

Findor et al.(2023) Replication

Asad Tariq and Isabella Mullen

March 15, 2025

Contents

1	Introduction	2
2	Original Paper	2
2.1	Design Overview: Study 3	2
2.2	Methodology	3
2.3	Replication Results	5
3	Additional Model	6
3.1	Model Design	6
3.2	Results	7
3.2.1	In-sample Predictive Performance	7
3.2.2	Out-of-sample Predictive Performance	10
3.2.3	Quantity of Interest	11
3.2.4	Preferred Model	14
4	Conclusion	14
4.1	Discussion	14
5	Appendix	16

1 Introduction

This project replicated Study 3 from *"Equality, Reciprocity, or Need? Mitigating Ethnocentric Bias in Policy Evaluation with Distributive Fairness"* by Andrej Findora, Matej Hrušák, Roman Hlatký, Tomáš Hrustič, and Zuzana Bošellová. The original study examined how different justifications for a social housing project influenced ethnocentric biases in policy opinions. The treatment conditions were three scenarios framed around need, equity, or reciprocity. The reciprocity condition elicited the most support for building the housing project, while the need condition evoked the most disagreement. This was statistically distinguishable from the effects associated with the other conditions (equality and need). This project aimed to replicate Table 3 from the original paper and to explore alternative models fit to the same dependent variable. Replicating Study 3 provided an opportunity to explore how the addition of supplementary covariates changed the analysis.

The original paper used an ordered logistic regression model to infer the effect of conditional framing of a social housing project on willingness to support the project. We constructed an alternate model that incorporated additional demographic covariates (age, education, sex, and region) to improve the prediction power of the model. When comparing the original and alternate models, the alternate model provides a better in-sample fit and marginally improved out-of-sample predictive accuracy. Incorporating demographic variables highlights additional contextual patterns in the data. Older individuals exhibited stronger support than younger individuals, and there were clear regional differences.

2 Original Paper

2.1 Design Overview: Study 3

For this project we chose to replicate Study 3 from the original paper (Findor et al. 2023). The experiment was between-subjects with a control group and three treatment conditions (equality, reciprocity, and need). Each participant read a control paragraph describing a fictional village with a Roma settlement where a new apartment building could be built. The scenario explained that the new housing was for the poorest inhabitants of the Roma settlement and that they would be required to pay rent. It also stated that 95% of the funding came from EU funds, while the remaining 5% came from the municipality. Participants in the treatment groups were then randomly assigned an additional paragraph corresponding to one of the treatment conditions: equality, reciprocity, or need. In the equality condition, the same amount of money was allocated to projects intended for the non-Roma community. In the reciprocity condition was that only those who worked on the construction of the building could live there. And in the need treatment was

that only those who needed the most assistance could live there. After reading the scenarios, participants answered whether they agreed or disagreed with the project on a scale from 1 (“strongly disagree”) to 4 (“strongly agree”). Additional questions were asked but they were not relevant to this replication project.

The unit of analysis was an individual-level respondent. The sample size for Study 3 consisted of 1,009 survey respondents. This sample was meant to be a nationally representative sample of Slovaks. The original researchers set quotas on gender, age, region, municipality size, and education. The final sample for Study 3 was 52% female, with a median age of 44.7. The respondents were randomly recruited through an online platform. The survey participants opted in to an online survey, were aware that they were participating in a research study, and were compensated for their time.

Because this was an online survey, neither we nor the original researchers could verify with complete certainty that each survey was taken independently. However, assuming that each participant needed an account to complete the survey and was exposed to only one randomly assigned condition, it was reasonable to assume independence within the sample. In addition to the randomization of conditions, the relatively large sample size and demographic quotas suggested that the sample was diverse and representative.

2.2 Methodology

The original authors used their model to investigate the effects of different framing conditions (equity, reciprocity, and need) on support for redistributive policies in Slovakia. They used ordinal logistic regression to analyze how framing a social housing policy with different distributive fairness principles affected individual support for the policy. They quantified the impact of each framing condition on personal support levels to determine which justifications were the most effective in reducing ethnocentric bias and increasing public support for policies that benefited marginalized groups. This model applied causal inference to assess how changes in the treatment conditions influenced support for the project. This was achieved by randomizing the condition for each participant and comparing responses between groups.

The authors used this model to explain the dependent variable and treated ‘group’ as the predictor.

$$\log \left(\frac{P(E3 \leq j)}{P(E3 > j)} \right) = \beta_0 + \beta_1 \text{Equality} + \beta_2 \text{Reciprocity} + \beta_3 \text{Need} \quad (1)$$

$$\text{polr}(E3 \sim \text{group}) \quad (2)$$

The dependent variable in Study 3 was the willingness to support the social housing project, referred to as ‘E3’ in the data. The dependent variable was an ordinal factor with four levels (1–4). Table 1 presents

the descriptive statistics for personal agreement with the social housing policy across four experimental conditions: control, equality, proportionality, and need. The proportionality condition had the highest mean agreement (2.83), while the need condition had the lowest mean (2.40), indicating that respondents were more supportive when fairness was framed around proportional contributions rather than need-based justifications.

Table 1: *Descriptive statistics - Personal agreement*

Group	N	Min	Q1	Median	Q3	Max	Mean	SD	Skew	Kurtosis
Control	252	1	1	3	3	4	2.50	1.08	-0.121	1.72
Equality	249	1	2	3	3	4	2.63	1.04	-0.328	1.93
Proportionality	254	1	2	3	4	4	2.83	1.06	-0.561	2.10
Need	254	1	1	3	3	4	2.40	1.07	-0.054	1.17

Note: We replicated the results from this table in the original paper (Findor et al. 2023) (Table A19), except for the kurtosis values. We assumed that this discrepancy was due to the use of a different R function or a different definition of kurtosis. Since this value was not relevant to the scope of this project, we left the values as is. The values included in Table 1 are our calculations.

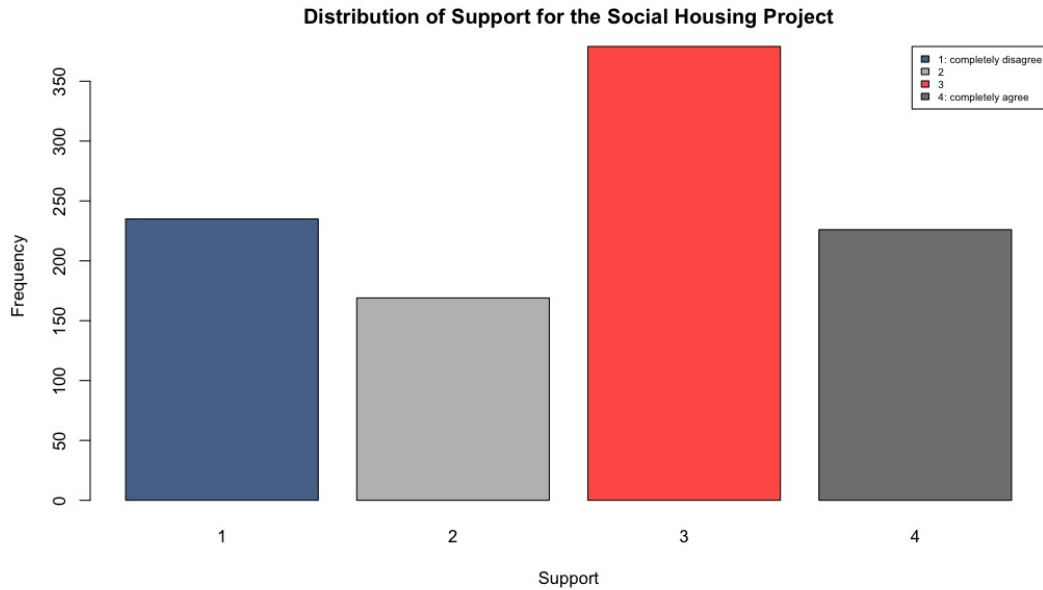


Figure 1: Distribution of the Dependent Variable

Figure 1 displayed the overall distribution of the dependent variable for all respondents. The x-axis repre-

sented levels of support, ranging from 1 ("completely disagree") to 4 ("completely agree"), while the y-axis represented the frequency of responses. The highest frequency was observed at level 3 ("agree"), followed by levels 1 ("completely disagree") and 4 ("completely agree"), with level 2 being the least common response.

Figure 2 provided a breakdown of the support for the housing project by condition (control, equality, proportionality, and need). Each bar represented the number of respondents who selected a given level of support, with the colors indicating the experimental group. The proportionality (reciprocity) condition appeared to generate the highest level of support compared to the other groups, while the need frame showed a higher proportion of disagreement than the other conditions. The control and equality conditions displayed relatively balanced distributions.

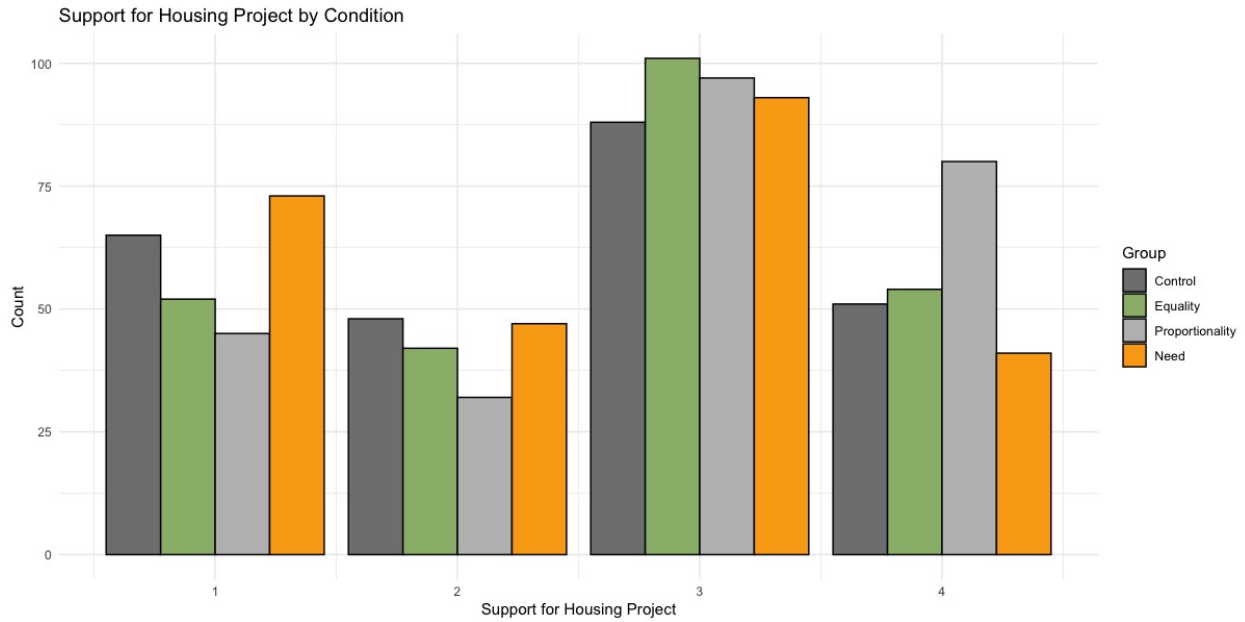


Figure 2: Distribution of the Dependent Variable by Group

In the data preparation step, the authors handled missing data by explicitly assigning NA in the income recoding process. The summary statistics code ensured that missing values were ignored, preventing distortions in calculations. However, in the ordinal logistic regression model, missing data was not explicitly handled because there were no missing data.

2.3 Replication Results

Table 2 (see on page 6) presented the results of an ordinal logistic regression model assessing support for building social housing under different distributive fairness conditions. The control group served as the baseline, with the effects of reciprocity, equality, and need frames estimated relative to it. The results indicated

that the reciprocity frame significantly increased support for social housing compared to the control condition, meaning that respondents exposed to this condition were 83% more likely to express greater support. The equality frame and the need frame did not show statistically significant differences from the control condition. When using reciprocity as the baseline, both the equality and need frames significantly reduced support. The original authors also analyzed the results with the reference category changed to proportionality (reciprocity) rather than control. This did not alter the results but provided a more straightforward interpretation. These findings highlighted how different fairness frameworks influenced attitudes toward redistribution, reinforcing the idea that reciprocity-based narratives resonated more strongly than need-based appeals.

Table 2: *Support for building social housing*

	Estimate	Std. Error	<i>p</i>	OR [95% CI]
Baseline = Control				
Reciprocity	0.60	0.16	<.001	1.83 [1.33, 2.52]
Equality	0.22	0.16	.172	1.25 [0.91, 1.71]
Need	-0.16	0.16	.323	0.85 [0.62, 1.17]
Baseline = Reciprocity				
Equality	-0.38	0.16	.018	0.68 [0.50, 0.94]
Need	-0.76	0.16	<.001	0.47 [0.40, 0.64]

Using the original authors’ data and the replication code, we successfully reproduced the results of table 3 in the original paper (see table 2) on page 6). Our replication confirmed that the statistical estimates, confidence intervals, and significance levels remained consistent with the published findings. This successful replication supported the robustness of the original analysis and reinforced the conclusions.

3 Additional Model

3.1 Model Design

After replicating the original model, we constructed an alternate ordered logistic model with a different linear predictor (with more than one independent variable). We chose **age** (AGE), **education** (EDU), **sex** (SEX) and **region** (REG) along with **group** (which was already present in the original model) to be part of the alternate model. By including independent variables that related to demographics, we believed that we

will have better holistic and contextual predictions of their willingness to support the social housing project.

The alternate model we considered was the following:

$$polr(E3 \sim group + AGE + EDU + SEX + REG) \quad (3)$$

We chose this model for two main reasons: we wanted to determine if the addition of demographic variables would incur a change in the predictions for the dependent variable, and whether a more complex model (with a greater number of predictors) would perform better than a simpler model (with just one predictor).

See the regression table (Table 4) on page 16 displays the coefficients and their standard errors for the alternate model.

3.2 Results

3.2.1 In-sample Predictive Performance

We began by comparing the two models using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) metrics, the results of which are shown in Table 5 on page 17. Using the AIC metric, we observed that the alternative model outperformed the original model, with an AIC value of 2644 compared to 2694 for the original. This indicated that it provides a better fit to the data while balancing the complexity of the model.

Next, a calibration plot (Figure 3) was made to compare the prediction probabilities of each model. The calibration plot visually represented how well the predicted probabilities of two models, 'alter' (red) and 'logit' (blue), align with the observed outcomes for the dependent variable. The 'logit' model was the original model. Each vertical strip on the x-axis corresponded to one of the four observed categories, while the y-axis represents the predicted probability of the observed outcome of each model. Ideally, a well-calibrated model should have assigned high probabilities (close to 1) to the correct category and low probabilities to the incorrect categories.

The smooth trend lines further highlighted the systematic differences between the models. The fact that the probabilities did not cluster near 1.0 for the observed outcomes suggested that both models did not predict the true category probabilities well. This may have indicated underlying issues such as class imbalance or over-regularization.



Figure 3: A multi-category calibration plot of the predicted probabilities of the two models for each category.

We then performed a likelihood ratio test to determine whether our alternate model provided a better fit to the data. The p-value obtained (as shown in Table 5) was extremely small, giving a strong indication that, despite having additional parameters and complexity, the alternate model is significantly better at fitting the data compared to the original.

Lastly, we plotted the ROC curves of the two models. Figure 4 representing the original model, showed ROC curves that closely follow the diagonal reference line. This indicated poor predictive performance. The model's ability to distinguish between different classes was weak, as the curves did not significantly rise above the diagonal.



Figure 4: An ROC plot for the original model.

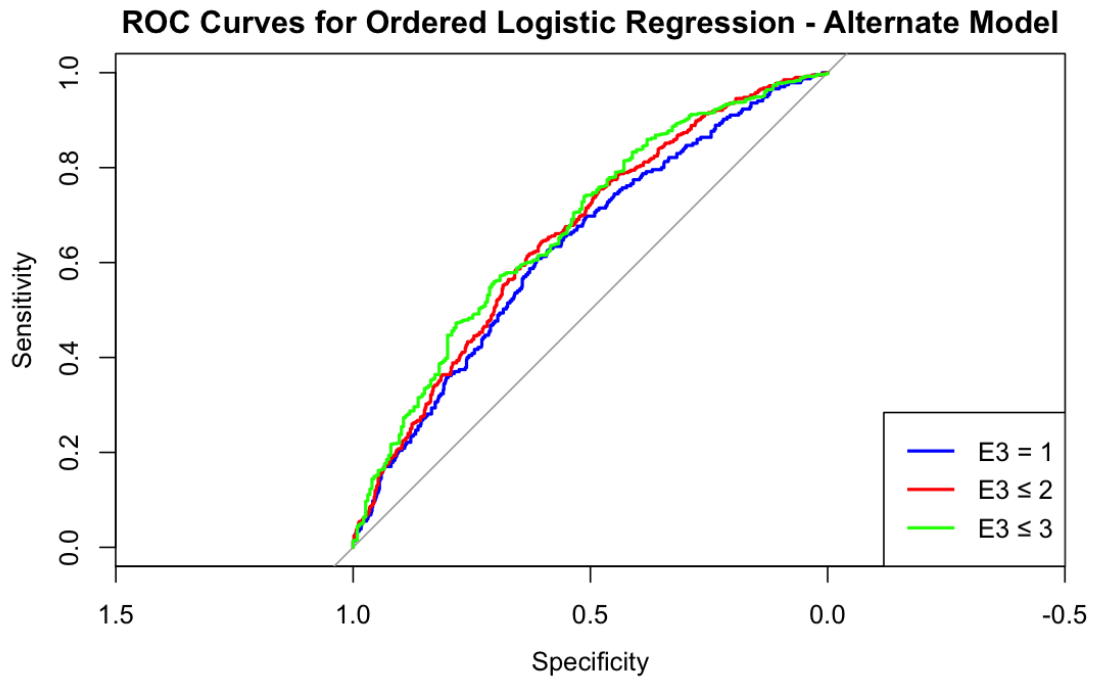


Figure 5: An ROC plot for the alternate model.

Comparatively, the ROC curves in Figure 5, for the alternate model, were further above the diagonal, indicating better sensitivity-specificity trade-offs. This suggested that the alternate model was stronger at discriminating, as its ROC curves are farther from the diagonal and have a larger area under the curve (AUC). The separation between the curves for different threshold values of

E3 were also more distinct in the alternate model, suggesting that it provided better differentiation between the ordered categories.

Overall, the alternate model demonstrated improved classification performance compared to the original model, and hence, the modifications made in the alternate model resulted in a better ability to predict the ordered outcomes.

3.2.2 Out-of-sample Predictive Performance

To evaluate the out-of-sample predictive performance of the two proportional odds logistic regression models, we performed a 10-fold cross-validation. To perform cross-validation the dataset was split into 10 equally sized folds (subsets). Each of these folds were used as a test set while the left over 9 folds were used in the training set. This process was repeated 10 times, ensuring that each observation appeared in the test set only once. The aim of both models was to predict the ordinal outcome variable E3. The results are shown in Table 3 below.

The models were then evaluated with each link function on the basis of their accuracy obtained from the cross-validation results. The purpose of checking the accuracy of cross-validation across different link functions was to determine which link function best fits the relationship between the predictor variables and the ordinal outcome (E3).

For the original model, the accuracy remained consistent at 0.376 across all five link functions: cauchit, cloglog, logistic, loglog, and probit. This indicated that there was no change in the predictions of the model regardless of which link function was chosen. On the other hand, our alternative model showed some improvements over the original, achieving better accuracy with the cloglog link function, which was chosen for the final model, yielding an accuracy of 0.390.

Table 3: Cross-Validation Results

	Model	Method	Accuracy	Kappa
1	Original	cauchit	0.376	0
2	Original	cloglog	0.376	0
3	Original	logistic	0.376	0
4	Original	loglog	0.376	0
5	Original	probit	0.376	0
6	Alternative	cauchit	0.376	0.052
7	Alternative	cloglog	0.390	0.071
8	Alternative	logistic	0.375	0.049
9	Alternative	loglog	0.372	0.041
10	Alternative	probit	0.375	0.047

3.2.3 Quantity of Interest

Before building a scenario by manually specifying the predictor values to investigate how one particular independent variable impacted our dependent variable, we wanted to look at how age and region in particular relate to the personal willingness to support the social housing project.

Figure 6 illustrated the effect of age on personal willingness to support the social housing project, measured at four levels of willingness ($E3 = 1$ to 4). Each panel represented a different level of support, with age on the x-axis, and the predicted probabilities on the y-axis.

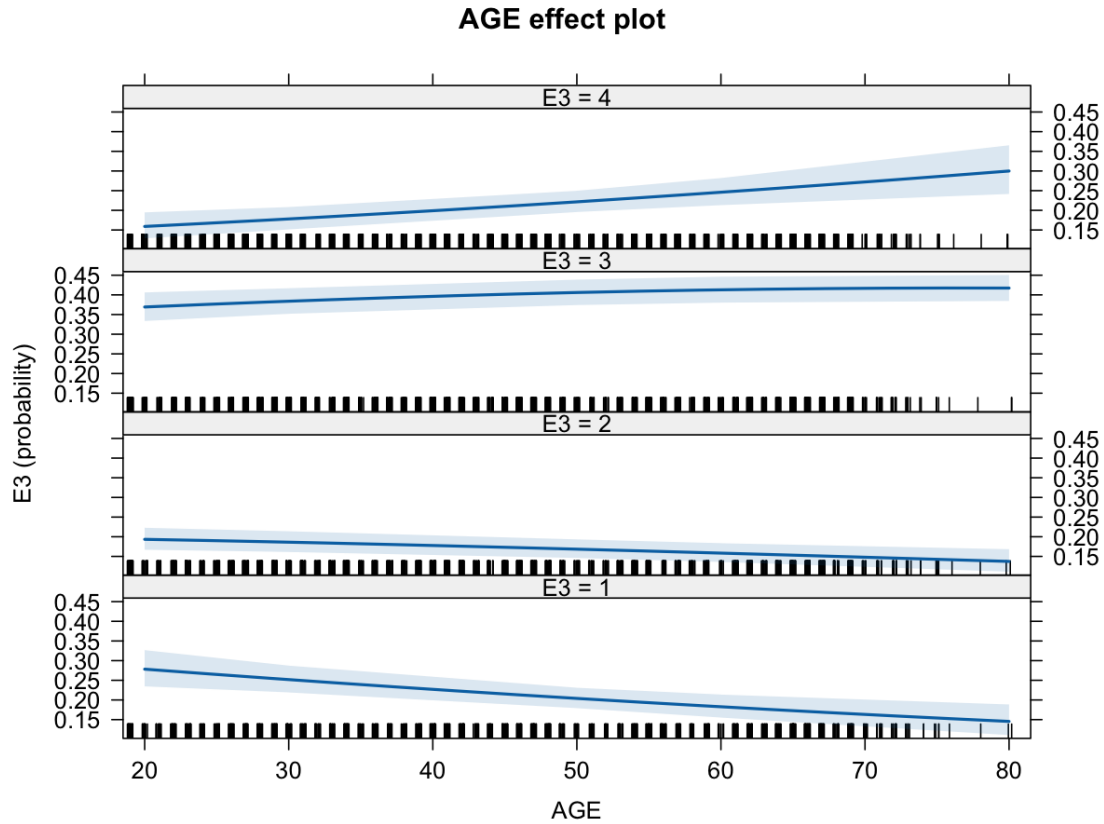


Figure 6: Effect of the age predictor on the personal willingness to support the social housing project.

From the figure, we can see that people who were most supportive of the project ($E3 = 4$) tended to show a slight increase in the probability of support as they aged. Which means that older individuals were more likely to express strong support. Conversely, individuals with lower levels of support ($E3 = 2$ and $E3 = 1$) showed a decrease in the probability of support with age. Which means that younger individuals were more likely to oppose the housing project compared to older individuals.

Figure 7 presented the estimated effect of region (REG) on personal willingness to support the social housing project (E3), with regions on the x-axis. Each panel corresponds to a different level of willingness (E3 = 1 to 4). The blue points represent predicted probabilities, while the accompanying orange error bars indicated confidence intervals, for the uncertainty in these estimates.

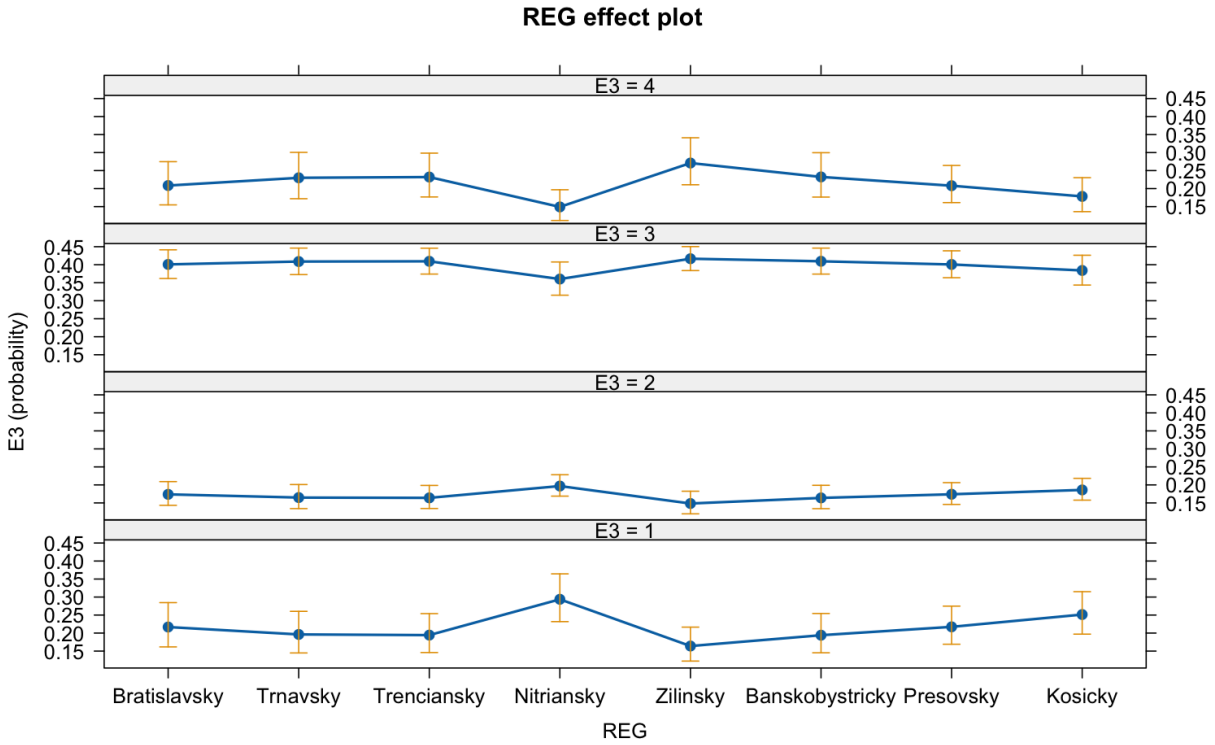


Figure 7: Effect of the region predictor on the personal willingness to support the social housing project.

We can see that the probability of strong support ($E3 = 4$) fluctuated between regions, with Zilinskii and Nitrianskii having relatively higher probabilities. Conversely, Kosicky had a lower probability of strong support. In contrast, opposition to the project ($E3 = 1$) followed an opposite pattern, with higher probabilities observed in Nitriansky and Kosicky. This suggests greater regional resistance to the initiative. The probability of moderate support ($E3 = 3$) remained largely stable across regions.

These findings suggested that regional context played a role in shaping attitudes toward social housing, potentially reflecting underlying socioeconomic, political, or cultural differences across regions.

Moving on to a particular scenario, the density plot below (Figure 8) visualized the implied effect of regional differences (Bratislavsky vs. Zilinsky) on personal agreement with the construction of an apartment (E3). The x-axis represented the change in predicted probability of different levels of agreement due to the regional difference, while the y-axis showed the density of these estimated probability shifts.

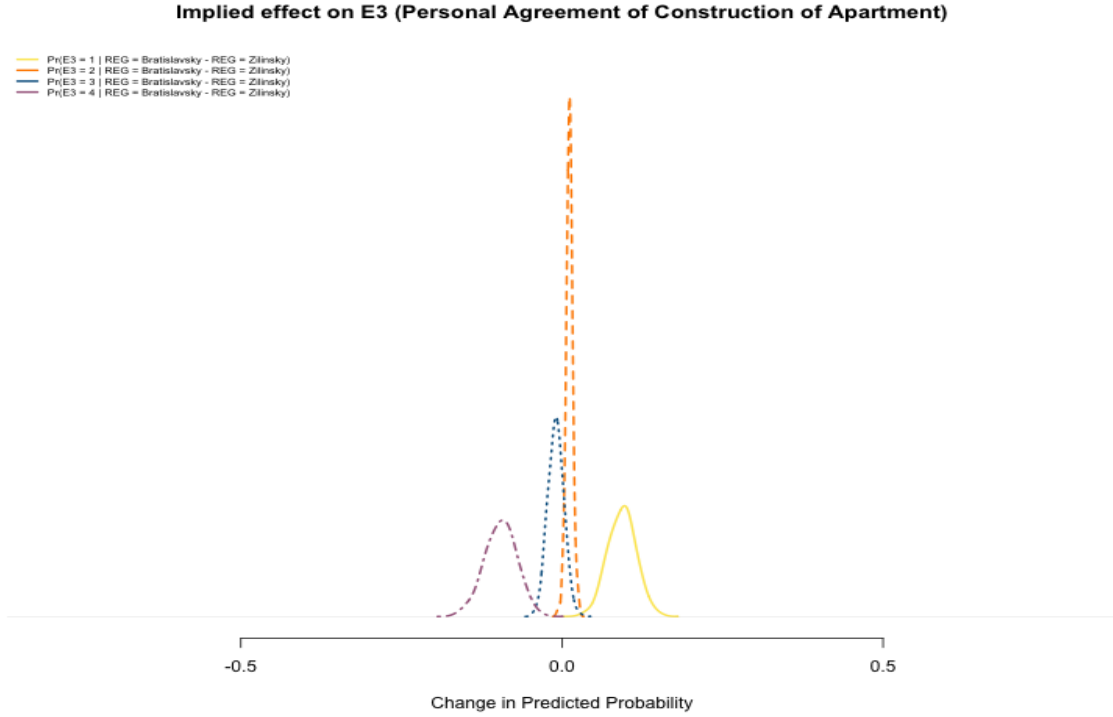


Figure 8: A density plot representing the quantity of interest (region) on personal agreement of Construction of Apartment.

The densities can be seen to be centered around zero, suggesting that the difference in predicted probabilities between Bratislavsky and Zilinsky was relatively small across all agreement levels. However, there was a bit of spread for $E3 = 1$ (strong opposition) and $E3 = 4$ (strong support). This indicated that regional differences might have had a modest influence on individuals with extreme opinions.

Specifically, the density for $E3 = 4$ (strong support) skewed slightly negative, implying that respondents from Zilinsky might have been less likely to strongly support the construction compared to those from Bratislavsky. On the other hand, $E3 = 1$ (strong opposition) had a slight positive shift, suggesting that opposition could have been a little bit higher in Zilinsky.

Overall, the magnitude of these changes were limited, as most density curves remained close to zero. This suggested that factors beyond just the region could also be playing a role in shaping individual agreement levels.

3.2.4 Preferred Model

Based on the in-sample and out-of-sample predictive performance of each model, we concluded that our alternative model performed better at fitting and predicting the data. Additionally, it included demographic predictor variables that provided contextual insight and an advantage over the original model of the authors.

4 Conclusion

This replication study of "Equality, Reciprocity, or Need? Mitigating Ethnocentric Bias in Policy Evaluation with Distributive Fairness" successfully reproduced the findings from Study 3, confirming that the reciprocity frame elicited the highest level of support for a social housing project, while the need condition generated the most disagreement. In addition to replicating the original results, this paper introduced an alternate ordered logistic regression model incorporated demographic covariates (age, education, sex, and region) to examine whether these additional factors improved model performance.

Through a comparison of in-sample and out-of-sample predictive performance using AIC, BIC, and cross-validation, we found that the alternate model provided a better fit to the data, albeit with marginal improvements in predictive accuracy. The inclusion of demographic factors captured additional contextual variation, particularly regarding age and regional differences in support for social housing. Older individuals were more likely to express strong support, while opposition was slightly higher among younger respondents. Regional differences also emerged, although their influence was relatively modest compared to other factors.

Overall, while the addition of demographic variables improved model performance, the impact of these factors on willingness to support the social housing project was not as pronounced as the influence of framing conditions. These findings suggested that while demographic and regional contexts contribute to shaping public attitudes, framing effects, particularly reciprocity-based narratives, remain the most influential in influencing support for redistributive policies. Future research could further explore the interaction between framing effects and socioeconomic conditions, as well as other attitudinal variables, to provide a more comprehensive understanding of how public support for social policies is shaped.

4.1 Discussion

The original authors' model effectively demonstrates the relationship between willingness to support the social housing project and treatment condition, providing strong evidence for how different fairness principles shape public attitudes toward redistribution. Our reanalysis builds on their work by incorporating additional demographic variables, revealing deeper patterns in the data and improving predictive accuracy. These

additions offer valuable insights into the role of individual characteristics in shaping policy support. It is important to note that our replication was limited to Study 3, excluding Studies 1 and 2, which focused on broader fairness perceptions and ethnocentric bias. Given our findings and the original authors' discussion, we believe that future research should explore how these predictive models could be applied to policy design. Specifically, using these predictions to identify communities most receptive to social housing initiatives could inform targeted interventions, ensuring that policies were framed in ways that maximized public support. Additionally, further cross-cultural research could test whether these framing effects held in different political and social contexts, as suggested in the original study's conclusions.

5 Appendix

Table 4: Regression Table for Alternate Model

	<i>Dependent variable:</i>
	E3
groupEquality	0.218 (0.163)
groupProportionality	0.586*** (0.165)
groupNeed	−0.162 (0.164)
AGE	0.014*** (0.004)
EDUSecondary (no diploma)	−0.493** (0.219)
EDUSecondary (complete)	0.222 (0.221)
EDUUniversity	0.471* (0.242)
SEXFemale	−0.075 (0.119)
REGTrnavsky	0.126 (0.254)
REGTrenciansky	0.138 (0.245)
REGNitriansky	−0.407* (0.238)
REGZilinsky	0.344 (0.242)
REGBanskobystricky	0.139 (0.245)
REGPresovsky	−0.003 (0.233)
REGKosicky	−0.193 (0.234)
Observations	1,009
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 5: Log likelihood, AIC, and BIC comparison between the Original and Alternate Model

	<i>Dependent variable:</i>	
	E3	
	(1)	(2)
groupEquality	0.220 (0.161)	0.218 (0.163)
groupProportionality	0.604*** (0.163)	0.586*** (0.165)
groupNeed	-0.159 (0.160)	-0.162 (0.164)
AGE		0.014*** (0.004)
EDUSecondary (no diploma)		-0.493** (0.219)
EDUSecondary (complete)		0.222 (0.221)
EDUUniversity		0.471* (0.242)
SEXFemale		-0.075 (0.119)
REGTrnavsky		0.126 (0.254)
REGTrenciansky		0.138 (0.245)
REGNitriansky		-0.407* (0.238)
REGZilinsky		0.344 (0.242)
REGBanskobystricky		0.139 (0.245)
REGPresovsky		-0.003 (0.233)
REGKosicky		-0.193 (0.234)
Log-Likelihood	-1341.27	-1304.2 (0.0000000000533)
Observations	1,009	1,009
Akaike Inf. Crit.	2,694.000	2,644.000
Bayesian Inf. Crit.	2,724.000	2,733.000
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

We used ChatGPT 4o LLM/AI tool in this report. We used this tool to understand the author’s original code. We also used ChatGPT to save time debugging our code, helping with latex formatting, and translating some of the original materials from Slovak to English. The tool was helpful and efficient for these tasks. ChatGPT was also used to help interpret the calibration plot and ROC curves we made for the in-sample predictive performance of the two models. It helped us understand how the models performed for in-sample predictions compared to each other.

Please click [here](#) and [here](#) to view the conversations we had with ChatGPT.

References

Findor, Andrej, Matej Hruška, Roman Hlatky, Tomáš Hrustič, and Zuzana Bošellová. 2023. “Equality, Reciprocity, or Need? Bolstering Welfare Policy Support for Marginalized Groups with Distributive Fairness.” *American Political Science Review* 117(3): 805–21. doi:10.1017/S0003055422001046.