

Vignette Experiment Code Example: Jurado (2022) Replication with Data Generated from ChatGPT

Ying Yan

March 25, 2025

1 Dataset and Code

The original survey data is taken from the supplementary material of Jurado (2022). The pseudo-survey dataset is generated by ChatGPT by describing exactly the same setup and sample size as the paper. The exact wording of this data generation prompt is shown in the Appendix. The data generation code generated by ChatGPT and the code for this replication and comparison are available at [Github](#).

2 Balance Tests

Table 1 summarize the result of balance test for sociographics in two datasets. Both datasets pass the balance test, though the original dataset obtained through survey is much balanced than the pseudo-survey data generated by ChatGPT.

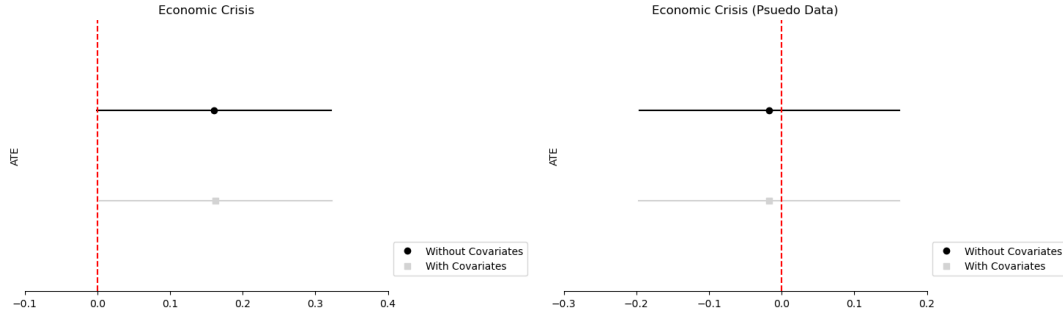
Variable	Original			ChatGPT		
	Treated	Control	p(diff=0)	Treated	Control	p(diff=0)
Female	1.50	1.51	0.79	1.51	1.50	0.92
Age	44.59	44.60	0.98	45.09	45.28	0.81
Education	2.08	2.09	0.68	2.11	2.14	0.40
Unemployed	0.09	0.08	0.79	0.10	0.09	0.77
Satisfaction with democracy	2.53	2.50	0.65	2.79	2.82	0.70
Feel European citizen	3.77	3.77	0.99	3.58	3.66	0.2

Table 1: Balance Test for Original and ChatGPT Data

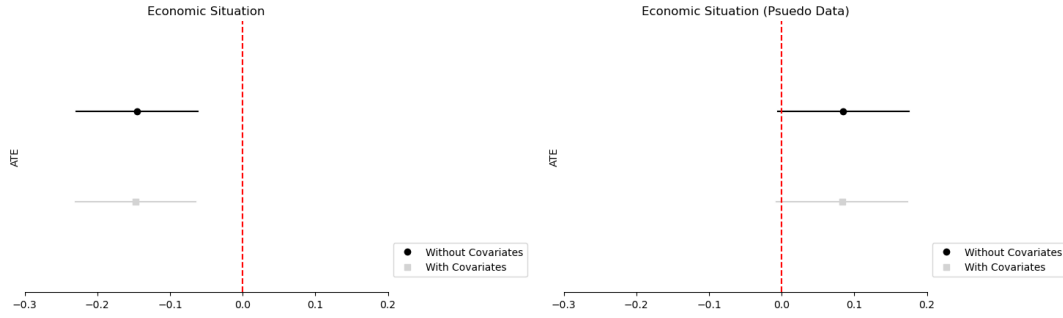
3 Manipulation Checks

This aims to check whether the treatment effect is actually driven by the perceptions on economic crisis. This is a replication of Figure A.1. in the paper appendix, in addition to the same regression result on ATE using the dataset generated by ChatGPT.

Figure 1 represents the result. Again, the original survey data shows more robust changes in the economic crisis perception of respondents who received the vignette. The ATEs estimated from both datasets are also in the opposite direction.



(a) Depth of Economic Crisis.



(b) Economic Situation.

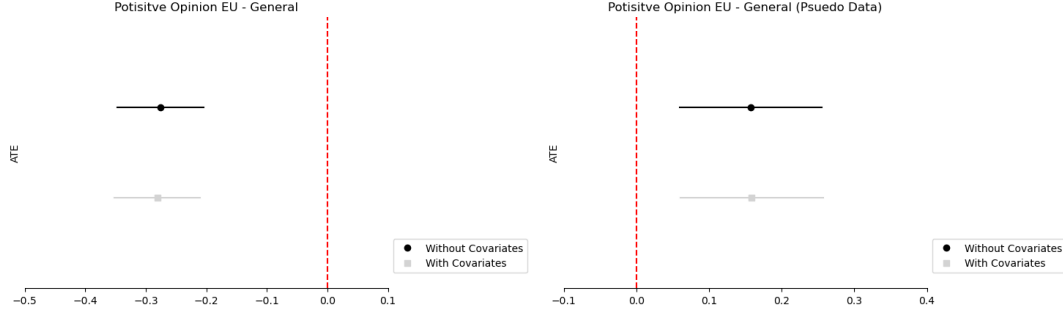
Figure 1: Manipulation Checks.

Bars denote 90% confidence intervals for each treatment effect.

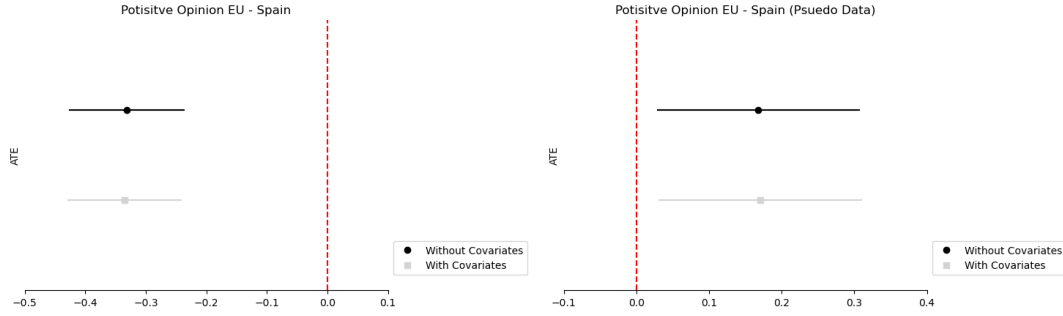
4 Regression Models

4.1 Positive Opinion EU

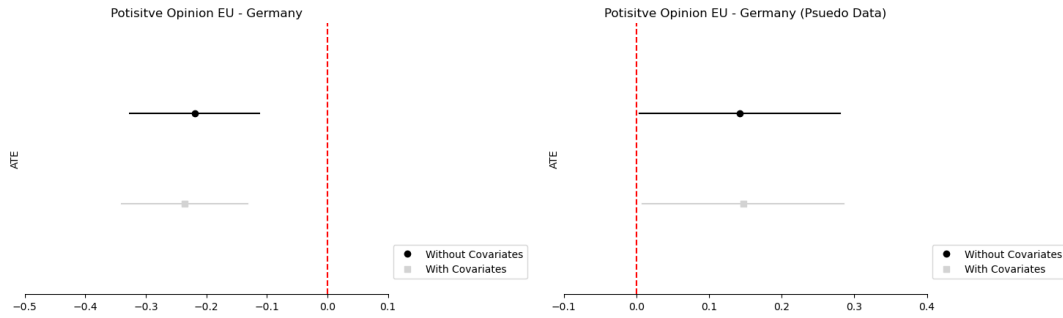
Figure 2 replicates Figure 1 in the paper, whose dependent variable is the opinions towards EU, comparing with the same regression result estimated based on ChatGPT's data. The ATEs estimated from both real and pseudo survey data are significant at 10% confidence interval, but in opposite directions.



(a) Total.



(b) in Spain.



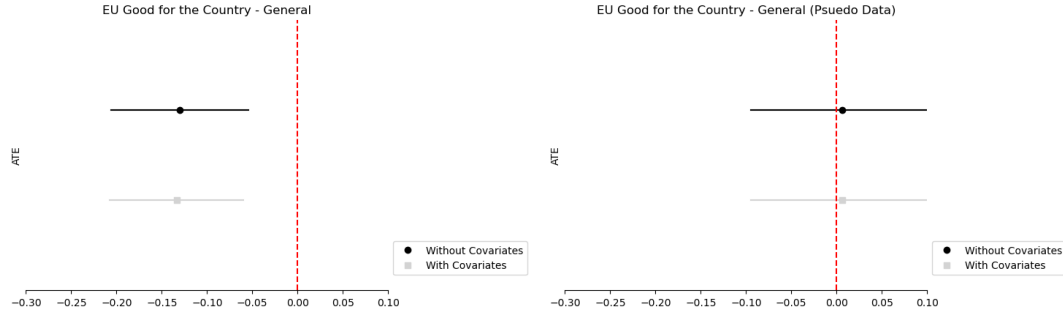
(c) in Germany.

Figure 2: ATEs on EU Opinion.

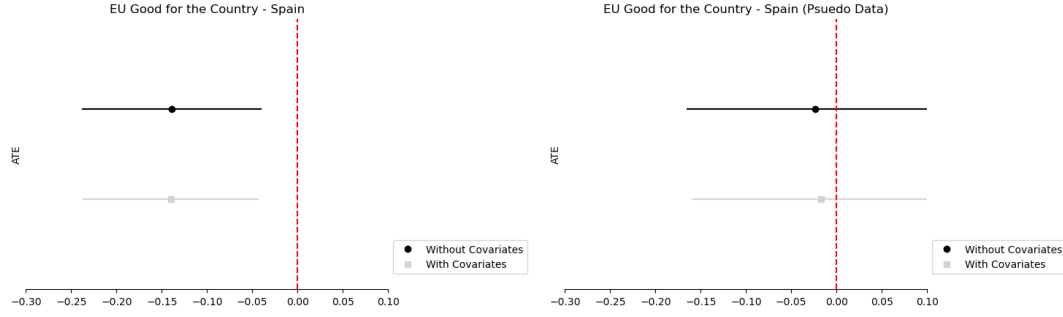
Bars denote 90% confidence intervals for each treatment effect.

4.2 EU Good for the Country

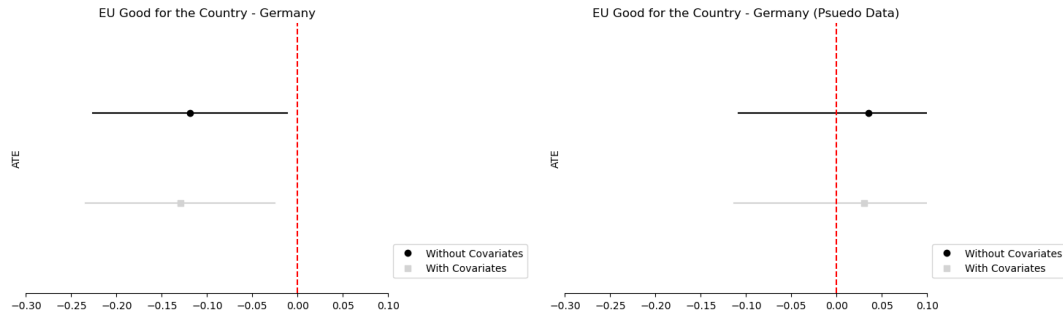
Figure 3 replicates Figure 2 in the paper, where the dependent variable is whether the EU is good for the country, compared with the same regression result estimated from ChatGPT data. For such an outcome variable, only ATEs estimated from real survey data are statistically significant.



(a) Total.



(b) in Spain.



(c) in Germany.

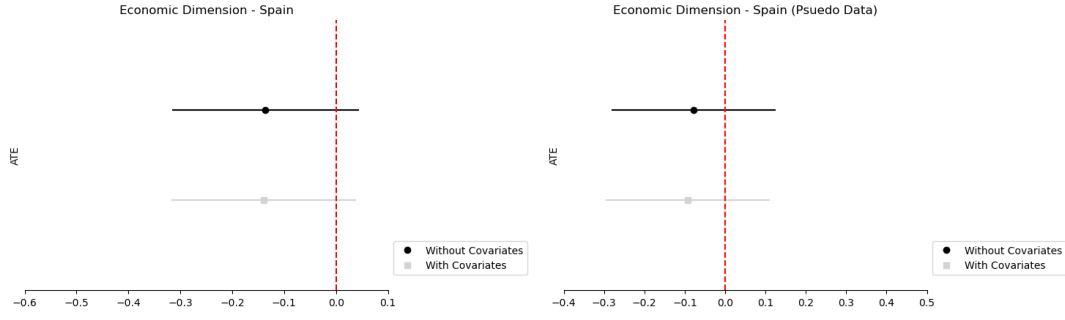
Figure 3: ATEs on EU good for the country.
Bars denote 90% confidence intervals for each treatment effect.

4.3 Perceptions over EU

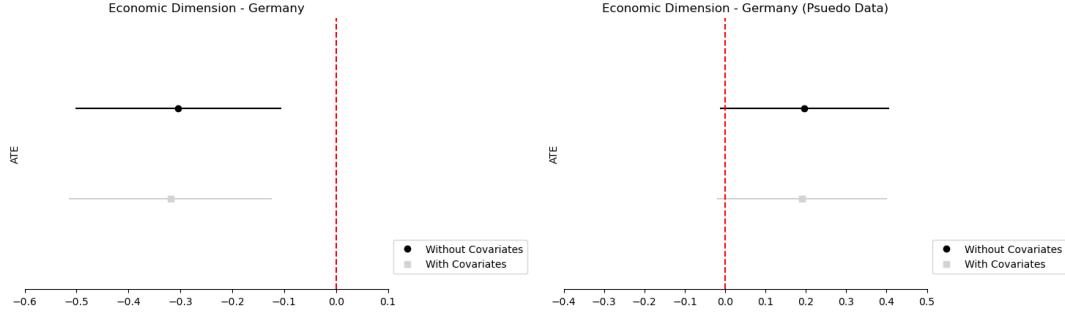
Figure 4 and 5 are replications of Figure 3 in the paper, compared with the same regression result estimated by data generated by ChatGPT.

References

Jurado, I. (2022). Economic perceptions and attitudes towards the european union: A survey experiment. *European Union Politics*, 23(4), 721–728. <https://doi.org/10.1177/14651165221107100>

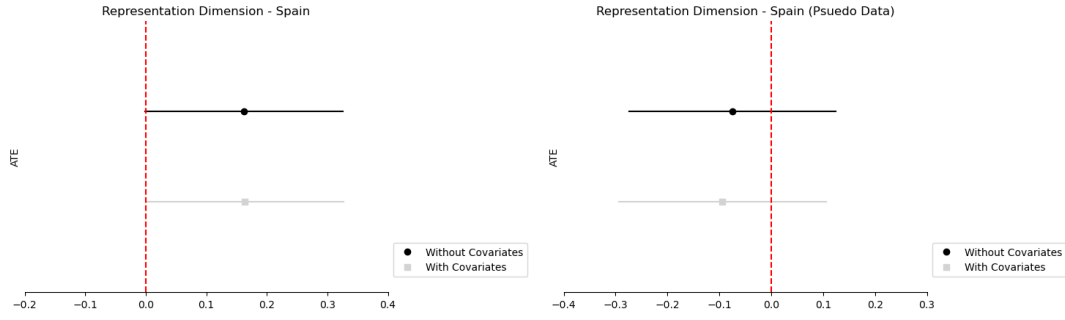


(a) Spain.

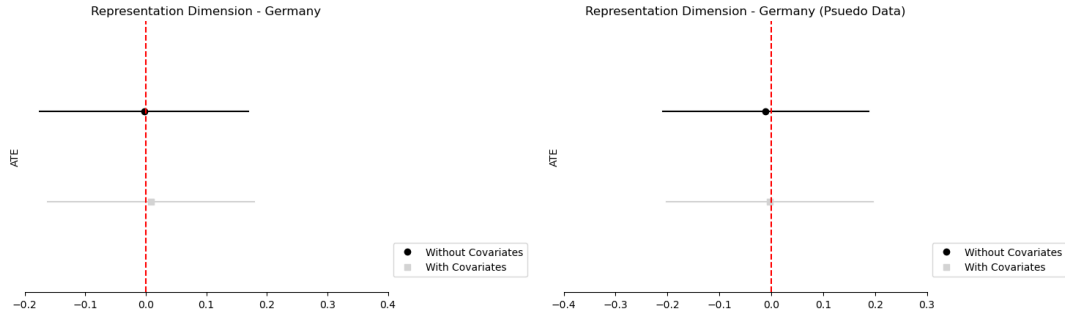


(b) Germany.

Figure 4: ATEs on Economic Dimension of the Crisis.
Bars denote 90% confidence intervals for each treatment effect.



(a) Spain.



(b) Germany.

Figure 5: ATEs on Representation Dimension of the Crisis.
Bars denote 90% confidence intervals for each treatment effect.

5 Appendix

The prompt to generate pseudo data

I would like you to generate two synthetic survey datasets for my survey experiment. Please simulate realistic response patterns and generate two CSV datasets that I can download. Please name the dataset "data_spain.csv" and "data_germany.csv". Please then briefly explain your process for generating the dataset. Also include the sources underlying the sociographic distributions. Please also provide the python code that you use to generate the data. Finally, please formulate the research question underlying this experiment.

Sample size: 1166 in simulated responses data_spain.csv and 1147 in data_germany.csv.

Distribution of the sample: Representative for Spain AND Germany in August 2018 separately according to

- Nationality (Spanish or German);
- Age (Min = 16; Max = 95);
- Gender (man = 1, woman = 2);
- Highest level of education achieved (1 = Primary education or less; 2 = Secondary education, 3 = Tertiary education);
- Employment status (1 = unemployed, 0 = otherwise);

The survey experiments are vignette experiments conducted in Spain and Germany simultaneously with the same questions and procedures. There are two scenarios. Which scenario each respondent receives is randomly distributed. The data set must contain a variable that indicates which scenario the person has seen. Please call this variable "Treatment". Before the scenario, the respondent have to answer two attitudinal questions, Question 1 and Question 2. Please call the first variable 'satdemo' and the later variable 'european_citizen'. After the person has seen the scenario, they are asked two manipulation questions, Question 3 and Question 4. Please call the question 3 variable "crisis_country" and the fourth variable "economicsituation". Afterwards, each respondent is asked Question 5 to Question 10 about their opinions regarding EU. Please call Question 5 'opinioneu'; Question 6 'countryeu'; Question 7 'eubenefic', Question 8 'euworth'; Question 9 'eurestric' and Question 10 'eunotall'. Please estimate the answers to questions 1 to 10 depending on the sociodemographic characteristics of the respondents.

Scenario 0: The respondents receive no vignette.

Scenario 1: The respondents receive are requested to read a vignette as follows:

From 2008 on, the European Union has suffered the deepest economic and debt crisis since its creation. As a consequence of the economic crisis, there has been a sharp increase in unemployment in European countries, where unemployment rose from 6.8% in 2008 to 11% in 2013 in the whole European Union. Public debt rocketed, rising from 57% of GDP in 2007 to 86% in 2014 to finance the public deficits of European countries. As of today, the levels of public debt and unemployment have not returned to pre-crisis levels.

Question 1: how satisfied are you with how democracy works in your country?

Answer option 1 is a scale from 0 ("Extremely unsatisfied") to 5 ("Extremely satisfied"), and 6 'I don't know/Prefer not to answer'

Question 2: Your degree of agreement with the statement "I feel an European citizen"

Answer option 2: 1 "Strongly disagree"; 2 "Somewhat disagree"; 3 "Neither disagree nor agree"; 4 "Somewhat agree"; and 5 "Strongly agree", and 6 'I don't know/Prefer not to answer'.

Question 3: How deep do you think the economic crisis has been?

Answer option 3 range from 0 "The economic crisis has been very shallow" to 10 "The economic crisis has been very deep" Depth of the economic crisis in the country: (0-10)

Question 4: How would you assess the economic situation in (Spain/Germany)?

Answer option 4: range from 1 ("Very good") to 5 ("Very bad").

Question 5: Overall, how is your opinion about the European Union?

Answer option 5: 1 'Very negative'; 2 'Somewhat negative'; 3 'Neither negative nor positive'; 4 'Somewhat positive'; 5 'Very positive'; 6 'I don't know/Prefer not to answer'

Question 6: And specifically for [Germany/Spain], do you think that being a Member State of the European Union has been...?

Answer option 6: 1 'Very negative for [Germany/Spain]'; 2 'Somewhat negative for [Germany/Spain]'; 3 'Neither negative nor positive for [Germany/Spain]'; 4 'Somewhat positive for [Germany/Spain]'; 5 'Very positive for [Germany/Spain]'; 6 'I don't know/Prefer not to answer'

Question 7: To which degree do you believe that 'the EU is a project that is beneficial from an economic perspective'?

Answer option 7: 1 'Completely disagree'; 2 'Somewhat disagree'; 3 'Neither disagree nor agree'; 4 'Somewhat agree'; or 5 'Completely agree'

Question 8: To which degree do you believe that 'From an economic perspective, it is worth for [Germans/Spaniards] to be part of the EU'?

Answer option 8: 1 'Completely disagree'; 2 'Somewhat disagree'; 3 'Neither disagree nor agree'; 4 'Somewhat agree'; or 5 'Completely agree'

Question 9: To which degree do you believe that 'the EU has restricted too much the set of policies that governments can implement'?

Answer option 9: 1 'Completely disagree'; 2 'Somewhat disagree'; 3 'Neither disagree nor agree'; 4 'Somewhat agree'; or 5 'Completely agree'

Question 10: To which degree do you believe that 'The EU does not allow national governments to implement the policies that citizens demand'?

Answer option 10: 1 'Completely disagree'; 2 'Somewhat disagree'; 3 'Neither disagree nor agree'; 4 'Somewhat agree'; or 5 'Completely agree'