

Effectiveness of Indirect Interventions for Reducing Mortgage Defaults*

Reproduction of Unemployment Insurance as a Housing Market Stabilizer (Hsu, Matsa, and Melzer 2018)

Roy Chan

27 April 2022

Abstract

Preventing home foreclosures due to their associated costs is a key concern for policymakers during periods of economic instability. (Hsu, Matsa, and Melzer 2018) investigate the impact of unemployment insurance (UI) to stabilize the housing market in the United States, evaluating that UI expansions during the Great Recession aided in preventing foreclosures and stabilized home values. This paper replicates the analysis of their study and investigates alternative factors that impacted foreclosure rates. **Keywords:** housing market stability, personal finance, great recession, unemployment insurance, foreclosures, economic policy, united states

1 Introduction

During periods of severe economic and housing market instability such as the time of the Great Recession, policymakers are concerned with the potential of a significant increase in the number of mortgage defaults, wherein a mortgage agreement is broken most often caused by the mortgagor (borrower in a mortgage agreement) failing to make payments. Mortgage defaults directly lead to the process of foreclosures which create additional cost for lenders who attempt to recoup the cost of the property. However, it is not always clear what type of intervention is most appropriate and/or feasible to relieve this issue provided the likelihood of high strain on the economy during such times.

In the paper *Unemployment Insurance as a Housing Market Stabilizer* (Hsu, Matsa, and Melzer 2018), it is discussed that there exists a debate on the reasoning for mortgage defaults rendering it difficult to effectively design an intervening policy. Some mortgagors may choose to purposefully default on their mortgages, in the cases where the value of the debt significantly outweighs the price of the property. Their paper focuses on the expansions of unemployment insurance and benefits during the Great Recession, and hypothesises that it resulted in a significant reduction of the number of foreclosures during that time, as well as dampening changes in housing prices as a result of labor market instability. In this paper, we will be replicating the analysis originally presented by (Hsu, Matsa, and Melzer 2018) as well as expanding on their analysis of housing price changes.

2 Data

We would like to investigate the relationship between changes in unemployment benefit generosity and the number of home foreclosures within the United States. The original analysis performed by (Hsu, Matsa,

*Code and data are available at: <https://github.com/chan-roy/HousingMarkets>

and Melzer 2018) is done in Stata. For the purposes of aiding reproducibility, we will be reproducing their scripts using **R** (R Core Team 2020), **dplyr** (Wickham et al. 2021) and converting data filetypes with **haven** (Wickham and Miller 2021). Data is retrieved for presenting within this report using **here** (Müller 2020). Plots are created using **ggplot2** (Wickham 2016). Tables are generated using **kableExtra** (Zhu 2021).

As a reproduction, we will be utilising the same data as the original paper, as made available and obtained from openICPSR (Hsu, Matsa, and Melzer 2019). There are five separate datasets used within the paper:

- Attributes relating to unemployment benefits for each state for the years of 1991-2010, obtained from the Bureau of Labor Statistics, Bureau of Economic Analysis, Department of Labor, and Hirsch and Macpherson (2003).
- Factors of mortgage borrowers relating to potential of delinquency from the Survey of Income and Program Participation (SIPP) years 1991, 1992, 1993, 1996, 2000, 2004, 2008 and supplemented by the corresponding unemployment insurance program characteristics obtained from the Department of Labor.
- Household data of mortgage borrowers from the NLSY79 survey years 2010 and 2012 for foreclosure related variables.
- Characteristics of mortgages and housing relating to foreclosure risk by state from NLSY79 2010 and 2012.
- Home price growth, unemployment rates, and unemployment insurance data of each state from 2008 to 2013, obtained from Zillow, Bureau of Labor Statistics, and Department of Labor respectively.

The following table, Table 1 is a reproduction of Table 1 from the original paper, showing summary statistics of the variables of interest for our analysis.

Table 1: Summary Statistics

	Mean	Median	SD
State characteristics (1991-2010, N = 1020)			
Unemployment insurance			
Max Benefit (\$ thousands)	8.45	7.88	3.08
Max Weekly Benefit (\$ thousands)	0.32	0.30	0.11
Max Regular Duration (weeks)	26.14	26.00	0.75
Real Max Benefit (2011 \$ thousands)	10.71	10.18	3.24
log of Max Benefit	8.99	8.97	0.33
Max Benefit / wages (% of semi-annual wages)	46.14	45.67	11.07
UI trust fund reserves (% of semi-annual wages)	1.60	1.45	1.18
UI trust fund reserve ratio < 0 (%)	4.80	0.00	21.40
Max EB EUC (\$ thousands, 2009, N = 51)	17.72	17.39	8.37
Max EB EUC Duration (weeks, 2009, N = 51)	40.02	46.00	11.95
Economic variables			
Unemployment rate (%)	5.47	5.20	1.82
log of real GDP per capita	10.46	10.47	0.31
Average annual wages (\$ thousands)	38.31	36.85	10.74
Union coverage (%)	14.13	13.55	5.85
SIPP household characteristics and delinquency data (1991-2010, N = 52,713)			
Mortgage and housing			
Delinquent prior 12 months (%)	5.40	0.00	22.61
Evicted prior 12 months (%)	0.15	0.00	3.89
Loan-to-value (%)	0.59	0.60	0.32
Negative equity indicator (%)	0.05	0.00	0.23
Mortgage payment (\$ per month)	900.87	766.50	595.07
Mortgage payment per week / max weekly benefit (%)	0.61	0.51	0.39
Employment, income, and assets			
Max Benefit per household (\$ thousands)	6.50	6.70	4.02
Layoff within household in prior 12 months (%)	0.15	0.00	0.36
Annual earnings (\$ thousands)	52.66	42.25	55.39
Liquid financial assets (\$ thousands)	0.43	0.04	8.98
Net worth (\$ millions)	0.19	0.09	0.95
Education (highest within household)			
No high school diploma (%)	0.06	0.00	0.23
High school diploma only (%)	0.21	0.00	0.41
Some college studies (%)	0.33	0.00	0.47
College degree (%)	0.23	0.00	0.42
Some graduate studies (%)	0.18	0.00	0.39
NLSY household characteristics and foreclosure analysis (2010 and 2012, N = 7689)			
Mortgage and housing			
Foreclosure initiation (%)	4.08	0.00	19.79
Foreclosure completion (%)	1.93	0.00	13.77
Loan-to-value, main property (% , N = 6064)	0.44	0.44	0.37
Loan-to-value, other properties (% , N = 658)	0.04	0.00	0.16
Negative equity indicator, main property (% , N = 6483)	0.03	0.00	0.17
Negative equity indicator, other properties (% , N = 1193)	0.00	0.00	0.05
Employment, income, and assets			
Layoff in prior 24 months (%)	0.11	0.00	0.31
Annual earnings (\$)	81512.43	66000.00	72523.39
Net worth (\$)	300998.91	141000.00	534014.04

3 Model

As we are primarily interested in how changes in unemployment benefits affect the number of mortgage defaults, a suitable model should encompass the relationship between these two factors. Thus, we will be using and replicating the results of model 2 in the original paper, which models the relationship between mortgage delinquency and unemployment benefits. An advantage of this model as compared to the others within the paper is that there exists a significant amount of variability between each state's unemployment benefits (Hsu, Matsa, and Melzer 2018), and so we want to account for economic differences between households by the inclusion of control variables. The model uses data from the SIPP years 1991-2010, and is as follows:

$$Delinquent_{is} = \alpha + \beta MaxEBEUC_s \times Layoff_i + \delta Layoff_i + \zeta X_i + \lambda_s + \epsilon_{is}$$

Layoff is a categorical variable for indicating if the respondent household has experienced a layoff, and thus eligible for unemployment benefits. It is interacted with *MaxEBEUC*, the product of the maximum amount of weekly unemployment benefits available comprised from extended benefits (EB) and emergency unemployment compensation (EUC), and the maximum number of weeks for which the benefits can be paid, *Layoff_i* is the indicator of whether a layoff occurred within the household in prior 12 months, and $\zeta X_i, \lambda_s$ represent the control household and state variables respectively.

4 Results

Firstly, we will replicate Figure 3 of the original paper, plotting mortgage delinquency against maximum unemployment benefits, to quickly visualise and validate that there is a relationship between them.

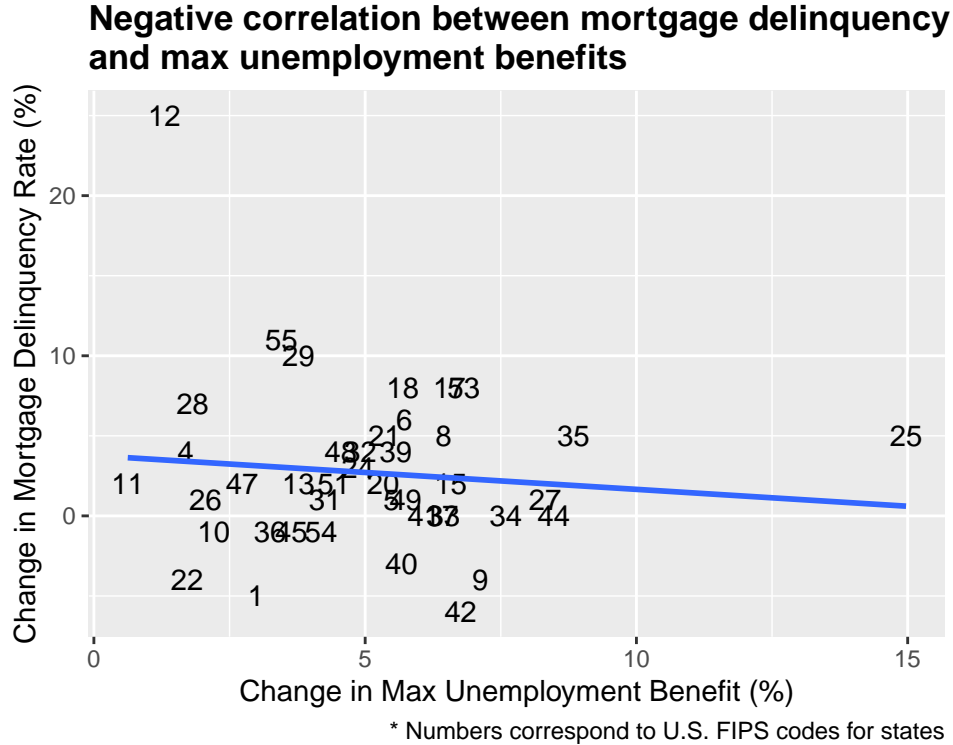


Figure 1: Correlation between change in mortgage delinquency and max benefit by state, for SIPP years 1991 and 2010

As we can discern from the plot, there exists a negative correlation between unemployment benefits and mortgage delinquency, as increases in benefits reflect a decrease in delinquency. At this point, we perform our linear regression using the aforementioned model. There are three variations of the model discussed within the original paper, the first being as originally shown, the second as a cubic regression for the variable of *MaxEBEUC*, and the third focusing on weeks of benefit duration instead of compensation directly. As an expansion to investigate additional factors not originally addressed, we create a fourth model using net worth as an indicator. We show the results of these three regression models in Table 2.

Table 2: Summary of Linear Regression Results

	Mortgage delinquency			
	1	2	3	4
Max EB EUC * Layoff	-0.246 (0.07)	-0.30 (0.07)		
Max EB EUC Duration * Layoff			-0.33 (0.08)	
Net worth * Layoff				-14.7 (0.849)
Layoff	10.2	10.1	9.95	9.65
Observations	12,606	12,606	12,606	12,606
R^2	0.03	0.03	0.03	0.02

Note:

Standard error of estimates stated in brackets.

Relative to the results obtained in the original paper, we are able to replicate similar results using the data and adapting the regression models from Stata to R. For all four variations, our estimated values are statistically significant at the $p(< 0.05)$ level. Thus, we confer the result that for every \$1000 increase in unemployment benefits, there is a corresponding -0.24% decrease in probability of mortgage delinquency. For model 4, we see a strong negative correlation between net worth and mortgage delinquency, with a -14.7% decrease in probability of delinquency per \$1 million in net worth.

5 Discussion

5.1 Limited effectiveness of unemployment insurance to prevent mortgage delinquency

Our analysis was able to determine a statistically significant decrease in mortgage delinquency, however the magnitude of the percentage probability change is small. We theorise that, as previously introduced, there exists a trend of strategic defaulting. During the period of the Great Recession, it is estimated that up to 17% of mortgages were strategically defaulted (Experian-Oliver Wyman 2011), owing to falling property prices outweighing the debt of a defaulted mortgage. This is compounded by the use of indirect intervention-individuals may not be concerned with prioritising mortgage payments in times of economic hardship.

5.2 Weaknesses and next steps

While this paper was successful in replicating aspects of the original paper, there are several weaknesses that limit its accuracy. As the original analysis was written and performed in Stata, I still have limited familiarity with the software, thus the code conversion to R may not have fully replicated original processes, with not all methods used in the original being available. With regards to the limitations of the data, the variable of house price growth was not made available, which required us to exclude it from our models. Weights were not applied for SIPP and NLSY analysis due to complications with applying different weights

per year, resulting in summary statistics figures and model estimates differing slightly for both total number of observations and calculated variables. An expansion of the analysis performed in this paper would ideally utilise differing datasets. Given that the data is localized for the United States but the Great Recession had global economic ramifications, it would be interesting to investigate whether the trends can be observed within other countries.

Appendix

A Enhancement

A.1 Datasheet for Dataset

Extract of the questions from (Geburu et al. 2021).

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
 - The dataset was created to investigate the potential effectiveness of indirect interventions in the housing market, specifically in relation to expansions in unemployment insurance benefits.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
 - Joanne W. Hsu, David A. Matsa, Brian T. Melzer on behalf of the American Economic Association.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
 - Information unavailable
4. *Any other comments?*
 - None

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
 - Employment data, survey responses from SIPP and NLSY79
2. *How many instances are there in total (of each type, if appropriate)?*
 - 5
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
 - Contains all possible instances, nationally representative for the United States, validated through survey design and data collection methodology
4. *What data does each instance consist of? “Raw” data (for example, unprocessed text or images) or features? In either case, please provide a description.*
 - Raw data, categorical variables
5. *Is there a label or target associated with each instance? If so, please provide a description.*
 - No

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
 - Yes, no recorded responses for certain individual variables in survey data instances
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
 - No
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
 - No
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
 - Redundancies of education level recorded twice
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
 - Self-contained
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
 - No, individual household price growth has been removed
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
 - No
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
 - No
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
 - No
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
 - Reveals financial data, specifically mortgage information, equity, income, employment status, net worth
16. *Any other comments?*
 - None

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*
 - Reported by subjects
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
 - Manual human curation, household visits and telephone interviews
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
 - N/A
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
 - U.S. government employees, compensation data unavailable
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
 - 1991-2010, dataset creation in 2018.
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
 - No
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
 - Obtained via third party, made accessible through ICPSR
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
 - N/A
9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
 - N/A
10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
 - N/A
11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

- No
12. *Any other comments?*
- None

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
 - Yes, log transformations and demeaning of certain variables
2. *Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data.*
 - Yes
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
 - Yes, <https://www.stata.com/>
4. *Any other comments?*
 - None

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
 - Yes, used in published paper
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
 - <https://www.aeaweb.org/articles?id=10.1257/aer.20140989>
3. *What (other) tasks could the dataset be used for?*
 - Exploring economic trends
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
 - No
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
 - No
6. *Any other comments?*
 - None

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
 - Yes, available through ICPSR
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
 - Website, DOI: 10.3886/E116160V1
3. *When will the dataset be distributed?*
 - Distributed on 2019-12-06
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
 - Modified BSD License, Creative Commons Attribution 4.0 International Public License. Terms stated in LICENSE.txt
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
 - No
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
 - No
7. *Any other comments?*
 - None

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
 - ICPSR
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
 - Can be contacted through ICPSR
3. *Is there an erratum? If so, please provide a link or other access point.*
 - N/A
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
 - Unknown
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
 - No

6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
 - Unknown
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
 - TBD
8. *Any other comments?*
 - None

References

- Experian-Oliver Wyman. 2011. “Experian-Oliver Wyman Market Intelligence Report Understanding Strategic Default in Mortgages Part i.” Experian, Oliver Wyman.
- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. “Datasheets for Datasets.” *Communications of the ACM* 64 (12): 86–92.
- Hsu, Joanne W, David A Matsa, and Brian T Melzer. 2018. “Unemployment Insurance as a Housing Market Stabilizer.” *American Economic Review* 108 (1): 49–81.
- . 2019. “Replication Data for: Unemployment Insurance as a Housing Market Stabilizer.” *American Economic Review*, December. <https://doi.org/10.3886/E116160V1>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. <https://CRAN.R-project.org/package=haven>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.