Power boost or source of bias?

Monte Carlo evidence on (machine learning) covariate adjustment in randomized trials in education

German Institute for
Adult Education
Leibniz Centre for
Lifelong Learning

Leibniz
Association

Introduction

- **Experimental research (RCT) has gained in importance in many economics and social science discipline.**
- **One prominent example: RCTs with pupils (educational RCTs) concerning**
  - educational decision-making (Barone et al., 2017; 2018; Ehlert et al., 2017; Finger et al. 2020; Piepenburg and Fervers, 2021),
  - cognitive skills (Lynch et al., 2022; Markovitz et al., 2022; Borman et al., 2020), or
  - health or psycho-social outcomes (Lazoswski and Hullemann, 2016; Murano, 2022; Taherkani, 2016), among others

Introduction & Contribution

- At the same time: comparatively little attention to estimation of treatment effects

- One controversy (mostly in mathematical statistics and biostatistics, EMA (2015)): Is it *better* to estimate effects by simple *difference-in-means comparison* or to use *regression adjustment* (i.e. regress outcome on treatment plus covariates)

Introduction & Contribution

- **My contribution: Conduct Monte Carlo simulation to compare performance of estimators**

- **Previous research**

  - mostly relies on simulated data (Miratrix et al., 2013; Kahan et al., 2016; Asafu-Adjei and Sampson, 2018; Tackney et al., 2022), exceptions are mostly from biostatistics or medical statistics (McHugh et al., 2010; Turner et al., 2012; Kahan et al., 2014; Morris et al., 2022)

  - rarely assessed ML techniques for variable selection (execptions include Wager et al., 2016 PNAS; Benkeser et al., 2019 JASA)

- **→ I use real-world data that mirrors data structures in educational RCTs (11 different outcomes, 9 different sample sizes ranging from 50 to 500)**

- **→ I employ machine-learning (LASSO) covariate adjustment**

Universität zu Köln

Statistical theory on covariate adjustment

- **Regression adjustment can increase efficiency (reduces error variance, corrects chance imbalance)**

- **Formally (Wager, 2020):** $AVar\left(\pi_{adj}\right) \leq A\mathrm{Var}\left(\pi_{unadj}\right)$**, even if functional form is misspecified (for simulation results see Kahan et al. 2016)**

Statistical theory on covariate adjustment

- **Disadavantages (Athey and Imbens, 2017; Freedman, 2008, Adv Appl Maths):**
  - Regression adjustment can introduce finite sample bias
  - In finite samples, regression adjustment can *hurt* efficiency and reduce power, especially when predictive power of covariates is low and/or the number of covariates is large relative to sample size (high-dimensional data)
  - In applied research, repeated model specification can lead to false-positives (Kahan et al., 2014)
- **→ CovAdjust has ambiguous effects on efficiency, possibly bias-efficiency trade-off**
- **→Disadvantages particularly relevant in high-dimensional data (which are the norm in educational trials!)**

Universität zu Köln

The role of machine learning

- Key idea: ML algorithms could be used to find ideal set of predictors

- Originally, ML techniques mostly applied for (out-of-sample) prediction or forecasting

- Causal machine learning literature has adapted ML algorithms as tool for variable selection in high-dimensional settings (originally in the non-experimental context; Belloni et al., 2012, 2014, 2016; Chernozhukov et al., 2018; Wager et al., 2016; Ahrens et al., 2020)

- → I borrow from the latter to find an ideal set of covariates to control for

Post-double selection LASSO algorithm

- **General Idea (Belloni et al., 2014; 2016): for outcome variable y, vector of controls x and a treatment dummy d:**
  - 1. Perform linear LASSO of y on $\boldsymbol{x}$, obtain set of predictors $\tilde{x}_y$
  - 2. Perform linear LASSO of d on $\boldsymbol{x}$, obtain set of predictors $\tilde{x}_d$
  - 3. Estimate treatment effect by regressing y on d, controlling for the union of $\tilde{x}_y$ and $\tilde{x}_d$
- **Intuition: we control for strong predictors of y as well as strongly imbalanced covariates**

The LASSO algorithm

- For a set of p regressors and n observations, LASSO solves the following minimization problem:

$$\hat{\beta}_{lasso}(\lambda) = arg \quad min \quad \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \lambda \sum_{j=1}^{p}|b_j|\gamma_j$$

(with λ: penalty parameter; γ_j factor loadings)

The LASSO algorithm

- **Intuition: Minimize squared prediction error by using as few variables as possible (note: it will shrink some coefficients to zero)**

- **One peculiarity: choosing lambda**
  - Traditional approach: cross-validation (optimizes out-of-sample prediction)
  - Alternative approach: plug-in formula (Belloni et al., 2012)

Implementation in STATA

- **Post-double LASSO can be implemented in STATA using** dsregress **(STATA corporation) or** pdslasso/lassopack (Ahrens et al., 2018; 2020)

- **For alternative algorithms:** ddml **for double-debiased machine learning (Chernozhukov et al., 2018; Ahrens et al., 2020)**

Universität zu Köln

Simulation study

- Data: NEPS, SC3 (5-graders)

- Simulated treatment between waves 2 and 3 (→ pre-info from waves 1 and 2, post info from waves 3-9)

- Outcome variables: cognitive skills (numeracy, literacy, reading speed, ICT, science), psycho-social constructs (satisfaction, math and German-related self-efficacy, motivation, self-esteem, reading behaviour

- 9 different sample sizes (between 50 and 500)

- Controls: pre-treatment outcomes, socio-demographics, social background, parenting styles, social and cultural activities (29 controls in total)

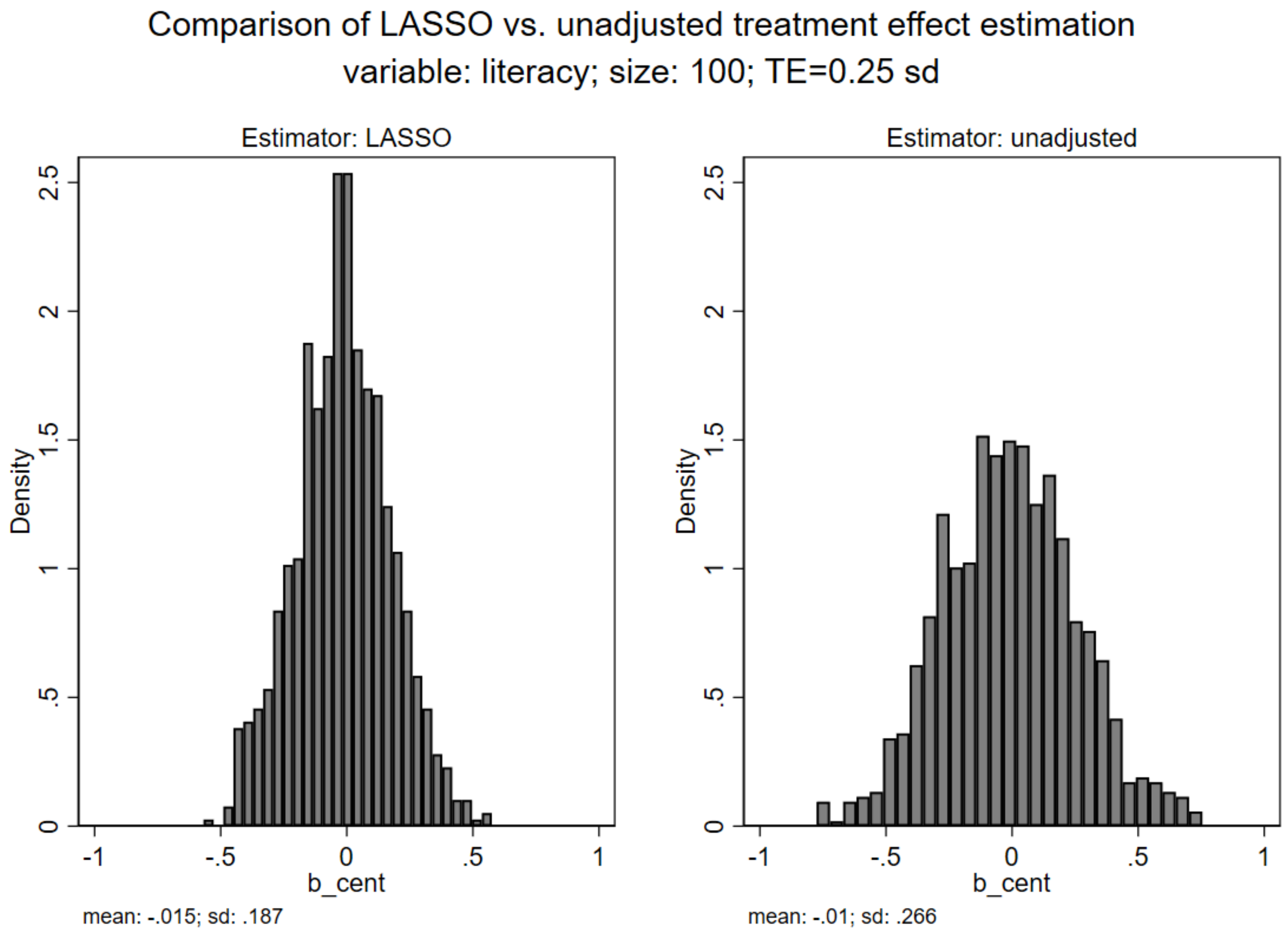
Simulation protocol

- **1. Draw sample of size** s
- **2. Assign units to treatment and control group with probability** $p = 0.5$
- **3. Simulate treatment with a size of** $\pi = 0.25 * sd(y_{pre})$
- **4. Estimate treatment effect** $\hat{\pi}$ **by**
  - a) unadjusted estimation/bivariate regression
  - b) unadjusted regression with change of outcome variable (pre-post as dependent variable)
  - c) post-double LASSO
- **5. Repeat steps 1-4c** $n = 1000$ **times**

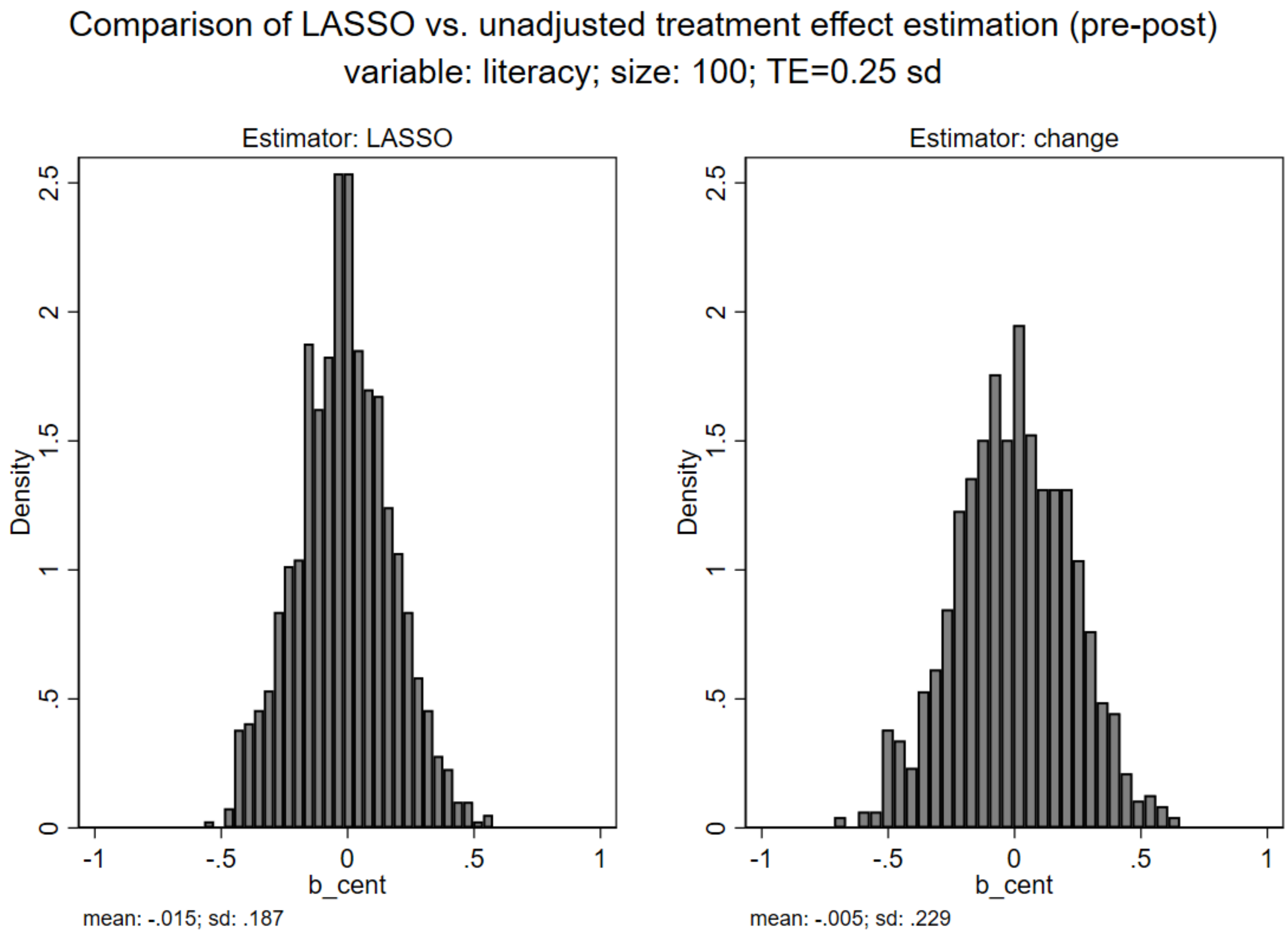
Performance measures

- Bias: $\theta = E(\hat{\pi}) - \pi$
- Variance: $Var = \sum_{i=1}^{n}(\hat{\pi}_i - E(\hat{\pi}))^2$
- $\mathbf{RMSE} = \sqrt{\sum_{i=1}^{n}}(\hat{\pi}_i - \pi)^2$
- **Power:** $p = \frac{1}{n}\sum_{i=1}^{n} 1(p_i < \alpha)$
- Coverage: $c = \frac{1}{n}\sum_{i=1}^{n} 1(lci_i < \pi < uci_i)$
- **For RMSE and power, I calculate relative differences**

Results – selected DGPs



Results – selected DGPs



Results – aggregate measures (all DGPs)

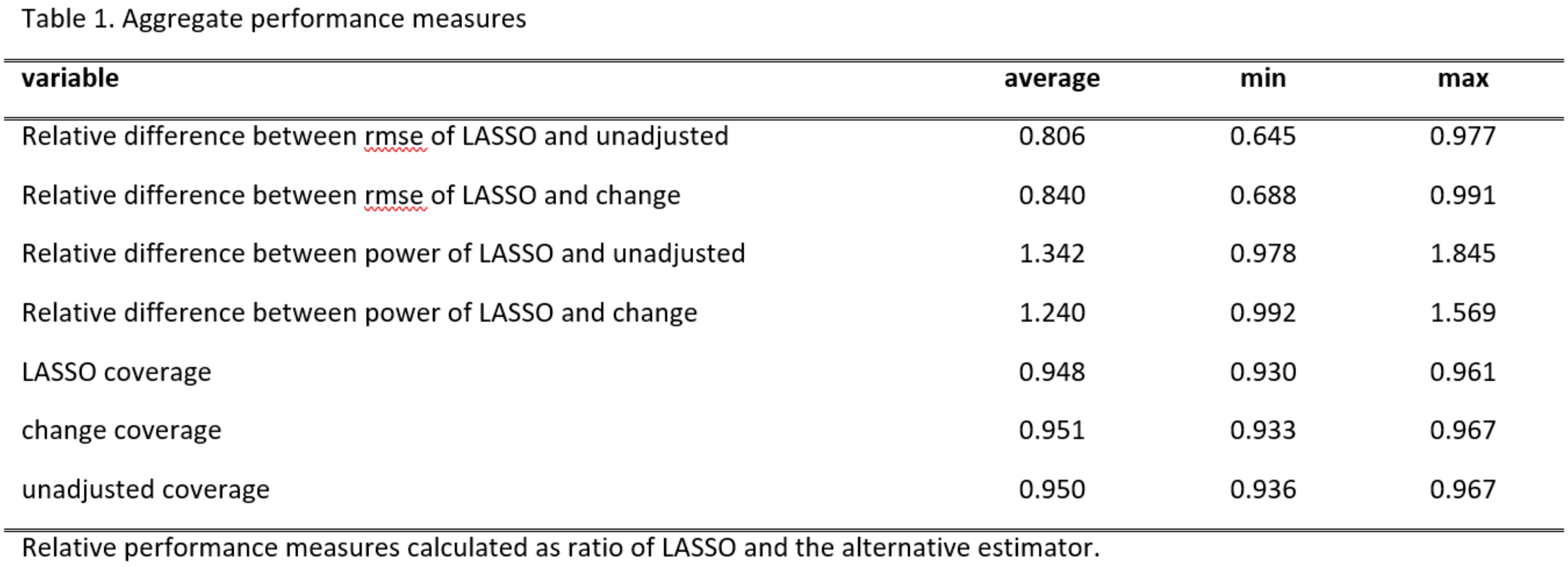


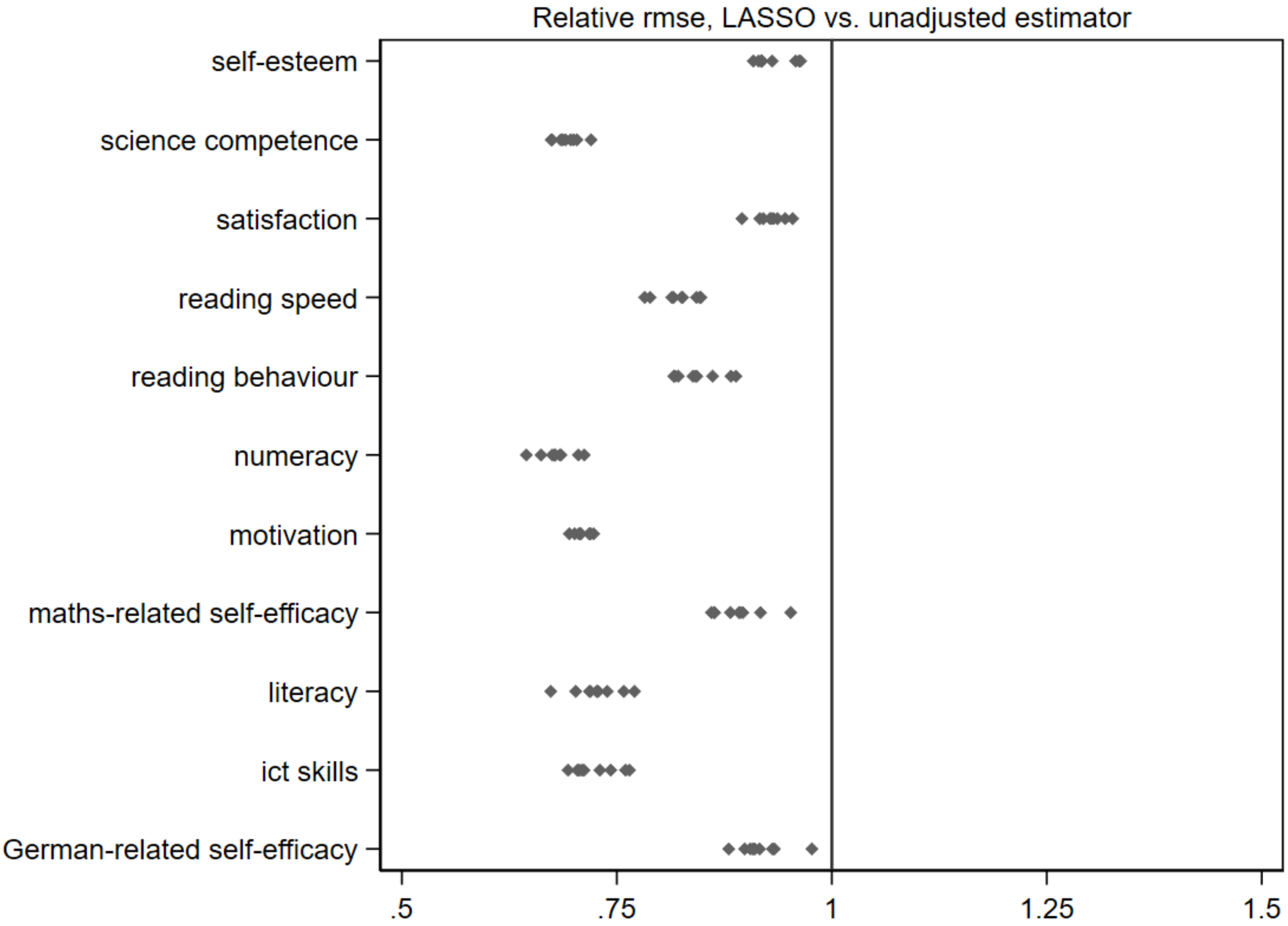
Table 1. Aggregate performance measures

variable	average	min	max
Relative difference between rmse of LASSO and unadjusted	0.806	0.645	0.977
Relative difference between rmse of LASSO and change	0.840	0.688	0.991
Relative difference between power of LASSO and unadjusted	1.342	0.978	1.845
Relative difference between power of LASSO and change	1.240	0.992	1.569
LASSO coverage	0.948	0.930	0.961
change coverage	0.951	0.933	0.967
unadjusted coverage	0.950	0.936	0.967

Relative performance measures calculated as ratio of LASSO and the alternative estimator.

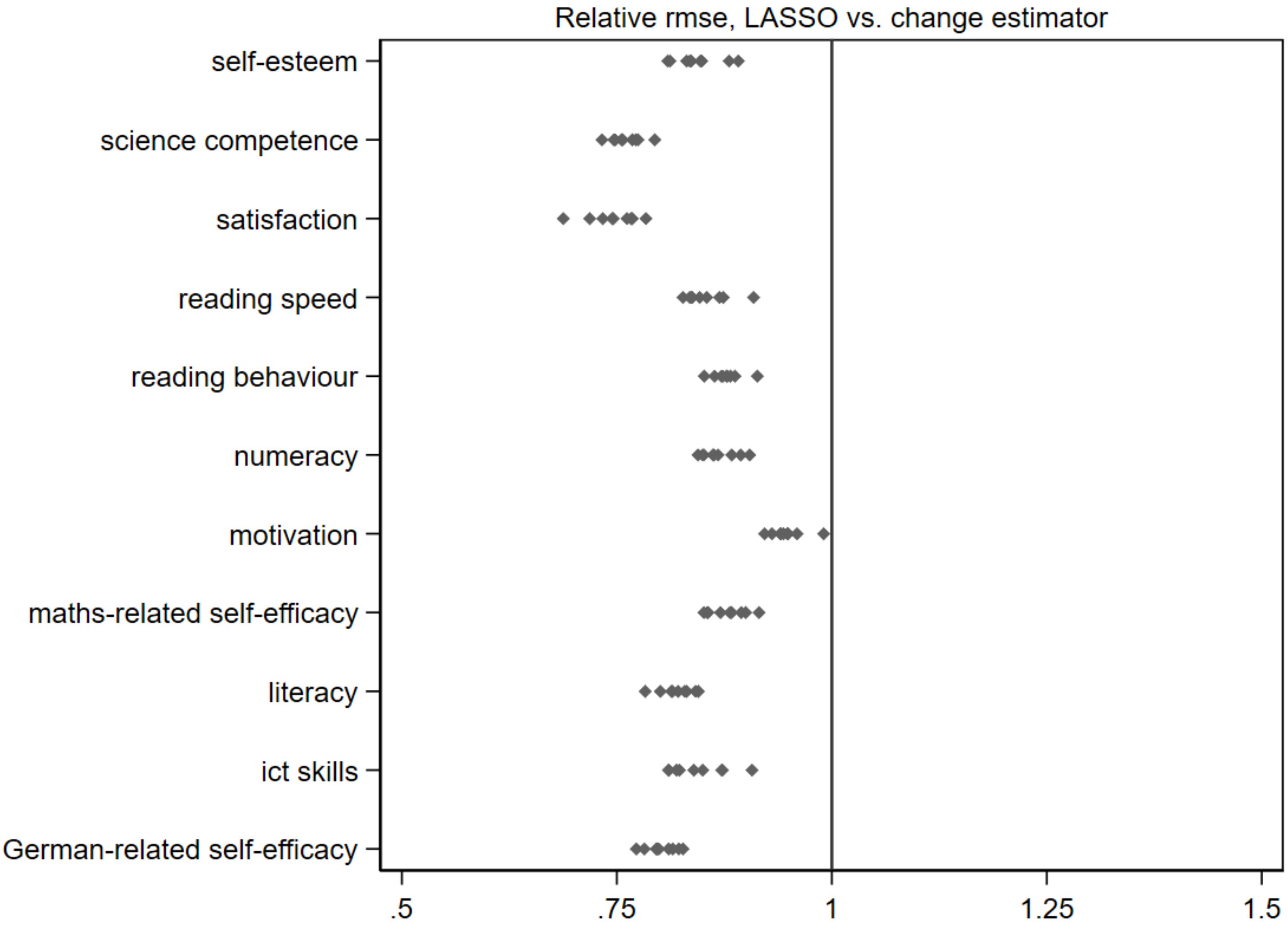
- ➢ LASSO estimation yields substantial increase of power and decrease of RMSE, (almost) never worse than unadjusted estimators
- ➢ Coverage is about 95% for all estimators
- ➢ In small samples, LASSO is (sometimes) very slightly downward biased

Results by outcome and sample size

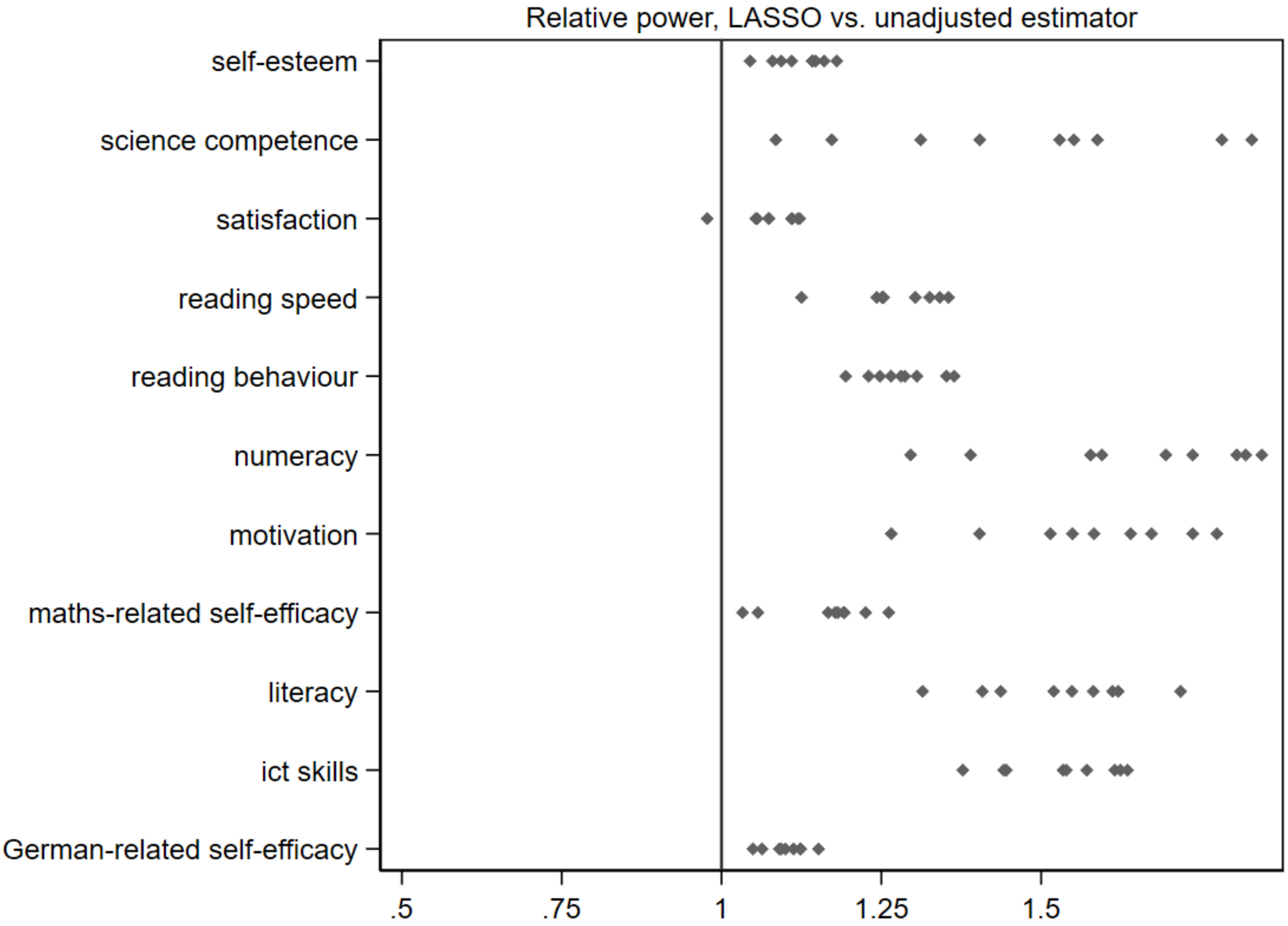


Universität zu Köln

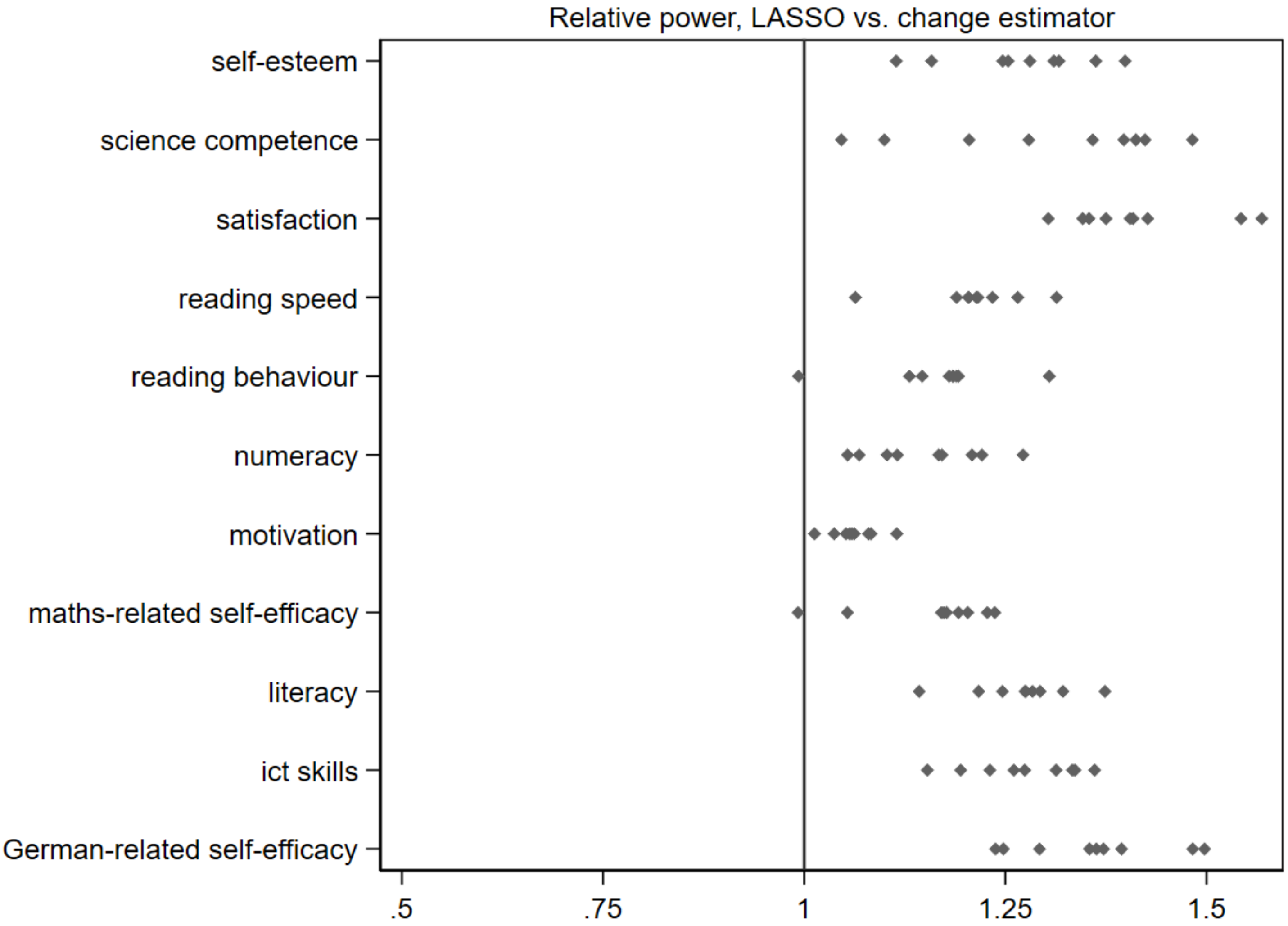
Results by outcome and sample size



Results by outcome and sample size



Results by outcome and sample size



Summary and conclusion

- **Hypothesis (following statistical theory and recent developments in the ML literature): post-double LASSO estimation can improve treatment effect estimation in educational RCTs**

- **Simulation evidence: substantial gains in power (in spite of slight downward bias) due to strong reduction of variance**

- **→ supports the more optimistic view on CovAdjust**

Outlook to future research

- Further robustness checks (e.g. different tuning parameters, different algorithms (especially double-debiased ML))
- Further evidence on different data structures (e.g. binary outcomes, survival data etc.)
- In general: recent developments in causal ML literature seem to open up a wide range of opportunities to improve estimation of treatment effects in experimental settings

Thank you for your attention!!