My Final College Paper

_____

A Thesis

Presented to

The Division of Mathematics and Natural Sciences

Reed College

_____

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

_____

Wenxin Du

May 2020

Approved for the Division
(Mathematics)

_____

Andrew Bray

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

Table of Contents

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.

Introduction

Chapter 1

SCA and Its Applications

It has come to researchers' attention that, the minor decisions made by researchers along the way of performing a data analysis can have a larger effect on the final research output than expected. A dataset can be analyzed in lots of different ways with the flexible decisions that a researcher can make: which statistical test to perform, which variables to include, which transformation to make on variables, etc. Certain "decisions" made, such as an action commonly called "multiple comparisons", ensures a statistically significant result to be found, even on data with absolutely no relationship. This problem is often called "p-hacking" or "researcher degrees of freedom", and it's been found to be one of the major reasons for the replication problem in some fields. A great number of studies have studied these problems closely. It is now commonly realized that simply following the traditionally "appropriate" way to conduct data analysis and look for a statistically significant p-value is no longer sufficient to produce reliable results.

Another term for the decisions made by a researcher while conducting a study is specifications. When conducting a single study, the specifications performed by a researcher are often a small subset of a much larger set of valid specifications. Thus there can be a great limitation on the conclusiveness of the results, as the results usually hinge on the selected specifications. Methods have been proposed to work around with the existence of such a problem, and one of the approaches taken in social science is to consider the robustness of models in response to alternative specifications. It considers that when alternative sets of specifications were performed by the researchers, how might the obtained results agree/differ with the original result. From there it is possible to get a sense of to which extends do the original result hinges on the choices of specifications. The one method that will be discussed in this thesis is the Specification-curve analysis (SCA), proposed by Simonsohn, Simmons, Nelson in 2015. SCA considers non-overlapping sets of reasonable specifications and the potential different conclusions that one can arrive on. The method provides a way to visualize the different results one can arrive based on the different choices of specifications and to have a general understanding of where the differences may originate from. Most importantly, it provides an assessment of a model's robustness in response to changes in specifications.

The few applications of SCA are all in the field of psychology. Several psychologists

have applied this method on controversial topics that gather global attention. However, some usage of the method seems to deviate from the original purpose of the method, and the conclusion drawn by the analysis remains questionable. One of the major application of SCA, the study of the association between adolescent well-being and digital technology use by Orben and Przybylski published on nature in 2019, is one of such applications. After a full replication of Orben's study, it's found that the main problems of this study lie in the misunderstanding of the type of specifications that SCA works with, and the inference of the SCA result. The replication and details of the problems will be discussed in Chapter 2. In the following sections, we will introduce in detail about SCA, its existing applications, and Orben's application.

1.1   Specification-Curve Analysis

Conducting a specification-curve analysis involves three steps: (1) Identifying the set of specifications, (2) Estimate all specifications and construct a descriptive specification curve, and (3) Conduct inferential analysis on a specification curve. This section discusses the details in each step, along with the important assumptions and concepts of the method.

1.1.1   Specifications

The first step of conducting a Specification-Curve Analysis is to enumerate the set of specifications to be considered. Before choosing our specifications, it's important to first understand the type of specifications an SCA will be working with. Specification-Curve Analysis focuses on a specific set of specifications: the set of specifications which are (1) consistent with the underlying theory, (2) expected to be statistically valid, (3) are not redundant with other specifications in the set. The specifications used in an SCA should be the valid and non-redundant specifications as considered by the researcher. Commonly, different researchers have disagreements over specifications, and when conducting an SCA, a researcher needs only to consider the set of valid specifications in their perspective. If there are lots of overlaps between two researchers' sets of valid specifications, the results of two SCA's should be similar. If the two sets hardly or even never overlap, the results of two SCA's would expectedly be very different. And when SCA's are applied appropriately, such a difference between analyses' results are likely not happening by chance but could be originated by something fundamentally different, maybe different underlying theory.

One important concept about the Specification-Curve Analysis is that the specifications considered in an SCA are all operationalization decisions, not theorizing decisions. Say we are conducting an SCA studying the relationship between Y and X. Some appropriate specifications to be used in an SCA can be, "Do a log transformation on variable X", "Exclude these three outliers", "Include variable K as control variable", "Add an intersection term between X and K", or "Do a logit model instead of a probit model". These are decisions researchers can make after a statistical hypothesis has been stated, or a theory to be studied is determined. These specifications all focus on

the type of operations researchers can do that does not change the main characters and background in the story, but may make small differences that can lead to a different story ending. Specifications that are based on different underlying theories are not the type of specifications that an SCA can work with. For example, say we want to study the relationship between class performance and hair color, where the hair color refers to the natural hair color that is determined by genes. Using a variable that also considers dyed hair color would not be appropriate since the action of dyeing hair and the choices of colors can reveal information regarding personalities. The relationship between class performance and this variable can be different than the story we want to tell. Thus, the variable "appearance hair color" will be an inappropriate specification to use for conducting an SCA on this research question.

1.1.2 Specification Curve

The next step will be building a specification curve. After determining the specifications, a set of combinations of the specifications can be determined, where each combination leads to a different model to be run. For example, say we determined the full list of specifications to be: 1) Use regression model A, 2) Use regression model B instead of A, 3) Use variable X as the independent variable, 4) Remove outliers from X and use the new variable X' as the independent variable. The specification models we can build are:

1. Model A with independent variable X

2. Model A with independent variable X'

3. Model B with independent variable X

4. Model B with independent variable X'

One can consider this step to be collecting all different combinations of specifications of different types.

When the dataset is large and lots of data-processing have to be performed, the list of specifications can be large, which makes the set of specification models to be huge and difficult to computationally work with. For example, say we are working on a dataset with 10 variables, and say we have: 1) 2 choices of regression model, 2) 2 ways of transforming each of the 10 variables, 3) 3 ways for each variable to deal with outliers, and 4) 10 ways of adding interaction terms, this will result in 12000 different specification models. Running all 12000 models can be computationally intense and maybe difficult to perform in real life. It is also not rare for the number of variables to be much larger than 10 in real life, and the model form can be much more complicated and computationally difficult. In this case, a random subset of the specification models can be used instead.

Now all the models have been determined, the next step is to run all the models and extract the point estimates from each of the models. In the case of linear regressions, the extracted point estimates are most likely to be the estimates of β from each

model. The estimates are then plotted as a curve, where the vertical axis refers to their numerical values, and the horizontal axis refers to the set of specifications that generated the specific model for this estimate.
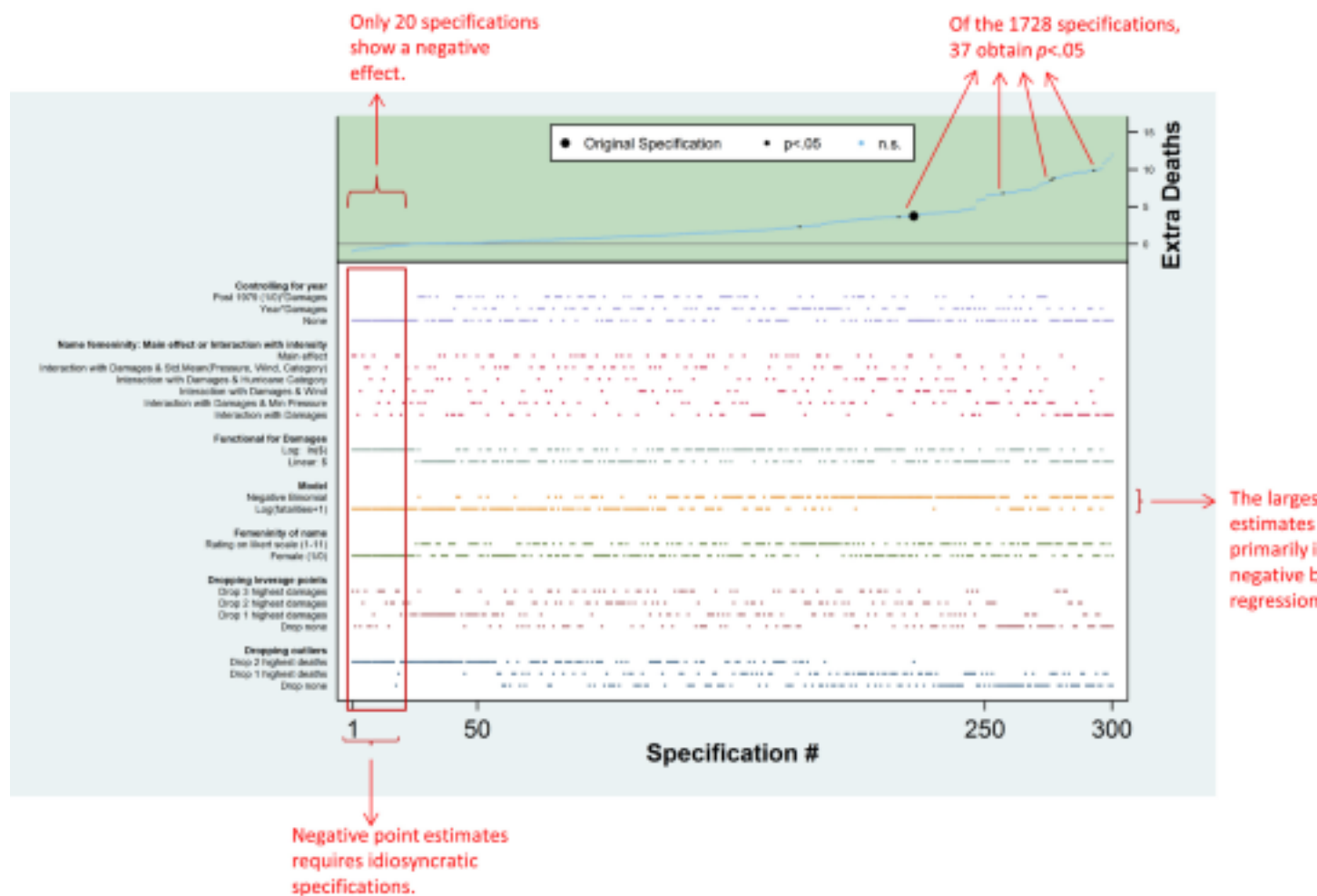

Figure 1.1: Specification Curve

As shown in Figure 1.1, a descriptive specification curve encompasses two parts: the top plot of a curve, and the bottom plot with lines and dots on it. In the top plot, the curve is the curve of the estimates from each of the models, ordered from lowest value to highest value. The vertical axis is the numerical value for the estimates, and the values on the horizontal axis represent the set of specifications used for this specific model, represented by dots in the bottom plot. In the bottom half of the plot, each dot represents the usage of a specification. The vertical axis is the name of the specifications. For example, the first dot on the curve is an estimate from a model

using the specifications: "No controlling for year", "Main effect and no interaction", "Log Damage instead of Damage (in \$)", "log-linear model instead of negative binomial model", "Use 0/1 for feminity instead of a 1-10 scaling", "Drop three hurricanes with highest damages as outliers", and "Drop two hurricanes with highest deaths as outliers".

Overall, it is possible to visualize from a specification curve plot if there exists a certain pattern relating to the choice of specifications and the corresponding estimation. For example, in the plot shown above, negative point estimates appear to require an idiosyncratic set of specifications. Also included in the plot is the indication of the models with statistically significant estimation. From the plot, it is possible to visualize if the statistical significance appears to be happening purely by chance, or if there appears to be some real relationship. For example, in this case, of the 1728 specification models, only 37 obtained statistically significant estimates. Overall, this specification curve may be suggesting that a non-statistically significant result is robust under alternative specifications.

1.1.3   Specification-Curve Analysis

The last step of an SCA is the statistical inference on the single specification curve result. The question for an inferential analysis, as stated by the authors, is "*Considering the full set of reasonable specifications jointly, how inconsistent are the results with the null hypothesis of no effect?*". Although more formal and detailed guidance of conducting an inferential analysis for a single specification curve is desired, the authors only provided suggestions on how an inferential analysis may be performed. It was suggested that using the technique of resampling, one can generate an expected distribution of the specification curves when the null hypothesis is true. The examples provided in the paper all used the permutation technique for resampling of the data, and it was suggested that a bootstrapping technique can be applied for studies without random assignment.

Once a distribution of specification curves is obtained, three test statistics are proposed to do the inferential analysis, and the authors did not decide on which ones to be more favored: 1) the median overall point estimate from the specification curve, 2) the share of estimates in specification curve that are of the dominant sign, 3) the share that is of the dominant sign and also statistically significant ($p < 0.05$). The dominant sign here refers to the sign of the majority of estimates. If the majority of the estimates in an SCA have a positive sign, then the dominant sign will be positive. Generally, we would not expect half the estimates to be positive and the rest to be negative, as the different models are not fundamentally different but rather similar at most places. This test statistic serves as a summary statistic of the entire specification curve. So instead of a "distribution of curves", the analysis works with the distribution of the test statistics from the curves. The p-value extracted, as claimed by the authors, will provide an answer to the proposed inferential question. One thing worth noting is the interpretation of the p-value. Although not specified in the main text, in the examples listed in the paper, the actual numerical value of the test statistic is not considered meaningful. For example, the authors did not use the magnitude of the

median estimate for inference on effect size. The p-values are used for indicating how robust the effects are in response to changes in specifications. A low p-value indicates that the effect is robust in response to changes in specifications. This suggests that the result is inconsistent with the null hypothesis of no effect, indicates a strong sign for the existence of a statistically significant relationship. A high p-value indicates consistency with the null hypothesis of no effect, suggesting the failure to reject the hypothesis that no relationship exists.

1.2   Applications of SCA

The paper that proposed SCA is not yet published, however, the method has already been widely applied in the field of Psychology. A few published studies have used SCA to study topics including the effect of social media on adolescent life satisfaction (Orben, Deinlin & Przybylski, 2019), relationship between adolescent mental health and technology use (Orben & Przybylski, 2019), association between digital-screen engagement & adolescent well-being (Orben & Przybylski), effect of birth-order position on personality (Rohrer, Egloff & Schmukle, 2017), etc. Among all the existing applications, Orben and her team have used the method most frequently, and some of their studies have received great attention. In this thesis, I will focus on their study on the association between adolescent well-being and digital technology use, published on Nature in 2019.

1.2.1   Orben's Application

This study conducted by Orben and Przybylski attempt to assess the association between digital technology use and adolescent well-being using 3 large-scale social datasets: Monitoring the Future (MTF), Youth Risk and Behaviour Survey (YRBS), and Millennium Cohort Study (MCS). The data were collected from studies of the same names, and encompass survey answers from adolescents and relatives on a variety of topics over a long period. For each of the three datasets, Orben identified a set of specifications and conducted SCA analysis for the research question of "the association between adolescent well-being and digital technology use". In this section, we summarize Orben's approach, main steps, and main findings. A detailed assessment and critique of the usage of SCA in this study will be provided in Chapter 2.

Identifying Specifications The first step to conduct an SCA analysis is to identify the set of reasonable specifications. In this study, there are mainly three types of specifications considered: 1. alternative variables representing adolescent mental well-being; 2. alternative variables representing digital technology use by individuals; 3. whether or not to include a set of predetermined control variables. The model is set in default to be linear regression. When the control variables are in use, the model will be multivariate linear regression, otherwise, it will be just simple linear regression. A table including all specifications determined by Orben in each dataset will be included in the Appendix. Here we provide an example: the list of alternative variables to represent "digital technology use" for MCS includes: "Whether or not

own a computer at home", "Hours of social media use on a normal weekday", "time on TV viewing on a weekday", etc. A total number of 372 specification models were determined for YRBS, 40,966 specification models were determined for MTF, and 603,979,752 specification models were determined for MCS. In the case of MCS, and a random subset of the specifications models with size 20,004 was used instead for computational purposes.

Single SCA and analysis After determining the set of specifications for each of the three datasets, three specification curves were generated. For each fitted specification model, the estimate of β on the variable representing "technology use" was collected and presented on the curve. Instead of focusing on the curves, Orben analyzed the summarized statistics from the specification curves. She focused on the sign and magnitude of the median β estimates and concluded that a small negative relationship is determined for each of the three datasets. A full table of results will be included.

Bootstrapping test and inference The last step is to conduct inference on the SCA result. Orben performed a bootstrapping test and generated 500 specification curves on bootstrapped data. The inference was performed using all three test statistics as suggested by Simonsohn et al. The p-values found were all approximately 0. As a result of the test, she concluded that evidence has been found supporting the negative relationship between digital technology use and adolescent well-being.

Others In addition to the SCA analysis on the research question, Orben performed additional SCA analyses on relationship between adolescent mental well-being and several other variables of interest, including binge-drinking, smoking marijuana, being bullied, arrested, perceived weight, eating potatoes, etc.. The mean estimates on technology use variables is compared with these results, and it's suggested that the small negative effect of technology use on adolescent mental well-being may be too small to warrant policy changes.

Overall, a small negative relationship between adolescent well-being and digital technology use was found, and the effect is suggested to be small enough such that no policy changes may be needed. However, several major issues exist in the application of SCA in this study, and the reliability of this result can be questionable. In Chapter 2, a description of the full replication of the study along with detailed assessments and critiques of this study will be provided.

Chapter 2

Replication and Evaluation

This section discusses the attempt to replicate Orben's study along with the assessment of the use of SCA in this study. We begin by introducing the three datasets used, which can all be found through public sources under permission. We then discuss in detail the attempt to replicate the study, including the obstacles to overcome during the replication process.

2.1 Data and Reprocessing

Three large-scale social datasets were used in Orben's study: Monitoring the Future (MTF) from the US [cite], Youth Risk and Behavior Survey (YRBS) from the US [cite], and Millennium Cohort Study (MCS) from the United Kingdom [cite]. The three datasets were all survey data obtained from the scientific study of the same name, and encompass survey answers from adolescents aged predominately 12-18 from 2007 to 2016. The datasets provided wide measures of adolescents' psychological well-being and digital technology use. A considerable number of psychology studies in the existing literature were conducted based on large-scale studies, which provided a wide selection of approaches to modeling and analysis based on the specific dataset. In this section, we discuss the background information of the three datasets and the reprocessing of the data obtained from public sources.

2.1.1 YRBS

The Youth Risk Behavior Surveillance was first launched in 1990, and it's a biennial survey of adolescents that reflects a nationally representative sample of students attending secondary schools. Orben's study focused on the data collected from 2007 to 2015, and the same set of data was obtained through (website name)[cite]. While Orben used data in SPSS format, we were only able to access the data through Microsoft Access. The datasets were extracted and saved under excel format. It was confirmed that the same number of observations were included in the obtained dataset as the data used by Orben, 37,402 girls and 37,412 boys from 2007 to 2015. It was also confirmed that all variables used in Orben's study are contained in the obtained

dataset. Most of the work in the preprocessing step for YRBS focused on transforming the characteristic values of the variables into corresponding numerical values.

One noticeable obstacle in this step was that, since the study is conducted annually and is still ongoing, the survey questions and indexings have been updated several times in recent years. The majority of the variables in the datasets are named after the survey questions indexes, and the recent updates in survey questions result in differences of indices for survey questions between the current survey and surveys conducted before 2015. This leads to mismatches between variable names in the incorporated dataset including data from the year of 2015 and prior–the one used by Orben–and the variable names in the dataset obtained for this study, including data from the year of 2017 and prior. Careful research and recoding are done to ensure the correct set of variables was used for the replication.

2.1.2   MTF

Monitoring the Future was first launched in the year of 1975, and it is an annual nationally representative survey of approximately 50,000 US adolescents in grades 8, 10 and 12. Surveys on adolescents in grade 12 were not used in the analysis since "many of the key items of interest cannot be correlated in their survey". Orben focused on the data collected from 2008 to 2016, which included 136,190 girls and 132,482 boys. The data are publicly accessible. In Orben's study, a merged dataset containing MTF data from 2008 to 2016 was used. While the MTF data for each year is publicly accessible, no access to a merged MTF dataset for the specified period have been found. From 2008 to 2016, the survey has been updated multiple times, along with one major change in data file format after RStudio's release in the year of 2011. Due to the frequent updates in the annual surveys and changes in data files, the variable names vary greatly among the available datasets. This brings excessive difficulties to obtain the same dataset as used in Orben's study for replication purposes.

2.1.3   MCS

The Millennium Cohort Study follows a specific cohort of children born between September 2000 and January 2001 and collects data from both the children and the caregivers. Orben's study focused specifically on the data collected in 2015 when the participated children were aged between 13 and 15. The sample included 5926 girls and 5946 boys along with 10605 caregivers. The same dataset as used by Orben was obtained. Access to the data is open to the public but requires specific permission. While Orben obtained data in CSV format, we were only able to obtain data in SPSS format. The same set of observations, with 5926 girls and 5946 boys born between September 2000 and January 2001, were included in the dataset, along with the same set of variables as used in Orben's study.

Unlike working with YRBS and MTF, the variable names in the obtained dataset matches well with the variable names in the dataset used by Orben. However, instead of using numerical indices to represent survey answers, in the dataset obtained, the variable values were all in characters. After careful reprocessing, all variable values

were transformed into the exact numerical indices matching with the values of the variables as were in Orben's study. However, two variables–one related to family incomes and one related to siblings–had only NA values in the obtained dataset. The omissions might be done for confidential purposes. The two variables were used as control variables in Orben's study. As we fail to obtain the two variables, they were removed for this attempt to replicate.

2.2 Replication

After obtaining the datasets we began the replication of Orben's study. The replication consists of two parts, the replication of generating a single specification curve for each dataset, and the replication of the inferential specification curve analysis, which assesses the significance of the single SCA result. In the following section, we discuss the procedure, obstacles, and specific resolutions to the obstacles of replicating the analysis.

2.2.1 SCA

The first part of the replication is to replicate the single SCA analysis for each dataset. All the replications in this section were done using the original code provided by Orben in the public GitHub repository. Due to the necessary reprocessings mentioned in the previous sections, slight modifications were made to the original code for smooth replication.

As mentioned in Chapter 1, three types of specifications were identified by Orben. Based on the public code, we were able to obtain the same set of specifications as used in Orben's study. A note-worthy obstacle is that, due to a large number of specification models determined for the MCS study, a random subset of 10,004 specification models was used instead. A seed is not provided by Orben for the random subset, thus we failed to obtain the same subset of specification models for this SCA analysis. We instead randomly generated our subset of 20,004 specifications. This randomness may result in a discrepancy in this specification curve. Considering that the random subset has a large size, we expect the degree of this discrepancy to be small. And this expectation is confirmed by replication result: while Orben obtained the median coefficient of the independent variable to be $\text{Median}(\beta) = -0.032$, our replication obtained $\text{Median}(\beta) = -0.0328$.

The problem does not exist for the studies YRBS and MTF. There were fewer variables available in the dataset relating to technology use and adolescent mental well-being. The number of specifications identified in the two studies is in a reasonable size, therefore the exact set of specifications was used for the replications. The result matched well with Orben's result. The median coefficient of the independent variable in the YRBS study was found to be $Median(\beta) = -0.035$ in Orben's study. The result obtained in this replication, when rounded to the same digits, is also -0.035.

2.2.2   Bootstrapping test

The next part of the replication is to replicate the inference of the single specification curves for each dataset. Orben chose to use a bootstrapping test on the median overall point estimate for the significance of the result. We will later assess the choice of the inference test and the correctness of the inference. For now, we focus only on replicating the test and the result.

500 specification curves were conducted in Orben's study on bootstrapped samples for each of the three datasets. It was found that the test statistic for the single specification curves were statistically significant in all three cases. The code to conduct the bootstrapping test and generate specification curves are all publically available on Orben's GitHub repository [cite]. The initial attempt of the replication was done using the original code. However, due to the large sizes of the three datasets and the great number of loops used in the R code, the replication process was extremely computationally intense. A single specification curve will take around 8 hours to be generated, and performing 500 specification curves will take nearly 24 weeks. An ARC computer cluster at Oxford was used by Orben to reduce running time, however, no access to such an advanced computer is available for this replication. Therefore, instead of using purely the original code, the code for this replication was rewritten for parallel running. The running time has been significantly reduced. The dataset YRBS has the least number of observations and specifications, and after the recoding, it now takes about 9 hours to generate a complete bootstrapping distribution of 500 specification curves on a Rstudio server with 8 cores. With access to an AWS server with 96 cores, the running time can be further reduced. More time will be needed for the other two datasets, as the number of observations and specifications can be much higher in those two cases, but still within a computationally reasonable time range.

[Results will be added once the full implementation of the bootstrapping test code is finished]

2.3   Evaluating Orben's work

A full replication allows a full understanding of Orben's approach and procedure. It is only when we have a full understanding of the work that our critiques and assessments on it will be responsible and reliable. In this section, we talk in detail about our critiques on the usage of the SCA method in this study, including some fundamental misunderstanding of the intentions and applicabilities of the SCA method, inappropriate choice of specifications, and misinterpretation of the SCA results.

2.3.1   "one-to-many" mapping from scientific to statistical hypotheses

We will start by assessing the research question of this study. The article is titled "The association between adolescent well-being and digital technology use". As addressed in the paper, the main focus of this paper is to study the association between

digital technology use and adolescent well-being. This is a broad topic to be studied. Intuitively, one could consider the different types of digital technologies to mean very different things for the adolescent as they may have very distinct functions. It would not be intuitively surprising to say that the use of social media by adolescents in the modern era is identical to the use of TV. The interactive nature of certain types of digital technologies makes a distinction between them and devices that only transmit and output information. Intuitively, it does not sound right to consider that the numerous distinct digital technologies in the modern era would have similar functions and effects on adolescent well-being for them to be considered as an integrated category. However, throughout Orben's study, a different type of digital technologies including "TV use", "Social Media use", "Time spent on electronic games" are considered alternative variables to use representing "digital technology use". With a glance, it seems like multiple related but different research questions have been collapsed into one. As discussed by Gelman and Loken (2013), the *(*scientific) hypotheses described here correspond to multiple *statistical hypotheses*. Whatever results conducted by this study, due to the broad research question, they would be fit into theories easily. This will result in the multiple comparison problem in the study, even if the scientists were not intended to do so.

It has been studied in the field of Psychology that categorizing certain types of digital technology use into a broader overarching category is inappropriate. Studies suggest, for example, that categorizing the different types of internet activities (such as interactive usage of social media and passive consumption of social media) into an overarching category is suboptimal. (Bessiere, Kiesler, Kraut, & Boneva, 2008; Burke, Kraut, & Marlow, 2011; Verduyn, Ybarra, Resibois, Jonides & Kross, 2017) It's also been found with an overall reviewing on the consequences of interacting with social network sites for subjective well-being that, "passively" using social networks result in a negative relationship with subject well-being, while "actively" using social networks has a positive relationship with subject well-being. If the interactive usage of social network has different relationship with subject well-being than passive use of social network, would it be appropriate to consider that the general usage of social media, which allows both interactive activities and passive browsing of information, would have identical relationship with subject well-being as the usage of television which provides mainly browsing of information, or electronic games which would have distinct functions depending on type of games?

2.3.2 Choice of Specifications

Another word to describe the problem mentioned above is that the specifications on the independent variables determined by Orben indeed are specifications in light of different underlying theories. While one specification suggests using the variable on TV used to represent general digital technology use, a different specification suggests using the variable on electronic games use to represent general digital technology use. The stories told by these different models generated by the different specifications can be very different. When performing SCA on such specifications, it's not only the impact of arbitrary operationalizations of the models that are moderated but also

the impact of non-arbitrary theorizing that's moderated. This conflicts with the true intention and appropriate usage of SCA.

It's also worth mentioning that the specifications determined by Orben in this study are all specifications relating to the inclusion/exclusion of variables. The determined specifications can be categorized into three types: 1. specifications on the choice of the dependent variable, 2. specifications on the choice of the independent variable, 3. whether or not to include a pre-determined list of control variables. However, as mentioned by Simonsohn et. al when proposing the method, the SCA should consider all operationalization decisions but not just those of variable selections. Important operationalization decisions, such as the recoding of the variables as performed by Orben in the data processing step before actual analysis, are decisions that can have an important effect on the result and are not being considered in this study.

[Tree plot in processing]

Using the idea of the garden of forking paths, it's possible to visualize this problem made by Orben in this study. We use the tree-like figure to represent the decisions scientists can make along the path to conduct scientific research. The scientists would first propose the research question of interest, determine hypothesis, choose data to work on/perform an experiment for the research question, collect data and organize data, choose the model and make operational decisions around the model. Each branch on the tree represents an alternative decision that could have been made along the way. The true SCA analysis would consider the set of specifications on the operationalization decisions for a given theory. On this plot, it will be the set of specifications at the bottom level of the trees with a single parent branch to be considered. However, the set of specifications considered by Orben includes subsets of specifications at the bottom level of the tree from several different parent branch. The story told by this study conducted by Orben is indeed not the type of story that should have been told from an SCA analysis. Analyzing the result using the approach of an SCA analysis is inappropriate in this case.

2.3.3   SCA interpretation

The last major problem of this study is the way Orben interprets the single SCA result. As discussed in the previous chapter, the single SCA generated is used for a descriptive curve that can provide information on whether or not the relationship appears to be happening by chance, and if a certain pattern of a true relationship is observed, if the relationship appears to be robust in response to changes in specifications. The single specification curve should not be used for any interpretation of the actual magnitude of the numerical values of the estimates. However in Orben's study, when interpreting the single generated specification curve, the median values of the β estimates were used and the magnitudes of the numerical value were considered. Here is a quote from the study:

> The SCAs showed that there is a small negative association between technology use and well-being, ...

The SCAs here refers to the single specification curve generated for each of the

three datasets, MTF, YRBS, and MCS. And the "small negative association" was concluded from the median estimate of the β's from models with changing specifications. Nowhere suggested by Simonsohn et. al. describes this interpretation of the numerical result from a specification curve. The three examples provided in the original paper describing the method do not make such conclusions from a single specification curve but only used it to assess if the relationship seems strong and which specifications appear to have the largest effect on the estimate.

Chapter 3

Tables, Graphics, References, and Labels

3.1 Tables

In addition to the tables that can be automatically generated from a data frame in **R** that you saw in [R Markdown Basics] using the `kable` function, you can also create tables using *pandoc*. (More information is available at `http://pandoc.org/README.html#tables`.) This might be useful if you don't have values specifically stored in **R**, but you'd like to display them in table form. Below is an example. Pay careful attention to the alignment in the table and hyphens to create the rows and columns.

Table 3.1: Correlation of Inheritance Factors for Parents and Child

Factors	Correlation between Parents & Child	Inherited
Education	-0.49	Yes
Socio-Economic Status	0.28	Slight
Income	0.08	No
Family Size	0.18	Slight
Occupational Prestige	0.21	Slight

We can also create a link to the table by doing the following: Table 3.1. If you go back to [Loading and exploring data] and look at the `kable` table, we can create a reference to this max delays table too: Table **??**. The addition of the (`\#tab:inher`) option to the end of the table caption allows us to then make a reference to Table `\@ref(tab:label)`. Note that this reference could appear anywhere throughout the document after the table has appeared.

3.2   Figures

If your thesis has a lot of figures, *R Markdown* might behave better for you than that other word processor. One perk is that it will automatically number the figures accordingly in each chapter. You'll also be able to create a label for each figure, add a caption, and then reference the figure in a way similar to what we saw with tables earlier. If you label your figures, you can move the figures around and *R Markdown* will automatically adjust the numbering for you. No need for you to remember! So that you don't have to get too far into LaTeX to do this, a couple **R** functions have been created for you to assist. You'll see their use below.

In the **R** chunk below, we will load in a picture stored as `reed.jpg` in our main directory. We then give it the caption of "Reed logo", the label of "reedlogo", and specify that this is a figure. Make note of the different **R** chunk options that are given in the R Markdown file (not shown in the knitted document).

```
include_graphics(path = "figure/reed.jpg")
```


Figure 3.1: Reed logo

Here is a reference to the Reed logo: Figure 3.1. Note the use of the `fig:` code here. By naming the **R** chunk that contains the figure, we can then reference that figure later as done in the first sentence here. We can also specify the caption for the figure via the R chunk option `fig.cap`.

Below we will investigate how to save the output of an **R** plot and label it in a way similar to that done above. Recall the `flights` dataset from Chapter **??**. (Note that we've shown a different way to reference a section or chapter here.) We will next explore a bar graph with the mean flight departure delays by airline from Portland for 2014. Note also the use of the `scale` parameter which is discussed on the next page.

```
flights %>% group_by(carrier) %>%
  summarize(mean_dep_delay = mean(dep_delay)) %>%
  ggplot(aes(x = carrier, y = mean_dep_delay)) +
  geom_bar(position = "identity", stat = "identity", fill = "red")
```
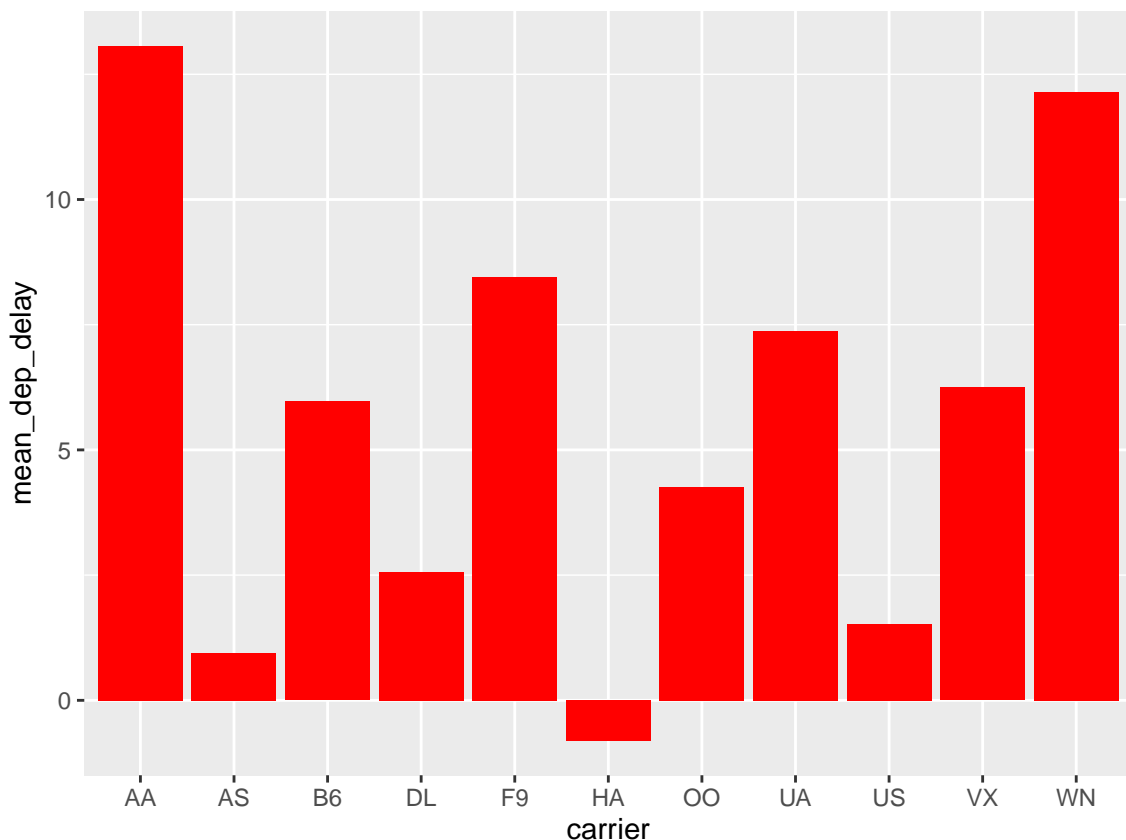

Figure 3.2: Mean Delays by Airline

Here is a reference to this image: Figure 3.2.

A table linking these carrier codes to airline names is available at `https://github.com/ismayc/pnwflights14/blob/master/data/airlines.csv`.

Next, we will explore the use of the `out.extra` chunk option, which can be used to shrink or expand an image loaded from a file by specifying `"scale= "`. Here we use the mathematical graph stored in the "subdivision.pdf" file.

Figure 3.3: Subdiv. graph

Here is a reference to this image: Figure 3.3. Note that `echo=FALSE` is specified so that the **R** code is hidden in the document.

More Figure Stuff

Lastly, we will explore how to rotate and enlarge figures using the `out.extra` chunk option. (Currently this only works in the PDF version of the book.)
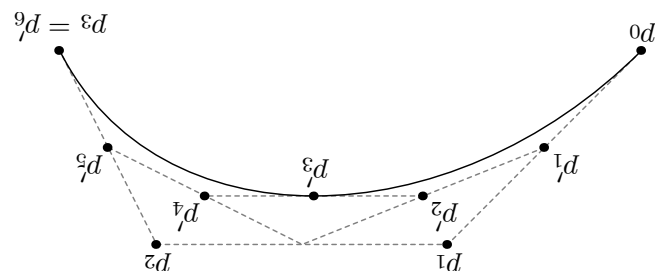

Figure 3.4: A Larger Figure, Flipped Upside Down

As another example, here is a reference: Figure 3.4.

3.3   Footnotes and Endnotes

You might want to footnote something.[1] The footnote will be in a smaller font and placed appropriately. Endnotes work in much the same way. More information can be found about both on the CUS site or feel free to reach out to `data@reed.edu`.

3.4   Bibliographies

Of course you will need to cite things, and you will probably accumulate an armful of sources. There are a variety of tools available for creating a bibliography database (stored with the .bib extension). In addition to BibTeX suggested below, you may want to consider using the free and easy-to-use tool called Zotero. The Reed librarians have created Zotero documentation at `http://libguides.reed.edu/`

---

[1]footnote text

`citation/zotero`. In addition, a tutorial is available from Middlebury College at `http://sites.middlebury.edu/zoteromiddlebury/`.

R Markdown uses *pandoc* (`http://pandoc.org/`) to build its bibliographies. One nice caveat of this is that you won't have to do a second compile to load in references as standard LaTeX requires. To cite references in your thesis (after creating your bibliography database), place the reference name inside square brackets and precede it by the "at" symbol. For example, here's a reference to a book about worrying: (Molina & Borkovec, 1994). This `Molina1994` entry appears in a file called `thesis.bib` in the `bib` folder. This bibliography database file was created by a program called BibTeX. You can call this file something else if you like (look at the YAML header in the main .Rmd file) and, by default, is to placed in the `bib` folder.

For more information about BibTeX and bibliographies, see our CUS site (`http://web.reed.edu/cis/help/latex/index.html`)[2]. There are three pages on this topic: *bibtex* (which talks about using BibTeX, at `http://web.reed.edu/cis/help/latex/bibtex.html`), *bibtexstyles* (about how to find and use the bibliography style that best suits your needs, at `http://web.reed.edu/cis/help/latex/bibtexstyles.html`) and *bibman* (which covers how to make and maintain a bibliography by hand, without BibTeX, at `http://web.reed.edu/cis/help/latex/bibman.html`). The last page will not be useful unless you have only a few sources.

If you look at the YAML header at the top of the main .Rmd file you can see that we can specify the style of the bibliography by referencing the appropriate csl file. You can download a variety of different style files at `https://www.zotero.org/styles`. Make sure to download the file into the csl folder.

Tips for Bibliographies

- Like with thesis formatting, the sooner you start compiling your bibliography for something as large as thesis, the better. Typing in source after source is mind-numbing enough; do you really want to do it for hours on end in late April? Think of it as procrastination.
- The cite key (a citation's label) needs to be unique from the other entries.
- When you have more than one author or editor, you need to separate each author's name by the word "and" e.g. `Author = {Noble, Sam and Youngberg, Jessica},`.
- Bibliographies made using BibTeX (whether manually or using a manager) accept LaTeX markup, so you can italicize and add symbols as necessary.
- To force capitalization in an article title or where all lowercase is generally used, bracket the capital letter in curly braces.
- You can add a Reed Thesis citation[3] option. The best way to do this is to use the phdthesis type of citation, and use the optional "type" field to enter "Reed thesis" or "Undergraduate thesis."

---

[2]Reed College (2007)
[3]Noble (2002)

3.5   Anything else?

If you'd like to see examples of other things in this template, please contact the Data @ Reed team (email `data@reed.edu`) with your suggestions. We love to see people using *R Markdown* for their theses, and are happy to help.

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

```r
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis.
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(thesisdown))
  devtools::install_github("ismayc/thesisdown")
library(thesisdown)
```

In Chapter 3:

```r
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis and also two functions
# used for labeling and referencing
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(dplyr))
    install.packages("dplyr", repos = "http://cran.rstudio.com")
if(!require(ggplot2))
    install.packages("ggplot2", repos = "http://cran.rstudio.com")
if(!require(ggplot2))
    install.packages("bookdown", repos = "http://cran.rstudio.com")
if(!require(thesisdown)){
  library(devtools)
  devtools::install_github("ismayc/thesisdown")
  }
```

```r
library(thesisdown)
flights <- read.csv("data/flights.csv")
```

Appendix B

The Second Appendix, for Fun

References

Angel, E. (2000). *Interactive computer graphics : A top-down approach with opengl.* Boston, MA: Addison Wesley Longman.

Angel, E. (2001a). *Batch-file computer graphics : A bottom-up approach with quicktime.* Boston, MA: Wesley Addison Longman.

Angel, E. (2001b). *Test second book by angel.* Boston, MA: Wesley Addison Longman.

Molina, S. T., & Borkovec, T. D. (1994). The Penn State worry questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 265–283). New York: Wiley.

Noble, S. G. (2002). *Turning images into simple line-art* (Undergraduate thesis). Reed College.

Reed College. (2007, March). LaTeX your document. Retrieved from `http://web.reed.edu/cis/help/LaTeX/index.html`