TERMÍN: 22.12.2020

xx25813xx
Recenzia B
Imrich Berta
imrich.berta1@gmail.com

Prosím nezasahujte do tejto tabuľky

RECENZENT/KA (meno a priezvisko, pozícia, inštitúcia)**: Imrich Berta**
NÁZOV MATERIÁLU: **Digitalization in Slovak Primary Schools: A Wasted Opportunity**
TYP VÝSTUPU*[1]: **Analýza**
(pri spoločných výstupoch uviesť aj typy individuálnych vkladov):
ANALYTICKÝ ÚTVAR, REZORT: **Ministerstvo práce, sociálnych vecí a rodiny SR - Inštitút sociálnej politiky**
AUTORI/KY: **Marcela Veselkova;**
SPOLUAUTORI/KY: - - ; - - ; - - ; - -
RECENZNÝ FORMÁT*[2]: **2**

PRIPOMIENKY:

P.č.	Pripomienka sa vzťahuje k (strana, odsek):	Text pripomienky*[3]	Odôvodnenie pripomienky	Vysporiadanie sa s pripomienkou*[4]
1	10,3	There is no explanation how the propensity score is calculated, and which variables are used in the process.		Accepted. Explanation was added to the methodological section: “The propensity score model should include all variables related to the outcome to decrease the

				variance of an estimated exposure effect without increasing bias (Brookhart et al. 2006). In this paper, variable selection for propensity score estimation was automated via a variable selection model (see Bon 2022; Hahn et al. 2020). I regressed outcome against covariates (excluding treatment variable) using R package “BART”, version 2.9 (McCulloch et al. 2021) to select a subset of covariates most associated with the outcome. Then I used these covariates to estimate propensity scores using covariate balancing propensity score (CBPS) methodology, which models treatment assignment while optimizing the covariate balance (Imai & Ratkovic 2014). The propensity

				scores were estimated using R package "WeightIt", 0.13.1 (Greifer 2022). Balance diagnostics is available in Appendix 1."
2	10,4	It is a good practice to include a simple convergence test in results. Default Geweke convergence diagnostic would be suitable.		Accepted. The stable Gelman-Rubin convergence diagnostic (Vats and Knudson 2018; Knudson and Vats 2019) was used to assess the chain convergence. In the final model, I ran 10 chains with 10 000 iterations each (in addition to 2 000 burn-in iterations). All chains converged. Explanation was added to the methodological section: "The chain convergence was assessed using the package "stableGR", version 1.1, which calculates stable Gelman-Rubin convergence diagnostic for Markov chain

				Monte Carlo (Knudson and Vats 2021; see also Vats and Knudson 2018). In all cases, the potential scale reduction factor was close to 1, indicating that the sample collected by the Markov chain has converged to the target distribution. Convergence was achieved also according to the function n.eff, which calculates effective samples size for a set of Markov chains using lugsail variance estimators. Further visual convergence diagnostics is available in Appendix 2."
3	10,4	Any paper with advanced modelling techniques, call it Bayesian machine learning in this case, should include some goodness-of-fit measures like R^2. The best practice is comparison to linear regression performance on the same variables, and cross-validated performance of the final model.		Accepted. Final model was cross-validated. Comparison of the BART fit to the linear regression fit was added. R-squared was calculated for both models. Explanation was added to the Results section: "The

				estimated BART model explains around 40% of the variability of mathematical reasoning, 48% of the variability of scientific reasoning and 47% of the variability of overall reading. For comparison, linear regression model explains around 31% of the variability of mathematical reasoning, 40% of the variability of scientific reasoning and 38% of the variability of overall reading. Comparison of BART fit to linear model is available in Appendix 5.”
4	14,3	Subject of the posterior histograms is not clear. Clear indication that the posterior histograms are differences between groups with maximum and minimum ATE might be helpful.		Not accepted. Subgroup differences were quantified by plotting the posterior histogram of the difference between the rightmost and leftmost nodes of the rpart tree. These do not necessarily have to be subgroups with maximum and minimum ATE,

				depending on the tree depth.
5	15, Figure 5	In the ‘rpart’ decision tree for reading lessons ATE, there is a branching on ‘ps’ variable. ● If it is the preschool variable, it should be clearly stated. ● In case it stands for the propensity score, it shouldn’t been included in variables for rpart regression.		Accepted. Propensity score was excluded from rpart regression.
6	16,1	Last sentence, 2915 is probably a typo.		Accepted.
7	16,2	CART is mentioned as a decision tree method but ‘rpart’ is used for the figures above.		Accepted. rpart is the R implementation of the CART model. Terminology in the paper was unified and the model is consistently referred to as an “rpart model”.
8	17,3	Statement ‘This paper shows that massive computerization of Slovak primary schools failed to improve student performance.’ is not supported by data or results in this paper.	In the last two decades, there might have been an overall positive effect on performance, as the ICT resources at home were lower. I wouldn’t recommend generalising the results of the paper based	Accepted. Text was revised as follows: “This paper examined the impact of computer availability in mathematics, science and reading lessons on performance of Slovak fourth-graders, using the data from the

			on 2015 data, to the whole history of computerization in modern Slovakia.	2019 round of eTIMSS and the 2016 round of PIRLS testing. The impact of computer availability in mathematics and science classes is statistically uncertain. However, computer availability in reading classes had a positive impact: it improved the overall reading score by 0.10 to 0.13 standard deviations. Based on the fingings of recent meta-analyses of education interventions, this can be considered a medium effect size (Kraft 2020)."
9	Overall	For future research, I would recommend comparing the residuals from the final model against geography of Slovakia. Just a choropleth map on NUTS3 or LAU 1 would be a great check.		Thank you for the suggestion.
10				

CELKOVÉ HODNOTENIE (recenzent/ka vyplní túto časť po vysporiadaní sa s pripomienkami analytickou jednotkou):

All my comments were sufficiently answered and resolved.

The author expertly uses advanced statistical methods for causal inference to approach the treatment causality, and its effect magnitude, on an investigated educational outcome from non-randomized cross-sectional data.

I appreciate the literature review and comparison of results, to those documented by other researchers, in the Conclusion section.

The methodology is thoroughly described in the paper and the results are well presented.

Insights from the subgroup treatment effects are without a doubt valuable for policy making.

---

[1] Výber medzi: 1. analýza (komplexný analytický materiál s návrhmi konkrétnych systémových opatrení); 2. komentár (rozsahovo menší analytický materiál venujúci sa konkrétnemu čiastkovému problému); 3. manuál (metodické usmernenie vyplývajúce z potreby zjednotenia procesov a postupov v konkrétnej oblasti).

[2] Formát 1 pre komentár/manuál (2 recenzenti bez povinného odborného workshopu); Formát 2 pre analýzu (3 recenzenti a povinný odborný workshop).

[3] Do tabuľky značiť pripomienky zásadného metodologického a obsahového charakteru (nie štylistické či gramatické opravy).

[4] Vyplní analytická jednotka: pripomienka bola akceptovaná / pripomienka nebola akceptovaná a zdôvodnenie / pripomienka bola čiastočne akceptovaná a zdôvodnenie.