediction error



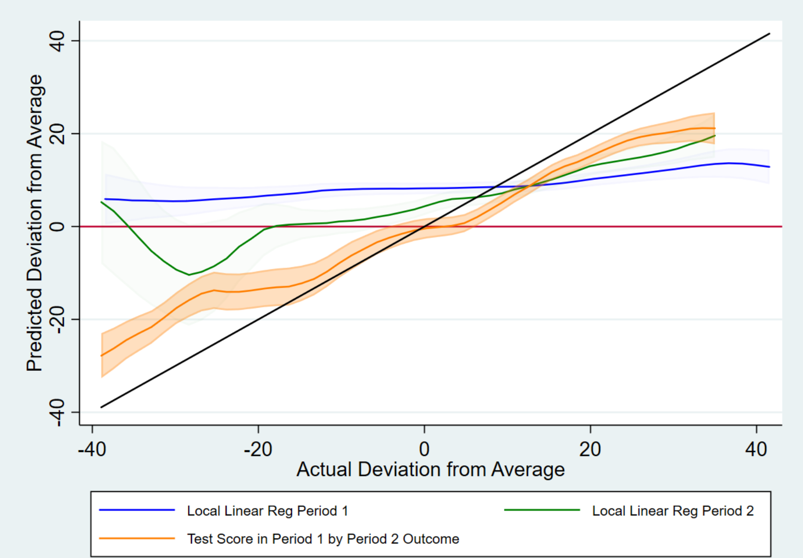
Belief Change



Belief Change



Belief Change



Descriptives

1. Large prediction mistakes, no improvement over time:
    - Absolute prediction mistakes of 15 and 16 on the first and final test

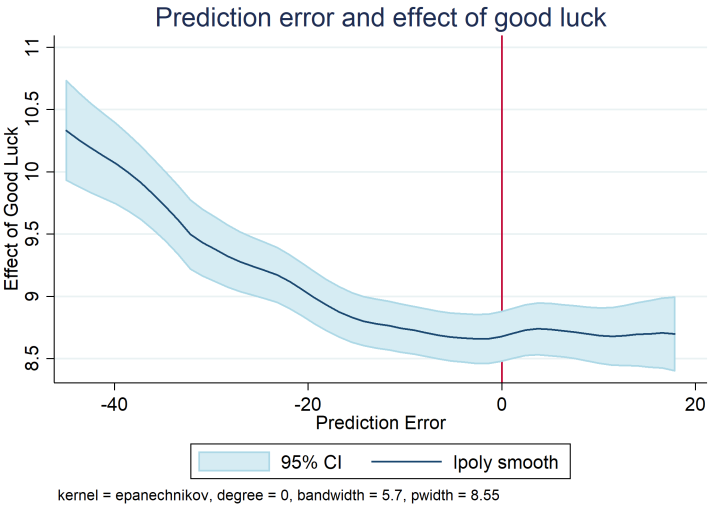
Descriptives

1. Large prediction mistakes, no improvement over time:
   - Absolute prediction mistakes of 15 and 16 on the first and final test
2. Students believe testing noise is high
   - Overestimate σ by a factor of 3

Descriptives

1. Large prediction mistakes, no improvement over time:
   - Absolute prediction mistakes of 15 and 16 on the first and final test
2. Students believe testing noise is high
   - Overestimate σ by a factor of 3
3. Higher belief of σ is correlated with higher prediction error
   - Belief in σ is stable

Belief Change



Randomized Control Trial

RCT Implementation

- We conduct an RCT: inform random students that testing is very informative

RCT Implementation

- We conduct an RCT: inform random students that testing is very informative
- Several technical and conceptual issues:
  1. Can only shock belief in testing noise - not ability

RCT Implementation

If your previous grades were determined largely by luck, then they may not be very helpful in predicting your future grades. However, if luck played only a small role, then your previous grades may be very helpful.

Using **statistics from last year,** we can see that **luck plays only a small role for most students.**

- Amongst students scoring **between 10 to 15 points below the average** across term tests - only **9%** scored **higher than 5 points above the average** on the final.
  - In comparison, **~45%** of all students scored **better than 5 points above the average** on the final exam.

- Amongst students scoring **between 10 to 15 points above the average** across term tests - only **9%** scored **worse than 5 points below the average** on the final.
  - In comparison, **~40%** of all students scored **worse than 5 points below the average** on the final exam.

Since term test grades predict the final exam grades fairly well, for most students, luck did not seem to play a big role in their grades.

Figure: Treatment Example

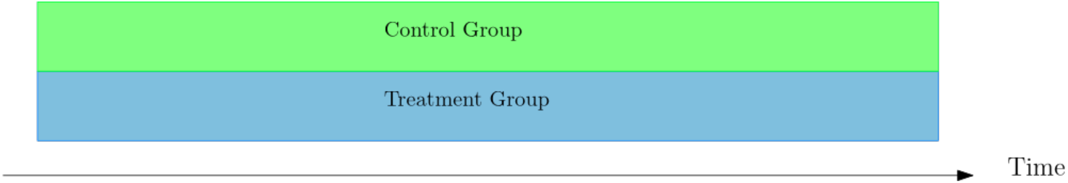
RCT Implementation

- We conduct an RCT: inform random students that testing is very informative
- Several technical and conceptual issues:
  1. Can only shock belief in testing - not ability

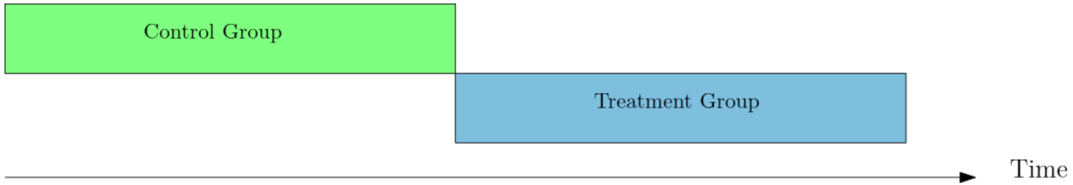
RCT Implementation

- We conduct an RCT: inform random students that testing is very informative
- Several technical and conceptual issues:
  1. Can only shock belief in testing - not ability
  2. Spillover effects and timing effects

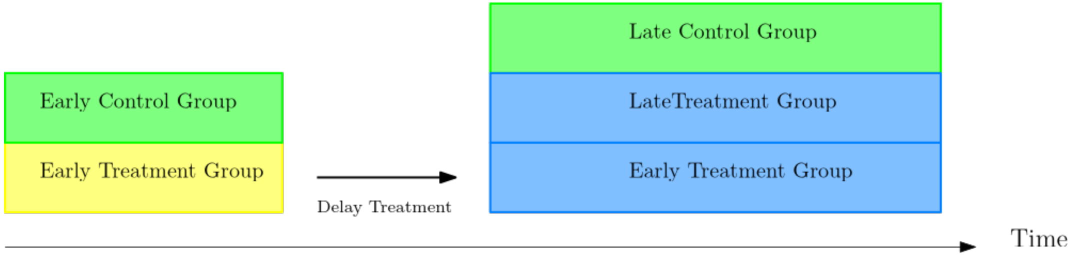
RCT Implementation



RCT Implementation



RCT Implementation



RCT Implementation

- We conduct an RCT: inform random students that testing is very informative
- Several technical and conceptual issues:
  1. Can only shock belief in testing - not ability
  2. Spillover effects and timing effects
  3. Must not induce behavioral changes - confounding channel

Balance Table (Grades Index to Test 1)

Table: Group Statistics

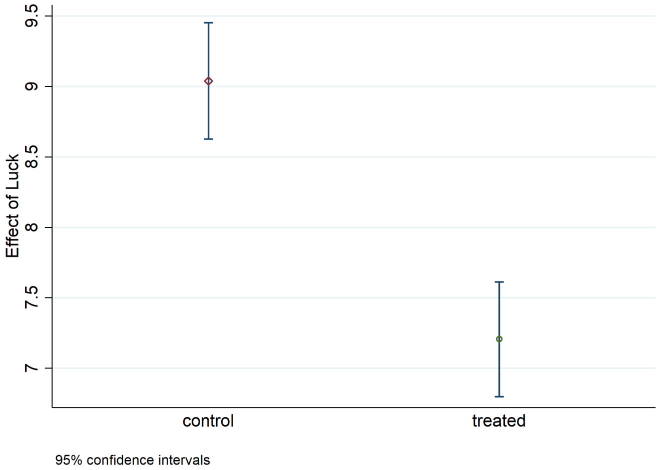
	Treatment Early		Control Early		Treatment Late		Control Late	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Test 1 grade g_{i1}	0.00	(1.33)	-1.96	(1.36)	-4.47	(0.94)	-5.56	(0.92)
Test 2 grade g_{i2}	0.00	(1.53)	0.04	(1.66)	-3.13	(1.13)	-4.01	(1.07)
Test 3 grade g_{i3}	0.00	(1.66)	0.53	(1.74)	-3.83	(1.18)	-5.00	(1.18)
Test 4 grade g_{i4}	0.00	(1.57)	-0.08	(1.77)	-6.61	(1.10)	-6.85	(1.04)
Male Prop.	0.61	(0.04)	0.72	(0.04)	0.59	(0.03)	0.61	(0.03)
First-gen Prop.	0.14	(0.03)	0.21	(0.03)	0.21	(0.02)	0.18	(0.02)
International Prop.	0.53	(0.04)	0.48	(0.04)	0.54	(0.03)	0.53	(0.03)
Pre-treatment expect grade $\hat{g}_{i5}$	65.35	(1.30)	66.35	(1.38)	63.53	(0.76)	62.56	(0.83)
Pre-treatment effect of luck $\hat{e}_{i5}$	9.55	(0.44)	9.14	(0.39)	8.84	(0.24)	8.88	(0.24)
Average past prediction error	-9.52	(0.79)	-10.19	(0.88)	-10.42	(0.63)	-11.10	(0.63)
Test 5 grade	0.00	(1.93)	1.17	(1.94)	-4.08	(1.26)	-5.22	(1.26)
Sample Size	146		153		364		385	

Note: Test grades are relative to the Treatment Early group.

Treatment Effects

- Result 1: Treated student lower their belief of the testing noise.

Treatment Effects - Belief about Effect of Luck



Treatment Effects

- Result 1: Treated student lower their belief of the testing noise.

Treatment Effects

- Result 1: Treated student lower their belief of the testing noise.
- Result 2: Treated students are 1.7 times more responsive to information

Responsiveness to Information

- To study responsiveness to information, need:
  - response measure
  - information measure

Responsiveness to Information

- To study responsiveness to information, need:
  - response measure
  - information measure
- Ideally:
  - response measure: the unit being adjusted by students
  - information: the unit students account for when adjusting the response

Responsiveness to Information

- To study responsiveness to information, need:
  - response measure
  - information measure

- Ideally:
  - response measure: the unit being adjusted by students
  - information: the unit students account for when adjusting the response

- One response and two information measures are natural:
  - Response: Grade prediction change
  - Information 1: Prediction error
  - Information 2: Deviation from average

Responsiveness to Information

- Response: $\Delta_i = \hat{g}_{i5} - \frac{1}{4}\sum_{t<5}\hat{g}_{it}$
- Information 1: $\Psi_i = \frac{1}{4}\sum_{t=1}^{4}(g_{it} - \hat{g}_{it})$
- Information 2: $\Phi_i = \frac{1}{4}\sum_{t=1}^{4}(g_{it} - \hat{\hat{g}}_{it})$

$$\Delta_i = \beta_0 + \beta_1\Psi_i + \beta_2 1_{\{i\in\mathcal{T}\}}\Psi_i + controls + \xi_i$$

$$\Delta_i = \beta_0 + \beta_1\Phi_i + \beta_2 1_{\{i\in\mathcal{T}\}}\Phi_i + controls + \xi_i$$

Responsiveness to Information

Table: Responsiveness to Information

	Ψ_i Prediction Error		Φ_i Average Deviation	
	Between	Within	Between	Within
β_1 control responsiveness	0.24^{***}	0.24^{***}	0.04	0.07
	(0.07)	(0.06)	(0.07)	(0.06)
β_2 treatment effect on responsiveness	0.16^{***}	0.16^{**}	0.10^{**}	0.09^{*}
	(0.07)	(0.07)	(0.05)	(0.05)
N	796	401×2	886	446×2

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Note: Robust standard errors in parentheses

Treatment Effects

- Result 1: Treated student lower their belief of the testing noise
- Result 2: Treated students are 1.7 times more responsive to information
- Result 3: Treated students lower prediction errors by 32%

Structural Estimation

Estimating the Impact of Misspecification

- The goal: provide an estimate of the effect of misspecification.

Estimating the Impact of Misspecification

- The goal: provide an estimate of the effect of misspecification.
- We recover:

Estimating the Impact of Misspecification

- The goal: provide an estimate of the effect of misspecification.
- We recover:
    1. Empirical prediction error E^e

Estimating the Impact of Misspecification

- The goal: provide an estimate of the effect of misspecification.
- We recover:
  1. Empirical prediction error E^e
  2. Objective model to get Bayesian + correctly specified error E^b

Estimating the Impact of Misspecification

- The goal: provide an estimate of the effect of misspecification.
- We recover:
  1. Empirical prediction error E^e
  2. Objective model to get Bayesian + correctly specified error E^b
  3. Subjective model to get Bayesian + misspecified error E^m

Estimating the Impact of Misspecification

- The goal: provide an estimate of the effect of misspecification.
- We recover:
  1. Empirical prediction error E^e
  2. Objective model to get Bayesian + correctly specified error E^b
  3. Subjective model to get Bayesian + misspecified error E^m
- Given $E^e \geq E^m \geq E^b$: $\Lambda = \frac{E^m - E^b}{E^e - E^b} \in [0, 1]$ as proportion of errors due to misspecification (lower bound!)

Structural Estimation

We estimate the following model both subjectively and objectively:

$$g_{it} = \theta_i + \eta_{it} + \bar{g}_t + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

- g_{it} - student grade
- θ_i - fixed ability, uncertain to the student
- η_{it} - student shock at time t, observed by student, autocorrelated
- ϵ_{it} - exogenous testing noise

Structural Estimation: Objective σ

- Challenge: η_{it} unobserved, residuals would overestimate σ

Structural Estimation: Objective σ

- Challenge: η_{it} unobserved, residuals would overestimate σ

Proposition

If η_{it} is a stationary AR(2) process, then σ is identified whenever N large and $t \geq 3$

Structural Estimation: Objective σ

- Challenge: η_{it} unobserved, residuals would overestimate σ

Proposition

If η_{it} is a stationary AR(2) process, then σ is identified whenever N large and $t \geq 3$

- We recover $\sigma = 3.75$ and $\mathbb{E}[\epsilon_{it} \mid \epsilon_{it} \geq 0] = 2.95$

Structural Estimation: Objective σ

- Challenge: η_{it} unobserved, residuals would overestimate σ

Proposition

If η_{it} is a stationary AR(2) process, then σ is identified whenever N large and $t \geq 3$

- We recover $\sigma = 3.75$ and $\mathbb{E}[\epsilon_{it} \mid \epsilon_{it} \geq 0] = 2.95$
- Goal: provide estimation of test correlation for Bayesian updating, $R^2 \approx 0.87$

Structural Estimation: Subjective σ

- We ask two questions about luck:
  1. Effect of Luck: $\mathbb{E}[\epsilon_{it} \mid \epsilon_{it} \geq 0]$
  2. % of prediction mistake due to luck: $r_{it} = \frac{\hat{\sigma}_{it}^2}{\text{Var}(\hat{g_{it}})}$

Structural Estimation: Subjective σ

- We ask two questions about luck:
  1. Effect of Luck: $\mathbb{E}[\epsilon_{it} \mid \epsilon_{it} \geq 0]$
  2. % of prediction mistake due to luck: $r_{it} = \frac{\hat{\sigma}^2_{it}}{\text{Var}(\hat{g_{it}})}$
- Yield very similar results and allow recovering $\hat{\theta}_i + \hat{\eta}_{it}$

Table: Summary Statistics of Estimated Subjective Variables

	Time Period				
Variable	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
$\mathbb{E}[\theta_i + \eta_{it}]$	10.91	7.99	8.14	7.37	7.19
$\hat{\sigma}_{it}$	10.67	10.99	11.55	11.54	11.33

Result

Using prior/posterior belief distributions and estimation of σ we recover:

Table: Prediction Errors - Empirical, Misspecified and Bayesian

	Test2	Test3	Test4	Test5
E_t^e	19.3	17.5	17.9	16.2
E_t^m	17.0	12.6	14.0	13.9
E_t^b	11.4	11.0	12.4	11.0
Λ_t	0.71	0.26	0.29	0.55

Conclusion

- Research Questions:
    1. Misspecification in the real world?

Conclusion

- Research Questions:
  1. Misspecification in the real world? - **Yes**
     - Overestimate testing noise by a factor of 3

Conclusion

- Research Questions:
  1. Misspecification in the real world? - **Yes**
     - Overestimate testing noise by a factor of 3
  2. Misspecification impacts updating?

Conclusion

- Research Questions:
  1. Misspecification in the real world? - **Yes**
     - Overestimate testing noise by a factor of 3
  2. Misspecification impacts updating? **Yes**
     - Causal evidence + structural model estimates half of errors due to misspecification

Conclusion

- Research Questions:
  1. Misspecification in the real world? - **Yes**
     - Overestimate testing noise by a factor of 3
  2. Misspecification impacts updating? **Yes**
     - Causal evidence + structural model estimates half of errors due to misspecification
  3. Improve updating by correcting misspecification?

Conclusion

- Research Questions:
  1. Misspecification in the real world? - **Yes**
     - Overestimate testing noise by a factor of 3
  2. Misspecification impacts updating? **Yes**
     - Causal evidence + structural model estimates half of errors due to misspecification
  3. Improve updating by correcting misspecification? **Yes**
     - Prediction errors lowered by 32% with a simple intervention