Critiquing: *Revisiting the labor market competition hypothesis in a comparative perspective: Does retirement affect opinion about immigration?*[*]

Johannes M. Halkenhäußer
johannes@halkenhaeusser.de [1]

[1]Hertie School, Berlin (GER)

June 8, 2022

Abstract

Any causal identification breaks down when biasing channels are not accounted for. In this critique of Jeannet (2018), I investigate how missingness and violations of the conditions of instrumental variables lead to a breakdown in causal methods. I try to fix the issue of missingness in the explanatory variables using predictive mean matching imputation but do not succeed due to non-random missingness. I elaborate on why the instrument is not valid and what could be done about it.[1]
Keywords: *Causal Inference; Missingness; Immigration; Public Opinion*

1 Introduction

Jeannet (2018) tests the labor market competition hypothesis using the heterogeneity of retirement ages in different European countries. The labor market competition hypothesis argues that those with stronger labor market competition due to immigration are more likely to oppose immigration than labor market participants with less competition. The study is based on survey data of men between the ages 50-69 that participated in the European Social Survey (ESS) [2]. Using retirement as a change in labor market competition, the author claims no causal connection between being retired and having negative sentiments wrt. immigration. The causal identification is based on the author's claim that government incentivized retirement ages can be used to instrument the retirement status of their observation.

The first part of the paper uses a tiered approach with a regression discontinuity design and incorporates multiple control variables, which appear in both steps of the estimation process. The DAG in Fig. 1 represents the full model according to the study. The instruments (above early retirement age and above full retirement age) are coded as dummy variables based on a participant's country and age. Whilst the Jeannet (2018) argues that the retirement ages are exogenous to a participant's answers, this is unlikely given that the exclusion restriction may be violated if respondents view immigration as a cause for retirement age.

[*]Final Project for Causal Inference with Prof. Will Lowe, Spring 2022

[1]View the code to this assignment on my Github

[2]Austria; Belgium; Denmark; Finland; France; Germany; Ireland; Netherlands; Portugal; Spain; Sweden; the United Kingdom; Norway; and Switzerland

In a second empirical exercise, the author tests whether the lack of an effect is driven by individuals' socio-tropic motivations. This means they find out whether the participant bases their opinion on immigration on an ego-centric framework or whether they arise on a belief of what is best for society. To answer the question, Jeannet (2018) uses a random experiment set up in the questionnaire, in which roughly half the participants were asked about their views on more immigration of either *professionals* or *unskilled laborers*. The difference between the control group (asked about unskilled laborers) and the treatment group (asked about professionals) is not dependent on whether the respondent is working or retired. Thus, the author concludes that the lack of a retirement effect on attitudes is based on the socio-tropic opinions of respondents.

In both empirical parts of the paper, the author assumes that non-response and missingness of responses are random. In the first section, observations with missing values are removed using row-wise deletion. In the second section, those who do answer "don't know" are also simply dropped. In an issue as charged as immigration, a non-response may indeed not be random, potentially containing fewer responses that demonstrate strong views against immigration in both cases. I show that missingness is particularly prevalent in the household income variable, skewing it to richer households. I then test an imputation of missing values using the MICE algorithm and replicate the study's model.

The author also restricts the result to males between 50 and 69. While the different bandwidths are tested, the restriction to males alters the determined quantity, and thus the claimed causal effect would (even if it were correct) only apply to males. **Overall, the study by Jeannet (2018) is causally flawed as it suffers from selection bias due to missingness and exclusion of the unemployed as well as a weak IV.**

2 Missingness

The author uses listwise deletion for the data set and excludes the unemployed, which biases the results. Missing data can alter how we go about study design as it may reduce the number of observations or induces selection bias. We distinguish between three types of missingness:

- *Missing completely at random (MCAR)*:
  When missingness is exogenous and not related to any variables in the DAG.

- *Missing at random (MAR)*:
  When missing is related to the explanatory variables but not the outcome variable. In Jeannet (2018) this could be if, for example, Education leads to a higher response rate.

- *Missing not at random (MNAR)*:
  When missingness is related to the outcome variable itself. Here this would be a specific opinion wrt. immigration leading to people choosing not to respond because they are ashamed/fear judgment etc.

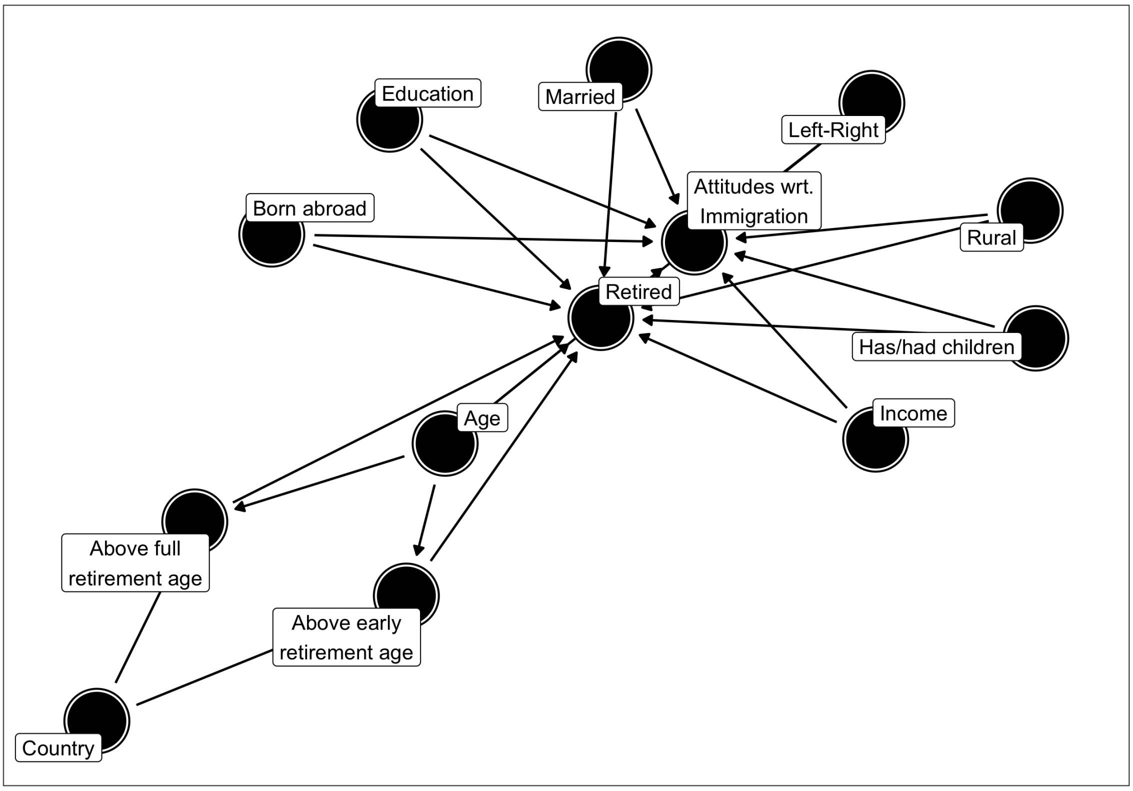

Figure 1: DAG derived from Jeannet (2018).
Note: I have excluded any edges between that were not between the independent/dependent variables and the controls. Adding a connection between, e.g., education and occupation, would have made the visualisation (even) more unclear and by virtue of controlling for them, their connection is not causally relevant any longer.

Analyzing the connection between the missingness in the variables, the variable with the most missing data is household income (Fig. 2). Next is the political opinion (LR-scale), followed by the three different outcome variables that are tested. imbgeco, imtcjob, and imbleco are the opinions on the impact of immigration on the economy, natives' job perspective, and taxes and services.
While the income variable is only an independent variable, it is tightly related to the causal hypothesis posed by the author. Further, the missingness in income does not appear to be at random. The survey data is coded between 1 and 10 and represents each country's income deciles (ESS Round 7 (2018)). Assuming that the survey has been carried out randomly, the distribution of the income variable suggests that the roughly 300 missing values are primarily from the lower tail of the distribution (Fig. 3). Thus, low-income households (those most exposed to labor market competition as these tend to be less specialized jobs) are excluded from the analysis. Even though the author controls for income in their analysis, excluding those with low income could lead to those respondents that were most affected by the shift into retirement being excluded. Such selection bias would lead to an underestimation of the effect of retirement.

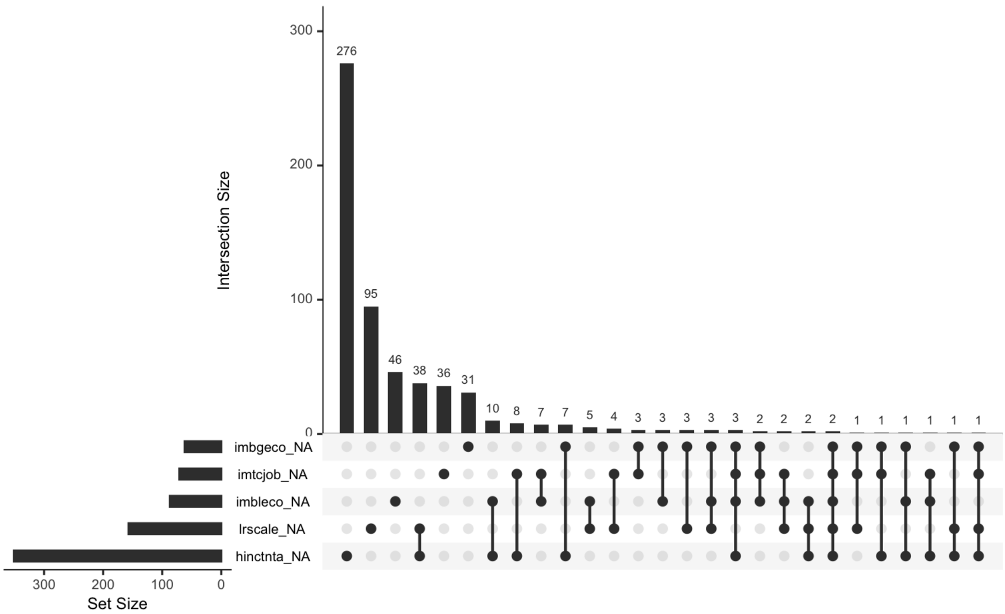

Figure 2: Frequencies of missingness for variables and in combinations of variables Tierney and Cook (2018). Only the top five missing variables and top 20 combinations of variables are shown.

2.1 Using MICE

A common way to deal with missing data is to use data imputation techniques such as the Multivariate Imputation by Chained Equations (MICE) algorithm (van Buuren and Groothuis-Oudshoorn (2011)). One of the most basic models is the predictive mean matching (pmm) approach which samples from a donor pool of similar units. The causal con-

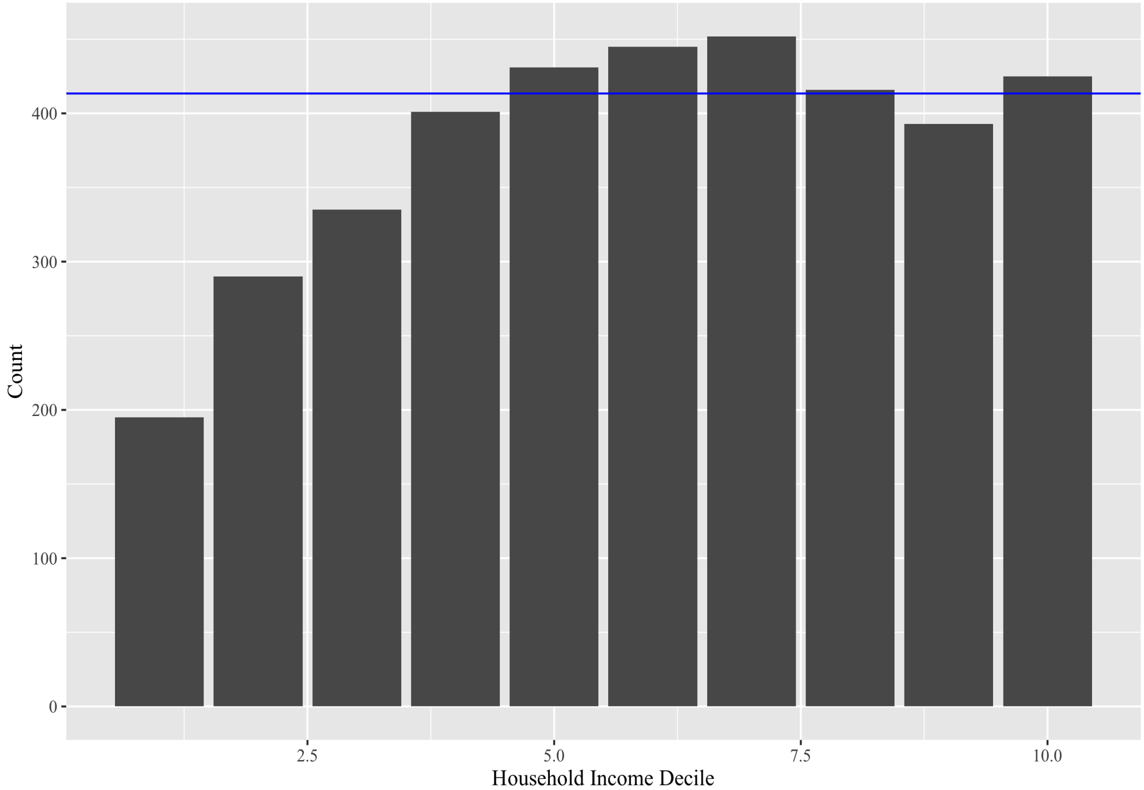

Figure 3: Income distribution of Jeannet (2018) dataset. The blue line indicates a completely even distribution of respondents into the deciles.

nection between the variables is somewhat irrelevant in this case as we only care about making accurate predictions, for which causality is helpful but not necessary.

After replicating the author's OLS and IV results [3], I use the MICE algorithm with five imputations and predictive mean matching to impute the missing data. I then run each model again, under the (strong) assumption that nothing else but missingness is wrong with the data (Tab. 1). I am able to replicate the fit statistics of the original study and have increased the sample size by roughly 500. Even though the imputed data is used, there is barely any difference in the outcome of the model. Only for the IV study that measures the attitudes towards immigration's impact on the economy is there a weak significant result at the 0.1 thresholds. There is hence no robust evidence that the imputation significantly alters the study's results.

A reason for this may be that the imputation does not actually alter the underlying distributions of the variables for which it imputes. For the problem-child-variable income, this means that the mean only shifted slightly downward, and the distribution looks almost identical. The distribution remains similar because it is not the other variables in the DAG that are affecting the missingness but the variable itself (we have MNAR). Hence, the main predictor of finding the right level of household income is missing. In fact, Pepinsky (2018) finds that using imputation can bias results more than using list-wise deletion. Adding a

[3] You may find the replication in the code

Table 1: Replicating the OLS and IV models from Jeannet (2018) with a dataset including imputations. Country and Occupation Dummies are reported in all models and can be reviewed in the R code.

	Econ		Jobs		Tax & Service	
	OLS	IV	OLS	IV	OLS	IV
Retired	−0.158	−0.595 +	−0.065	−0.336	−0.078	−0.258
	(0.108)	(0.306)	(0.099)	(0.287)	(0.103)	(0.295)
Age	0.017 *	0.041 *	0.002	0.017	0.008	0.017
	(0.009)	(0.018)	(0.008)	(0.017)	(0.008)	(0.017)
Age^2	-	-	-	-	-	
Education	0.073 ***	0.072 ***	0.060 ***	0.059 ***	0.049 ***	0.048 ***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Household Income Decile	0.080 ***	0.074 ***	0.052 **	0.049 **	0.043 **	0.041 *
	(0.017)	(0.018)	(0.017)	(0.018)	(0.016)	(0.017)
Has/had children	−0.125	−0.121	−0.041	−0.039	−0.179 *	−0.178 +
	(0.092)	(0.092)	(0.086)	(0.086)	(0.091)	(0.091)
LR-Scale	−0.141 ***	−0.140 ***	−0.042 *	−0.041 *	−0.143 ***	−0.142 ***
	(0.017)	(0.017)	(0.016)	(0.016)	(0.019)	(0.019)
Born abroad	0.764 ***	0.735 ***	0.688 ***	0.670 ***	0.869 ***	0.857 ***
	(0.129)	(0.131)	(0.122)	(0.124)	(0.126)	(0.127)
Constant	5.373 ***	5.779 ***	4.400 ***	4.652 ***	4.366 ***	4.534 ***
	(0.361)	(0.448)	(0.339)	(0.423)	(0.350)	(0.435)
Num.Obs.	4134	4134	4134	4134	4134	4134
Num.Imp.	5	5	5	5	5	5
R2	0.160	0.156	0.132	0.130	0.102	0.102
R2 Adj.	0.154	0.150	0.126	0.124	0.096	0.095

[a] p-values: + = .1, * = .05, ** = .01, *** = 0.001

variable that indicates whether income was missing did not alter the results.[4]

Another possible solution would be to try and make the distribution look uniform by randomly imputing about 2/3 of the missing values into the lowest income group and another 1/6 to the second and third lowest group, respectively. However, this would introduce a lot of bias into the data too. Thus, I decided against such a brute force approach.

2.2 Exclusion of groups

The author excludes those outside the ages of 50-69, women, and those unemployed after the age of 50 from the sample. While the restriction of age bands and including women may be statistically justifiable, the exclusion of the unemployed induces bias at the heart of the causal mechanism. The unemployed (similar to those of low income) are most likely to have been impacted by labor market competition. Hence, they are unemployed. Excluding

[4]Please refer to the code to obtain these results.

those most affected skews the sample to respondents who have not suffered from severe competition in the labor market for the last years of their professional life. It downward biases the estimands.

3 Instrumental Variables

The instrumental variable approach uses an exogenous variable to generate predictions of the treatment variable that are independent of the outcome variable Hernán and Robins (2006). This is commonly done by first constructing a model for the relationship between Z and X. Then, we predict $\hat{X}$ and use these predictions to build a second stage model for the outcome. By making $\hat{X}$ independent of all other factors than Z, we avoid unobserved confounding. Hence, if we have access to a valid instrument, we can find a causal estimand. For the instrument to be valid, three conditions need to be satisfied (Hernán and Robins (2006)):

1. Z needs to have a causal effect on X
2. Z can only affect the outcome Y through the treatment X. That means there is no direct or indirect effect through another unaccounted variable. This is known as the *exclusion restriction*.
3. There are no common causes of Y and Z. This means there are no backdoor paths that open up a confounding channel between Y and Z.

The first assumption we can test by and argue with theory. The latter assumptions are unverifiable and have to be argued using data, as we can never be sure that there is no unobserved path front- or backdoor path leading from instrument to outcome. Hence, we prefer instruments that are completely random.

In the study, the instrument is described as the government incentivized retirement age in each of the countries studied. However, in their estimation, the author uses whether an individual is over the early or the full retirement age. The indicator variables are hence merely a compound of the respondent's country and their age 1. Going through the conditions for an instrument, we can take a closer look at its validity.

3.1 Verifying the basic assumptions

3.1.1 Causal effect

It is highly plausible that the retirement age of a country has a causal effect on whether someone is actually retired at a given age. Jeannet (2018) also shows this in their paper and reports strong F-statistics.

3.1.2 Exclusion Restriction

Cunningham (2021) suggests a good instrument's effect on the outcome should feel weird when spoken aloud: "The retirement age affects an individual's attitudes towards immigration." While this certainly feels weird at first, one can imagine that retirement legislation would change one's opinion on immigration when one believes that immigrants exploit the welfare state. That will be the case if the retirement age is interpreted by a respondent to be a result of increased immigration. This could, for example, be the case if a politician justifies an increase in retirement age with an increase in public spending on immigrants, as we have seen in the US by Trump (Benkler, Faris, and Roberts (2018)). The third outcome variable question on the impact of taxes and services probes exactly for such an opinion ESS Round 7 (2018). This direct causal pathway would violate the exclusion restriction and make the instrument invalid. While the Left-Right Scale might be able to measure such attitudes, they are certainly more complex than a simple Left-Right distinction.

3.1.3 Common Causes

Age, immigration status, and the country link the retirement age/status of someone being above or below the ages to their attitudes wrt. immigration. Age obviously determines whether someone is above or below the government retirement age. It also affects respondents' opinions on immigration as older people are, on average, more conservative. A person's immigration status determines whether it affects the country they are in and thus affects whether they are above or below retirement age. Immigrants might have a more positive attitude towards other immigrants. Conversely, they could see them as competition. In either case, their attitudes wrt. immigration are affected. The country of a respondent is a bundle of sticks affecting retirement age as well as immigration attitudes through demographics, local culture (press, media, history, religious make-up), and the other control variables (such as education, income, left-right scale, rural). However, by controlling for country and the individualized control variables, the author is able to close these backdoor paths.
The beauty of the IV approach is that with a perfectly random IV, we would not need to worry about the other variables in the system. While the country variable allows us to absorb a lot of common causes, it makes the instrument far from randomly assigned. This renders the IV approach here to nothing but a sheer reliance on knowing the entire DAG to validate that no connection into the country that could confound the relationship is missed.

I can, hence, conclude that even though condition (1) seems to be fulfilled, the instrument does not necessarily satisfy condition (2) and is quite weak in terms of condition (3). This poses a serious threat to the causal identification method posed by Jeannet (2018).

3.2 Improvements

To fix the problem of the weak IV, will have to either make a sufficient explanation about how and why all backdoor paths have been closed, besides using the country effects. Further, the violation of the exclusion restriction needs to be addressed and sufficiently improved upon. Violations in the (2) and (3) are the most common critiques of any IV approach as they cannot be tested for and hence require an everlasting effort by the author to defend.

A second option would be to find a new instrument. This is notoriously hard to do as random assignment without any other factors is rarely found. However, one approach could be to study the effects of lottery winners or recipients of universal basic income on immigration attitudes. Like retirement, winning the lottery would similarly lead to an exit from the labor market (or at least encourage it) and hence activate the causal mechanism to reduce sentiment against immigration. However, participating in a lottery would cause selection bias and is likewise not an ideal instrument.

Another approach would be to change the entire empirical approach to the method and switch to matching or inverse probability of treatment weighting (IPTW). Matching would find a non-retired respondent that mimics a retired respondent by using the nearest-neighbor method or calculating the propensity score. IPTW also uses the propensity score and assigns weights to each respondent in the sample based on the propensity score. Both methods balance the observable covariates but rely on an entirely correct DAG.

4 Critique for Part II

The author similarly causes bias through missingness in the second part of the study when estimating whether the non-effect is due to sociotropic opinions. Responses are randomly assigned to answer a question about whether more or fewer professionals (Control) or low-skilled (Treatment) are supposed to be let into the country. In their study, the author simply codes all "Don't know" responses as missing values and continues with list-wise deletion. This again leads to a selection bias so common it has its own name: non-response bias. Similar to the missingness in Part I, the missingness is likely MNAR, as those with strong opinions for or against may feel like they are not responding in the socially accepted way. When trying to verify if there is an imbalance in non-response, I can not find any non-response entries in the data set and observe the same number as NA values in the treatment assignment. This means that either there are no non-responses or that the NA values for non-response were likely coded to NA before the publication of the replication data. The earlier is unlikely, given that the author explicitly mentions them in their paper. In the case with no non-response, the outcome would be causal, given that there was random assignment between the question groups. If there were non-responses that cannot be reviewed given the data, then a heckit model could be used. The heckit approach allows us to model missingness in the response variable and is able to adjust for the selection bias given a suitable model Heckman (1976).

In summary, the second part of the paper rests on the much stronger bedrock of randomization. Nevertheless, it suffers is likely to also suffer from the missingness that is MNAR and would thus be biased.

5 Conclusion

Overall, the study by Jeannet (2018) exhibits multiple flaws pertaining to selection bias and the validity of its instrument. The selection bias is induced by dropping the unemployed and using list-wise deletion. Because the missingness in income skews the distribution toward higher incomes, the effect of retirement on attitudes towards immigration is downward bias. Using MICE, the problems of MNAR cannot be fixed. I review the instrument on the three main conditions and find that it is a weak instrument, as the retirement age is unlikely to satisfy the exclusion restriction and is not randomly assigned. The combination of these flaws casts doubt over the internal validity of the study.

To improve upon the study design, an imputation method that allows for a change in the distribution of the income distribution would be needed. Further, the unemployed (those willing and able to work who are not working) should be included in the study design, especially if there is a change in benefits with the transition into retirement.

Instead of using the retirement age, a different instrument such as a random lottery or a different empirical method should be used. The earlier would suffer from selection bias in the lottery and the latter from unverifiable assumptions.

In further steps, one could request the full raw data from the author and build a heckit model for the second part of the paper.

References

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in american politics*. Oxford University Press.

Cunningham, S. (2021). Causal inference. In *Causal inference.* Yale University Press.

ESS Round 7. (2018). *European social survey: Ess-7 2014 documentation report. edition 3.2.* Bergen, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4* (pp. 475–492). NBER.

Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 360–372.

Jeannet, A.-M. (2018). Revisiting the labor market competition hypothesis in a comparative perspective: Does retirement affect opinion about immigration? *Research & Politics*, *5*(3), 2053168018784503.

Pepinsky, T. B. (2018). A note on listwise deletion versus multiple imputation. *Political Analysis*, *26*(4), 480–488.

Tierney, N. J., & Cook, D. H. (2018). *Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations.* arXiv. Retrieved from `https://arxiv.org/abs/1809.02264` doi: 10.48550/ARXIV.1809.02264

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from `https://www.jstatsoft.org/index.php/jss/article/view/v045i03` doi: 10.18637/jss.v045.i03