DISSERTATION

THREE ESSAYS ON WELFARE, WELL-BEING, AND LABOR

Submitted by

Sayorn Chin

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Sammy Zahran
Co-Advisor: David Mushinski

Ray Miller
Jude Bayham

ABSTRACT

THREE ESSAYS ON WELFARE, WELL-BEING, AND LABOR

This dissertation explores several topics in welfare, well-being, and labor economics, with a focus on: (1) health, wealth, and racial and ethnic welfare inequality; (2) the natural environment and well-being; and (3) whether labor markets place a wage premia for jobs that require workers to consume disamenities. To achieve these goals, the study utilizes three distinct datasets and applies a range of machine learning and econometric techniques, including natural language processing algorithms, as well as dynamic panel data estimators, natural experiments, and microsimulations.

In Chapter 1, titled *"Beyond Income: Health, Wealth, and Racial/Ethnic Welfare Gaps Among Older Americans"*, we estimate racial and ethnic disparities in well-being among the older U.S. population using an expected utility framework that incorporates differences in consumption, leisure, health, mortality, and wealth. We use longitudinal data from the Health and Retirement Study (HRS) supplemented with data from the Consumption and Activities Mail Survey (CAMS). Together, these provide a long and rich panel (1992-2016) for our analysis. Our measure broadly indicates that racial and ethnic inequality is larger than suggested by other welfare metrics such as income or consumption. We also find health, mortality, and wealth gaps are important in explaining the level of racial and ethnic welfare inequality among the older Americans in our sample, with leisure playing a comparatively minor role. Our decomposition exercises show that a majority of the estimated welfare gaps are determined by age sixty initial conditions as opposed to racial and ethnic differences in dynamic processes after age sixty. Our morbidity counterfactuals further suggest that eliminating common heath risk factors such as hypertension or diabetes in late-life only marginally closes overall welfare gaps. These

simulations suggest that policies aimed at closing racial and ethnic gaps in late-life may be more successful and efficient if targeted earlier in the life-cycle. In other words, outside of direct wealth transfers, it may largely be too late to target such interventions directly at older populations.

In Chapter 2, titled *"The Morning Advantage: Differential Returns to Sunlight Exposure on Well-Being"*, we estimate the effect of sunlight exposure on well-being by mimicking a natural experiment that utilizes the transition to daylight savings time as an external shock to the reallocation of sunlight between the morning and evening induced by differences in sunrise and sunset times across space, and time. We combine a collection of geolocated and timestamped tweets from Twitter with Natural Language Processing algorithms to create a comprehensive panel dataset of well-being (2014-2022) for the United States. Our findings show that the returns to sunlight on sentiment are stronger in the morning than in the evening. These results contribute significantly to the ongoing debate about whether to continue or abandon the practice of daylight savings. Specifically, the positive turn of sentiment in the morning highlights the underappreciated benefits to human well-being. Therefore, the potential shifting to darker mornings and brighter evenings following the proposed Sunshine Protection Act may do more harm than good.

In Chapter 3, titled *"The Compensation of Conscience: Evidence from the U.S. Labor Market"*, we investigate compensating differentials in the U.S. labor market related to the degree of moral compromise required in different occupations. Specifically, we explore whether jobs that require workers to compromise their moral values offer higher compensation to compensate for the disamenities that contradict their moral beliefs. To conduct our analysis, we utilize data from the National Longitudinal Survey of Youth 1997 (NLSY97) and supplement it with data from the Occupational Information Network (O*NET) job descriptor, which allows us to develop a continuous measure of moral index across occupations. This data provides a rich and extensive panel spanning from 1997 to 2017 for our analysis. Our findings, obtained through the use of two-ways fixed-effects and first-difference models, indicate that jobs that require workers to compromise their moral principles are associated with higher compensation. This suggests that there is indeed a compensating differential for engaging in disamenities that conflict with a worker's moral values. Additionally, we observed that workers with a college education receive higher pay in jobs that require moral compromise, indicating that individuals with a college degree may have more employment opportunities and greater bargaining power, influencing their compensation preferences. Furthermore, we discovered evidence supporting an asymmetric relationship between changes in the occupational moral index and total hourly compensation. This relationship appears to be responsive to the intensity of moral compromise in the job.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my doctoral committee, which consists of Drs. Sammy Zahran, David Mushinski, Ray Miller, and Jude Bayham, for their invaluable guidance and expertise that has equipped me with the knowledge and skills required to excel as an economics producer.

In particular, I extend special thanks to my advisor, Dr. Sammy Zahran. From the moment I met you in your Economics 698 (Economic Science Replication Workshop) class, I knew I wanted to work with you. Your conduct, speech, and creative thinking are truly remarkable, but what stands out the most is your unwavering loyalty to those you invest your precious time and effort in. Your unwavering support, patience, and guidance have been instrumental in my success, and I am deeply grateful for every moment we spent together. Our collaboration, from the late nights and weekends spent working together to the brainstorming sessions, problem-solving, and data analysis, has been a true highlight of my journey. These moments represent the depth of our collaboration, determination, and mutual respect, and they are memories that I will always cherish. Your commitment to excellence has been a constant inspiration to me, and your guidance has extended far beyond the technical aspects of my research. Your generosity, kindness, and warmth have created a safe and nurturing environment that allowed me to grow not only as a researcher but also as a person. Your endless encouragement and support have given me the confidence to push my limits and believe in my abilities. I appreciate the times when you went above and beyond, providing me with the support and feedback I needed to achieve my goals, even when it required extra effort from you. Your mentorship helped me develop critical skills and habits that will continue to benefit me for years to come. Looking back on this journey, I realize that I could not have accomplished what I did without your guidance and support. You have been a beacon of hope, a source of inspiration, and a

model of excellence. I will always be grateful for everything you have done for me, and I hope that I can continue to make you proud.

To Dr. Ray Miller, during my time at Colorado State University, you ARE the best hire that the Economics Department has ever made. I consider myself incredibly fortunate to have had the opportunity to work with you. Thank you for taking me under your wing and demonstrating how to conduct excellent and impactful research. Your guidance has been invaluable, and I am grateful for the time and effort you invested in helping me develop my research skills. Any graduate student would be incredibly fortunate to have you as their advisor. I have learned a great deal from you, and I will carry these lessons with me throughout my career. Your willingness to assist others is one of the many qualities I admire most about you. Even when I frequently sought your help or interrupted you in your office while you were busy, you never hesitated to help me or simply chat about random topics. I always felt comfortable approaching you with questions or concerns. Your unwavering belief in me gave me the confidence to persevere, even when I considered giving up. Whenever we worked on a problem together, you carefully worked through every step with patience and care. You have gone above and beyond what is expected of a mentor, and your mentorship has made a lasting impact on my life and career.

To Dr. David Mushinski, I remember the invaluable encouragement you gave me when I was struggling to choose my dissertation topics. I am deeply grateful for your support and guidance. Our conversations have been stimulating, as you consistently challenge me to think deeply about the underlying econometric problems, and how they are connected to the research questions and economic theories.

To my Mountain Data Group family, including Drs. Salvador Lurbe, Chris Slootmaker, Christopher Keyes, and Ryan Levitt, I want to express my heartfelt appreciation. It was a privilege of a lifetime to work and learn alongside all of you. Thank you for taking me under your wing and providing me with incredible mentorship. I became the skilled programmer I am today because of your guidance. Watching your passion for uncovering

hidden patterns in data has been an enriching experience, and I will always be grateful for that.

To my dearest friends in graduate school, Drs. Kenese Io, Tao Lu, Fatih Kirsanli, Wisnu Nugroho, Manuel David Cruz, and Ashish Sedai, I want to express my deepest gratitude for the countless hours we have spent collaborating and sharing our lives together. These moments have been some of the most joyful and fulfilling experiences of my academic journey. Your unwavering support and encouragement have helped me navigate the challenges of graduate school, and your friendship has been a constant source of comfort and inspiration. I cherish the memories we have created together and look forward to continuing our bond beyond our academic pursuits.

I am deeply grateful to those individuals and organizations who appreciate diversity, equity, and inclusion. Without their support, I would not be where I am today. Firstly, I would like to express my gratitude to the Centre for Children's Happiness. During my formative years, this incredible home gave me the opportunity to be acknowledged and heard for the first time. Secondly, I would like to thank the United World College of Southeast Asia, where I gained a deeper understanding of multiculturalism and global citizenship. Additionally, I am indebted to my other family members, Yvonne Rees, Malcolm Rees, Zachary Rees, and Luke Rees, who welcomed me warmly when I first arrived in Singapore. I will always be grateful for their love and support. Moreover, I would like to express my appreciation for Colorado College, my third home. The robust liberal arts education it offered allowed me to explore my interests and passions. In addition, I found a close-knit family in Toan Luong, Turner Williams, Kali Place, and Sasha Sapp. Lastly, I am grateful to Colorado State University, my fourth home. I not only met incredibly intelligent and kind people there, but I also found the love of my life, Alexandra Brito-Amador.

Alexandra, your endless support and encouragement were instrumental in completing this dissertation. Thank you, my queen!

DEDICATION

I dedicate this dissertation to my beloved mother, Sharon Gregory. From the moment we met, I knew that I had found my best friend in you. Our love runs deep, deeper than bones, and I cannot express enough gratitude for all the love you have shown me. We have been through both good and tough times, but we have always stood by each other's side. You are a true blessing that God has bestowed upon me, and I love you unconditionally. Writing this dedication brings to mind the unwavering support you have given me throughout my academic journey. You have been my rock, my guide, and my source of inspiration, and I am forever grateful for everything you have done for me. This dissertation is a testament to our bond and to the love that we share.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Beyond Income: Health, Wealth, and Racial/Ethnic Welfare Gaps Among Older Americans

1.1 Introduction

Racial and ethnic inequality remains large and persistent in many social and economic domains (e.g., Darity Jr and Myers Jr, 1998; Pager and Shepherd, 2008; Margo, 2016). Income and consumption have traditionally been the chosen metrics for examining racial and ethnic economic disparities in the United States. However, additional factors have been more closely examined in recent years. For example, a persistent wealth gap has been identified between White, Black, and Hispanic Americans (Smith et al., 1997; Shapiro and Kenty-Drane, 2005; Aliprantis et al., 2019; Ashman and Neumuller, 2020; Conley, 2000; Bhutta et al., 2020). Importantly, these alternate metrics provide somewhat different pictures of racial and ethnic inequities. For instance, studies have found that income inequality across racial and ethnic groups is usually lower than wealth inequality, implying some underestimation of the broader racial and ethnic well-being gap when only income is considered (Bhutta et al., 2020).

When alternate metrics are broken down by age cohort, the differences in captured inequality are even greater. In particular, research has indicated that wealth inequality may be a significantly better measure than income when examining welfare disparities at older ages (Smith et al., 1997; Bhutta et al., 2020; Ozawa and Tseng, 2000). Other studies have cited inequality in lifespan, health outcomes, and even leisure as major underlying factors of welfare disparity among older populations (Benhabib et al., 2017; Manton, 1987; Lynch, 2008; Adams et al., 2011; Steptoe et al., 2015; Adams et al., 2011; Hribernik and Mussap, 2010; Han and Patterson, 2007; Pollack et al., 2007; Shea et al., 1996; Smith and Egger, 1993;

Miller and Bairoliya, 2022; Miller et al., 2022). That health disparities matter a great deal at older ages is perhaps unsurprising given that most population level health differences are concentrated in late-life (Deaton and Paxson, 1998; Minkler et al., 2006). The question then remains around the appropriate use of a single metric such as income, wealth, or life expectancy to analyze welfare gaps across racial and ethnic lines. While each such variable individually contributes to the gaps in racial and ethnic well-being, it remains unclear if the adoption of such narrowly defined metrics can adequately capture the true welfare inequality between racial and ethnic groups (e.g., Patton et al., 2016; Lepinteur, 2019; Strife and Downey, 2009). Accounting for the underlying factors contributing to welfare may reveal patterns of inequality that conflict with well-established estimates.

The use of a multidimensional approach to measuring welfare has been adopted by some social scientist when measuring inequality (Maasoumi and Nickesburg, 1983; Rohde and Guest, 2013; Maasoumi, 1986; Maasoumi and Nickelsburg, 1988; Goetz, 1991). Similar to other money metrics of inequality, multidimensional measures create an index based on aggregating attributes of welfare using a social welfare function. This composite measure of welfare combines indicators in their original form that are weighted based on their contribution to overall welfare (Maasoumi, 1986; Manduca, 2018). Individual utility functions are used when creating the aggregate inequality index and the decomposition of these aggregate measures allows for the estimation of the relative contribution of each measure to total welfare inequality.

Aggregate inequality measures have been found to be more informative than the unitary analysis, and more successfully reflect the distribution changes within and between demographic groups in the Unites States (Maasoumi and Nickelsburg, 1988; Rohde and Guest, 2013). These measures, however, fail to account for dynamic spillovers across indicators, which would not be captured with the ad hoc aggregation of individual welfare indicators. Furthermore, the choice of weights applied to each indicator is subjective to the researcher and is required to be sample specific. That is, it is difficult to unambiguously determine how important one indicator is relative to another and how much a surplus on one criterion should be used to compensate for a shortfall in another.

The aim of this chapter is to estimate racial and ethnic welfare inequality among the older U.S. population using an expected utility framework that incorporates differences in consumption, leisure, health, wealth, and mortality. We take a life-cycle approach to better quantify aggregate inequality by incorporating contemporaneous and dynamic spillovers across all modeled outcomes at the individual level. This is an important departure from estimates derived using aggregate models as they may fail to capture the inter-linkages among these factors. For example, if economic and health outcomes are strongly correlated, racial and ethnic disparity measures based on cross-sectional income or consumption might underestimate the aggregate racial and ethnic welfare inequality and would only be presenting a part of the bigger story. Furthermore, the share of Americans over age 65 is projected to reach 20% by 2030 and continue to rise thereafter (Vespa et al., 2018). This highlights the importance of understanding the underlying factors of inequality among older Americans. Our measure of inequality is constructed using a similar framework as Miller and Bairoliya (2022). Specifically, we propose a panel vector autoregressive (VAR) model to approximate the joint late-life evolution of consumption, leisure, health, mortality, and wealth (valued as bequests at death). Throughout the chapter, we will use the terms: wealth and bequest interchangeably, but they convey the same meaning. We estimate parameters of the model using longitudinal data from the Health and Retirement Study (HRS) supplemented with data from the Consumption and Activities Mail Survey (CAMS). Together, these provide a long and rich panel (1992-2016) for our analysis. We then use the estimated system to simulate potential outcome paths by race/ethnicity for a sub-sample of HRS respondents starting from age sixty. Finally, these paths are embedded in a simple expected utility framework to compute a forward-looking ex-ante metric of welfare (measured in consumption equivalents) for each individual in our sample at age sixty. As our measure incorporates individual expectations about outcomes over the entirety of remaining life, it provides a useful single metric of ex-ante well-being at older ages.

Based on the data available in the HRS, we estimate welfare gaps among study participants who self-reported as non-Hispanic Black (hereafter, Black), Hispanic, and non-Hispanic White (hereafter, White). Our main findings can be summarized as follows:

1. Ex-ante age sixty welfare was significantly higher among White HRS respondents. Mean welfare for Black respondents was 38% that of White respondents (Black-White welfare ratio of 0.38). The analogous estimate for Hispanic compared to White respondents was 34% (Hispanic-White welfare ratio of 0.34).

2. Expected annual consumption gaps over remaining life explain the largest share of the welfare gaps between races/ethnicities, accounting for roughly 60-70% of the overall gaps. The mean Black-White welfare ratio based only on consumption was estimated to be 0.62 (or 62%). The analogous estimate for the Hispanic-White ratio was 0.51 (or 51%).

3. Black and Hispanic respondents retired earlier than White respondents overall, but these differences had only small effects on our aggregate measure of racial and ethnic welfare gaps.

4. Health and longevity (life expectancy) were important for overall welfare gaps. Accounting for longevity differences was more important for Black participants, decreasing the estimated mean Black-White welfare ratio by 12 percentage points (pp). In contrast, the welfare cost of living in poor health was more important for Hispanic participants, decreasing the estimated Hispanic-White welfare ratio by 7 pp.

5. Smaller financial bequests (or wealth at death) are nearly as important to estimated welfare gaps as health and longevity. Adjusting for bequests lowers the Black-White welfare ratio an additional 10 pp and the Hispanic-White ratio an additional 9 pp.

Further simulations in which the most racially and ethnically dispersed health risk factors (hypertension and diabetes) are counterfactually eliminated in late-life only marginally closes overall welfare gaps. Moreover, decomposition exercises show that a majority of the estimated welfare gaps are determined by age sixty initial conditions as opposed to racial and ethnic differences in dynamic processes after age sixty. This suggests that policies aimed at closing racial and ethnic gaps in late-life may be more successful and efficient if targeted earlier in the life-cycle. In other words, outside of direct wealth transfers, it may largely be too late to target such interventions directly at older populations.

This study makes several contributions to the existing literature on measuring racial and ethnic inequality. First, most previous studies carried out estimation in a cross-sectional or clinical setting (Aliprantis et al., 2019; Rohde and Guest, 2013; Maasoumi, 1986; Maasoumi and Nickelsburg, 1988). Our study employs a longitudinal panel that captures both contemporaneous and dynamic spillover effects across several economic and health outcomes. This allows for a more comprehensive measure that incorporates the cumulative contribution of each factor to welfare. Our use of microsimulations from a model of life-cycle dynamics also allows us to construct a measure at the individual level within a larger representative sample, so we can examine the entire distribution of welfare. Our forward-looking framework also incorporates differences in the uncertain evolution of outcomes over remaining life, providing a more complete measure of racial and ethnic welfare inequality when compared to other multidimensional measures (Maasoumi, 1986; Maasoumi and Nickelsburg, 1988; Rohde and Guest, 2013). We also use a broader indicator of health, incorporating several morbidities and physical limitations, in addition to self-reported health.

Finally, we contribute to the literature that more specifically focuses on racial and ethnic inequality among older populations. Existing studies in this area have generally focused on a single metric like wealth (Smith et al., 1997; Ozawa and Tseng, 2000; Williams et al., 2001; Martin and Soldo, 1997). We add to this line of research by examining racial and ethnic inequality among older Americans using a dynamic and multi-dimensional metric. Our simulations also shed light on how successful early versus late-life interventions may be in impacting racial and ethnic welfare gaps at older ages.

The rest of this chapter is divided into four sections. First, we will examine the data and statistical methods used in our study in Section 1.2. Second, we will discuss the construction of our welfare measure in Section 1.3. Third, in Section 1.4, we will present the results of our welfare analysis. Finally, we will conclude and discuss policy implications while also addressing limitations in Section 1.5.

1.2 Data and Methods

1.2.1 Data

We utilized data from the Health and Retirement Study (HRS), which is a national biennial longitudinal survey tracking individuals aged 50 and above in the United States across multiple cohorts. The HRS data includes seven birth cohorts, namely the initial HRS cohorts (born between 1931 and 1941), the Study of Assets and Health Dynamics Among the Oldest Old (AHEAD) cohort (born before 1924), the Children of Depression (CODA) cohort (born between 1924 and 1930), the War Baby (WB) cohort (born between 1942 and 1947), and the Early, Mid, and Late Baby Boomer cohorts (born after 1947). Our main data source was the publicly available 2016 RAND HRS Longitudinal File which includes data from 1992 to 2016. The file provided us with cleaned data on various individual characteristics such as race/ethnicity, health, mortality, economic outcomes, age, education, gender, birth cohort, region, and occupation. In the following section, we provide more detailed information on the variables employed in our analysis.

Race/Ethnicity Variables

In the HRS survey, respondents were asked two questions about their race/ethnicity: "Do you consider yourself Hispanic or Latino?" and "Do you consider yourself primarily

White or Caucasian, Black or African American, American Indian or Asian, or something else?" For our analysis, we categorized race/ethnicity into three groups: White, non-Hispanic; Black, non-Hispanic; and Hispanic, based on their answers. We excluded American Indian or Alaskan Native, Asian or Pacific Islander, and Unknown categories from the analysis, as they are not representative.

Health Outcomes

Importantly for older populations, our model incorporates data on comorbidities. Specifically, we include binary indicators for doctor's diagnosis of eight specific health problems as well as an indicator for ever reported difficulties with activities of daily living (ADLs). ADLs include activities such as bathing, getting dressed, walking across the room, and toileting. The health problems included are: (1) high blood pressure and hypertension; (2) diabetes; (3) cancer or any kind of malignant tumor, excluding melanoma; (4) chronic lung disease excluding asthma, chronic bronchitis or emphysema; (5) heart attack, coronary heart disease, angina, congestive heart failure or other heart related problems; (6) stroke or transient ischemic attack; (7) emotional, nervous or psychiatric problems; and (8) arthritis or rheumatism. These health metrics are arguably more objective measures of health. However, self-rated health outcomes, where individuals rank their health on a five-point scale from poor (one) to excellent (five), have also been shown to be a good predictor of mortality even after controlling for other health conditions, health behavior, and socioeconomic characteristics (Idler and Benyamini, 1997). Therefore, we include self-rated health status in our model to test if people have significant private information about their health beyond diagnosis given by a doctor or other observable indicators of health.

Economic Outcomes

Annual hours worked was estimated using self-reported data on weekly hours and number of weeks worked. For the purposes of this study, retired individuals are defined as those with less than 500 hours of work per year. To estimate individual consumption, we use data provided by the Consumption Activities Mail Survey (CAMS), which was sent to a sub-sample of HRS respondents on off years of the core survey. The 2017 RAND CAMS data file provides a constructed estimate of total household consumption derived from household spending data on durables, non-durables, transportation, and housing collected from 2001-2015. We subtracted out-of-pocket health spending from total household consumption and then divided by the number of household members to derive our individual consumption measure. We merge consumption data from CAMS with data from the previous core HRS wave. CAMS data is available for about 20% of HRS respondents from 2000-2014. To address missing consumption data, we follow Miller and Bairoliya (2022) and apply the multiple imputation method, proposed by Honaker and King (2010) for cross-sectional time-series data, which relies on closely related available data such as wealth and income (see Online Appendix for details). Finally, we estimate expected bequests using estimated household asset wealth from the RAND HRS data file. These assets include financial, housing, and other durable wealth (e.g. vehicles, jewelry, etc).

1.2.2 Simulation Model and Estimation

We adapt the panel vector autoregressive (VAR) model of Miller and Bairoliya (2022) to estimate the joint evolution of consumption, leisure, health, mortality, and wealth (valued as bequests at death) across different racial/ethnic groups in late-life. Our proposed model enables us to: (1) accurately measure the racial/ethnic disparities in welfare within a given population; and (2) explore the extent to which these disparities could potentially be reduced through various counterfactual scenarios. The dynamics of the life-cycle are represented as a statistical process and estimated directly from the data. While explicitly modeling the maximization of lifetime utility would enable better policy analysis, it involves solving a complex intertemporal structural model that considers endogenous savings, labor supply, and multiple morbidity and health outcomes. Given that the primary goal of this paper is to develop a welfare measure that accurately reflects population well-being, we believe that a data driven statistical approach is more appropriate in this context.

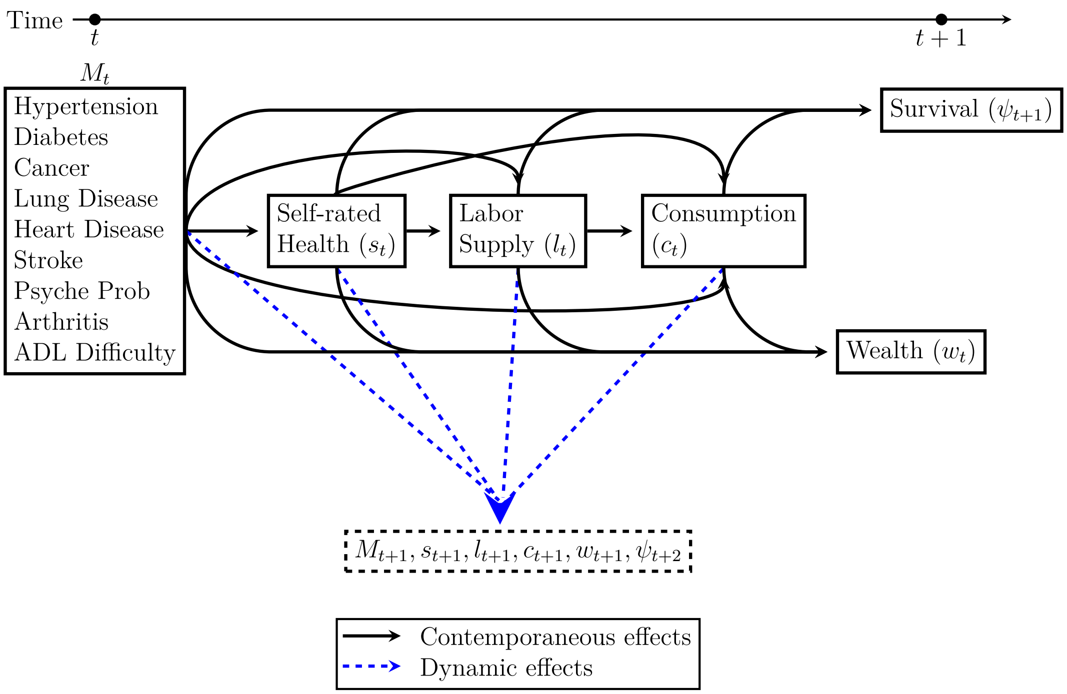


Figure 1.1: Simulation Model With One Period Lag

The core structure of the simulation model is illustrated in Figure 1.1. At the beginning of each time period, morbidity status is updated based on random shocks and exogenous characteristics of an individual. The individual then updates their self-rated health, which affects their labor supply (i.e., their decision to retire) and, in turn, impacts consumption, wealth, and the likelihood of survival to the next time period. Note that the model allows both direct and indirect contemporaneous effects. For example, a stroke may influence retirement directly or through a change in self-rated health. Finally, general lagged effects are also included in the model (e.g., hypertension this period can impact the chance of heart disease next period). An important aspect of including lagged effects is that it allows for more recent diagnoses of a morbidity to have a different impact on health and economic changes than long-standing diagnoses.

Panel VAR Representation

While we allow for higher order lags in estimation, the following VAR(1) demonstrates the relevant structure of the model. In this model, Y_{it} represents a vector of outcomes for an individual i at time t. This vector includes log consumption c, retirement indicator r, self-rated health s, cube root of wealth w, and $n = 9$ morbidity states which are given by the $n \times 1$ vector M. We model each morbidity as an absorbing state to be consistent with the HRS data (e.g., ever diagnosed with hypertension). For simplicity, we also model retirement as an absorbing state (e.g., once retired always retired). We further include a $k \times 1$ vector of fixed individual characteristics X_{it} as exogenous predictors in our model.

Conditional on survival, the outcomes evolve according to the structural VAR(1) model:

$$AY_{it} = BY_{it-1} + CX_{it} + \epsilon_{it}. \tag{1.1}$$

where ϵ is a vector of independent and identically distributed (iid) shocks with zero mean, and the diagonal elements of matrix A are scaled to one. All parameters in the model are identical across individuals and time (e.g., $A_{it} = A$ for all i and t).

The model is estimated in five "blocks" of outcomes: morbidities, self-rated health, retirement, consumption, and wealth blocks. Setting aside the exogenous vector X_{it} for exposition, the VAR(1) model can be written in the following block matrix form:

$$
\begin{array}{r}
n\{ \\
\\
4\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right.
\end{array}
\overset{\overbrace{\quad}^{n}\;\overbrace{\qquad\qquad\qquad\qquad}^{4}}{\left[\begin{array}{c|cccc}
-A_{11} & -A_{12} & -A_{13} & -A_{14} & -A_{15} \\
\hline
-A_{21} & 1 & -a_{23} & -a_{24} & -a_{25} \\
-A_{31} & -a_{32} & 1 & -a_{34} & -a_{35} \\
-A_{41} & -a_{42} & -a_{43} & 1 & -a_{45} \\
-A_{51} & -a_{52} & -a_{53} & -a_{54} & 1
\end{array}\right]}
\left[\begin{array}{c}
M_{it} \\
\hline
s_{it} \\
r_{it} \\
c_{it} \\
w_{it}
\end{array}\right]
=
\overset{\overbrace{\quad}^{n}\;\overbrace{\qquad\qquad\qquad\qquad}^{4}}{\left[\begin{array}{c|cccc}
B_{11} & B_{12} & B_{13} & B_{14} & B_{15} \\
\hline
B_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\
B_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\
B_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\
B_{51} & b_{52} & b_{53} & b_{54} & b_{55}
\end{array}\right]}
\left[\begin{array}{c}
M_{it-1} \\
\hline
s_{it-1} \\
r_{it-1} \\
c_{it-1} \\
w_{it-1}
\end{array}\right]
+
\left[\begin{array}{c}
\epsilon_{1,it} \\
\hline
\epsilon_{2,it} \\
\epsilon_{3,it} \\
\epsilon_{4,it} \\
\epsilon_{5,it}
\end{array}\right],
$$

where $n \times n$ matrix A_{11} has diagonal terms scaled to one. As illustrated in Figure 1.1, we assume the contemporaneous causal pathway runs from morbidities to self-rated health to retirement to consumption to wealth. This assumption is represented in the VAR(1) model by setting $A_{12} = A_{13} = A_{14} = A_{15} = 0$ in the morbidity block, $a_{23} = a_{24} = a_{25} = 0$ in the self-rated health block, $a_{34} = a_{35} = 0$ in the retirement block, and $a_{45} = 0$ in the consumption block. Note that health outcomes and retirement are allowed to affect all future outcomes through general lagged effects. We further allow lagged consumption to impact future wealth, but consumption and wealth are otherwise assumed not to have lagged effects.[1] By applying such block triangulation of the system, we eliminate simultaneity across blocks and allow for block-by-block estimation.

Exogenous characteristics X_{it} include a linear trend for calendar year and dummies for age, education, gender, census division, census occupation code, birth cohort and a post-2008 indicator to account for the great recession. We also include a time invariant individual fixed effect in the consumption equation (π^c) and in the wealth equation (π^w). The unobserved fixed effect helps maintain the appropriate variance in consumption and wealth across time by acting as a person specific drift in the autoregressive process. The entry of exogenous characteristics in the VAR(1) can be explicitly written as:

$$CX_{it} = \begin{array}{c} n\left\{\vphantom{C_{11}}\right. \\ \\ \\ \\ \\ \end{array} \underbrace{\left[\begin{array}{ccccccccccc} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & C_{18} & C_{19} & 0 & 0 \\ \hdashline c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} & c_{27} & c_{28} & c_{29} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} & c_{37} & c_{38} & c_{39} & 0 & 0 \\ c_{41} & 0 & 0 & 0 & 0 & 0 & 0 & c_{48} & c_{49} & c_{410} & 0 \\ c_{51} & 0 & 0 & 0 & 0 & 0 & 0 & c_{58} & c_{59} & 0 & c_{511} \end{array}\right]}_{(n+4) \times k} \underbrace{\begin{bmatrix} Age_{it} \\ Education_i \\ Gender_i \\ Race_i \\ Division_i \\ Occupation_i \\ Cohort_i \\ Year_t \\ Post_t \\ \pi_i^c \\ \pi_i^w \end{bmatrix}}_{k \times 1}.$$

[1]i.e. $B_{14} = B_{15} = b_{24} = b_{25} = b_{34} = b_{35} = b_{45} = 0$

Here we have excluded time invariant regressors from the consumption and wealth equations due to colinearity with the fixed effects. Time invariant socioeconomic characteristics are used instead of fixed effects in the health and retirement equations because absorbing states and ordinal models raise challenges in estimating dynamic panel models with fixed effects. Moreover, the model does well in replicating the dynamics of health and retirement even without unobserved fixed effects (see Appendix A for more details). Finally, note that we normalize c_{410} and c_{511} to one to allow identification of the unobserved fixed effects in the consumption and wealth blocks.

Morbidities

The system's block triangulation does not allow for the direct identification of the structural parameters in the morbidity block since there are nine separate outcomes. Therefore, the morbidity block is estimated as a reduced form VAR. To obtain the reduced form system, the structural system block is pre-multiplied by the inverse of matrix A_{11} as follows:

$$M^*_{it} = -A_{11}^{-1}B_{11}M_{it-1} - A_{11}^{-1}B_{12}s_{it-1} - A_{11}^{-1}B_{13}r_{it-1} - A_{11}^{-1}\left[C_{11}, \ldots, C_{19}\right]X_{it} - A_{11}^{-1}\epsilon_{1,it}.$$

Denoting $-A_{11}^{-1}B_{1j} = \hat{B}_j$, $-A_{11}^{-1}\left[C_{11}, \ldots, C_{19}\right] = \hat{C}$ and $-A_{11}^{-1}\epsilon_{1,t} = e_t$ yields the following reduced form system:

$$M^*_{it} = \hat{B}_1 M_{it-1} + \hat{B}_2 s_{it-1} + \hat{B}_3 r_{it-1} + \hat{C}X_{it} + e_{it}.$$

In the reduced form VAR, all right-hand side variables are predetermined at time t, and morbidity states do not have a direct contemporaneous effect on each other. However, there could be a potential correlation across morbidity states given that the error terms e_t are composites of morbidity-specific structural shocks (i.e., cov$(e_{it}, e'_{it}) \neq 0$). This allows for contemporaneous correlation in the probability of morbidity states. We assume that contemporaneous morbidity shocks follow a standard multivariate normal distribution with an $n \times n$ covariance matrix given by $\sum$.

Morbidity outcomes are binary, and forecasting of the measures is not a true linear VAR process. Therefore, we assume that a continuous latent variable m^* underlies each observed outcome such that:

$$m_{j,it} = 0 \quad if \quad m^*_{j,it} \leq 0$$
$$m_{j,it} = 1 \quad if \quad m^*_{j,it} > 0$$

for $j = 1 \ldots n$. We then have the following model:

$$\begin{bmatrix} m^*_{1,it} \\ \vdots \\ m^*_{n,it} \end{bmatrix} = \begin{bmatrix} \hat{b}_{11} & \cdots & \hat{b}_{1n} \\ \vdots & \ddots & \vdots \\ \hat{b}_{n1} & \cdots & \hat{b}_{nn} \end{bmatrix} \begin{bmatrix} m_{1,it-1} \\ \vdots \\ m_{n,it-1} \end{bmatrix} + \hat{B}_2 s_{it-1} + \hat{B}_3 r_{it-1} + \hat{C} X_t + \begin{bmatrix} e_{1,it} \\ \vdots \\ e_{n,it} \end{bmatrix}. \quad (1.2)$$

It is important to note that the determination of each latent morbidity variable relies on lagged values of the other observed self-rated health and morbidity states. The morbidity block of equations takes the form of a multivariate probit model.

Self-Rated Health

Self-rated health is evaluated using a five-point scale. Therefore, similar to morbidity outcomes, predicting this measure is not a linear VAR process. We assume a continuous latent variable, denoted as s^*, underlies the observed self-rated health state. Accordingly, the relevant equation given in system (1) can be explicitly written as follows:

$$s^*_{it} \quad = \quad A_{21} M_{it} + B_{21} M_{it-1} + b_{22} s_{it-1} + b_{23} r_{it-1} + [c_{21}, \ldots, c_{29}] \, X_{it} + \epsilon_{2,it}. \quad (1.3)$$

The observed health state is defined by the following equation:

$$s_{it} = \delta \;\; if \; \kappa_{\delta-1} < s^*_{it} < \kappa_{\delta} \;\; for \; \delta = 1, \ldots, 5.$$

Here, δ = 1 represents the poorest health state (poor), while $\delta = 5$ represents the best health state (excellent). To account for the persistence of general health shocks over the life-course, we assume that the latent self-rated health depends on the lagged value of the observed self-rated health category. We also assume that ϵ_2 is an iid shock with a standard normal distribution. Consequently, the evolution of self-rated health follows an ordered probit structure. Unlike the morbidity block, this equation may be estimated independently of other outcome blocks, with all structural parameters identified.

Retirement

We assume that retirement is a binary outcome, and that there is a continuous latent variable, denoted by r^*, which underlies the observed outcome. Specifically, we define r_{it} as follows:

$$r_{it} = 0 \quad if \quad r^*_{it} \leq 0$$

$$r_{it} = 1 \quad if \quad r^*_{it} > 0.$$

Assuming that the individual worked during the previous period (and setting $b_{33} = 0$), the retirement model, as defined in system (1), can be expressed as follows:

$$r^*_{it} \;\; = \;\; A_{31} M_{it} + a_{32} s_{it} + B_{31} M_{it-1} + b_{32} s_{it-1} + [c_{31}, \ldots, c_{39}]\, X_{it} + \epsilon_{3,it}. \qquad (1.4)$$

Here, retirement is influenced by both current and lagged values of self-rated health and specific morbidities, as well as exogenous individual characteristics. We assume that ϵ_3 is an iid shock with a standard normal distribution, which implies that the retirement model has a standard probit structure.

Consumption and Wealth

The equation for consumption forecasting given in system (1) can be explicitly written as follows:

$$
\begin{aligned}
c_{it} = {} & A_{41}M_{it} + a_{42}s_{it} + a_{43}r_{it} + B_{41}M_{it-1} + b_{42}s_{it-1} + b_{43}r_{it-1} \\
& + b_{44}c_{it-1} + c_{41}Age_{it} + c_{48}Year_t + c_{49}Post_t + \pi_i^c + \epsilon_{4,it}.
\end{aligned} \tag{1.5}
$$

Similarly, the equation for wealth can be given as:

$$
\begin{aligned}
w_{it} = {} & A_{51}M_{it} + a_{52}s_{it} + a_{53}r_{it} + a_{54}c_{it} + B_{51}M_{it-1} + b_{52}s_{it-1} + b_{53}r_{it-1} \\
& + b_{54}c_{it-1} + b_{55}w_{it-1} + c_{51}Age_{it} + c_{58}Year_t + c_{59}Post_t + \pi_i^w + \epsilon_{5,it}.
\end{aligned} \tag{1.6}
$$

Both of these equations are standard linear dynamic panel data models with a lagged dependent variable and individual-level fixed effects (π). These equations can also be estimated independently of other blocks with all structural parameters identified, including the variance of ϵ_4 and ϵ_5.

Mortality

The last process to model is the survival from one life period to the next. Mortality probabilities are estimated separately from the VAR system mentioned earlier, as all other outcomes described are dependent on survival. Given that an individual is alive at time $t-1$, the survival to the next life period is modeled using the following equation:

$$
\psi_{it} = I\left(\sum_{k=1}^{K}\left[\gamma_k^M M_{it-k} + \gamma_k^s s_{it-k} + \gamma_k^r r_{it-k}\right] + \delta X_{it} + u_{it} > 0\right) \tag{1.7}
$$

Here, $\psi = 1$ indicates survival, X is the vector of previously defined observed individual characteristics, and u_{it} is an iid random shock with a standard normal distribution. The specification allows K lags of morbidity states, self-rated health, and retirement to affect the probability of survival.

Simulations

Our empirical analysis involves three steps, which utilize our forecasting model. Firstly, we estimate the parameters of the model using data from the HRS. The data includes all individuals aged fifty and older from all available waves of the HRS from 1992-2016, amounting to 40,708 unique individuals and 238,091 total individual-year observations. Additional details on the model estimation procedures and results can be found in Appendix A.

Secondly, we simulate remaining life-cycle paths for mortality, health, consumption, wealth, and leisure for a sub-sample of the HRS respondents using the estimated parameter values and age sixty data as initial conditions. The simulation sample consists of all individuals in the initial HRS cohort with age sixty data and the requisite lagged data for simulations. Further information on initial condition descriptives, sampling weights and representativeness, and simulation procedure is also provided in Appendix A.

Finally, we use our expected utility framework, detailed in the following section, to embed the simulated data and construct a measure of ex-ante welfare at age sixty for each individual in our simulation sample by racial/ethnic group.

1.3 Welfare Measure

We extend and modify the measure proposed by Miller and Bairoliya (2022) to include the potential gains in welfare from leaving bequests. We begin by defining expected (remaining) lifetime utility at age j for individual i as:

$$U_{ij} = E\left[\sum_{a=j}^{J} \psi_{ia}\beta^{a-j}\phi(h_{ia})[\bar{u} + log(c_{ia}) + v(l_{ia})] + (1-\psi_{ia})\beta^{a-j}\zeta(b_{ia})\right]$$

Here, c represents consumption (in thousands of dollars), l represents leisure, h represents health, b represents bequests, and Ψ is a survival indicator. We assume log utility over consumption and additive separability with leisure, allowing for a simple decomposition of results. We also report robustness checks where we relax these assumptions. The health measure h is a vector of indicators for each modelled morbidity and self-rated health. We assume that utility from consumption and leisure is scaled by the health function $\phi(h) \in [0, 1]$. Note that $\phi(h) = 1$ represents the utility for a person in perfect health, and $\phi(h) = 0$ represents the utility for a person who is dead. By combining the survival indicator with the health function, we obtain a measure of quality-adjusted life years (QALYs). For example, $\psi\phi(h) = 1$ represents a year of life with no adverse health conditions. Furthermore, we consider the potential welfare gains from leaving bequests, as it could quantitatively contribute to driving inequalities across racial and ethnic groups, since bequests can be significant and are likely correlated with health and consumption.

We define welfare using a consumption-equivalent variation measure. In particular, we define welfare for an individual i at age j to satisfy the following condition:

$$U_{ij} = E\left[\sum_{a=j}^{J} \psi_{ma}\beta^{a-j}\phi(h_{ma})[\bar{u} + log(\lambda_{ij}) + v(l_{ma})] + (1-\psi_{ma})\beta^{a-j}\zeta(b_{ma})\right]$$

Here, ψ_m, h_m, l_m, and b_m are reference levels of survival, health, leisure, and bequests chosen by the individual. The welfare λ_{ij} is defined as the fixed annual consumption that, when combined with the reference health, leisure, survival, and bequest profiles, yields the same expected lifetime utility as the outcome profiles of the individual. For instance, if $\lambda_{ij} = 20$, it means that the individual would be indifferent between receiving their own stochastic outcome profiles moving forward or receiving $\$20,000$ in annual consumption together with the reference profiles for health, leisure, bequests, and survival.

The welfare condition can be rearranged to yield an additive decomposition:

$$log\left(\lambda_{ij}\right) = \tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left[E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]E_{\psi}\left[log\left(c_{ia}\right)\right]+\Phi\right] \tag{1.8}$$

$$+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]\left(E_{\psi}\left[\nu\left(l_{ia}\right)\right]-E_{\psi}\left[\nu\left(l_{ma}\right)\right]\right) \tag{1.9}$$

$$+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left(E\left[\psi_{ia}\right]-E\left[\psi_{ma}\right]\right)E_{\psi}\left[\phi\left(h_{ma}\right)\right]E_{\psi}\left[u_{ia}\right] \tag{1.10}$$

$$+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}\left(E_{\psi}\left[\phi\left(h_{ia}\right)\right]-E_{\psi}\left[\phi\left(h_{ma}\right)\right]\right)E\left[\psi_{ia}\right]E_{\psi}\left[u_{ia}\right] \tag{1.11}$$

$$+\tilde{\psi}\sum_{a=j}^{J}\beta^{a-j}E\left[(1-\psi_{ia})\zeta(b_{ia})-(1-\psi_{ma})\zeta(b_{ma})\right] \tag{1.12}$$

where, Φ is defined as follows:

$$\Phi = \left(E\left[\psi_{ia}\phi\left(h_{ia}\right)u_{ia}\right]-E\left[\psi_{ia}\phi\left(h_{ia}\right)\right]E_{\psi}\left[u_{ia}\right]\right)$$
$$-\left(E\left[\psi_{ma}\phi\left(h_{ma}\right)\nu\left(l_{ma}\right)\right]-E\left[\psi_{ma}\phi\left(h_{ma}\right)\right]E_{\psi}\left[\nu\left(l_{ma}\right)\right]\right)$$

and $\tilde{\psi}$ is the reciprocal of the reference discounted quality-adjusted life expectancy, and E_{ψ} denotes expected values conditional on survival.

The first term in equation (1.8) represents expected utility from consumption weighted by the reference quality-adjusted life expectancy. The Φ term is an adjustment for uncertainty over the life cycle. Together, these terms provide an individual's consumption-equivalent welfare before adjusting for expected leisure, life expectancy, health, or bequests. The term in equation (1.9) is the welfare adjustment for leisure, which represents the expected utility difference in leisure weighted by the reference quality-adjusted life expectancy. The correction term in equation (1.10) is the difference in life expectancy weighted by how much a life year is worth, which represents the expected flow utility from outcome bundles of individual i. The term in equation (1.11) corrects for expected health differences between individual i and the reference over remaining life. Finally, the term in equation (1.12) adjusts welfare for differences in expected bequests.

1.3.1 Calibration

To calibrate the preference parameters, we assume that the health utility is directly proportional to the health state vector, represented as $\phi(h_t) = \gamma h_t$. To determine the utility weights vector γ, we follow the methodology of Miller and Bairoliya (2022) and utilize the Health Utilities Index Mark 3 (HUI3) instrument. Further details regarding the calibration process can be found in Appendix A. The HUI3 has been extensively employed in the health utility literature (Furlong et al., 1998; Feeny et al., 2002; Horsman et al., 2003), and data from the year 2000 collection of a subset of Health and Retirement Study (HRS) respondents were used.

For retired individuals, we normalize leisure time to one, while for workers, we set leisure time to 0.66, assuming an endowment of $5,840$ hours per year (16 hours a day $\times$ 365 days), where workers supply $2,000$ hours of labor. We define preferences over leisure time using the function $v(l) = -\frac{\theta\epsilon}{1+\epsilon}(1-l)^{\frac{1+\epsilon}{\epsilon}}$, where ϵ represents the constant Frisch elasticity of labor supply. In line with Jones and Klenow (2016), we set $\epsilon = 1$ and derive a benchmark disutility weight of $\theta = 7.82$, such that the marginal cost of leisure is equated to the marginal benefit for the median individual in our sample.

Furthermore, we choose a discount factor of $\beta = 0.98$, which corresponds to an annual discount rate of one percent, in line with previous studies (De Nardi, 2004). We define preferences for bequests using the function $\zeta(b) = \Phi_1 \left(1 + \frac{b}{\Phi_2}\right)^{1-\sigma}$, where Φ_1 reflects the strength of the bequest motive and Φ_2 measures the extent to which bequests are a luxury good. Consistent with De Nardi (2004), we set $\Phi_1 = -9.5$, $\Phi_2 = 11.6$, and $\sigma = 1.5$ for our benchmark calibration.

With the preferences defined above, a retired individual will prefer life to death as long as the flow intercept $\bar{u}$ plus log consumption is positive. We set $\bar{u} = -log(2)$, which implies that \$2,000 of consumption is needed for a retiree to maintain positive flow utility. This is approximately 10% of the mean annual consumption in our sample, which has been argued to be a reasonable parameterization of the flow intercept (Murphy and Topel, 2006). This value of $\bar{u}$ also yields a median value of remaining life for sixty-year-olds of about \$60,000 per QALY in our simulation sample, which falls within the range of typical values reported in the literature (Ryen and Svensson, 2015; Kaplan and Bush, 1982). For more details, see Appendix A.

1.3.2 Reference Outcomes

To calculate welfare, we need to define reference profiles that will be used for all individuals. For leisure, we choose retirement by age sixty as our reference, meaning full leisure from age sixty onward. For health-adjusted welfare equivalents, the standard approach is to use a notion of "normal" or "good" health as the reference. This allows us to compare individuals based on their consumption differences. We follow the approach of Miller and Bairoliya (2022) and use a constant reference health level of $\phi(h_{ma}) = 0.8$ and a reference sixty-year-old life expectancy of 24 years. To ensure the robustness of our analysis, we also conduct a sensitivity analysis using a longer reference life expectancy. Finally, we choose a reference bequest level of $\$500,000$. In summary, we assume that we can compare the welfare of age 60 retirees who expect to live to age 84 in "good" health and leave a bequest of $\$500,000$ solely based on expected consumption profiles.

1.4 Welfare Results

Our presentation of welfare analyses across racial groups includes (1) age sixty descriptive statistics, (2) model estimates, (3) mean outcomes and welfare measures, (4) decomposition exercise, (5) selected morbidity counterfactuals, and (6) robustness and sensitivity. We discuss these results using the EHRS cohort as our benchmark group as it is the earliest of the seven and contains the longest panel of available data.

Table 1.1: Simulation Sample Age Sixty Descriptive Statistics by Race/Ethnicity

	White	Black	Hispanic
Individuals	2,339	536	235
Hypertension (%)	35.27	59.85	37.76
Diabetes (%)	10.00	22.79	19.30
Cancer (%)	7.05	5.48	5.07
Lung disease (%)	7.66	4.84	3.63
Heart disease (%)	14.04	13.03	10.23
Stroke (%)	2.66	5.53	1.63
Psyche problem (%)	7.18	6.36	12.24
Arthritis (%)	44.81	47.19	41.30
Difficulty with ADLs (%)	9.85	19.89	25.26
Self-rated health (%)			
Poor	5.75	13.51	19.11
Fair	12.96	25.71	31.64
Good	28.00	29.83	29.34
Very good	34.42	20.17	13.12
Excellent	18.87	10.78	6.78
Retired (%)	50.25	55.55	60.24
Annual consumption ($1000s, mean)	30.19	18.35	14.77
Male (%)	47.09	45.08	37.49
Education (%)			
<HS	23.92	48.03	70.99
HS	36.01	27.05	16.51
Some College	20.45	16.37	7.25
College	19.61	8.54	5.25

Notes: Respondents from the initial HRS cohort. Estimates using base year respondent analysis weights. Consumption is reported in real 2010 dollars. Source: HRS.

Table 1.1 summarizes the initial conditions at age sixty in the simulation sample, grouped by race/ethnicity. The prevalence of hypertension, diabetes, stroke, arthritis, and difficulty with activities of daily living (ADLs) was higher among Black and Hispanic respondents than White respondents. However, stroke and arthritis were exceptions for Hispanic respondents. For instance, the reported incidence of diabetes was 23% for Black respondents and 19% for Hispanic respondents, compared to only 10% for White respondents. This represents a 2.3-fold and 1.9-fold difference, respectively. In terms of self-reported health, 14% of Black respondents and 19% of Hispanic respondents reported poor health status, whereas only 6% of White respondents did so. Black and Hispanic respondents, on average, retired earlier than White respondents. Specifically, 56% of Black respondents and 60% of Hispanic respondents retired by age sixty, compared to only 50%

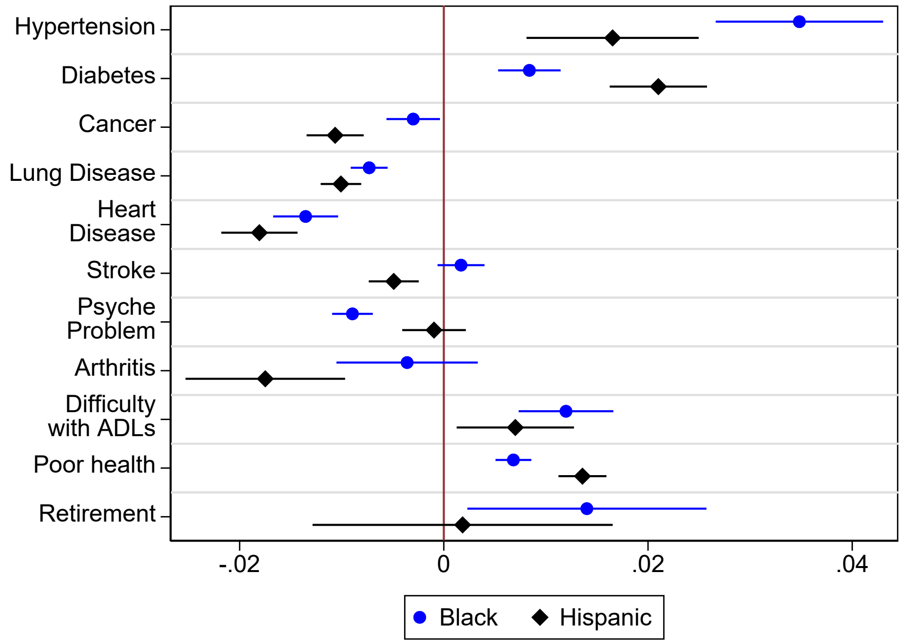
of White respondents. Additionally, cross-sectional consumption at age sixty averaged $18,350 for Black respondents and $14,770 for Hispanic respondents, as opposed to $30,190 for White respondents. These differences correspond to a 1.6-fold and 2-fold difference, respectively. Finally, at age sixty, 48% of Black and 71% of Hispanic respondents reported less than a high school education, while only 24% of White respondents did so.

1.4.1 Model Estimates

This section presents some selected results from our simulation model aimed at gaining a better understanding of the correlation between race and other outcomes in the data. In particular, we use Figure 1.2 to illustrate the estimated average marginal effects of race on various health and retirement indicators. Our findings reveal that, in comparison to White respondents, Black and Hispanic respondents have a higher likelihood of experiencing health problems such as hypertension, diabetes, stroke, difficulty with ADLs, self-rated poor health, and early retirement (although stroke is an exception for Black respondents). For instance, compared to White respondents, Black and Hispanic respondents have a marginal increase in the probability of hypertension by about 2.8 and 1.8 percentage points (pp), respectively. We also discovered that race, in relation to morbidities, is associated with self-rated health. For example, the average marginal increase in the probability of reporting poor health is approximately 0.8 pp for Black respondents and 1.7 pp for Hispanic respondents.

1.4.2 Welfare Gaps in EHRS Cohort

This analysis examines the mean outcomes and distribution of a welfare measure across racial and ethnic groups of sixty-year-olds from the EHRS cohort, as presented in Table 1.2. Panel A displays the mean consumption, retirement, life expectancy, QALE, and expected bequests at age sixty, while Panel B shows the cumulative contribution of each factor to the welfare measure. Additionally, the mean Black-White and Hispanic-White outcome and welfare ratios are presented.



Notes: Dependent variables across rows. White non-Hispanics are the reference group. Spikes indicate 95% confidence intervals.

Figure 1.2: Average Marginal Effect of Race on Health and Retirement Probabilities

In Panel A, it is observed that the annual consumption for Black respondents at age sixty is approximately 61% of that for White respondents (Black-White ratio of 0.61). The corresponding estimate for Hispanic respondents is around 49% (Hispanic-White ratio of 0.49). Black and Hispanic respondents retire earlier than White respondents overall, with a Black-White and Hispanic-White ratio of about 1.11 (or 110%) and 1.20 (or 120%), respectively. Moreover, White respondents have higher life expectancy, QALE, and financial bequests. For instance, the life expectancy of Black respondents averages around 18.6 years compared to 22 years for White respondents. On the other hand, the difference in life expectancy between Hispanic and White respondents is less than a year. However, the difference is more significant in QALE, with about 14 years for Hispanic respondents compared to 17 years for White respondents. Additionally, the expected financial bequests for Black and Hispanic respondents are approximately 26% and 24% of that for White respondents, respectively.

Table 1.2: Outcomes and Welfare by Race/Ethnicity

Measure	White	Black	Mean Hispanic	Black-White Ratio	Hispanic-White Ratio
Panel A: Outcomes					
Consumption	30.190	18.346	14.773	0.608	0.489
Retired	0.502	0.555	0.602	1.106	1.199
Life Exp.	21.540	18.556	21.137	0.861	0.981
QALE	16.866	13.572	14.296	0.805	0.848
Bequests	396.459	101.327	95.810	0.256	0.242
Panel B: Welfare					
Consumption	23.817	14.752	12.120	0.619	0.509
Leisure	22.046	13.853	11.426	0.628	0.518
Life Exp.	22.677	11.476	11.399	0.506	0.503
Health	20.178	9.583	8.636	0.475	0.428
Bequests	18.322	6.946	6.340	0.379	0.346

Notes: Estimates use base year respondent analysis weights. Consumption and welfare reported in \$1000s. Life expectancy and QALE reported in years. Retired is an indicator. Panel B presents cumulatively adjusted welfare estimates.

Even if differences in expected leisure, life expectancy, health, and financial bequests are ignored, Panel B shows a substantial overall welfare gap between races/ethnicities. The "consumption" Black-White welfare ratio is approximately 0.62 (or 62%), while the Hispanic-White welfare ratio is 0.51 (or 51%). Furthermore, the average expected consumption for Black and Hispanic respondents of the welfare distribution is only \$14,752 and \$12,120, respectively, compared to an average of \$23,817 for White respondents. Slight adjustments for lost leisure due to working past age sixty slightly decrease the welfare gap, increasing the Black-White welfare ratio by an additional 1 pp and the Hispanic-White welfare ratio by 1 pp. This is because Black and Hispanic respondents are expected to retire earlier than White respondents overall, but these differences have only small effects on the fully-adjusted measure of racial and ethnic welfare gaps. In other words, adjusting welfare for later retirement lowers average welfare by \$900 $(14,753 - 13,853)$ for Black and \$694 $(12,120 - 11,426)$ for Hispanic respondents. This implies that Black and Hispanic respondents would be willing to give up an average of \$900 and \$694 in expected annual consumption to retire at age sixty, respectively.

Health and life expectancy are essential for overall welfare gaps. Further adjusting for life expectancy differences is more important for Black respondents, decreasing the estimated mean Black-White welfare ratio by 12 pp. In contrast, the welfare cost of living in poor health is more important for Hispanic respondents, decreasing the estimated Hispanic-White welfare ratio by 7 pp. The last row of Panel B displays adjustments for leaving financial bequests, yielding the fully-adjusted welfare measure. Smaller financial bequests are almost as important to estimated welfare gaps as health and longevity. Adjusting for bequests lowers the Black-White welfare ratio by an additional 10 pp and the Hispanic-White ratio by an additional 8 pp.

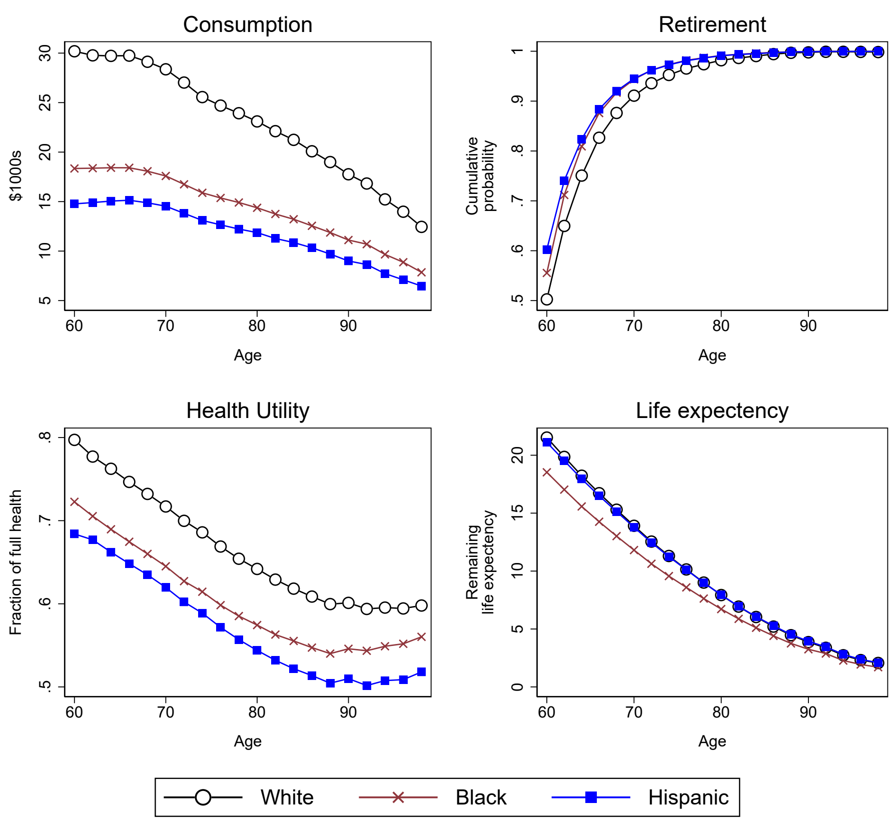


Figure 1.3: Average Life Cycle Profiles by Race/Ethnicity

Figure 1.3 presents the average expected life-cycle profiles, which can help us understand the differences between racial and ethnic groups. The mean gaps in consumption, retirement, and life expectancy are most significant at age sixty and decrease gradually as individuals age. However, consumption gaps still exist even into the nineties. On the other hand, health gaps are significant at age sixty and persist throughout the remaining life. Our welfare results indicate that consumption, health, life expectancy, and bequests are crucial factors that contribute to racial and ethnic welfare inequality, while earlier retirement and leisure have a relatively minor impact.

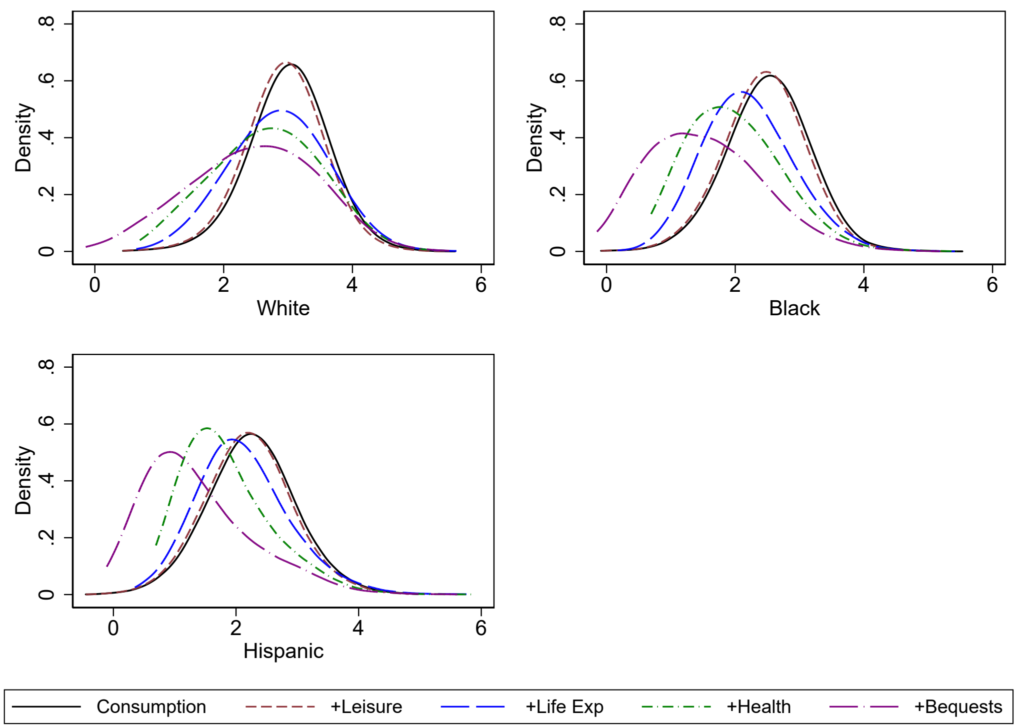


Figure 1.4: Cumulative Change in Distribution of Log Welfare by Race/Ethnicity

Figure 1.4 illustrates the cumulative change in the distribution of log welfare at age sixty across racial and ethnic groups in greater detail. Adjusting for leisure, life expectancy, health, and bequests has a greater negative impact on the welfare distribution of Black and Hispanic respondents than White respondents. It is worth noting that the adjustments cause inequality *within* the Black and Hispanic respondent populations to increase more than the White population (i.e., the left tail of the welfare distribution becomes fatter).

This is consistent with existing evidence on inequality, which shows that relative income disparity between the top and bottom 10 percent is particularly acute for Black Americans. Pew Research Center reported in 2016 that the 90th percentile of Black households earned nearly ten times as much as the 10th percentile (Pew Research Center, 2018).

1.4.3 Decomposition

In our estimates, the increase in welfare inequality across racial and ethnic groups can be attributed to two potential factors: (1) differences in the distribution of initial conditions at age sixty across races and ethnicities and/or (2) differences in the stochastic processes experienced by each racial and ethnic group after age sixty. The main question we aim to answer in this analysis is: to what extent do initial conditions at age sixty versus differences in outcome dynamics after age sixty explain the racial and ethnic welfare gaps? To address this question, we conduct several experiments to estimate the impact of initial differences at age sixty as well as differential outcome evolutions across racial and ethnic groups after age sixty. In all our experiments, we eliminate disparities by assigning initial conditions or late-life transitions of White participants to Black and Hispanic participants. Our main results are presented in Table 1.3, where we report the Black-White and Hispanic-White ratios for quality-adjusted life year (QALE), expected lifetime consumption (ELC), and our fully-adjusted welfare measure at age sixty.

Table 1.3: Decomposition

	QALE ratio		ELC ratio		Welfare ratio	
Experiment	Black-White	Hispanic-White	Black-White	Hispanic-White	Black-White	Hispanic-White
Baseline	0.809	0.846	0.559	0.491	0.399	0.355
Transitions	0.847	0.806	0.578	0.459	0.432	0.337
Initial conditions	0.953	1.039	0.965	1.053	0.901	1.075

Notes: Estimates use base year respondent analysis weights.

In our first round of experiments, we assigned transition probabilities of White participants after age sixty to Black and Hispanic groups to investigate how the evolution of outcomes after sixty affects gaps in QALE, ELC, and welfare. However, as displayed in the second row of Table 1.3, the differences in the evolution of outcomes can only account for a small portion of the racial and ethnic welfare gaps. For instance, assigning White transition probabilities to Black participants only increases the QALE ratio by 3.8 pp, ELC ratio by 1.9 pp, and fully-adjusted welfare by 3.3 pp. Surprisingly, outcomes for Hispanic respondents become slightly worse when given White transition probabilities, with the QALE ratio decreasing by 4 pp, ELC ratio by 3.2 pp, and the fully-adjusted welfare ratio by 1.8 pp.

We then shifted our focus to the role of age sixty differences in explaining the estimated racial and ethnic welfare gaps. Our previous experiment only changed the evolution of outcomes after age sixty, while keeping the initial distribution of outcomes the same for each racial and ethnic group. As indicated in the last row of Table 1.3, when we instead assign the initial conditions of White respondents to Black and Hispanic groups, the estimated Black-White and Hispanic-White ratios in QALE, ELC, and fully-adjusted welfare measures increase significantly. Equating initial conditions enhances the Black-White welfare ratio by 50 pp and the Hispanic-White ratio by 70 pp. Our decomposition exercises indicate that the majority of the estimated welfare gaps are determined by age sixty initial conditions rather than racial and ethnic differences in dynamic processes after age sixty.

1.4.4 Morbidity Counterfactuals

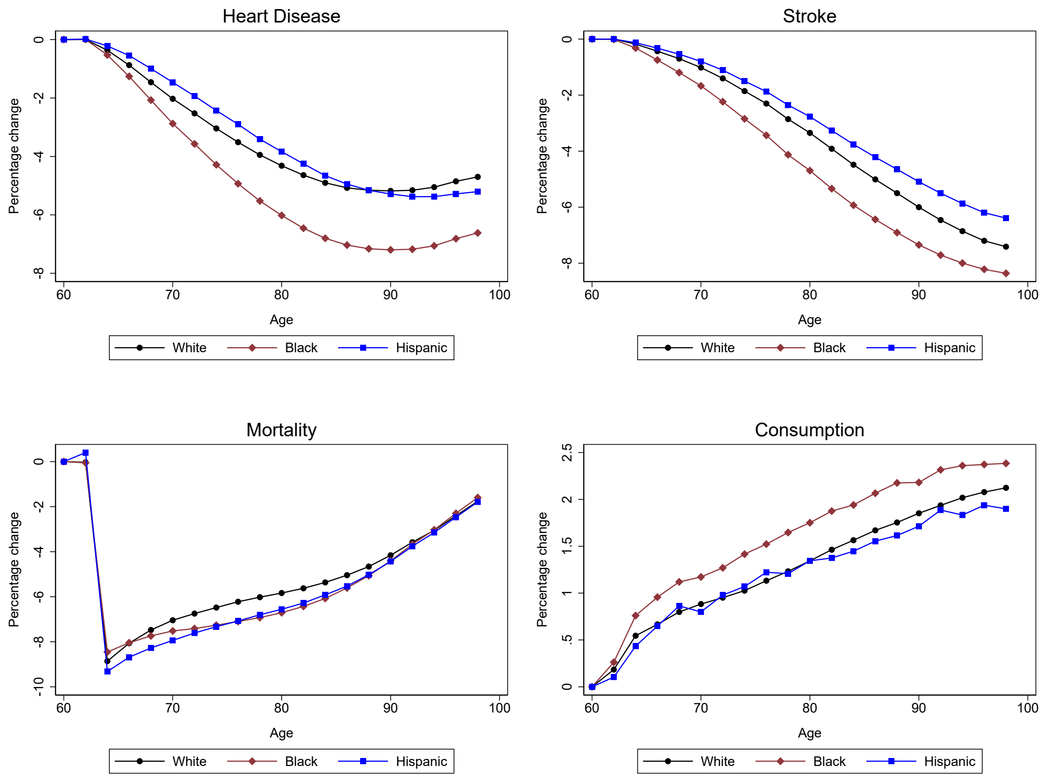
This section aims to investigate how morbidities affect outcomes and welfare across different racial and ethnic groups. Table 1.4 displays the increase in quality-adjusted life expectancy (QALE) and expected lifetime consumption (ELC), the loss in bequests, and the Black-White/Hispanic-White welfare ratios when all hypertension or diabetes cases are eliminated after age sixty. We selected hypertension and diabetes since our model's estimates (refer to Figure 1.2) indicate that, among other health measures and exogenous characteristics, these two conditions have the most significant racial and ethnic disparities. Additionally, hypertension and diabetes are established risk factors for various downstream health issues, including stroke, ischemic heart disease, renal dysfunction, kidney failure, and other medical problems (e.g., Lewington, 2002; Rapsomaniki et al., 2014; Huang et al., 2014; Kokubo and Iwashima, 2015; Raghavan et al., 2019).

Table 1.4: Eliminating Late-life Hypertension and Diabetes by Race/Ethnicity

	Hypertension			Diabetes		
Outcomes and Welfare	White	Black	Hispanic	White	Black	Hispanic
QALE gain	1.241	1.462	1.409	0.699	1.040	1.251
ELC gain	32.152	26.594	19.484	17.926	19.788	17.016
Bequest loss	13.947	5.390	5.116	5.499	2.385	3.460
Welfare ratio	–	0.384	0.341	–	0.391	0.356
Baseline ratio	–	0.379	0.346	–	0.379	0.346

Notes: Estimates use base year respondent analysis weights. Consumption and bequests reported in $1000s. QALE reported in years. Welfare ratio is measured as Black-White and Hispanic-White.

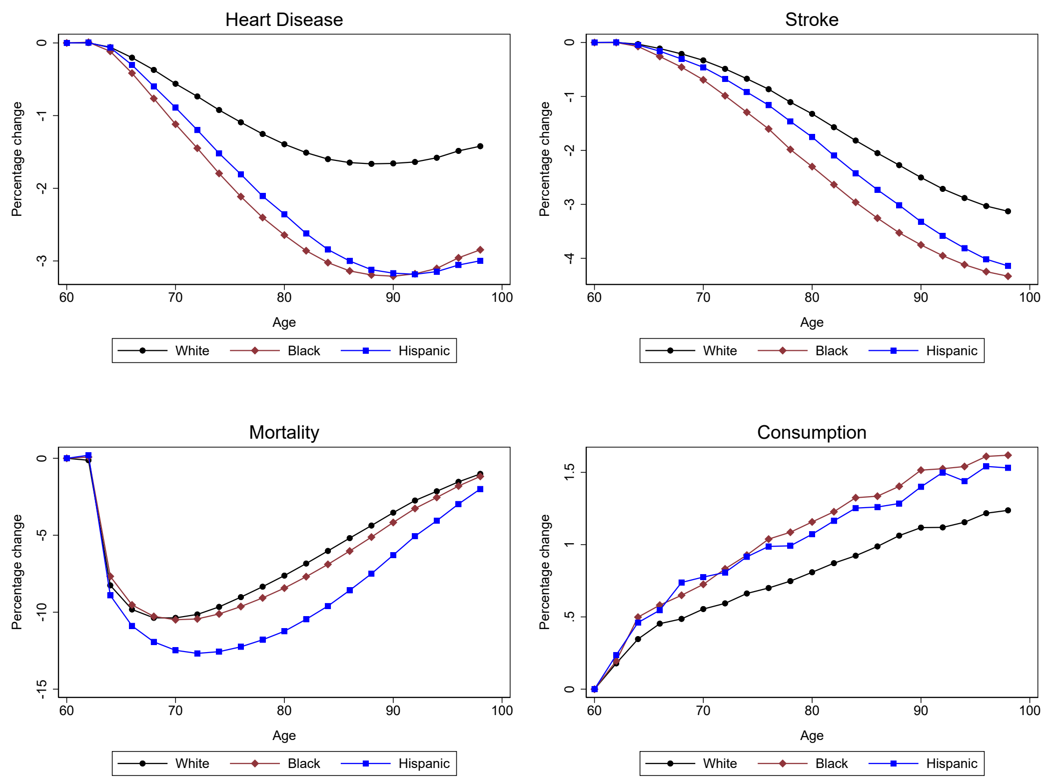
Table 1.4 shows that eliminating hypertension resulted in Black and Hispanic respondents gaining slightly more QALE than White respondents at age sixty. Specifically, Black and Hispanic respondents gained about 1.5 and 1.4 years, respectively, compared to 1.2 years for White respondents. However, White respondents had a higher gain in ELC of $32,152 compared to $26,594 for Black and $19,484 for Hispanic respondents due to their larger annual consumption. But this gain in lifetime consumption was partially offset by a larger decline in bequests for White respondents ($13,947) compared to Black ($5,390) and Hispanic ($5,116) respondents. Eliminating late-life diabetes had similar patterns but with smaller effects. Black respondents gained more lifetime consumption than White respondents, and Hispanic respondents had higher QALE gains (and bequest losses) than Black respondents. However, overall, the counterfactual welfare ratios suggest that eliminating these diseases only marginally closes overall welfare gaps. Eliminating hypertension after age sixty increases the Black-White welfare ratio by 0.005 pp but lowers the Hispanic-White ratio by 0.005 pp. Eliminating diabetes saw slightly larger improvements, with the Black-White welfare ratio increasing by 0.012 pp and Hispanic-White welfare ratio by 0.01 pp.



Notes: Results plot percentage difference in expected outcomes with the exogenous elimination of hypertension after age sixty relative to baseline. Sample includes all individuals in the simulation sample from the EHRS cohort. Expected outcomes are conditional on survival.

Figure 1.5: Impulse Response to Elimination of Hypertension after Age 60

To better understand how morbidities influence the dynamics of other outcomes in the system across racial and ethnic groups, Figures 1.5 and 1.6 illustrate the average percentage change in several expected outcomes with the exogenous elimination of hypertension and diabetes after age sixty. Eliminating hypertension after age sixty reduces the average probability of developing heart disease and stroke for all races and ethnicities, with the strongest changes for Black respondents. For example, Black respondents experienced a decreased probability of heart disease of about 6% by age eighty compared to approximately 4% for White and Hispanic respondents. Similarly, the probability of stroke by age eighty decreased by about 5% for Black respondents compared to 3% for White and Hispanic respondents. Interestingly, although Black respondents saw the largest gains in consumption, we see similar mortality gains for Hispanic respondents.



Notes: Results plot percentage difference in expected outcomes with the exogenous elimination of diabetes after age sixty relative to baseline. Sample includes all individuals in the simulation sample from the EHRS cohort. Expected outcomes are conditional on survival.

Figure 1.6: Impulse Response to Elimination of Diabetes after Age 60

Compared to eliminating hypertension, general patterns are similar in the diabetes experiment, with the main difference being that effects are relatively stronger for Hispanic respondents. For example, Hispanic and Black respondents see similar improvements in heart disease incidence and consumption when eliminating late-life diabetes. However, Hispanic respondents clearly have the largest mortality gains. These patterns are consistent with the very strong association with increased diabetes risk among Hispanic respondents from our model estimates shown in Figure 1.2.

1.4.5 Robustness

We estimated our main results under a variety of alternate modeling assumptions from our benchmark to gauge the sensitivity of our findings. These included using a race and ethnicity specific forecasting model, a higher reference life expectancy and reference bequests, and alternate preference parameter values. Summary results are presented in Table 1.5. While welfare levels are somewhat sensitive to robustness specifications, the Black-White ratio remains in the range of 0.36-0.45 and the Hispanic-White ratio in the range of 0.33-0.41.

Table 1.5: Sensitivity of Mean Welfare by Race/Ethnicity

	White	Black	Hispanic	Black-White Ratio	Hispanic-White Ratio
Benchmark	18.322	6.946	6.340	0.379	0.346
Race specific forecast	18.392	6.741	6.691	0.367	0.364
Reference life expectancy	11.655	5.221	4.744	0.448	0.407
Reference bequests	17.644	6.690	6.106	0.379	0.346
$\bar{u}$ = -log(1.5)	18.448	6.714	6.211	0.364	0.337
β = 0.90	17.675	7.713	6.756	0.436	0.382
ϵ = 0.5	19.622	7.364	6.728	0.375	0.343
ϵ = 2	16.510	6.357	5.800	0.385	0.351
θ = 16	16.999	6.517	5.944	0.383	0.350
Φ_1 = -5	19.111	8.036	7.284	0.421	0.381
Φ_2 = 6	18.473	7.095	6.444	0.384	0.349
σ = 2	18.448	7.100	6.415	0.385	0.348
Health utility weights	18.602	7.148	6.589	0.384	0.354

Notes: Estimates use base year respondent analysis weights. Welfare reported in $1000s.

Race/Ethnicity Specific Simulation Model

In our benchmark simulation model, we allowed dynamics to vary across race/ethnicity through a race/ethnicity intercept (or individual fixed effect for consumption and wealth). However, we assumed that other model parameters were the same for all racial and ethnic groups. For instance, we assumed that the direct effect of diabetes on self-rated health was identical for White, Black, and Hispanic respondents. In contrast, the "race-specific forecast" results in Table 1.5 were obtained by separately estimating a forecasting model for each of the three groups. This approach has the disadvantage of a loss in precision and fewer observations, especially for the Hispanic sample. However, we found that mean welfare only slightly increased for White and Hispanic respondents and slightly decreased for Black respondents when using this approach. As a result, the Black-White welfare ratio decreased by only 1 pp, and the Hispanic-White ratio increased by 2 pp compared to our benchmark results.

Reference Life Expectancy and Bequests

The third row of Table 1.5 shows sensitivity of results when we increase the reference age sixty life expectancy from 24 to 30 years. As is clear from equation (1.10), increasing reference life expectancy is more costly to log welfare for those with higher flow utility. Thus we see larger mean declines in welfare for White respondents, with a corresponding increase in the Black-White welfare ratio of 8 pp and in the Hispanic-White ratio of 4 pp. The next row in Table 1.5 provides results when the reference bequest level is increased from $\$500,000$ to one million dollars. Quantitatively, this has a much smaller effect on mean welfare than reference life expectancy, and welfare ratios are unchanged compared to the benchmark.

Preference Parameters

The remainder of Table 1.5 presents sensitivity results for our choice of calibrated preference parameter values. We first check the sensitivity of results to our choice of flow intercept $\bar{u}$. Specifically, we set $\bar{u} = -log(1.5)$, implying that $\$1,500$ of consumption is needed for a retiree to maintain positive flow utility compared to our benchmark value of $\$2,000$. The change has only a small impact on estimated welfare inequality, decreasing both reported ratios by about 1 pp. With a lower time discount rate $\beta = 0.9$, anticipated gaps in future consumption and health are less important for welfare. As such, the Black-White welfare ratio increases about 6 pp and the Hispanic-White ratio 4 pp. The welfare ratios increase by a similar magnitude when we decrease the strength of the bequest motive Φ_1 by roughly half compared to the benchmark. Changes in Frisch elasticity of labor supply ϵ, disutility weight on labor supply θ, and the other bequest parameters Φ_2 and σ, each have very small impact on inequality results. Lastly, in our benchmark estimates we calibrated health utility weights by assuming that consumption and leisure were conceptualized as fixed across health states by HRS respondents that completed the HUI3 (see Appendix A for full discussion on this assumption and how it can be relaxed). The last row of Table 1.5 shows that results are largely insensitive to relaxing this assumption.

Consumption and Leisure Utility

In our study, we also investigate the reliability of our findings using a more general form of flow utility for consumption and leisure, represented by the following equation:

$$\phi(h)\left[\frac{c^{1-\gamma}}{1-\gamma}\left(1-(1-\gamma)\frac{\theta\epsilon}{1+\epsilon}(1-l)^{\frac{1+\epsilon}{\epsilon}}\right)^{\gamma}-\frac{\bar{u}^{1-\gamma}}{1-\gamma}\right] \tag{1.13}$$

When $\gamma = 1$, this formula is equivalent to our benchmark case. However, when $\gamma > 1$, the curvature over consumption increases. This creates several challenges. Firstly, it becomes impossible to determine welfare for individuals at the very top of the health distribution, as increasing their consumption would never provide the same expected life-time utility as the reference life expectancy. Therefore, we report the median welfare instead of the mean welfare in Table 1.6. Secondly, higher curvature creates another issue: as γ increases, the implied value of life rises steeply. As shown in the first column of Table 1.6, the median value of life per QALY is $\$178,000$ per QALY when $\gamma = 2$, which is high but still reasonable. The estimated median Black-White and Hispanic-White welfare ratios also increase modestly to 0.39 and 0.36, respectively. When $\gamma = 3$, the value of life reaches about $\$557,000$ per QALY, and the welfare ratios increase more substantially to 0.58 and 0.54. However, only three out of 23 value of life studies surveyed by Ryen and Svensson (2015) estimated a mean value of life over $\$150,000$. Therefore, caution should be exercised when interpreting robustness results with high curvature values, as the value of life may be overstated. Nonetheless, higher curvature values provide an understanding of the robustness of key results.

Table 1.6: Sensitivity for Higher Curvature–Median Welfare by Race/Ethnicity

	VOL	White	Black	Hispanic	Black-White Ratio	Hispanic-White Ratio
$\gamma = 1.0$	59.805	12.231	4.277	3.223	0.350	0.264
$\gamma = 1.5$	104.601	7.416	2.421	1.969	0.326	0.265
$\gamma = 2.0$	178.162	3.729	1.472	1.336	0.395	0.358
$\gamma = 3.0$	557.168	1.756	1.012	0.951	0.576	0.541

Notes: Estimates use base year respondent analysis weights. Welfare reported in $1000s.

1.5 Conclusion

We propose and estimate an individual measure of welfare incorporating heterogeneity and uncertainty in future consumption, leisure, health, wealth and mortality at age sixty. Our measure broadly indicates that racial and ethnic inequality is larger than suggested by other welfare metrics such as income or consumption. We also find health, mortality, and wealth gaps are important in explaining the level of racial welfare inequality among the older Americans in our sample, with leisure playing a comparatively minor role.

Our decomposition exercises show that a majority of the estimated welfare gaps are determined by age sixty initial conditions as opposed to racial/ethnic differences in dynamic processes after age sixty. Our morbidity counterfactuals further suggest that eliminating common heath risk factors such as hypertension or diabetes in late-life only marginally closes overall welfare gaps. These simulations suggest that policies aimed at closing racial/ethnic gaps in late-life may be more successful and efficient if targeted earlier in the life-cycle. In other words, outside of direct wealth transfers, it may largely be too late to target such interventions directly at older populations.

Our approach is not without limitations. We do not explicitly account for morbidity spillover effects such as the cost of caregiver time and the numerous costs associated with the loss of a spouse. Likewise, we abstract from other potentially important inputs into late-life welfare such as social interactions and end-of-life care. We assume institutions and relevant policies remain fixed moving forward and past trends in late-life health, retirement, and consumption continue into the future. For example, significant anticipated changes to Social Security or Medicare programs or exponential advances in medicine could alter the distribution of our welfare measure. Nonetheless, our framework provides important insights into the sources and scope of racial/ethnic welfare gaps.

Chapter 2

The Morning Advantage: Differential Returns to Sunlight Exposure on Well-Being

2.1 Introduction

For a long time, Gross Domestic Product (GDP) has been widely regarded as a welfare metric of well-being due to its ability to monetarily value a variety of goods and services, its linear methodology, objectivity, and clarity, as well as its usefulness in international comparisons. However, there is increasing agreement among economists and policy makers that the relationship between GDP and standard of living indicators is not universal and differences in income account for only a small portion of differences in happiness among people (Frey and Stutzer, 2002). In fact, GDP has been criticized for being a weak indicator of social welfare and for potentially misleading public policy decisions (Fleurbaey, 2009). As such, Richard Easterlin in what came to be known as the "Easterlin paradox", noted that beyond a certain level of income, increases in economic prosperity do not lead to increases in happiness. For example, countries with higher GDP per capita may also have higher levels of inequality, which can negatively impact well-being (Oishi et al., 2011). This perspective challenges the traditional economic focus on GDP as a measure of a society's well-being and suggests that other factors, such as health, longevity, and social interactions should also be taken into account (Easterlin, 1995; Diener and Biswas-Diener, 2011; Miller and Bairoliya, 2022).

Well-being, however, is a multifaceted concept influenced by a variety of factors. In addition to health, longevity, and social interactions, chronobiological factors are critical to our overall well-being. Sunlight is a significant contributor to chronobiology, influencing our mood and energy levels (Blume et al., 2019). Several mental health benefits are associated with exposure to sunlight, including increased serotonin production and improved symptoms of seasonal affective disorder (SAD) (Lambert et al., 2002; Kent et al., 2009; Gloth 3rd et al., 1999). Vitamin D, primarily obtained through sunlight exposure, is essential for bone health and immune system function, with vitamin D deficiency linked to an increased risk of depression (Anglin et al., 2013). Furthermore, sunlight exposure can improve sleep quality and patterns, resulting in an overall better quality of life (Takasu et al., 2006; Mead, 2008). Exposure to morning sunlight, in particular, has been linked to better sleep quality and mood (Figueiro et al., 2017; Mead, 2008).

Exposure to sunlight has been linked to reduced risk of various diseases such as cardiovascular disease and diabetes (Wallis et al., 2008; Exebio et al., 2016). This means that individuals who are exposed to sunlight may potentially save money on healthcare costs in the long term by reducing their risk of these diseases. Preventing or managing these diseases can be costly, but taking preventative measures such as getting enough sunlight exposure can result in cost savings for individuals and society as a whole. Therefore, understanding the importance of sunlight exposure for overall well-being is essential for promoting healthy lifestyles and potentially reducing healthcare costs.

In this chapter, we investigate whether exposure to sunlight increases well-being. We construct a spatially rich dataset on expressed sentiment, or emotional state from the online social media platform Twitter, and estimate the relationship between sentiment and sunlight exposure in the United States (hereafter, U.S.). However, estimating the causal effect of sunlight exposure on sentiment is empirically challenging. Most literature on the subject lack proper identification strategies needed to vouch for causal interpretation. Moreover, most of the evidence is based on descriptive studies or laboratory experiments (Lambert et al., 2002; Anglin et al., 2013; Figueiro et al., 2017; Veleva et al., 2018; Shah and Gurbani, 2019). Since randomization is unfeasible, the natural experiment induced by daylight savings time (hereafter, DST) serves as an alternative, potentially as good as randomization to identify the effect of interest. DST refers to the practice of adjusting clocks by moving them forward by an hour in spring, commonly known as "Spring Forward", and back by an hour in fall, commonly known as "Fall Back". At present, virtually all nations in the European Union, the vast majority of states in the U.S. and Canadian provinces, and 40 additional countries, such as Mexico, Chile, Israel, and Iran, implement DST.[2]

Our identification strategy focuses on the quasi-experimental nature of the DST transition. Specifically, we examine the Spring Forward transition, when an hour of sunlight is moved from the morning to the evening, and the Fall Back transition, when an hour of sunlight is moved from the evening to the morning. These transitions act as external shocks to the allocation of sunlight between morning and evening, which varies across space and time due to differences in sunrise and sunset times. Our results show a significant increase in sentiment during the morning hour after the Fall Back transition, when there is additional sunlight. In contrast, we observe a significant decrease in sentiment during the morning hour after the Spring Forward transition, when there is a loss of sunlight. However, we find that the variations in sentiment during the evening hour are relatively minor. Next, we estimate differential returns to sunlight by morning versus evening while netting out seasonal, weekend, and holiday effects. Our results indicate that the returns to sunlight on sentiment is stronger in the morning compared to the evening. Specifically, the positive effect of sunlight on sentiment in the morning highlights the underappreciated benefits to human sentiment. Therefore, the potential shifting to darker mornings and brighter evenings following the proposed Sunshine Protection Act may do more harm than good.

[2]DST was initially introduced in the U.S. as a measure to save energy during wartime. Since the Uniform Time Act of 1966 was passed, DST has been a recurring practice in most U.S. states. This legislation granted the states the freedom to choose whether to implement DST, while mandating uniform start and end dates for those who opted to observe it. Over time, Congress has made two significant changes to the DST transition dates. In 1986, the spring transition was moved from the final Sunday in April to the first Sunday in April, effective starting in 1987, by amending the Uniform Time Act. More recently, the Energy Policy Act of 2005 changed both transition dates. DST now starts on the second Sunday of March and continues until the first Sunday of November, providing a three to four-week extension in the spring and a one-week extension in the fall, beginning in 2007.

This chapter contributes to the well-being literature in economics, which has gained prominence in recent times. Traditionally, economics has been focused on economic growth and material wealth, but there is now a growing recognition of the crucial role played by non-monetary factors in shaping people's well-being. These factors include health, longevity, social interactions, leisure, political and natural environments, and others that have all been linked to individual well-being (Easterlin, 1995; Diener and Biswas-Diener, 2011; Miller and Bairoliya, 2022). In addition to these factors, it is essential to consider sunlight when studying well-being, as it plays a critical role in regulating circadian rhythms and affecting mood. Different regions experience varying amounts of sunlight due to differences in latitude and seasonal changes. Therefore, sunlight can be a significant determinant of well-being and should be considered when assessing overall well-being in a population.

This chapter is also related to studies that have used DST as an empirical identification strategy. These studies have shown that the Spring Forward transition affects various factors, including crime rates (Doleac and Sanders, 2015), traffic accidents (Smith, 2016; Bünnings and Schiele, 2021), energy demand (Kotchen and Grant, 2011; Sexton and Beatty, 2014), myocardial infarction (Toro et al., 2015), and our well-being (Kountouris and Remoundou, 2014; Kuehnle and Wunder, 2016). In addition, another study has demonstrated that the Fall Back transition reduces hospital admissions in both Germany and the U.S. from an hour of additional sleep (Jin and Ziebarth, 2020).

We also contribute to the methodology of evaluating overall well-being in economics. One established approach for assessing overall well-being is through subjective well-being (SWB) surveys. However, self-reported surveys are known to be unreliable sources of information (Schwarz, 1999). This unreliability is due to various factors such as the respondent's mood, the context of the survey, previous questions on the survey that can temporarily highlight certain aspects of life (Schwarz and Strack, 1999), regional dialects and norms, and the interaction between the interviewer and interviewee (Clark, 2003). Additionally, the quality of answers can be impacted by the specific questions asked or the order in which they are presented. Surveys are also resource-intensive, time-consuming, and more susceptible to small-sample bias. To overcome these limitations, we use sentiment analysis borrowed from Computer Science. Specifically, we utilize Natural Language Processing (NLP) algorithms to examine unstructured words in tweet contents using Twitter data and assign them an expressed sentiment or emotional state. Expressed sentiment can be considered an estimate of "experienced utility", which is a momentary measure of pleasure and pain, similar to subjective well-being (Baylis, 2020).

The chapter proceeds as follows. Section 2.2 outlines our testing hypotheses. Section 2.3 describes the data and discusses the empirical framework and identification strategy. Section 2.4 presents results from our regression analyses. Section 2.5 estimates the differential returns to sunlight. Section 2.6 provides a number of sensitivity checks validating the robustness of our results. Section 2.7 concludes the chapter with a summary of the main findings and further policy implications.

2.2 Hypotheses

It is important to keep in mind that DST affects sentiment in two ways. Firstly, it reallocates sunlight between morning and evening, and secondly, it disrupts sleep schedules. Sleep is a crucial determinant of both physical and mental health, and inadequate sleep has been linked to various negative health outcomes such as increased risk of depression, mood disorders, cardiovascular, and metabolic diseases (Ford and Cooper-Patrick, 2001; Tobaldini et al., 2017). Figure 2.1 illustrates the effects of DST on sunrise and sunset times throughout the year. During the spring transition, clocks move forward from 2 am to 3 am, causing a one-hour delay in both sunrise and sunset times. On the other hand, during the fall, clocks are set back by an hour, making the transition back to Standard Time more comfortable and moving sunrise and sunset times earlier by one hour.

Our first hypothesis about the effect of DST transitions on sentiment is based on the combined effects of sunlight exposure and sleep during the first week after the transition dates. For instance, on the morning after the fall transition, we anticipate an improvement in sentiment because individuals get an extra hour of sleep and experience an earlier sunrise during the first week, as shown by a red square in Figure 2.1. Conversely, on the morning after the spring transition, we expect a decline in sentiment because individuals lose an extra hour of sleep and experience a late sunrise during the first week, as shown by a purple square in Figure 2.1.

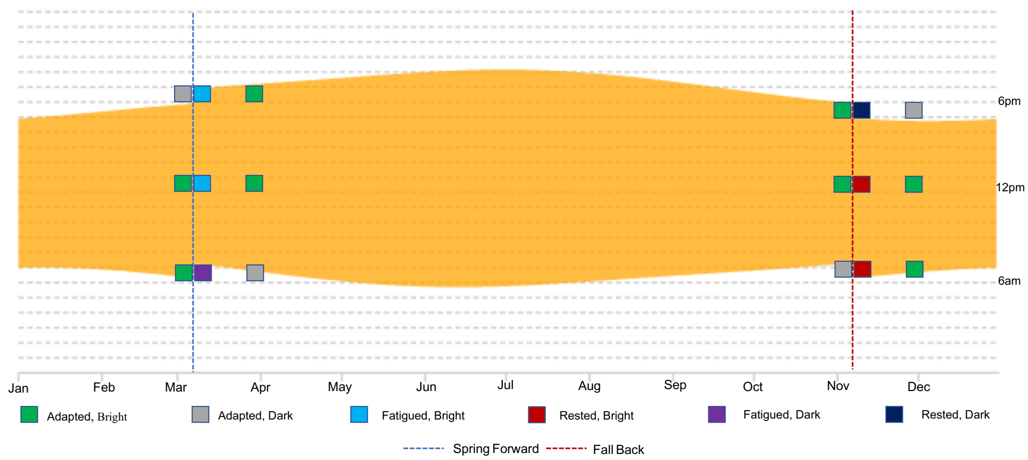


Figure 2.1: The Influence of Daylight Savings Time on Sunlight

The second hypothesis explores the impact of DST transitions on evening sentiment. It is unclear how these transitions will affect sentiment during the first week after the change, as sunlight and sleep operate in opposite directions. In Figure 2.1, the fall transition is represented by a dark-blue square because of an earlier sunset and an additional hour of sleep, while the spring transition is represented by a light-blue square because of a later sunset but a loss of an additional hour of sleep.

Finally, our third hypothesis assumes that sleep patterns will return to normal by the fourth week after the transition dates. For example, Barnes and Wagner (2009) used the American Time Use Survey to report that Americans sleep approximately 40 minutes less on the night of the spring transition. The impact of this transition on sleep patterns can last anywhere from two days to two weeks, depending on the individual (Valdez et al., 1997), with an average duration of about one week (Harrison, 2013). Given that sunlight matters independently, and sleep patterns are expected to be restored by week four, we predict an improvement in sentiment in the morning during the fall transition. This improvement will capture a moment of increased sunlight and adapted sleep patterns (represented by a green square in Figure 2.1). In contrast, we predict a decline in sentiment in the evening during the fall transition as this moment will reflect reduced sunlight and adapted sleep patterns (represented by a gray square in Figure 2.1). On the other hand, during the spring transition, we anticipate a decrease in sentiment in the morning due to reduced sunlight and adapted sleep patterns (represented by a gray square in Figure 2.1). However, we expect an increase in sentiment in the evening during the spring transition, reflecting a moment of increased sunlight and adapted sleep patterns (represented by a green square in Figure 2.1).

2.3 Data and Identification Strategy

2.3.1 Data

Conducting a survey to determine hourly and daily sentiment across the U.S. would be too expensive and time-consuming, but using publicly available historical and real-time sentiment on Twitter is a more affordable option. Using a collection of geolocated and time-tamped tweets with NLP algorithms that identify sentiment, we are able to analyze hourly and daily fluctuations in sentiment throughout the U.S. The data for this study was collected from Twitter[3] using the Twitter's Academic Research Application Programming

[3]Twitter was founded in 2006. At the time, it was a simple platform that allowed users to send short messages, or "tweets", of up to 140 characters. The goal was to create a platform where users could quickly share updates, thoughts, and information with their followers. Initially, Twitter was not very popular, with only a few thousand users in the first year. However, it quickly gained traction and by 2009, there were over 10 million active users on the platform. As the platform continued to grow, so did the number of users. By 2012, there were over 500 million users on Twitter.

Interface (API). Twitter's Academic Research API allows researchers to access a subset of Twitter's historical and real-time data for academic research purposes. The API provides access to a sample of approximately 1% of all tweets that are posted on the platform, which is approximately 500 million tweets per month. We started gathering geolocated and timestamped tweets from within the U.S. We collected a large number of tweets that met our requirements and were created within the time period starting in 2014 and ending in 2022.

Geolocated tweets are tweets that have been tagged with a specific location that the users have consented to have the location information shared for the posts. This location can be the exact latitude and longitude coordinates of the tweet, or it can be the name of a city, state, or country. Timestamped tweets are tweets that have a specific time associated with them. This time can be the time the tweet was posted, or it can be the time the tweet was created. The tweet's date and time are recorded in Coordinated Universal Time (UTC). However, we are interested in the local time of day that the users made the tweet at their specific location. Unfortunately, Twitter removed time zone information from its API in 2018. As a result, we wrote a Python code to determine the timezone for each tweet. We found that 98% of the locations that were successfully geocoded to a state were also successfully geocoded to exactly one timezone.[4]

The concern regarding the Twitter data revolves around a selected sample in two ways: firstly, these users are actively engaged on Twitter, and secondly, their tweets must be geolocated for us to determine their local timezone. It is possible that these users may differ fundamentally from both the overall population and/or the population of Twitter users who opt not to geolocate their tweets. A study conducted by Baylis (2020) investigated temperature preferences using a sample of Twitter users and found that the population of users who choose to geolocate their tweets does not exhibit significant differences from the

[4]Instructions on how to access Twitter's Academic Research API and gather tweets containing geolocation information and timestamps are provided in Appendix B.

general Twitter population. Furthermore, Baylis (2020) demonstrates that when comparing Twitter users to the larger population, Twitter users are likely to be representative of the overall population. Therefore, while we acknowledge the limitations of our selected Twitter sample, it is important to consider that previous research suggests the potential representativeness of Twitter users in relation to the broader population.

To determine the sunrise and sunset times for each tweet made by the users, we used latitude and date information. The Earth rotates at an angular velocity of 15° per hour, so we will need to use a formula that takes into account the location and day of the year to accurately calculate these times. Sunrise time is the time when the sun first becomes visible or when daylight arrives. Sunset time, on the other hand, is the moment when the sun goes below the horizon. We can estimate the sunrise/sunset times for a given location on the day of the year by using the solar declination, δ. Following Almorox et al. (2005),

$$\delta = (180/\pi) \cdot (0.006918 - 0.399912 \cdot cos\Gamma + 0.070257 \cdot sin\Gamma - 0.006758 \cdot cos2\Gamma + 0.000907 \cdot sin2\Gamma - 0.002697 \cdot cos3\Gamma + 0.00148 \cdot sin3\Gamma)$$

where the day angle Γ is given by:

$$\Gamma = 2\pi \cdot (n - 1)/365$$

and n is the number of the day of the year, starting from the first of January.

The hour angle of the Sun (w_s) changes based on the specific location and time of the year. This is due to different interpretations of sunrise and sunset and if twilight is included in the definition. The average hour angle in a flat area, referred to as the theoretical sunrise or sunset, happens when the Sun is at the lowest point in the sky. This can be calculated in degrees from the center of the Sun.

$$w_s = cos^{-1}[(sin\Psi - sin\lambda \cdot sin\delta)/(cos\lambda \cdot cos\delta)]$$

Ψ is a measure of the Sun's position in the sky relative to the observer's meridian. We follow Almorox et al. (2005) and choose $\Psi = -0.8333$ (e.g. the top of the Sun that is apparently even with horizon). λ is the latitude. Thus, the sunrise/sunset times can be calculated as follow:

$$\textit{Sunrise Time} = 12 - acos(w_s)/(15/360 \cdot 2 \cdot \pi)$$
$$\textit{Sunset Time} = 12 + acos(w_s)/(15/360 \cdot 2 \cdot \pi)$$

We use machine learning algorithms to detect emotions in individuals based on their tweet contents. This process, called NLP, involves converting unstructured text data into quantitative data. Within NLP, a specific set of techniques called "sentiment analysis" is used to evaluate the expressed sentiment of individual posts. Sentiment analysis has been widely used in economics and other social sciences for various purposes, such as measuring the impact of environmental threats on well-being (e.g. extreme ambient temperatures or air pollution) (Baylis, 2020; Zheng et al., 2019), explaining financial market volatility (Boudoukh et al., 2019), or well-being during COVID-19 pandemic (Wang et al., 2022). There are more than fifty publicly available algorithms used to conduct sentiment analysis (Baylis, 2020). Because the method by which these measures are constructed can differ substantially, analyses using expressed sentiment should ideally demonstrate reasonable consistency across multiple measures. In this study, we translate tweet content into three measures of expressed sentiment derived from prior work: AFINN (Nielsen, 2011), VADER (Hutto and Gilbert, 2014), and Bidirectional Encoder Representation from Transformers (BERT) (Devlin et al., 2018).

Both AFINN and VADER use a word list or dictionary to determine the sentiment of English-language words. To measure the sentiment of a given piece of text, the average of all scored words within that text is calculated. AFINN uses an expert-created dictionary to map words to emotional states. Specifically, the AFINN-165 dictionary consists of

3,382 words with sentiment scores ranging from -5 to 5, where -5 represents a negative emotional state and 5 represents a positive emotional state. VADER, on the other hand, is a sentiment analysis tool that is specifically designed to detect sentiments expressed in social media and is well-suited to analyze texts from other domains. VADER is an open-source, normalized, weighted composite score. The lexicon used by VADER is created by aggregating ratings from ten independent human raters, and the candidate words for the lexicon are selected from existing measures of sentiment and augmented with lexical features commonly found in online contexts, such as emoticons (e.g., "):", ":") and slang (e.g., "nah", "lmao"). The VADER measure also incorporates information about the order of words and the use of intensifiers in a sentence to determine the valence of a particular sentiment. For example, "very good" would be scored higher than "good".

The BERT algorithm, developed by Google, is used to identify the emotions conveyed in a text by taking into account the context of the words within the sentence. BERT can analyze the relationships between words in both the forward and backward directions, enabling it to understand the meaning of words in their context. To determine the sentiment of a post, BERT assigns five scores ranging from zero to one based on the probability of the post being positive, negative, or neutral. We use a pre-trained language model of BERT, called BERTweet, which has demonstrated excellent performance in analyzing English-language tweets (Nguyen et al., 2020). Additionally, BERT has been shown to outperform dictionary-based models, such as AFINN and VADER (Nandwani and Verma, 2021). To ensure a consistent interpretation of expressed sentiment across all measures, we standardize each measure across the entire universe of tweets, setting the mean to zero and the standard deviation to one.[5] Throughout the chapter, we will use the terms: BERT and BERTweet interchangeably, but they convey the same meaning.

Table 2.1 shows the correlations between the three measures. All of the measures are strongly positively correlated with each other. To ensure the validity of sentiment analysis,

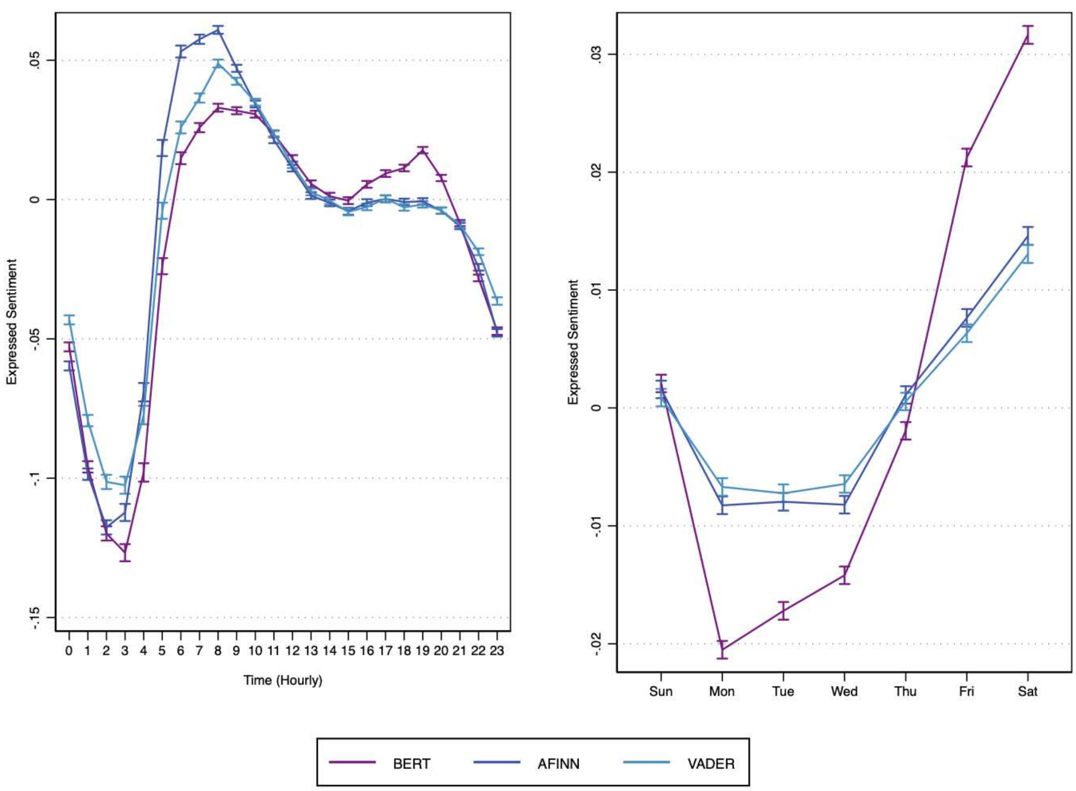
[5]Additional information on the construction of these measures can be found in Appendix B.

Table 2.1: Correlations of Expressed Sentiment Measures

	BERT	VADER	AFINN
BERT	1.000		
VADER	0.613	1.000	
AFINN	0.528	0.726	1.000

Notes: Pairwise correlations of means of measures of standardized expressed sentiment.

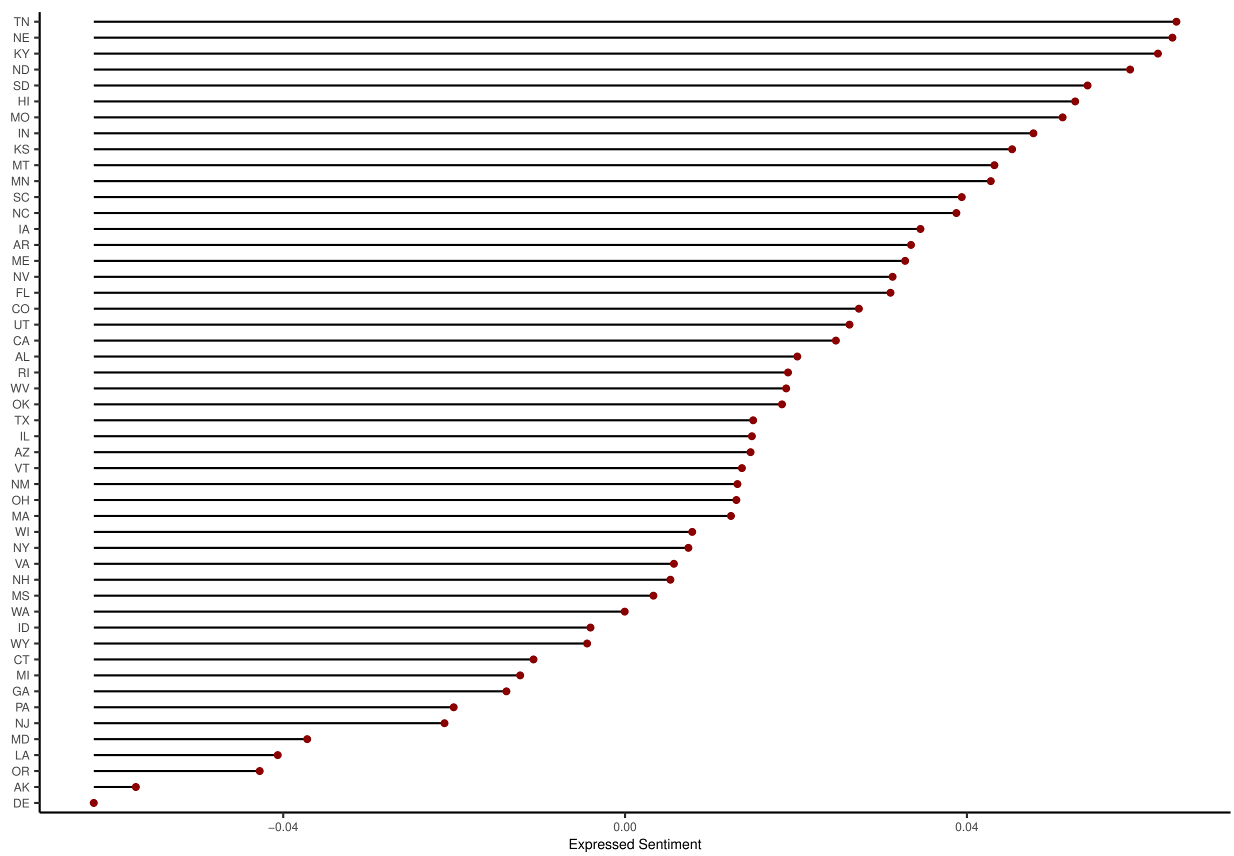
we also conduct a validation exercise that examines how expressed sentiment changes hourly and over the course of the days of the week. Figure 2.2 displays the fluctuation of mean standardized expressed sentiments hourly and throughout the week.



Notes: Hourly variation (left panel) and day of week variation (right panel) of mean measure of standardized expressed sentiment measures. Spikes indicates 95% confidence intervals.

Figure 2.2: Expressed Sentiment Measures by Hour and Day of Week

The hourly and weekly patterns observed in this study are consistent with previous research (Dodds et al., 2011; Helliwell and Wang, 2014; Yang and Srinivasan, 2016). Specifically, individuals tend to experience a more positive outlook on Fridays and weekends in comparison to other weekdays, with Monday having the lowest and Saturday showing the highest positive sentiment. Additionally, the sentiment expressed hourly shows a peak around 8 am and a low point around 3 am. Figure 2.3 illustrates averages standardized sentiment measure by state, which are calculated using pre-trained BERTweet. There is considerable variation in sentiment across states, with Tennessee appearing to be the most positive and Delaware the least positive in our sample.



Notes: The vertical axis ordered from highest to lowest, where highest indicates the most positive sentiment.

Figure 2.3: Expressed Sentiment by State

2.3.2 Identification Strategy

Short-Run Model

We compare average expressed sentiment in the morning and the evening, both one week before and one week after the transition out of DST for states that have adopted it. Our identification strategy depends on the exogenous reallocation of sunlight between the morning and the evening induced by differences in sunrise and sunset times across space and time.[6] To identify the responsiveness of expressed sentiment to changes in sunlight exposure and sleep, we estimate regressions based on the following equation:

$$\begin{aligned} S_{ijt} = \; & \alpha_0 + \; \alpha_1 POST_{it} + \; \alpha_2 E_{it} + \; \alpha_3 \left(POST_{it} \times E_{it} \right) \\ & + \; \Gamma_1 W_{it} + \; \Gamma_2 H_{it} + \; \Gamma_3 LT_{ij} + \; \Gamma_4 LG_{ij} + \; \Phi V_{ijt} + \; \lambda_j + \; \delta_{it} + \; \varepsilon_{ijt} \end{aligned} \tag{2.1}$$

Let i, j, and t index sampled tweet, state, and time. S_{ijt} is one of the three measures of expressed sentiment. We standardized the outcome measures to have a mean of zero and a unit standard deviation, so the point estimates, denoted by α_1 and α_3, represent the change in the conditional mean of expressed sentiment, measured in standard deviations. $POST_{it}$ assumes a value of one if tweet i is made in the week after the fall transition, and a value of zero if the tweet is made in the week before. Upon leaving DST in the fall, an hour of sunlight is removed from the evening and returned to the morning.[7] To take advantage of the variation in sunrise and sunset times across space, we break the sample into morning hour ($+/-$ one hour from the sunrise time in each location of tweet i) and evening hour ($+/-$ one hour from the sunset time in each location of tweet i). Thus, E_{it} equals one if tweet i is made in the evening hour, and zero if tweet i is made in the morning hour. Figure

[6] Doleac and Sanders (2015) and Smith (2016), for example, consider law changes to DST policy in 2007 to account for endogeneity, since DST occurs simultaneously across 48 states (Arizona and Hawaii do not observe DST) and at approximately the same time each year.

[7] Upon entering DST in the spring, an hour of sunlight is removed from the morning and returned to the evening.

2.4 illustrates the variation across space in the influence of DST on sunrise and sunset times throughout the year. What can be seen immediately are two sources of variation in sunlight: changes in the timing of sunrise and sunset times over the course of the year and by space.

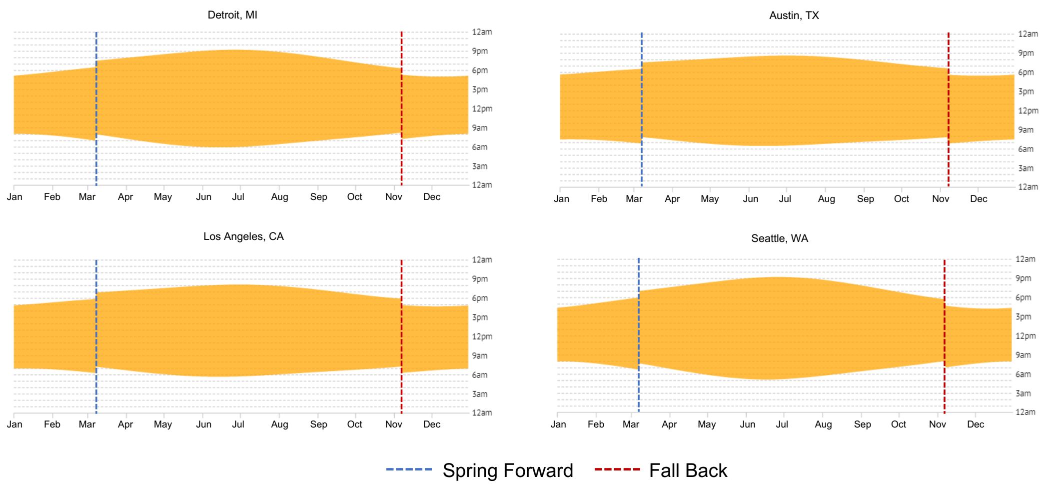


Figure 2.4: The Variation in the Influence of Daylight Savings Time on Sunlight

The coefficients of interest are α_1 and α_3, corresponding to $POST_{it}$ and $POST_{it} \times E_{it}$, respectively. α_1 gives us the estimated effect of fall transition on expressed sentiment for tweets that post in the morning hour, while α_3 measures the difference in the effect of the fall transition on expressed sentiment between tweets posted during the morning and tweets posted during the evening. $\alpha_1 + \alpha_3$ measures the total effect of the fall transition on the expressed sentiment for tweets made during the evening hour. We expect α_1, capturing a moment of added sunlight and rest, to be positive and α_3 to be unknown because it captures a moment of additional sleep but a loss of sunlight. LT_{ij} and LG_{ij} correspond to the latitude and longitude of the tweet location. LT_{ij} and LG_{ij} allow us to account for potential regional variations in expressed sentiment that are not captured by other variables in the model. They can capture the geographical location of the tweet, which can be associated with regional variations in culture, politics, climate, and other factors that can affect expressed sentiment.[8] V_{ijt} is volume of tweets occurring in the state and on the same day as tweet i. By controlling for tweet volumes, we can ensure that any observed changes in expressed sentiment scores during the fall transition or between different times of day are not simply a result of changes in overall tweet activity. ϵ_{ijt} is the random error term associated to the observed S_{ijt}.

We also include a set of additional temporal controls: year (δ_{it}), weekend (W_{it}), and holiday (H_{it}) fixed effects. Year fixed effects account for potentially correlated trends in expressed sentiment in the sample, while weekend and holiday fixed effects remove statistical noise related to within-week and by-holiday variation in expressed sentiment. On the other hand, the inclusion of λ_j accounts for non-time varying unobservable factors which may influence tweet sentiment that are common to sampled tweets within the same state.

The combination of these fixed effects defines the identification strategy: we assume that deviations in sunlight exposure induced by differences in sunrise and sunset times across space and time are as good as random after accounting for unobserved variation by state and year. We argue that it is a relatively clean setting without severe confounding factors. We apply the analogous procedure to the spring transition.[9]

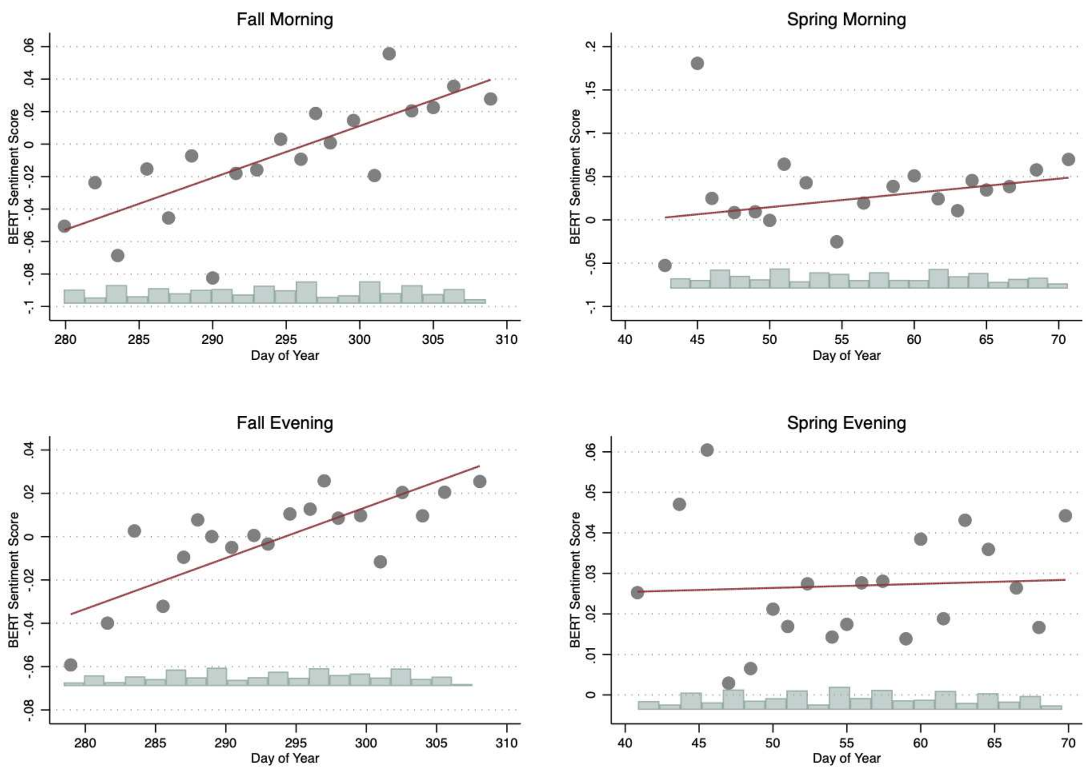
Long-Run Model

To investigate the long-term effects of DST, we use the definition of terms from Equation 2.1, but with a modification for $POST_{it}$. In this case, $POST_{it}$ takes a value of one if tweet i is made in the fourth week after the fall transition, and zero if tweet i is made in the week before. Our assumption is that sleep patterns have been restored by the fourth

[8]For example, tweets posted in the South may have different expressed sentiment scores compared to tweets posted in the North, even if they are posted during the same time of day and week of the fall transition. By controlling for latitude and longitude, we can capture these regional differences and ensure that the estimated coefficients for other variables are not biased by unobserved regional factors.

[9]In the spring transition, we expect α_1, capturing a moment of reduced sunlight and fatigue, to be negative and α_3 to be unknown because it captures a moment of sleep loss but added sunlight.

week after the transition into and out of DST, which is consistent with previous research.[10] Considering that sunlight matters independently and sleep patterns have been restored by week four, we expect α_1 to be positive, capturing a moment of added sunlight and adapted sleep pattern. Conversely, we expect α_3 to be negative, capturing reduced sunlight and the restoration of the sleep pattern.[11]



Notes: The panels located at the top and bottom left compare standardized BERT measurements for fall mornings and evenings, taken four weeks before the transition out of DST. Meanwhile, the panels at the top and bottom right compare BERT measurements for spring mornings and evenings, four weeks prior to the transition into DST.

Figure 2.5: Pre-Period Trend

[10]For example, Barnes and Wagner (2009) used the American Time Use Survey to report that Americans sleep approximately 40 minutes less on the night of the spring transition. This transition could impact sleep patterns for anywhere from two days to two weeks, depending on the individual (Valdez et al., 1997), with an average duration of about one week (Harrison, 2013).

[11]During the spring transition, we anticipate that α_1, which represents a moment of reduced sunlight and adapted sleep patterns, will be negative, and α_2, which reflects increased sunlight and the restoration of sleep patterns, will be positive. This assumption is based on the premise that sunlight has an independent effect, and by the fourth week, sleep patterns will have returned to normal.

When comparing the effect on expressed sentiment in morning and evening tweets four weeks after the transition, we encounter a potential threat to our identification. The reason for this is that there may be unobservable factors that differ between morning and evening sentiment four weeks after the transition. While the weekend and holiday fixed effects can account for within-week and holiday variation in expressed sentiment, the estimates may still be biased by confounders that affect this outcome differently between morning and evening hours. However, Figure 2.5 provides reassuring evidence that this is not the case. The pre-period trend four weeks before the transition between morning and evening hours is relatively similar up until the transition dates. Additionally, we include a histogram underneath each plot to demonstrate the distribution of the day of the year.

2.4 Results

This section presents the estimation results from Equation 2.1. Tables 2.2 and 2.3 display the short-run and long-run effects of Fall Back and Spring Forward on BERT, respectively. As previously mentioned, we examine several expressed sentiment measures: BERT, VADER, and AFINN. For both the Fall Back and Spring Forward samples, we provide separate regressions in all cases. Because the findings are consistent across measures, the remainder of the chapter focuses on results obtained using BERT.[12]

2.4.1 Short-Run Estimates

To begin, we will discuss our findings regarding the short-term impact of Fall Back on BERT. As shown in Table 2.2, we observe no significant differences between expressed sentiment in the morning hour (α_1). However, we have discovered that there is a significant negative impact on expressed sentiment during the evening hour compared to the morning hour in the week after the fall transition. Specifically, tweets posted in the evening show

[12] Our analysis reveals that the various measures of expressed sentiment yield remarkably similar results. More details are in Appendix B.

a 0.029 (or 0.022 in total effect) standard deviation reduction in expressed sentiment compared to those posted in the morning.

Table 2.2: The Short-Run Effects of Fall Back and Spring Forward on BERT

Variables	Fall Back	Spring Forward
POST (α_1)	0.007	-0.040***
	(0.007)	(0.005)
EVENING (α_2)	-0.001	0.007
	(0.008)	(0.005)
POST $\times$ EVENING (α_3)	-0.029***	0.023***
	(0.008)	(0.006)
$\alpha_1 + \alpha_3$	0.022***	-0.017***
	(0.004)	(0.003)
Constant	0.055	-0.112
	(0.105)	(0.089)
Observations	383,653	527,567
R-squared	0.003	0.002

Notes: Dependent variable: Standardized BERT. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10.

We now analyze the effects of the Spring Forward. We follow the same methodology and present results in the third column of Table 2.2. In contrast to the fall transition, however, we have observed significant differences in the expressed sentiment of morning tweets one week before and one week after the DST transition. We found that the spring transition into DST is associated with a 0.040 standard deviation decrease in expressed sentiment (α_1) during the morning hour. However, tweets posted in the evening hour show a 0.023 standard deviation smaller decrease in expressed sentiment, which results in a total effect of a 0.017 standard deviation decrease.

The direction of α_1 for both the fall transition and the spring transition is consistent with our hypothesis. We expect to see an improvement in sentiment on the morning after the fall transition because individuals get an extra hour of sleep and experience an earlier sunrise during the first week. Conversely, on the morning after the spring transition, we expect to see a decline in sentiment because individuals lose an extra hour of sleep and experience a later sunrise during the first week.

Overall, we can infer that the direction of sentiment is more negative during the evening hour in the week following the fall transition and during the morning hour in the week after the spring transition.

2.4.2 Long-Run Estimates

Our estimates in Table 2.2 reflect the aggregate effect, assuming both the effects of sleep and sunlight are present in the short-run. However, after four weeks from the transition dates, we assume that sleep patterns have been restored, and only the effect of sunlight is acting independently on sentiment. With this in mind, we will now discuss the long-term impact of the Fall Back transition on BERT. Our findings, as observed in Table 2.3, show a significant difference in the expressed sentiment of morning tweets one week before and four weeks after the DST transition. Specifically, we found that the Fall Back transition is associated with a 0.045 standard deviation increase in sentiment for morning tweets, while there is virtually no effect on the sentiment of evening tweets (a total effect of 0.001).

Our findings on the long-term impact of the Spring Forward transition on BERT reveal contrasting results, as expected. Specifically, we observed a significant difference in the expressed sentiment of morning tweets one week before and four weeks after the transition. The spring transition into DST is associated with a reduction of 0.023 in sentiment during the morning hour. On the other hand, tweets posted in the evening hour show a *small* increase in sentiment of 0.007 ($\alpha_1 + \alpha_3$) four weeks after the transition date, assuming sleep patterns have been restored. These findings support our hypothesis that the loss of morning sunlight following the spring transition attenuates sentiment, while the addition of evening sunlight improves it, albeit only marginally.

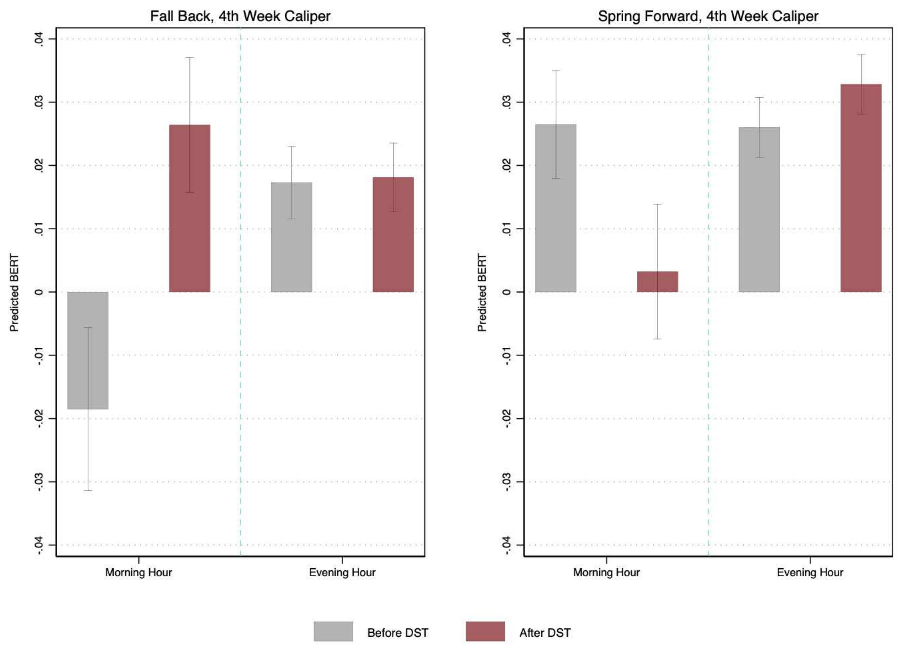
Table 2.3: The Long-Run Effects of Fall Back and Spring Forward on BERT

Variables	Fall Back	Spring Forward
POST (α_1)	0.045***	-0.023***
	(0.007)	(0.006)
EVENING (α_2)	0.036***	-0.000
	(0.008)	(0.005)
POST $\times$ EVENING (α_3)	-0.044***	0.030***
	(0.009)	(0.007)
$\alpha_1 + \alpha_3$	0.001	0.007**
	(0.004)	(0.003)
Constant	0.242**	-0.118
	(0.110)	(0.084)
Observations	353,394	516,819
R-squared	0.002	0.002

Notes: Dependent variable: Standardized BERT. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10.

To gain a deeper understanding of the results presented in Table 2.3, we plotted the predicted estimates for both the morning and evening hours before and after DST. The left panel of Figure 2.6 shows that sentiment during the morning hour following the fall transition increased significantly, by approximately 242%. Furthermore, there was a modest increase in sentiment of approximately 4% during the evening hour. On the other hand, the right panel of Figure 2.6 indicates that sentiment during the morning hour in the fourth week after the spring transition decreased significantly, by about 112%. In contrast, there was a modest increase in sentiment of about 26% during the evening hour.

To summarize, it appears that there is a more noticeable direction of sentiment during the morning hours. Specifically, there is a significant increase in sentiment during the morning hour following the fall transition when there is additional sunlight, but a significant decrease in sentiment during the morning hour in the fourth week following the spring transition when there is a loss of sunlight. Conversely, the variations in sentiment during the evening hour are comparatively minimal.



Notes: Dependent variable is standardized BERT. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Spikes indicate 95% confidence intervals.

Figure 2.6: Predicted BERT by Fall Back and Spring Forward in the Long-Run

2.5 Differential Returns to Sunlight

Our long-term findings suggest that sunlight has a more significant impact on sentiment during morning hours in comparison to evening hours. This observation raises a crucial policy question: how should sunlight be distributed? To answer this question, we estimate the following equation:

$$S_{ijt} = \beta_0 + \beta_1 POST_{it} + \beta_2 SP_{it} + \beta_3 BR_{it} + \beta_4 E_{it} + \beta_5 (BR_{it} \times E_{it})$$
$$+ \Gamma_1 W_{it} + \Gamma_2 H_{it} + \Gamma_3 LT_{ij} + \Gamma_4 LG_{ij} + \Phi V_{ijt} + \lambda_j + \delta_{it} + \varepsilon_{ijt} \quad (2.2)$$

In this case, the variable $POST_{it}$ serves as an indicator for transition dates during both the Fall Back and Spring Forward moments, which refer to the week before and the four weeks after the transitions out of and into DST. On the other hand, BR_{it} takes a value of one if tweet i occurs during a moment of sunlight with an adapted sleep pattern and zero if tweet i occurs during a moment of darkness with an adapted sleep pattern. The coefficients of interest are β_3 and β_5, which correspond to BR_{it} and $BR_{it} \times E_{it}$, respectively. Our focus is on testing whether the effect of added sunlight on sentiment varies based on whether it is during the evening or morning hours, given that we restrict to conditions where the effect of sleep is held constant. The variable SP_{it} takes a value of one if tweet i is in the spring months and zero if tweet i is in the fall months. Although we control for within-week (W_{it}) and by-holiday (H_{it}) variation in sentiment, we include SP_{it} to control for seasonal variation in daylight hours. The spring and fall months may have different sentiment patterns due to reasons unrelated to sunlight exposure, such as the onset of allergy season or changes in weather patterns. Lastly, the remaining terms are the same as in Equation 2.1.

Table 2.4: Differential Returns to Sunlight

Variables	BERT
BRIGHT (β_3)	0.035***
	(0.005)
EVENING (β_4)	0.024***
	(0.006)
BRIGHT $\times$ EVENING (β_5)	-0.015**
	(0.007)
$\beta_3 + \beta_5$	0.020***
	(0.006)
Constant	-0.017
	(0.073)
Observations	728,885
R-squared	0.002

Notes: Regression specification includes fixed effects for weekends, holidays, year, and state. Furthermore, there are additional independent variables in the regression, namely coordinates (latitude and longitude), tweet volumes, and an indicator for spring and fall months. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10.

We will now proceed to discuss our results obtained from Equation 2.2, as presented in Table 2.4. Our findings demonstrate that exposure to evening sunlight only leads to a minor increase of only 0.02 ($\beta_3 + \beta_5$) standard deviation in sentiment. Conversely, exposure to morning sunlight results in a more significant rise of 0.035 standard deviations in sentiment. As a result, our results imply that morning sunlight is more beneficial for enhancing sentiment compared to evening sunlight.

2.6 Robustness

In this section, we will conduct further robustness checks to confirm the primary assumption of our identification strategy. Our assumption is that the transition week has no correlation with the unobserved determinants of sentiment. To achieve this, we will revisit our section on differential returns to sunlight and estimate alternative specifications with varying sets of fixed effects. These checks will bolster the credibility and dependability of our findings.

The analysis comprises five distinct models presented in Table 2.5. Model I considers only state and year fixed effects. The findings indicate a positive and statistically significant relationship between sunlight and sentiment during the morning hours, leading to an increase of 0.031 standard deviations. In contrast, exposure to evening sunlight only leads to an increase of 0.002 ($\beta_3 + \beta_5$) standard deviations in sentiment.

Model II incorporates coordinates and volumes to account for potential regional disparities in sentiment and overall tweet activity. However, the estimated effects remain nearly identical. Model III adds weekend and holiday fixed effects to capture weekly variations in sentiment and changes related to holidays. However, the estimated effects remain largely unchanged. Model IV and Model V, the fourth and fifth models, respectively, include seasonal indicators and a continuous day of the year variable to account for seasonal trends. The point estimates demonstrate a positive and statistically significant relationship between sunlight and sentiment during the morning hours and a negative and

Table 2.5: Sensitivity of Differential Returns to Sunlight

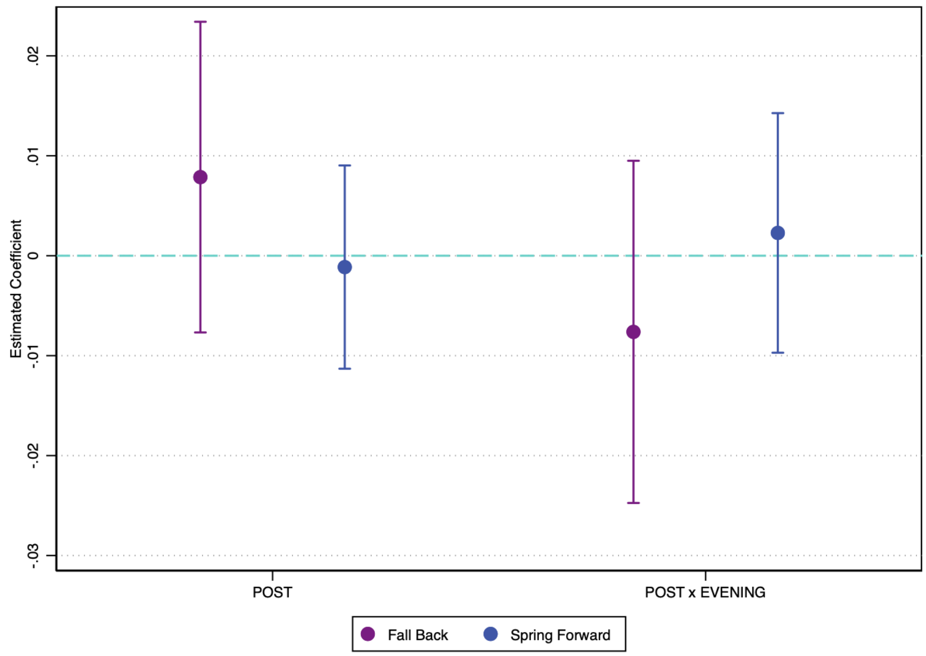
Variables	Model I	Model II	Model III	Model IV	Model V
BRIGHT (β_3)	0.031***	0.035***	0.035***	0.035***	0.035***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
EVENING (β_4)	0.019***	0.030***	0.030***	0.024***	0.023***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
BRIGHT $\times$ EVENING (β_5)	-0.029***	-0.026***	-0.026***	-0.015**	-0.015**
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Constant	-0.034***	-0.025	-0.025	-0.017	-0.032
	(0.006)	(0.073)	(0.073)	(0.073)	(0.073)
Observations	728,885	728,885	728,885	728,885	728,885
R-squared	0.002	0.002	0.002	0.002	0.002
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Weekend FE	No	Yes	Yes	Yes	Yes
Holiday FE	No	No	Yes	Yes	Yes
Season	No	No	No	Yes	No
Day of Year	No	No	No	No	Yes
Coordinates	No	Yes	Yes	Yes	Yes
Volumes	No	Yes	Yes	Yes	Yes

Notes: The dependent variable in our model is the standardized BERT. As the seasonal indicator and day of the year variables are correlated, we included them separately in the model. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10. Model IV is our main specification

statistically significant relationship between sunlight and sentiment during the evening hours, consistent with Models I, II, and III results.

Overall, the results presented in Table 2.5 demonstrate that our primary findings remain robust with the inclusion of additional fixed-effects. Therefore, any potential overlaps between the weekend, holidays, seasonal trends, and the transition do not bias our conclusions.

In the realm of causal inference, it is often advisable to estimate the impact of a treatment that is assumed to have no effect under the identification hypothesis. This estimation can lend support to the identifying assumption (Imbens, 2004), but does not necessarily prove identification. To avoid the possibility that our findings are merely a statistical fluke, we generate a placebo treatment by pretending that the transition took place one week



Notes: Dependent variable is standardized BERT. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Spikes indicate 95% confidence intervals.

Figure 2.7: The Effect of Pseudo-DSTs on BERT

prior to and after the actual transition date. According to our hypothesis, this placebo treatment should have no effect on sentiment during either the morning or evening hours. If the placebo estimates are statistically significant, then it is probable that our estimates are capturing a systematic correlation between the unobserved determinants of sentiment and time. One possible unobserved determinant of sentiment that may also correlate with the transition week is the weather (Connolly, 2013; Kämpfer and Mutz, 2013). Importantly, estimating the effect of a placebo treatment will also reveal any systematic weather effects, if they exist. The outcomes of the placebo regression analysis presented in Figure 2.7 provide robust backing for our identification strategy. The estimates for both morning (α_1) and evening (α_3) periods indicate that the placebo treatment is significantly indistinguishable from zero.

2.7 Conclusion

This chapter leverages the quasi-experimental nature of DST to evaluate the impact of sunlight exposure on sentiment. We find that an increase in sunlight exposure in the morning significantly enhances sentiment during the fall transition. However, in contrast, a significant decrease in sentiment is observed during the morning hour following the spring transition, when there is a loss of sunlight. In addition, we find that morning sunlight exposure has a greater beneficial effect on sentiment than exposure to sunlight in the evening.

Our study is one of the few causal studies that examine the impact of sunlight exposure on sentiment, as exogenous shifters of sunlight are rare in real-world settings. To identify the effects, we have constructed a dataset that is spatially rich and includes sentiment data from the online social media platform Twitter in the U.S. Investigating the impact of sunlight reallocation between morning and evening on sentiment necessitates having statistically powerful data. These data are crucial for estimating econometric specifications that are rich and consider both weekend and holiday effects as well as specific seasonal adjusters.

Our findings have significant implications for public policy as they reveal that the positive impact of sunlight on sentiment is greater in the morning than in the evening. This aligns with previous research showing that exposure to an hour of natural light in the morning can help improve sleep quality. Sunlight plays a vital role in regulating our circadian rhythm by signaling the body to increase or decrease melatonin levels. Thus, increasing daylight exposure can improve the body's ability to produce melatonin when it is time to sleep, thereby enhancing overall well-being (Figueiro et al., 2017; Mead, 2008). Our results are relevant to the ongoing debate on whether to retain or abolish daylight savings. The positive effect of morning sunlight on sentiment highlights the overlooked benefits of human well-being. Therefore, shifting to darker mornings and brighter evenings, as proposed in the Sunshine Protection Act, may do more harm than good.

The primary objective of this chapter is to provide evidence that supports a causal relationship between DST transitions, well-being, and sunlight. However, the chapter is not intended to draw any definitive conclusions about the general welfare effects of DST. Understanding the regional disparities in the benefits of morning sunlight can help us identify which areas experience greater advantages than others. For example, recent research suggests that individuals residing in areas with shorter days and longer nights during winters are more prone to mood disorders such as depression, as compared to those residing in sunnier regions (Van de Vliert and Rentfrow, 2021). Thus, the benefits of morning sunlight may be particularly vital in these regions than in those where sunlight is abundant all year round. Further research can delve into this inquiry to gain deeper insights.

Chapter 3

The Compensation of Conscience: Evidence from the U.S. Labor Market

3.1 Introduction

> Honour makes a great part of the reward of all honourable professions. In point of pecuniary gain, all things considered, they are generally under-recompensed . . . Disgrace has the contrary effect. The trade of a butcher is a brutal and an odious business; but it is in most places more profitable than the greater part of common trades. The most detestable of all employments, that of public executioner, is, in proportion to the quantity of work done, better paid than any common trade whatever.
>
> **Adam Smith**

> Finally there came a time when everything that men had considered as inalienable became an object of exchange, of traffic and could be alienated. This is the time when the very things which till then had been communicated, but never exchanged, given but never sold, acquired but never bought: virtue, love conviction, knowledge, conscience– when everything passed into commerce.
>
> **Karl Marx**

Compensating wage differentials are an important concept in labor economics that help to explain why some workers receive higher wages than others for jobs that require similar skills and productivity. These wage differences are often due to differences in job disamenities, such as unpleasant working conditions, long hours, or exposure to hazards. Workers who face these disamenities may require higher wages to compensate them for the discomfort or risk associated with performing such jobs.

The labor theory of compensating differentials, which traces its roots back to Adam Smith's seminal work, *"An Inquiry into the Nature and Causes of the Wealth of Nations"*, is built on the standard neoclassical framework. This theory highlights the influence of job disamenities on wage determination. In a competitive labor market, wages are determined by workers' marginal productivity. If two workers have the same productivity, but one faces more disamenities than the other, that worker may require a higher wage to accept the job. From the standpoint of the employer, a higher wage offer might be necessary to induce a worker to assume an undesirable job. This higher wage acts as a compensating differential to offset the negative aspects of the job and ensure that workers are indifferent between different job opportunities.

Several studies have found that disamenities such as physical hazards, exposure to dangerous materials, irregular shift schedules, job stress, and dirty jobs are associated with relatively high levels of monetary compensation to offset their negative impact on a worker's overall utility (Smith, 1979; Olson, 1981; Arnould and Nichols, 1983; Leeth and Ruser, 2003; Garen, 1988; Lanfranchi et al., 2002; French and Dunlap, 1998; Villanueva, 2007). In addition to physical and material disamenities, moral or ethical disamenities can also affect workers' utility and job satisfaction. For example, a Catholic nurse who holds strong beliefs that contraception is morally wrong, being asked to oblige a prescription for contraception may create psychological or moral discomfort. The nurse may feel that they are violating their conscience and acting against their moral beliefs. This can lead to feelings of guilt, conflict, and emotional distress. Like physical disamenities, moral disamenities may lead to compensating wage differentials, where workers who face moral disamenities receive higher wages to compensate them for the psychological discomfort or emotional distress they experience.

Prior research has focused mainly on physical and material disamenities and has not explored the role of moral or ethical disamenities in the labor market (e.g., Smith, 1979; Olson, 1981; Arnould and Nichols, 1983; Leeth and Ruser, 2003; Garen, 1988; Lanfranchi et al., 2002; French and Dunlap, 1998; Villanueva, 2007). In this study, we aim to address this gap by examining whether jobs that demand workers to compromise their moral values offer higher compensation to offset the disamenities that contradict their moral beliefs. Our study uses data from the National Longitudinal Survey of Youth 1997 (NLSY97) and the Occupational Information Network (O*NET) job descriptor to develop a continuous measure of a moral index across occupations. Our analysis benefits from a long and rich panel from 1997-2017, which allows us to obtain a comprehensive understanding of the relationships between moral disamenities and compensating wage differentials.

By shedding light on this important issue, this chapter aims to provide insights on the price of conscience. Compensating wage differentials reflect the economic reality that workers who face disamenities in their jobs, including moral disamenities, may require higher wages to compensate them for the negative aspects of their work. However, from a philosophical perspective, the idea of compensating someone for a moral disamenity raises questions about the nature and value of morality. If morality is seen as an inherent aspect of human dignity or a fundamental principle of justice, then it may seem inappropriate to reduce it to a monetary value. Furthermore, the practice of compensating workers for moral disamenities may contribute to a broader trend of treating ethical values as commodities to be bought and sold. This commodification of morality may be seen as a threat to the intrinsic value of moral principles and the integrity of the individual who holds them. Our study is not intended to provide a definitive answer to these philosophical questions. Instead, our aim is to contribute to the empirical literature on compensating wage differentials by examining the relationship between moral disamenities and wages. Our hope is that this research will stimulate further discussion and debate about the nature and value of morality in the labor market and society at large.

Based on our findings, it appears that jobs requiring workers to compromise their moral sense of right and wrong come with higher hourly compensation. This implies that there is a compensating differential to make up for the disamenities of going against one's moral values. Additionally, we noticed that individuals with a college education tend to earn more in these types of jobs, indicating that they may have greater bargaining power and more employment opportunities that influence their compensation preferences. Moreover, our research provides evidence for an asymmetric relationship between changes in the occupational moral index and total hourly compensation. This relationship seems to be sensitive to the degree of moral compromise required by the job.

The chapter is structured in the following manner: In Section 3.2, we delve into the conceptual model. In Section 3.3, we outline our data and empirical strategy. In Section 3.4, we present and discuss our empirical results and include several sensitivity checks that confirm the robustness and consistency of our findings, along with an analysis of heterogeneous effects. Additionally, in Section 3.5, we provide a detailed examination of the main results, specifically looking at the asymmetric effects of the relationship between moral disamenities and compensations. Finally, Sections 3.6 and 3.7 discuss the limitations of our study and summarize the main findings, respectively. Moreover, Section 3.7 will offer potential avenues for future research.

3.2 Conceptual Model

This section discusses the theoretical foundation of our reduced form empirical model, which takes into account the factors that influence a worker's labor supply decision, including the wage offered and the presence of non-pecuniary disamenities in a job. In our case, non-pecuniary disamenities refer to tasks that may require workers to compromise their moral values, leading to psychological discomfort and a sense of moral conflict in our context.

Following Villanueva (2007), we assume the presence of search frictions in the job market, workers cannot observe all available job positions. Therefore, we define each job based on two factors: wage level (w) and the level of non-pecuniary disamenity (M). A job with $M = 1$ involves the highest level of moral compromise, while a job with $M = 0$

is considered a "no moral compromise" job. In our model, a worker has a probability of receiving a job offer (w, M) with an exogenous probability of arrival. The preferences of a worker, indexed by i, are represented by the following utility function:

$$u_i(w, M) = w - Z_i M$$

where Z denotes the worker-specific marginal willingness to pay to avoid consuming jobs that require compromising their moral sense of right and wrong. Workers with a lower tolerance for moral compromise tasks tend to have a higher value for Z.

We make the assumption that the offers received by each worker are independent of their marginal willingness to avoid disamenities. This assumption is supported by the fact that firms cannot observe Z and therefore cannot target their offers to a specific type of worker.[13] In our context, we follow Villanueva (2007) by introducing an unobserved, individual-specific component (α). Thus, the equilibrium wage can be represented as:

$$w = \beta M + \alpha + \epsilon \tag{3.1}$$

where β represents the market's monetary return for the existence of a disamenity, while ϵ represents the difference between the productivity of a worker at a specific job and the average wage earned by workers with similar unobserved factors (e.g., ability, job condition). It is defined that $E(\epsilon|M, \alpha) = 0$. Assuming that job offers are unrelated to Z, it follows that $E(\Delta\epsilon|Z_i = 0)$ (the number of job offers received by a worker is not influenced by their level of tolerance).

The parameter, denoted by β, measures the market's return for the presence of a disamenity on a job. Specifically, this chapter examines the wage premiums that relate to

[13] Hwang et al. (1998) demonstrated that in a labor market with search frictions, a constant arrival rate of offers, and heterogeneity in the firm's cost of providing amenities, an equilibrium distribution of offer wages and disamenities exists in the market. This equilibrium distribution includes a distribution of wages for each level of amenities offered by firms.

workplace disamenities, which refer to the psychological discomfort experienced by workers when they are asked to compromise their moral sense of right and wrong. It is crucial to evaluate whether wage differentials competitively compensate for the disamenities that workers experience on the job, and β plays a critical role in this analysis. It is important to note that this chapter does not attempt to estimate Z_i.

3.3 Data and Empirical Strategy

3.3.1 Data

We utilized data from the National Longitudinal Survey of Youth 1997 (NLSY97), specifically from the 1997-2017 waves. The NLSY97 is a nationally representative sample of 8,984 individuals who were born between 1980 and 1984. This means that respondents were between the ages of 13 and 17 during the initial round of interviews in 1997, and between 33 and 37 years old during the most recently observed round in 2017. We record and track all instances of employment for each individual in the NLSY97 using unique job identifiers, which enables us to capture multiple employment observations within a single year. The average number of job-year observations per individual is approximately 19 while the average number of unique jobs held per individual is 9.

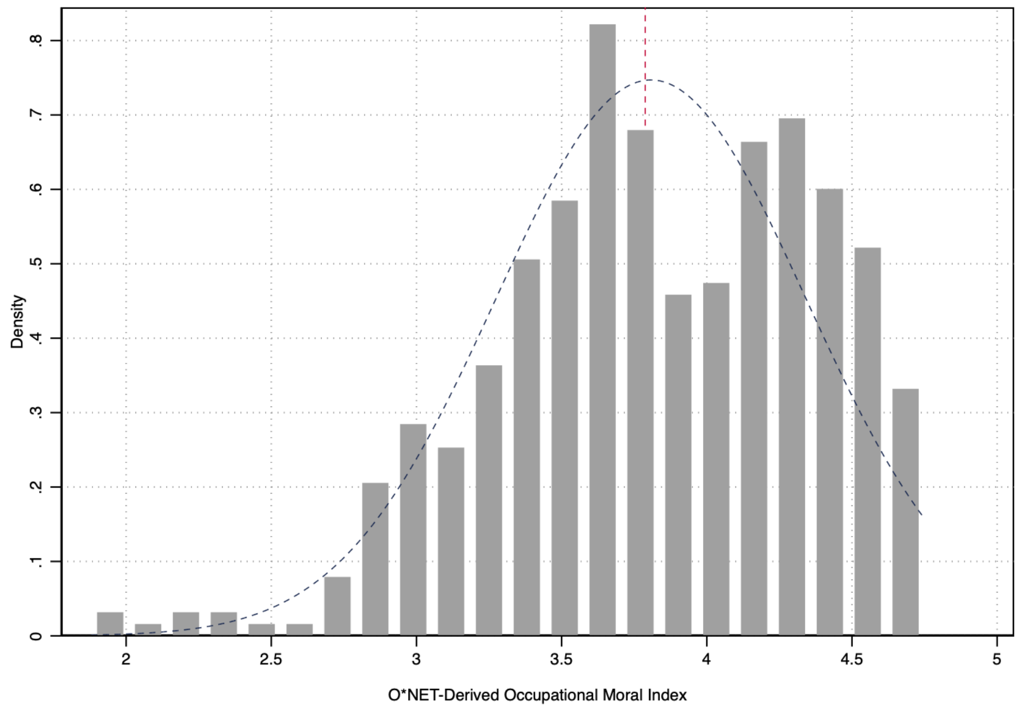
There are several advantages to using this data. Firstly, the NLSY97 dataset offers an advantage over other longitudinal samples, such as the Current Population Survey Outgoing Rotation Group (CPS-ORG) monthly earning files, as it allows for precise observation of the labor supply and employment history of individuals over their entire lifespan. Another advantage is the ability to track all instances and histories of employment through unique job identifiers.

The variable of interest for the outcome is the natural logarithm of real hourly compensation, which includes overtime, tips, bonuses, and other forms of compensation, in addition to the worker's hourly pay rate. The explanatory variable of interest is the moral index across occupations, which we measure using the O*NET Database 12.0 job descriptor data. We calculate a continuous moral index measure across occupations using the working definition that asks incumbents of occupations: "*Workers on this job are never pressured to do things that go against their sense of right and wrong.*"[14] One significant advantage of our NLSY97/O*NET dataset is that it provides a continuous measure of the moral index for all occupations, along with detailed job tasks, skill requirements, and working conditions. This comprehensive data enables us to obtain more accurate estimates of compensation differentials associated with the moral demands of occupations in the U.S. labor market.

Currently, the O*NET data align with the 2018 Standard Occupational Classification (SOC) occupations and codes, whereas the occupation data in the NLSY97 is coded based on the 2002 Census classification. To combine the O*NET-derived moral index with the NLSY97 sample, we utilize the crosswalk provided by the Integrated Public Use Microdata Series (IPUMS), as described by Ruggles et al. (2020). After utilizing the crosswalk to align 2002 Census occupations with 2018 SOC O*NET occupations, the resulting set is consolidated based on 2002 Census occupations through mean assignment. For instance, the category "Physicians and Surgeons" is assigned the mean of O*NET indicators across various specialties such as Neurologists, Obstetricians, Pediatricians, Hospitalists, Urologists, and others.

The density plot, shown in Figure 3.1, illustrates the distribution of 485 unique occupations in the NLSY97 sample across the domain of the O*NET-derived moral index. The peak of the distribution is located slightly to the right of center, with some clustering towards the higher end of the moral index domain. Table 3.1 shows the top and bottom 25

[14] O*NET does not explicitly define the word "*pressure*". Therefore, the word "*pressure*" in this context could potentially be ambiguous. It is used to describe the situation where workers are asked to do things that go against their sense of right and wrong. However, the exact nature of this pressure is not explicitly defined. The term "*pressure*" can encompass various scenarios, such as explicit coercion, subtle influence, or even internal psychological conflict. Without further clarification or specific examples, it is challenging to ascertain the precise nature and degree of pressure that workers may experience in different occupations. The ambiguity surrounding the term "*pressure*" could impact the interpretation and generalizability of our findings. Different individuals may have different thresholds for what they consider as pressure, and their responses to the moral index measure may vary accordingly. Furthermore, the interpretation of the relationship between compensation differentials and moral demands may depend on how individuals perceive and respond to different types and levels of pressure.



Notes: Raw scores of O*NET-Derived Occupational Moral Index.

Figure 3.1: Distribution of NLSY97 Occupations Across the O*NET-Derived Moral Index

NLSY97 occupations ranked by O*NET-derived moral index. Raw scores range from 1 to 5, with 5 meaning a typical worker is never asked to compromise their sense of right and wrong. The observed average moral index is 3.72 and the standard deviation is 0.49.

We incorporate five additional control variables, obtained from the O*NET data, along with the moral index, to account for various occupation-specific factors that may impact hourly compensation.[15] The first control variable is the care index, derived through Principal Component Analysis (PCA), which encompasses multiple indicators that assess the level of care required in an occupation. Specifically, we collect ratings for the importance of the activities labeled "assisting and caring for others" and "concern for others". The second control variable is the hazard index, also derived through PCA, which encompasses several indicators that measure the risk of bodily harm, such as exposure to contaminants,

[15]For the correlations between the moral index and these variables, please see Appendix C.

Table 3.1: Top and Bottom 25 NLSY97 Occupations Ranked by O*NET-Derived Moral Index

Rank	Occupation	Score	Occupation	Score	Rank
1	Network Systems and Data Analysts	1.88	Crushing, Grinding, Polishing, Mixing, and Blending Workers	4.58	461
2	Network and Systems Administrators	1.88	Heat Treating Equipment Setters, Operators, and Tenders	4.58	462
3	Lawyers	2.12	Textile, Apparel, and Furnishings Workers	4.59	463
4	Agents and Business Managers of Performers	2.25	Molders and Molding Machine Setters, Operators, and Tenders	4.62	464
5	Gaming Services Workers	2.25	Paper Goods Machine Setters, Operators, and Tenders	4.62	465
6	Public Relations Managers	2.37	Textile Cutting Machine Setters, Operators, and Tenders	4.62	466
7	Public Relations Specialists	2.37	Food Cooking Machine Operators and Tenders	4.62	467
8	Social Workers	2.50	Extruding and Drawing Machine Setters, Operators	4.62	468
9	Gaming Cage Workers	2.56	Miscellaneous Assemblers and Fabricators	4.62	469
10	News Analysts, Reporters and Correspondents	2.69	Tire Builders	4.62	470
11	Editors	2.75	Lathe and Turning Machine Tool Setters, Operators	4.62	471
12	Urban and Regional Planners	2.75	Pile-Driver Operators	4.62	472
13	Construction and Building Inspectors	2.75	Milling and Planning Machine Setters, Operators, and Tenders	4.62	473
14	Private Detectives and Investigators	2.75	Textile Knitting and Weaving Machine Setters, Operators	4.62	474
15	Credit Counselors and Loan Officers	2.87	Cutting, Punching, and Press Machine Setters, Operators	4.62	475
16	Bill and Account Collectors	2.87	Textile Winding, Twisting, and Drawing Out Machine Setters	4.62	476
17	First-Line Supervisors of Gaming Workers	2.87	Cleaning, Washing, and Metal Pickling Equipment Operators	4.62	477
18	Gaming Managers	2.87	Grinding, Lapping, Polishing, and Buffing Machine Tool Setters	4.62	478
19	Advertising and Promotions Managers	2.87	Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	4.66	479
20	Door-to-Door Sales Workers and Vendors	2.87	Plating and Coating Machine Setters, Operators, and Tenders	4.72	480
21	Telemarketers	2.87	Extruding and Forming Machine Setters, Operators	4.75	481
22	Tax Examiners and Collectors	2.87	Shoe Machine Operators and Tenders	4.75	482
23	Financial and Investment Analysts	2.87	Multiple Machine Tool Setters, Operators, and Tenders	4.75	483
24	Clergy	2.87	Data Entry Keyers	4.75	484
25	Chief Executives	2.88	Sewing Machine Operators	4.75	485

hazardous conditions, hazardous equipment, high places, minor burns, cuts, bites, or stings, and whole body vibration. The third control variable is the skills index, derived through PCA, which captures the basic skill competencies required across different occupations, including reading comprehension, active listening, writing, speaking, mathematics, and science. The fourth control variable is the social index, derived through PCA, which includes indicators that capture occupations where workers perform tasks for other people or have personal connections with others on the job, such as social service and social orientation. Lastly, the autonomy index measures worker supervision. To aid in the interpretation of our empirical estimates, we scale the generated indices from zero to one using min-max standardization.[16]

In Table 3.2, the means, standard deviations, minimums, and maximums for the major variables of the NLSY97 sample are presented. The average raw real compensation is approximately $18.3 per hour, with the lowest recorded at around $3.5 per hour and the highest at approximately $78.2 per hour. Additionally, around 13.1% of the respondents in

[16]Formula: $x' = \frac{x-\min(x)}{\max(x)-\min(x)}$, where x is an original value and x' is the normalized value.

our sample reported being enrolled in K-12, college, or a graduate program. The average years of employment experience is roughly 7.25 years, and the average cumulative number of 6+ week not-in-the-labor-force (NILF) spells is about 10. About 9.7% of the respondents reported being in a union, and 3.3% reported being self-employed. Lastly, our sample comprises approximately 57.6% men and 24.5% blacks.

Table 3.2: Sample Characteristics for Major Variables

	Mean	SD	Min	Max	N
Compensation					
Hourly Compensation (Real 2020 $)	18.336	10.003	3.507	78.270	72,588
Moral Measure					
Occupational Moral Index	3.716	0.488	1.880	4.750	72,544
Human Capital and Labor Supply					
Enrollment Status	0.131	0.338	0.000	1.000	72,588
Highest Grade Completed	13.058	2.699	0.000	20.000	72,588
Degree					
None	0.129	0.335	0.000	1.000	72,588
GED	0.107	0.309	0.000	1.000	72,588
HS Diploma	0.508	0.500	0.000	1.000	72,588
Associate	0.052	0.223	0.000	1.000	72,588
Bachelor	0.165	0.371	0.000	1.000	72,588
Master	0.031	0.174	0.000	1.000	72,588
PhD	0.002	0.045	0.000	1.000	72,588
Professional (DDS/JD/MD)	0.006	0.077	0.000	1.000	72,588
Number of 6+ Week NILF Spells	10.040	5.070	0.000	23.000	72,588
Years of Employment Experience	7.253	4.540	0.000	22.500	72,588
Other Job Characteristics					
Job Tenure (in Years)	2.245	2.699	0.000	27.596	72,588
Union Status	0.097	0.295	0.000	1.000	72,588
Self-Employed	0.033	0.179	0.000	1.000	72,588
Usual Hours per Week	42.736	7.603	40.000	168.000	72,588
Skills Index	0.263	1.443	-2.554	4.036	72,588
Hazard Index	0.388	2.445	-2.515	8.443	72,588
Care Index	-0.135	1.654	-4.607	4.768	72,588
Social Index	-0.197	1.331	-3.617	3.219	72,588
Autonomy Index	2.888	0.788	1.370	4.750	67,823
Demography					
Men	0.576	0.494	0.000	1.000	72,588
Black	0.245	0.430	0.000	1.000	72,588

Notes: Raw scores of moral, skills, hazard, care, social, and autonomy indices are presented. For the regression analysis, we will scale theses indices from zero to one using the minimum value.

3.3.2 Empirical Strategy

3.3.3 Level Compensation

Our goal is to estimate the parameter β outlined in Equation 3.1. However, in Equation 3.1, we have defined M as an indicator variable, while in our dataset, our measure of M is a continuous variable. As a result, lower values of M indicate higher levels of moral disamenity. Therefore, our research question can be restated as follows: What are the compensating wage differentials for various levels of occupational moral index, from the occupation with the highest level of moral disamenity to the occupation with the lowest level of disamenity? In essence, we are interested in determining how compensating wage differentials differ across occupations with varying levels of moral disamenity. To achieve this goal, we exploit the panel structure of the NLSY97 dataset by reformulating Equation 3.1 and estimating a semi-log model, which is a common practice in the compensating wage differentials literature.

Consider the simplest approach to examining the relationship between hourly compensation and the occupational moral index:

$$lnW_{ijt} = \beta M_{ijt} + \epsilon_{ijt} \tag{3.2}$$

Here, lnW_{ijt} represents the natural log of real hourly compensation for workers aged 18 years and older, denoted by i in job j at time t. M_{ijt} describes the occupational moral index. However, this regression may not yield a consistent estimate of the effect of the occupational moral index if a worker's characteristics are determinants of both hourly compensation and the propensity to select a job with a low or high moral index. Notably, observable worker's characteristics, such as work experience and education, are likely to influence both the type of occupation and a worker's hourly compensation. Failure to control for these omitted worker's characteristics may result in detecting a relationship between the occupational moral index and hourly compensation, even when no actual effect is present. To partially resolve this issue, a regression that includes observed worker's characteristics (X_{ijt}) can be used:

$$lnW_{ijt} = \beta M_{ijt} + X'_{ijt}\gamma + \epsilon_{ijt} \tag{3.3}$$

In this model, X_{ijt} is a vector of time- and job-varying controls that attempt to capture other potential changes in observables that may co-determine hourly compensation. These controls include measures of human capital, labor supply, and other job characteristics such as job tenure, union status, self-employment, usual hours per week, percentile rank on the ASVAB standardized test, and occupation-specific care, hazard, skills, and social indices as specified in Table 3.2. However, even with Equation 3.3, a consistent estimate of the effect of the occupational moral index on hourly compensation may not be generated if unobserved worker's characteristics are determinants of both hourly compensation and the occupational moral index. Specifically, there may be concerns that a worker's unobserved propensity to select a job with a low or high moral index may underlie the observed change in hourly compensation. To address this issue, we include unobserved worker's characteristics (α_i) and estimate the following equation:

$$lnW_{ijt} = \beta M_{ijt} + X'_{ijt}\gamma + \alpha_i + \epsilon_{ijt} \tag{3.4}$$

We further include a set of survey year dummies (δ_t) to control for time-varying shocks that are common across all workers, such as potential changes in the federal minimum wage or wage shocks resulting from large-scale economic downturns (Clemens and Wither, 2019). Lastly, λ_{ijt}, σ_{ijt}, and ω_{ijt} denote industry, region, and urban dummies, respectively. The inclusion of industry, region, and urban dummies in the model serves several purposes. First, it helps to control for unobserved heterogeneity across different industries. By including a set of industry dummies (λ_{ijt}), the model accounts for any industry-specific factors that may affect hourly compensation, such as differences in labor market conditions, demand for certain skills, or industry-specific regulations. This allows for a more accurate estimation of the relationship between hourly compensation and the occupational moral index, as it helps to isolate the effects of occupation-specific moral considerations from industry-level factors. Second, a set of region dummies (σ_{ijt}) controls for regional heterogeneity that may affect compensation. Factors such as cost of living, regional labor market conditions, and local economic conditions can vary across different regions, and including these dummies helps to account for these regional differences, allowing for a more precise estimation of the relationship between hourly compensation and the occupational moral index. Finally, a set of urban dummies (ω_{ijt}) controls for differences in compensation between urban and rural areas. Urban areas may have higher wages due to factors such as higher cost of living, higher demand for labor, or differences in labor market institutions. By including these dummies, the model accounts for these differences and helps to isolate the effects of the occupational moral index on hourly compensation, while controlling for urban-rural wage differences. We assume the error term (ϵ_{ijt}) is independent and identically distributed across workers and independent across survey waves. Therefore, our preferred specification can be formally represented as follows:

$$lnW_{ijt} = \beta M_{ijt} + X'_{ijt}\gamma + \delta_t + \alpha_i + \lambda_{ijt} + \sigma_{ijt} + \omega_{ijt} + \epsilon_{ijt} \tag{3.5}$$

3.3.4 Change in Compensation

In our sample, we observed that workers switch occupations and industries, and as a result, the impact of X_{ijt} on wage changes could vary for job switchers and non-switchers. As a robustness check for our main specification in Equation 3.5, we first-difference Equation 3.5 with respect to time. By doing so, we eliminated the effect of unobserved worker characteristics (α_i), resulting in the following estimation equation:

$$\Delta lnW_{ijt} = \Delta M_{ijt}\beta + \Delta X'_{ijt}\gamma + \Delta\epsilon_{ijt} \tag{3.6}$$

Here, Δ is the time difference operator. Note that the first-difference operator removes both the unobserved and all other invariant characteristics. Therefore, if we are to use this estimation strategy to identify the effect of occupational moral index on hourly compensation, then the data must exhibit time series variation in a worker's moral index across occupations. Since occupational moral index do not change[17], such time series variation in the level of occupational moral index can only arise if workers switch jobs. Provided enough workers switch and that initial and final occupational moral index levels are sufficiently different, then "switchers" will generate sufficient time series variation to identify the effect of occupational moral index on hourly compensation.

Equation 3.6 provides a useful robustness check on the results obtained in Equation 3.5 by limiting our sample exclusively to job switchers (workers changing both occupation and industry). However, it is important to note that with Equations 3.5 and 3.6, we cannot rule out bias due to possible correlation between time-varying unobservable heterogeneity and occupational moral index.

3.4 Results

To begin, we use Equation 3.5 to estimate the relationship between hourly compensation and the occupational moral index for all workers, as shown in Table 3.3. We present three different specifications: (1) without controlling for observed and unobserved characteristics of workers; (2) controlling only for observed characteristics of workers; and (3) controlling for both observed and unobserved characteristics of workers. Additionally, for specifications (2) and (3), we include a set of year, region, and urban dummies, as explained previously.

Next, we use a sample that identifies changes in compensation exclusively among job switchers (workers changing both occupation and industry) by estimating Equation

[17] O*NET updates ratings for occupations on a rolling basis; it takes several years for all occupations to have revised ratings.

3.6 as presented in Table 3.4. We provide three different specifications: (1) without any controls; (2) with all controls except age and race/ethnicity variables; and (3) including all controls. The next two subsections check the robustness of our results and examine the heterogeneous impacts among workers with and without college education.

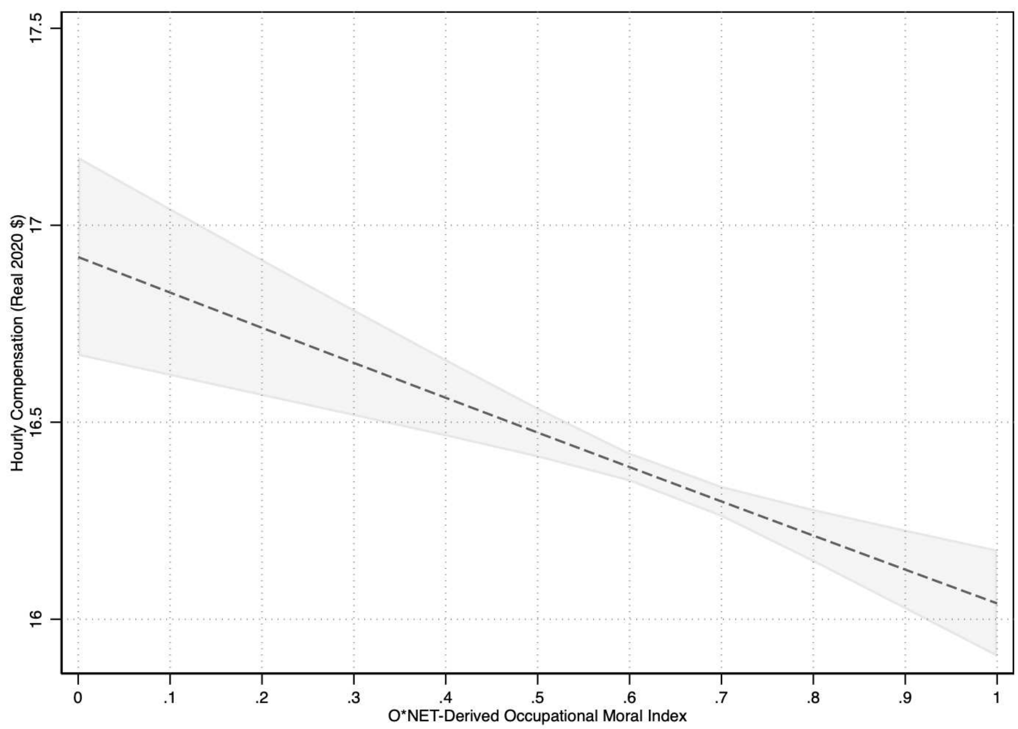
3.4.1 Level Compensation

Table 3.3: Parameter Estimates of Occupational Moral Index vis-a-vis Natural Log of Total Hourly Compensation

Variables	(1)	(2)	(3)
Occupational Moral Index	-0.519***	-0.115***	-0.053***
	(0.010)	(0.011)	(0.012)
Constant	3.126***	2.106***	2.145***
	(0.007)	(0.032)	(0.030)
Observations	72,544	72,544	72,544
R-squared	0.037	0.452	0.679
Year FE	No	Yes	Yes
Region FE	No	Yes	Yes
Urban FE	No	Yes	Yes
Industry FE	No	Yes	Yes
Demography Controls	No	Yes	No
Human Capital and Labor Supply Controls	No	Yes	Yes
Job Characteristics Controls	No	Yes	Yes
Person FE	No	No	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable is ln(real hourly compensation). Occupational Moral Index is scaled from zero to one using min-max standardization.

The presence of a compensation penalty is evident in Table 3.3. The estimated coefficient of the occupational moral index, which varies from 0 to 1, indicates the impact on hourly compensation when workers shift from jobs that frequently require them to compromise their moral compass to jobs that are less likely to require such compromise. We interpret these estimates as a movement across the interdecile range of the occupational moral index, specifically from 0.41 to 0.85.[18] For example, in column (3), when accounting for observed and unobserved worker characteristics, non-moral-related job characteristics, and fixed-effects such as year, region, urban, and industry, the coefficient estimate of -0.053 indicates that a movement across the interdecile range of the occupational moral index is associated with a 2.3 percent reduction in hourly compensation, holding all other factors constant.[19]



Notes: The shaded area on the graph represents the 95% confidence interval for the predicted hourly compensation. The estimates from column (3) in Table 3.3 are used for this prediction. Occupational Moral Index is scaled from zero to one using min-max standardization.

Figure 3.2: Predicted Association between Occupational Moral Index and Total Hourly Compensation

We utilized Equation 3.5 to make predictions regarding hourly compensation, using the estimates from column (3) of Table 3.3. Our analysis, as depicted in Figure 3.2, demon-

[18]The interdecile range is the difference between the first and the ninetieth deciles (10% and 90%).

[19]Percent Difference = (Coefficient Estimate × Interdecile Range of Occupational Moral Index) × 100.

strates that predicted hourly compensations decrease for occupations with lower moral compromise, assuming that other factors remain constant. This suggests that jobs involving moral compromise offer higher compensation compared to those with less moral compromise.

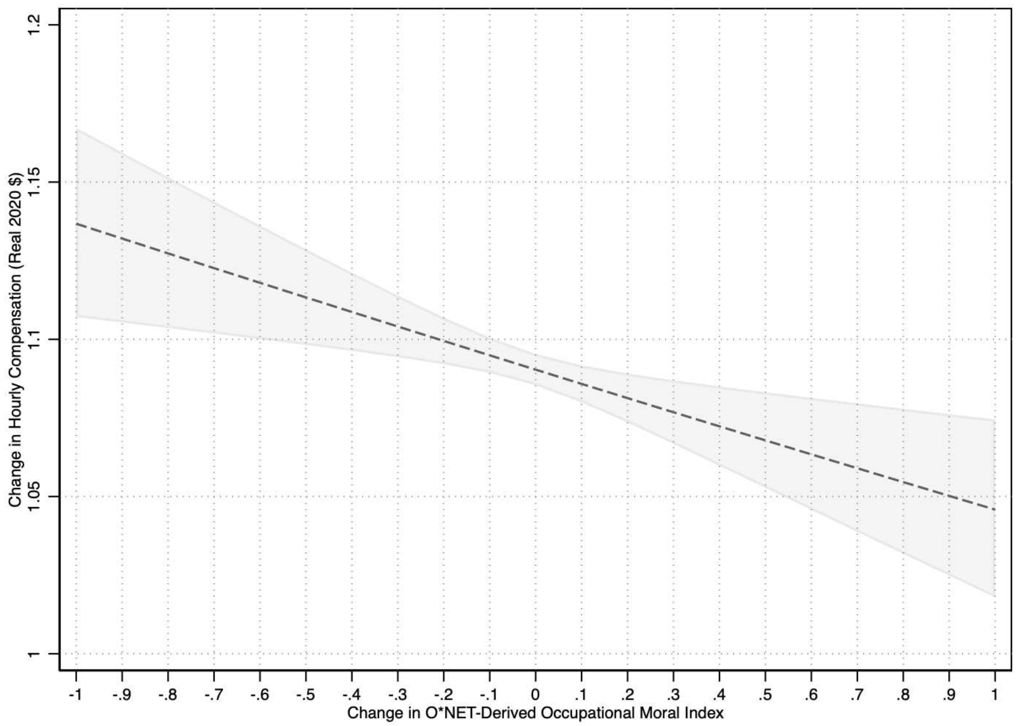
3.4.2 Change in Compensation

Table 3.4: Parameter Estimates of Change in Occupational Moral Index vis-a-vis Change in Natural Log of Total Hourly Compensation—Industry/Occupation Switchers Only

Variables	(1)	(2)	(3)
Δ Occupational Moral Index	-0.031**	-0.042***	-0.042***
	(0.012)	(0.013)	(0.013)
Constant	0.087***	-0.014	-0.003
	(0.002)	(0.061)	(0.061)
Observations	29,762	29,762	29,762
R-squared	0.000	0.085	0.086
Year FE	No	Yes	Yes
Region FE	No	Yes	Yes
Urban FE	No	Yes	Yes
Industry FE	No	Yes	Yes
Demography Controls	No	No	Yes
Δ Human Capital and Labor Supply Controls	No	Yes	Yes
Δ Job Characteristics Controls	No	Yes	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. Δ Human Capital and Labor Supply Controls consist of changes in college educated, union membership, self-employment status, years of employment experience. Δ Job Characteristics Controls include changes in O*NET-derived indices. Dependent variable is Δln(real hourly compensation). Δ Occupational Moral Index is scaled from zero to one using min-max standardization.

The results of the analysis on the change in compensation, as presented in Table 3.4, closely align with our findings from the analysis on compensation level, as shown in Table 3.3. By comparing the change in the coefficient of the occupational moral index in column (3) of Table 3.3 with those in all columns of Table 3.4, we observe a similar magnitude. Specifically, across the interdecile range of change in the occupational moral index (-0.16 to 0.13), the coefficient estimate of -0.042 in column (3) of Table 3.4 indicates that moving from an occupation where workers are often asked to compromise their sense of right and wrong to an occupation where workers are less likely to do so is associated with a 1.22% reduction in hourly compensation change, holding all other factors constant.



Notes: The shaded area on the graph represents the 95% confidence interval for the predicted change in hourly compensation. The estimates from column (3) in Table 3.4 are used for this prediction. Δ Occupational Moral Index is scaled from zero to one using min-max standardization.

Figure 3.3: Predicted Association between Change in Occupational Moral Index and Change in Total Hourly Compensation

Figure 3.3 shows that the predicted change in hourly compensation decreases as occupations become progressively less morally compromised, assuming all other factors remain constant. This suggests that workers who transition from highly morally compromised occupations to less compromised occupations experience a compensation penalty. Thus, workers who are asked to compromise their sense of right and wrong in their job are often offered higher compensation as an incentive. Conversely, workers who choose to work in morally upright occupations may face lower compensation as a result.

This trend implies that there is a wage premium associated with making moral compromises in the workplace, which is consistent with previous studies on other disamenities such as physical hazards, exposure to dangerous materials, irregular shift schedules, job stress, and dirty jobs (Smith, 1979; Olson, 1981; Arnould and Nichols, 1983; Leeth and Ruser, 2003; Garen, 1988; Lanfranchi et al., 2002; French and Dunlap, 1998; Villanueva, 2007). Workers who overlook moral concerns or engage in unethical practices may be financially rewarded for their actions, compensating for the psychological discomfort they may experience. Conversely, workers who choose to work in morally upright occupations may face lower compensation as a result.

3.4.3 Robustness

In this subsection, we conduct sensitivity tests on the previously discussed results, focusing on fours aspects: (1) incorporating an occupational autonomy index derived from O*NET; (2) limiting the sample to workers who have completed educational training; (3) restricting the sample to workers who are 24 years of age or older; and (4) clustering standard errors by occupation. All of these tests are performed using Equation 3.5.[20] In all tests, we also consider real hourly compensation as an outcome variable.

First, the autonomy index captures the extent of autonomy or decision-making authority that workers have in their jobs, which can be an important factor affecting hourly compensation. Workers in occupations with higher autonomy may have more control over their work environment and may be able to negotiate higher wages without compromising their sense of right and wrong in their jobs. Moreover, empirical evidence suggests that autonomy in the workplace can significantly impact various outcomes, including job satisfaction, job performance, and compensation (Arai, 1994; Carr and Mellizo, 2013). Column (1) in Table 3.5 indicates a compensation penalty even with the inclusion of the

[20]The findings of these sensitivity tests using Equation 3.6 can be found in Appendix C.

O*NET-derived occupational autonomy index, and the magnitude of the coefficient is similar to that of column (3) in Table 3.3.

Table 3.5: Sensitivity of Parameter Estimates of Occupational Moral Index vis-a-vis Natural Log of Total Hourly Compensation

Variables	(1)	(2)	(3)	(4)
Occupational Moral Index	-0.049***	-0.033**	-0.054***	-0.053**
	(0.013)	(0.013)	(0.016)	(0.024)
Constant	2.140***	2.278***	2.317***	2.145***
	(0.031)	(0.037)	(0.044)	(0.040)
Observations	67,823	60,040	46,662	72,544
R-squared	0.680	0.709	0.731	0.679
All Controls	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes
Autonomy Index	Yes	No	No	No
Education Restriction	No	Yes	No	No
Age Restriction	No	No	Yes	No
Occupation Cluster	No	No	No	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is ln(real hourly compensation). All controls consist of a set of year, region, urban, and industry dummies, as well as human capital and labor supply variables, and other job characteristics as specified in Table 3.2. Occupational Moral Index is scaled from zero to one using min-max standardization.

Second, as mentioned earlier, the NLSY97 dataset comprises relatively young individuals, with ages ranging from 33 to 37 years in the most recent observed year (2017). Consequently, many of the person-year-job observations used in the analysis thus far pertain to workers who are still enrolled in education. To address this, we subset the NLSY97 data to include only observations where workers have completed their educational training, in an effort to capture more permanent employment endeavors. This restriction allows us to test the sensitivity of the results to the potential presence of uncharacteristically low or high compensation observations. Despite this restriction, we still observe a compensation penalty across the range of the occupational moral index, as shown in column (2) of Table 3.5.

Third, in order to mitigate potential bias that may arise from including younger workers who are still in the early stages of their careers and may have different compensation patterns compared to more experienced workers, we restrict the sample to workers who are 24 years old and older. This is because our analysis so far has looked at all workers who are 18 years and older. This test allows us to examine whether the relationship between the occupational moral index and hourly compensation holds true across different age groups. Again, we observed a compensation penalty and observe a similar coefficient of magnitude (column (3) of Table 3.5) to the estimate in column (3) of Table 3.3.

There are two limitations of using O*NET job skill/task measures that deserve mentioning. Firstly, the O*NET values assigned to each occupation remain fixed over time. Secondly, the value of each O*NET attribute does not vary among workers within the same occupation. While we are not concerned about the first issue due to the gradual changes in relative occupational differences in attributes, the measurement of job attributes at the occupation level rather than the individual worker level is a more serious concern. This is because there is heterogeneity in job characteristics within detailed occupations, and these characteristics may vary to some extent across workers. It is unclear whether and to what extent measurement error in O*NET job attributes is addressed through a set of person fixed-effects and/or differencing in panel analysis. Because job attributes are measured at the occupation level rather than the individual level, our final sensitivity test is to cluster standard errors by occupation. As expected, the coefficient of magnitude in column (4) of Table 3.5 is exactly the same as that of column (3) in Table 3.3, with the significant level reduced to under 5%. However, we still observe a compensation penalty even with this inclusion.

Finally, Table 3.6 performs identical tests to those presented in Table 3.5, but with the outcome variable of interest being real hourly compensation instead of the natural logarithm. Once again, we observe a consistent compensation penalty across the occupational moral index, with similar magnitudes of coefficients.

Table 3.6: Sensitivity of Parameter Estimates of Occupational Moral Index vis-a-vis Total Hourly Compensation

Variables	(1)	(2)	(3)	(4)
Occupational Moral Index	-1.974***	-1.216***	-1.763***	-1.904***
	(0.296)	(0.314)	(0.411)	(0.542)
Constant	8.497***	10.746***	7.549***	8.383***
	(0.650)	(0.785)	(1.021)	(0.821)
Observations	67,823	60,040	46,662	72,544
R-squared	0.641	0.679	0.703	0.639
All Controls	Yes	Yes	Yes	Yes
Person FE	Yes	Yes	Yes	Yes
Autonomy Index	Yes	No	No	No
Education Restriction	No	Yes	No	No
Age Restriction	No	No	Yes	No
Occupation Cluster	No	No	No	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is real hourly compensation. All controls consist of a set of year, region, urban, and industry dummies, as well as human capital and labor supply variables, and other job characteristics as specified in Table 3.2. Occupational Moral Index is scaled from zero to one using min-max standardization.

3.4.4 Heterogeneous Effects

In this subsection, we investigate if the previously estimated compensation penalties are consistently applicable to workers with and without a college education. Thus, we estimate the following equation:

$$\begin{aligned} lnW_{ijt} = \beta M_{ijt} + \pi COLLEGE_{it} + \psi(M_{ijt} \times COLLEGE_{it}) \\ + \; X'_{ijt}\gamma + \delta_t + \alpha_i + \lambda_{ijt} + \sigma_{ijt} + \omega_{ijt} + \epsilon_{ijt} \end{aligned} \tag{3.7}$$

The terms are the same as in Equation 3.5. However, the coefficients of interest are β and ψ. The estimated main effect (β) captures the association between the occupational moral index and compensations among workers without a college education, while the estimated interaction effect, combined with the estimated main effect ($\beta + \psi$), captures the association between the occupational moral index and compensations among college-educated workers.

We present the results in Table 3.7. Specifically, we focus on column (3), which considers both observed and unobserved worker characteristics, as well as a set of year, region, and urban dummies. The results in column (3) indicate that jobs with lower levels of moral compromise intensity are associated with lower compensations among workers without a college education, holding all other factors constant. The coefficient estimate of -0.022 shows that moving across the interdecile range of the occupational moral index, from 0.41 to 0.85, is associated with a 0.97 percent reduction in total hourly compensations.

On the other hand, among workers with a college education, jobs with lower moral compromise are associated with even lower compensations, all else being equal. Combining the main and interaction effects in column (3) suggests that a similar movement across the interdecile range of the occupational moral index is associated with a 7.3 percent reduction in total hourly compensations for workers with a college education. We depict these findings graphically in Figure 3.4.

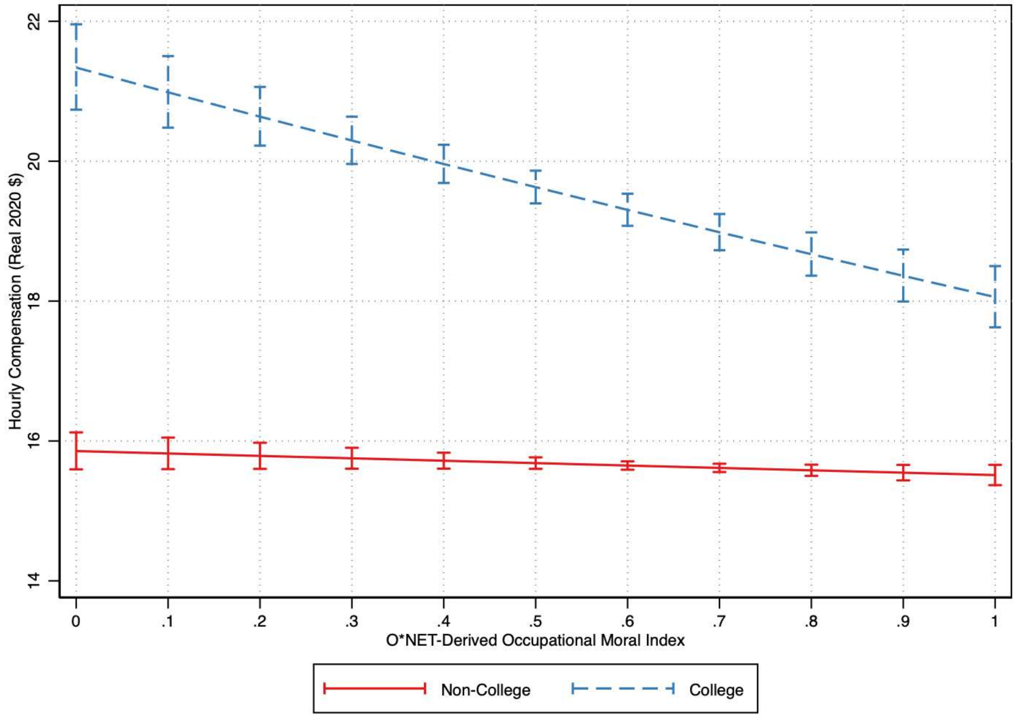
As depicted in Figure 3.4, the predicted real hourly compensation for workers without a college education remains relatively constant across the range of the occupational moral index, with a compensating difference of about \$0.31 $(15.82 - 15.51)$ per hour between the most and least morally compromising occupations. However, for college-educated workers, the compensating difference is about \$3.28 $(21.34 - 18.06)$ per hour. Nevertheless, due to the relatively large estimated compensation penalty among college-educated workers, the predicted hourly compensations for workers with and without a college education appear to converge in occupations with least moral compromise, suggesting that the compensation gap between college-educated and non-college-educated workers may be smaller in occupations with lower levels of moral compromise, when all other factors are equal.

Table 3.7: Parameter Estimates of Occupational Moral Index by College Education vis-a-vis Natural Log of Total Hourly Compensation

Variables	(1)	(2)	(3)
Occupational Moral Index	-0.170***	-0.066***	-0.022*
	(0.011)	(0.012)	(0.013)
College	0.543***	0.308***	0.297***
	(0.016)	(0.014)	(0.016)
Occupational Moral Index × College	-0.269***	-0.199***	-0.145***
	(0.027)	(0.023)	(0.026)
Constant	2.822***	2.074***	2.124***
	(0.008)	(0.032)	(0.030)
Observations	72,544	72,544	72,544
R-squared	0.144	0.453	0.679
Year FE	No	Yes	Yes
Region FE	No	Yes	Yes
Urban FE	No	Yes	Yes
Industry FE	No	Yes	Yes
Demography Controls	No	Yes	No
Human Capital and Labor Supply Controls	No	Yes	Yes
Job Characteristics Controls	No	Yes	Yes
Person FE	No	No	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable is ln(real hourly compensation). Occupational Moral Index is scaled from zero to one using min-max standardization.

The disparity in compensation between workers with a college education and those without, as evidenced by Table 3.7 and Figure 3.4, may be explained by two well-known labor theories. Human Capital Theory, proposed by Becker (1993) suggests that skills acquired through higher education can impact labor market outcomes, including compensation. This implies that college-educated individuals may have more choices when it comes to job selection without having to compromise their moral principles, and therefore may require higher compensation. Additionally, Job Market Signaling Theory, as proposed by Spence (1978), posits that education serves as a signal of an individual's ability and productivity to potential employers. As a result, employers may be aware of the alterna-



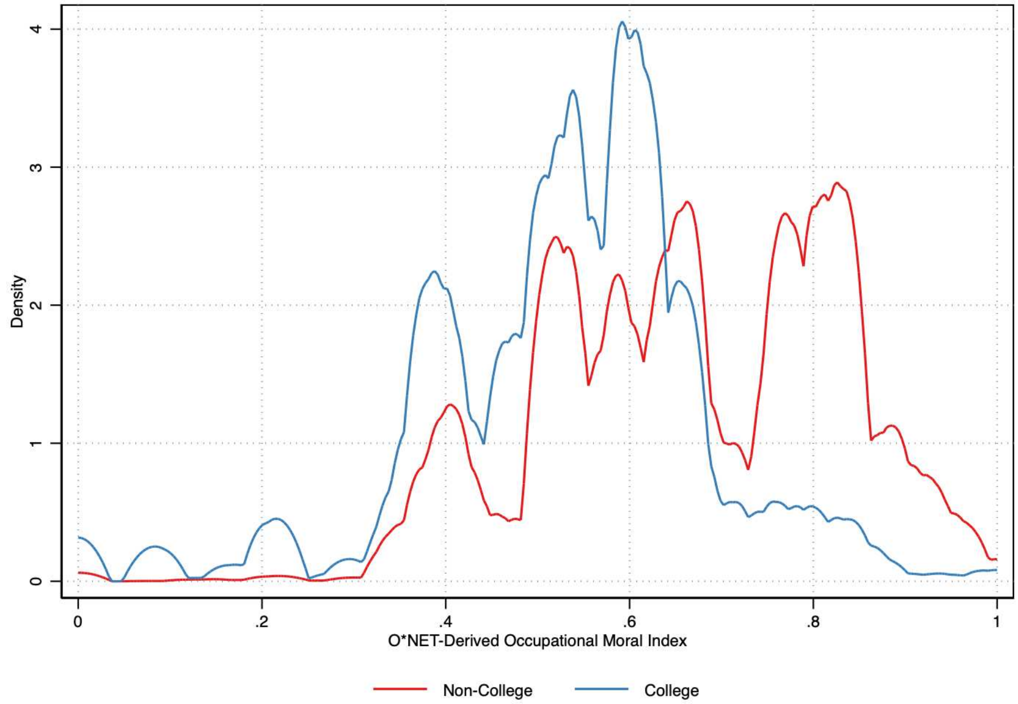
Notes: Spikes represent the 95% confidence interval for the predicted total hourly compensation. The estimates from column (3) in Table 3.7 are used for this prediction. O*NET-Derived Occupational Moral Index is scaled from zero to one using min-max standardization.

Figure 3.4: Predicted Association between Occupational Moral Index and Total Hourly Compensation by College Education

tive job options available to college-educated workers, and higher compensation may be necessary to incentivize these workers to compromise their moral values.

Although our NLSY97 sample does not allow for direct testing of these theories, they provide insights into why there may be a compensating difference between workers with and without a college education in jobs that require compromising one's moral compass. College-educated individuals may have different outside options and higher bargaining power that influence their compensation preferences in the presence of psychological discomfort caused by conflicting moral values.

Figure 3.5 displays the representation of NLSY97 sample distribution in relation to non-college and college-educated workers across the O*NET-derived occupational moral index. It is important to note that there is an uneven distribution of non-college educated



Notes: O*NET-Derived Occupational Moral Index is scaled from zero to one using min-max standardization.

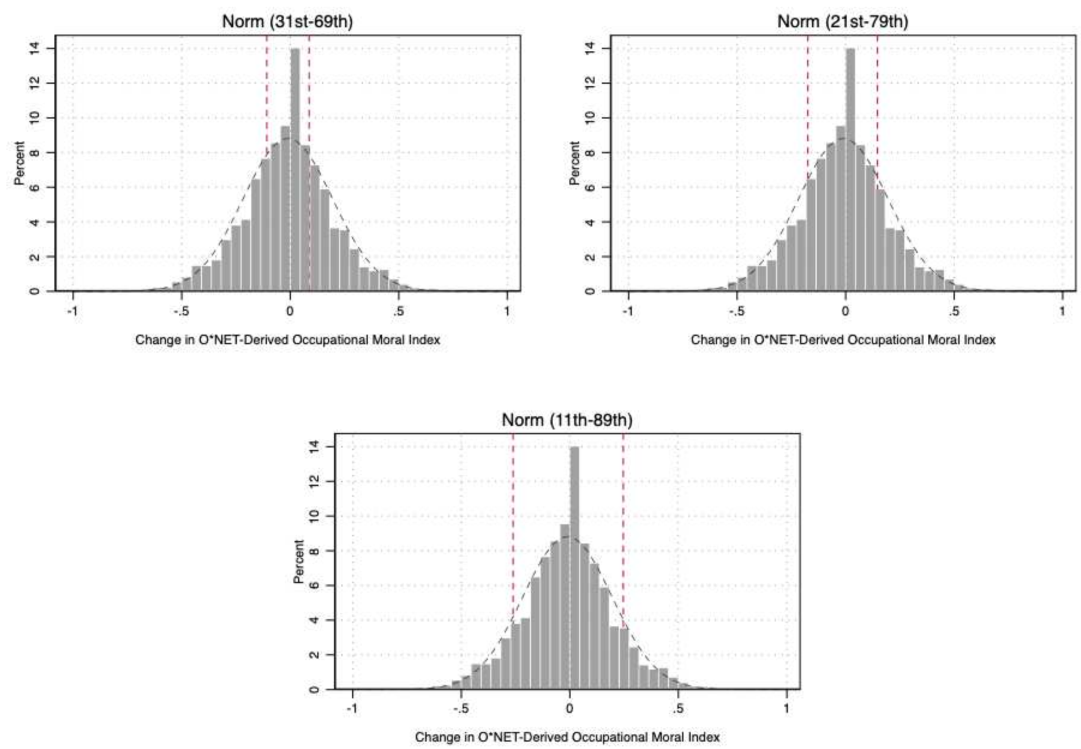
Figure 3.5: Kernel Density Distribution of NLSY97 Occupations Across the O*NET-Derived Moral Index by Non-College and College Educated Workers

workers and college-educated workers across occupations that often require workers to compromise their moral sense of right and wrong. While the error bars in Figure 3.4 for the predicted hourly compensation for non-college educated workers are relatively small, we should still exercise caution in drawing conclusive findings from these results, due to the uneven distribution between non-college educated workers and college-educated workers in occupations that often require compromising of their moral compass.

3.5 Extended Analysis

3.5.1 Asymmetric Effects

Until now, our empirical models have assumed a symmetrical relationship between the occupational moral index and hourly compensation. In this section, we will present empirical test results that relax this assumption. Specifically, we will focus on a sample of switchers who have changed occupations/industries and examine different movements across the change in the occupational moral index. To conduct our tests, we will refer to Figure 3.6 and utilize Equation 3.6. By doing so, we aim to gain a deeper understanding of the true relationship between the occupational moral index and hourly compensation.



Notes: Change in O*NET-Derived Occupational Moral Index is scaled from zero to one using min-max standardization.

Figure 3.6: Distribution of Change in O*NET-Derived Occupational Moral Index from Job Switching

We generated three categorical variables: low, medium, and high occupational moral index. These variables are displayed in Figure 3.6. For example, the third panel of Figure 3.6 compares workers in occupations that belong to the lowest 10 percentile of the occupational moral index (which indicates the highest level of moral compromise) to those in the normal range of the 11th-89th occupational moral index. In addition, the third panel of Figure 3.6 also compares workers in occupations that are in the top 10 percentile of the occupational moral index (which indicates the lowest level of moral compromise) to those in the normal range of the 11th-89th occupational moral index.

Table 3.8 presents evidence supporting an asymmetric relationship between changes in the occupational moral index and total hourly compensation. This relationship is influenced by the degree of moral compromise present in the job. Specifically, when workers switch to jobs with higher levels of moral compromise, they tend to experience an increase in hourly compensation. Conversely, when they transition to jobs with lower levels of moral compromise, the decrease in compensation is more modest.

However, the magnitude and statistical significance of this relationship vary depending on the percentile of the change in the occupational moral index. For instance, in column (1) of Table 3.8, an increase in hourly compensation is observed for workers who move from jobs with lower moral compromise to those with higher moral compromise. However, there is no statistically significant decrease in hourly compensation for workers who move from jobs with higher moral compromise to those with lower moral compromise. Additionally, in columns (2) and (3), the increase in hourly compensation is larger and statistically significant for workers who move to jobs with higher levels of moral compromise within the lower percentiles of the change in the occupational moral index. Conversely, the decrease in hourly compensation for workers moving to jobs with lower levels of moral compromise within the higher percentiles of the change in the occupational moral index is smaller and not statistically significant.

Table 3.8: Parameter Estimates of Association between Change in Occupational Moral Index Categories and Change in Natural Log of Total Hourly Compensation

Variables	(1)	(2)	(3)
Down $\leq$ 30th	0.016***		
	(0.006)		
Up $\geq$ 70th	-0.003		
	(0.006)		
Down $\leq$ 20th		0.019***	
		(0.007)	
Up $\geq$ 80th		-0.012*	
		(0.006)	
Down $\leq$ 10th			0.020**
			(0.008)
Up $\geq$ 90th			-0.005
			(0.009)
Constant	0.002	0.006	0.004
	(0.062)	(0.062)	(0.062)
Observations	29,769	29,769	29,769
R-squared	0.085	0.086	0.085
Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Urban FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Demography Controls	Yes	Yes	Yes
Δ Human Capital and Labor Supply Controls	Yes	Yes	Yes
Δ Characteristics Controls	Yes	Yes	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable is ln(real hourly compensation). The base category for column (1) is: Δ Norm (31st-69th); the base category for column (2) is: Δ (21st-79th); and the base category for column (3) is: Δ Norm (11th-89th). The term "$Down$" refers to workers transitioning from jobs that require less compromise of moral sense of right and wrong to jobs that require more compromise of moral sense of right and wrong. Conversely, the term "Up" refers to workers transitioning from jobs that require more compromise of worker moral sense of right and wrong to jobs that require less compromise of worker moral sense of right and wrong. Change in O*NET-Derived Occupational Moral Index is scaled from zero to one using min-max standardization.

3.5.2 Quadratic Transformation of Occupational Moral Index

In this section, and informed by results in the previous section pertaining to asymmetric effects, we make an assumption about the quadratic nature of the occupational moral index. We estimate the equation below:

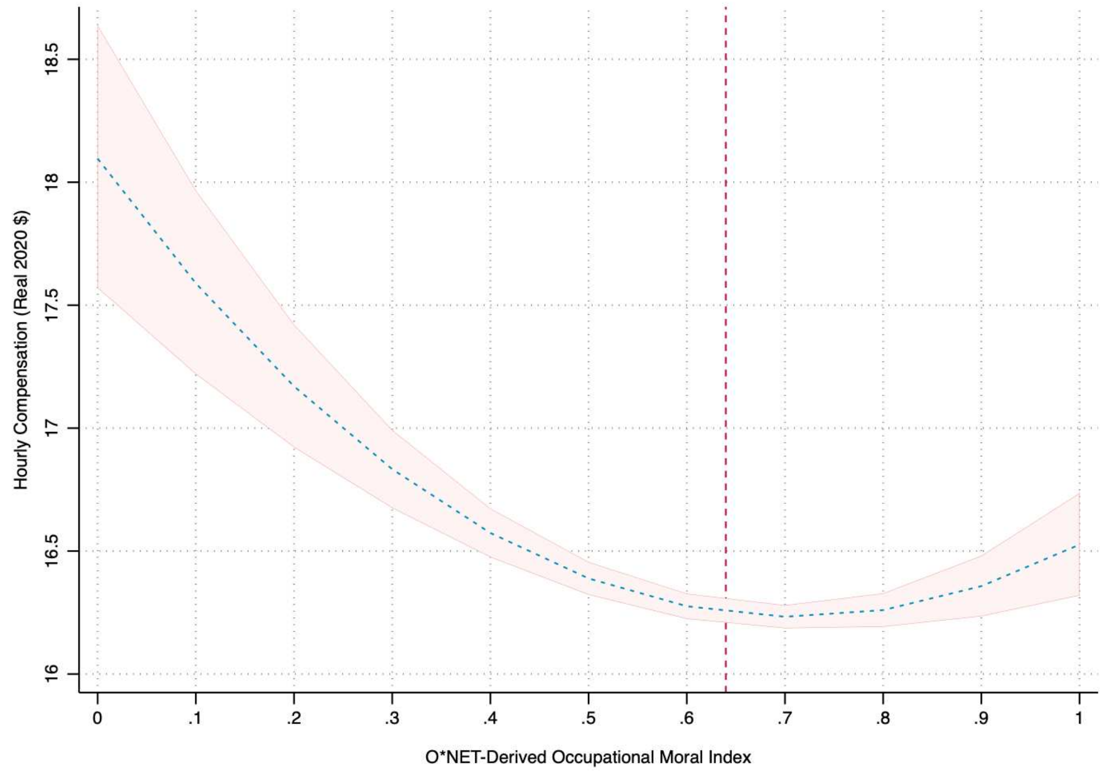
$$lnW_{ijt} = \beta_0 M_{ijt} + \beta_1 M_{ijt}^2 + X_{ijt}'\gamma + \delta_t + \alpha_i + \lambda_{ijt} + \sigma_{ijt} + \omega_{ijt} + \epsilon_{ijt} \tag{3.8}$$

We use the same terms defined in Equation 3.5, but add the term M_{ijt}^2. In this case, we assume that the O*NET-derived occupational model index is quadratic. Our decision to assume a quadratic relationship between moral compromise across occupations and hourly compensation is based on our asymmetric analysis. The findings suggest that the relationship between these two variables may not be linear. By assuming a quadratic relationship between moral compromise and compensation, we can explore whether a threshold effect exists and, if so, at what point it occurs. This can provide valuable insights into how occupational moral compromise affects compensation.

In Figure 3.7[21], we can see that there is a wage premium for occupations that require workers to compromise their moral values, while there is no such premium for occupations that are at the mean. The reason for the wage premium in high-compromise occupations could be that these jobs are emotionally and psychologically more demanding, resulting in stress and discomfort for the workers. As a result, workers may experience high turnover, absenteeism, and reduced job satisfaction, negatively impacting their productivity and work quality. Employers may offer higher wages to attract and retain workers in these positions. This is because, in a competitive labor market, higher wages can offset some of the negative effects associated with working in these jobs.

Conversely, the absence of a wage premium for average jobs implies that workers in these positions do not face the same level of moral compromise as those in high-

[21] Estimates from Equation 3.8 is available in Appendix C.



Notes: The shaded area in the table represents the 95% confidence interval for the predicted total hourly compensation. The vertical dotted line indicates the mean O*NET-Derived Occupational Moral Index. O*NET-Derived Moral Index is scaled from zero to one using min-max standardization.

Figure 3.7: Predicted Association between Quadratic Occupational Moral Index and Total Hourly Compensation.

compromise jobs. Such jobs may be more routine and require less emotional labor, making it unnecessary for employers to offer higher wages to retain employees. However, it is important to note that these explanations are general observations, and other factors may also contribute to the observed wage premiums or lack thereof.

3.6 Study Limitations

There are a number of important limitations to our study that need to be emphasized. First, as previously mentioned, solely using person fixed-effects is insufficient to adjust for any time-varying residual confounding linked to job change selection (i.e., endogenous job change), which can lead to biased longitudinal estimates. To ensure the reliability of our findings derived from person fixed-effects, we focused exclusively on occupation switchers and used the first-differencing method as a robustness check. Nevertheless, even though the magnitudes of the two approaches are similar, there may still be other time-varying confounders that should be considered. Specifically, job changes are likely to be endogenous rather than random, and may be associated with changes in compensation. Endogenous selection is likely to influence our coefficients in different directions based on whether workers move to occupations with lower or higher levels of moral compromise. Consequently, we acknowledge that our estimates may, at best, be conservative.

Second, although longitudinal data enables us to observe and control for labor supply and employment history, a downside of these data is that they often represent relatively small samples that may not be occupationally representative of the broader economy. In other words, we cannot state with certainty whether we can observe the same results when considering workers who are 38 years and older and not included in our study. However, despite these caveats, our findings provide novel evidence on another form of job disamenity, specifically conscience, which has received limited attention in studies of compensating wage differentials.

3.7 Conclusion

In our analysis, we investigate the existence of compensating differentials in the U.S. labor market concerning the degree of moral compromise required in various occupations. Specifically, we examined whether jobs that demand workers to compromise their moral values offer higher compensation to offset the disamenities that contradict their moral beliefs. We controlled for time-invariant heterogeneity and found that jobs that require moral compromise indeed offer higher compensation. However, our heterogeneity analysis uncovered two key findings: (1) we observed that workers with a college education tend to receive higher pay in jobs that require moral compromise compared to those without a college education; (2) we found that workers without a college education do not receive a compensation premium for moral compromise. The stronger association between compensation and moral compromise among college-educated workers may be attributed to their greater bargaining power and more abundant employment opportunities, which influence their compensation preferences (Spence, 1978; Becker, 1993).

Moreover, we discovered an asymmetric relationship between changes in the occupational moral index and total hourly compensation, indicating that the relationship is responsive to the intensity of moral compromise in the job. It is important to note the issue of endogeneity concerning the level of moral compromise required in various occupations. To control for selection bias, we captured unobserved heterogeneity with fixed effects. Future studies should explore the role of time-varying heterogeneity as data allows.

Bibliography

Adams, K., Leibbrandt, S., and Moon, H. (2011). A Critical Review of the Literature on Social and Leisure Activity and Wellbeing in Later Life. *Ageing & Society*, 31(4):683–712.

Aliprantis, D., Carroll, D., and Young, E. (2019). The Dynamics of the Racial Wealth Gap. *Federal Reserve Bank of Cleveland Working Paper No. 19-18*.

Almorox, J., Hontoria, C., and Benito, M. (2005). Statistical validation of daylength definitions for estimation of global solar radiation in Toledo, Spain. *Energy Conversion and Management*, 46(9-10):1465–1471.

Anglin, R. E., Samaan, Z., Walter, S. D., and McDonald, S. D. (2013). Vitamin D deficiency and depression in adults: systematic review and meta-analysis. *The British Journal of Psychiatry*, 202(2):100–107.

Arai, M. (1994). Compensating Wage Differentials Versus Efficiency Wages: An Empirical Study of Job Autonomy and Wages. *Industrial Relations: A Journal of Economy and Society*, 33(2):249–262.

Arnould, R. J. and Nichols, L. M. (1983). Wage-Risk Premiums and Workers' Compensation: A Refinement of Estimates of Compensating Wage Differential. *Journal of Political Economy*, 91(2):332–340.

Ashman, H. and Neumuller, S. (2020). Can Income Differences Explain the Racial Wealth Gap: A Quantitative Analysis. *Review of Economic Dynamics*, 35:220–239.

Bansal, A., Choudhry, A., Sharma, A., and Susan, S. (2022). Adaptation of domain-specific transformer models with text oversampling for sentiment analysis of social media posts on Covid-19 vaccines. *arXiv preprint arXiv:2209.10966*.

Barnes, C. M. and Wagner, D. T. (2009). Changing to daylight saving time cuts into sleep and increases workplace injuries. *Journal of Applied Psychology*, 94(5):1305.

Baylis, P. (2020). Temperature and Temperament: Evidence from Twitter. *Journal of Public Economics*, 184:104161.

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Benhabib, J., Bisin, A., and Luo, M. (2017). Earnings Inequality and Other Determinants of Wealth Inequality. *American Economic Review*, 107(5):593–97.

Bhutta, N., Chang, A. C., Dettling, L., and Hsu, J. (2020). Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances. Technical report, FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Blume, C., Garbazza, C., and Spitschan, M. (2019). Effects of light on human circadian rhythms, sleep and mood. *Somnologie*, 23(3):147–156.

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2019). Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies*, 32(3):992–1033.

Bünnings, C. and Schiele, V. (2021). Spring forward, don't fall back: the effect of daylight saving time on road safety. *Review of Economics and Statistics*, 103(1):165–176.

Carr, M. D. and Mellizo, P. (2013). The relative effect of voice, autonomy, and the wage on satisfaction with work. *The International Journal of Human Resource Management*, 24(6):1186–1201.

Clark, A. (2003). *Inequality-aversion and income mobility: A direct test*. Citeseer.

Clemens, J. and Wither, M. (2019). The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers. *Journal of Public Economics*, 170:53–67.

Conley, D. (2000). The Racial Wealth Gap: Origins and Implications for Philanthropy in the African American Community. *Nonprofit and Voluntary Sector Quarterly*, 29(4):530–540.

Connolly, M. (2013). Some like it mild and not too wet: The influence of weather on subjective well-being. *Journal of Happiness Studies*, 14:457–473.

Darity Jr, W. A. and Myers Jr, S. L. (1998). *Persistent Disparity: Race and Economic Inequality in the United States since 1945*. Edward Elgar Publishing.

De Nardi, M. (2004). Wealth Inequality and Intergenerational Links. *The Review of Economic Studies*, 71(3):743–768.

De Vos, I., Everaert, G., Ruyssen, I., et al. (2015). Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *Stata Journal*, 15(4):986–1018(33).

Deaton, A. S. and Paxson, C. H. (1998). Aging and Inequality in Income and Health. *American Economic Review*, 88(2):248–253.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Diener, E. and Biswas-Diener, R. (2011). *Happiness: Unlocking the mysteries of psychological wealth*. John Wiley & Sons.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS One*, 6(12):e26752.

Doleac, J. L. and Sanders, N. J. (2015). Under the Cover of Darkness: How Ambient Light Influences Criminal Activity. *Review of Economics and Statistics*, 97(5):1093–1103.

Easterlin, R. A. (1995). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization*, 27(1):35–47.

Everaert, G. and Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*, 31(4):1160–1184.

Exebio, J. C., Zarini, G. G., Ajabshir, S., Antwi, J., and Huffman, F. G. (2016). Validation of a sun exposure questionnaire among subjects with type 2 diabetes residing in south florida. *Journal of Immigrant and Minority Health*, 18:228–233.

Feeny, D., Furlong, W., and Torrance, G. W. (2018). What were they thinking when providing preference measurements for generic health states? the evidence for hui3. *Health and Quality of Life Outcomes*, 16(1):166.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M., and Boyle, M. (2002). Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Medical Care*, 40(2):113–128.

Figueiro, M. G., Steverson, B., Heerwagen, J., Kampschroer, K., Hunter, C. M., Gonzales, K., Plitnick, B., and Rea, M. S. (2017). The impact of daytime light exposures on sleep and mood in office workers. *Sleep Health*, 3(3):204–215.

Fleurbaey, M. (2009). Beyond GDP: The Quest for a Measure of Social Welfare. *Journal of Economic Literature*, 47(4):1029–75.

Ford, D. E. and Cooper-Patrick, L. (2001). Sleep disturbances and mood disorders: An epidemiologic perspective. *Depression and Anxiety*, 14(1):3–6.

French, M. T. and Dunlap, L. J. (1998). Compensating Wage Differentials for Job Stress. *Applied Economics*, 30(8):1067–1075.

Frey, B. S. and Stutzer, A. (2002). What Can Economists Learn from Happiness Research? *Journal of Economic Literature*, 40(2):402–435.

Furlong, W., Feeny, D., Torrance, G., Goldsmith, C., DePauw, S., Zhu, Z., Denton, M., Boyle, M., et al. (1998). Multiplicative Multi-Attribute Utility Function for the Health Utilities

Index Mark 3 (HUI3) System: A Technical Report. Technical report, Centre for Health Economics and Policy Analysis (CHEPA), McMaster University.

Garen, J. (1988). Compensating Wage Differentials and the Endogeneity of Job Riskiness. *the Review of Economics and Statistics*, pages 9–16.

Gloth 3rd, F., Alam, W., and Hollis, B. (1999). Vitamin D vs broad spectrum phototherapy in the treatment of seasonal affective disorder. *The Journal of Nutrition, Health & Aging*, 3(1):5–7.

Goetz, S. J. (1991). Single-and Multi-Dimensional Theil Measures of Inequality: An Application to Farm Households. Technical report.

Han, J. and Patterson, I. (2007). An Analysis of the Influence that Leisure Experiences Have on a Person's Mood State, Health and Wellbeing. *Annals of Leisure Research*, 10(3-4):328–351.

Harrison, Y. (2013). The impact of daylight saving time on sleep and related behaviours. *Sleep Medicine Reviews*, 17(4):285–292.

Helliwell, J. F. and Wang, S. (2014). Weekends and subjective well-being. *Social Indicators Research*, 116:389–407.

Honaker, J. and King, G. (2010). What to Do about Missing Values in Time-Series Cross-Section Data. *American Journal of Political Science*, 54(2):561–581.

Honaker, J., King, G., Blackwell, M., et al. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

Horsman, J., Furlong, W., Feeny, D., and Torrance, G. (2003). The Health Utilities Index (HUI®): Concepts, Measurement Properties and Applications. *Health and Quality of Life Outcomes*, 1(1):1–13.

Hribernik, J. and Mussap, A. J. (2010). Research Note: Leisure Satisfaction and Subjective Wellbeing. *Annals of Leisure Research*, 13(4):701–708.

Huang, Y., Cai, X., Zhang, J., Mai, W., Wang, S., Hu, Y., Ren, H., and Xu, D. (2014). Prehypertension and Incidence of ESRD: A Systematic Review and Meta-Analysis. *American Journal of Kidney Diseases*, 63(1):76–83.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Hwang, H.-s., Mortensen, D. T., and Reed, W. R. (1998). Hedonic wages and labor market search. *Journal of Labor Economics*, 16(4):815–847.

Idler, E. L. and Benyamini, Y. (1997). Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *Journal of Health and Social Behavior*, pages 21–37.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Jin, L. and Ziebarth, N. R. (2020). Sleep, health, and human capital: Evidence from daylight saving time. *Journal of Economic Behavior & Organization*, 170:174–192.

Jones, C. I. and Klenow, P. J. (2016). Beyond GDP? Welfare Across Countries and Time. *American Economic Review*, 106(9):2426–57.

Kämpfer, S. and Mutz, M. (2013). On the sunny side of life: Sunshine effects on life satisfaction. *Social Indicators Research*, 110(2):579–595.

Kaplan, R. M. and Bush, J. W. (1982). Health-Related Quality of Life Measurement for Evaluation Research and Policy Analysis. *Health Psychology*, 1(1):61.

Kent, S. T., McClure, L. A., Crosson, W. L., Arnett, D. K., Wadley, V. G., and Sathiakumar, N. (2009). Effect of sunlight exposure on cognitive function among depressed and non-depressed participants: a REGARDS cross-sectional study. *Environmental Health*, 8(1):1–14.

Kokubo, Y. and Iwashima, Y. (2015). Higher Blood Pressure as a Risk Factor for Diseases other than Stroke and Ischemic Heart Disease. *Hypertension*, 66(2):254–259.

Kotchen, M. J. and Grant, L. E. (2011). Does daylight saving time save energy? Evidence from a natural experiment in Indiana. *Review of Economics and Statistics*, 93(4):1172–1185.

Kountouris, Y. and Remoundou, K. (2014). About time: daylight saving time transition and individual well-being. *Economics Letters*, 122(1):100–103.

Kuehnle, D. and Wunder, C. (2016). Using the life satisfaction approach to value daylight savings time transitions: evidence from Britain and Germany. *Journal of Happiness Studies*, 17:2293–2323.

Lambert, G. W., Reid, C., Kaye, D. M., Jennings, G. L., and Esler, M. D. (2002). Effect of sunlight and season on serotonin turnover in the brain. *The Lancet*, 360(9348):1840–1842.

Lanfranchi, J., Ohlsson, H., and Skalli, A. (2002). Compensating Wage Differentials and Shift Work Preferences. *Economics Letters*, 74(3):393–398.

Leeth, J. D. and Ruser, J. (2003). Compensating Wage Differentials for Fatal and Nonfatal Injury Risk by Gender and Race. *Journal of Risk and Uncertainty*, 27(3):257–277.

Lepinteur, A. (2019). The Shorter Workweek and Worker Wellbeing: Evidence from Portugal and France. *Labour Economics*, 58:204–220.

Lewington, S. (2002). Prospective Studies Collaboration. Age-Specific Relevance of Usual Blood Pressure to Vascular Mortality: A Meta-Analysis of Individual Data for One Million Adults in 61 Prospective Studies. *Lancet*, 360:1903–1913.

Lynch, S. M. (2008). Race, Socioeconomic Status, and Health in Life-Course Perspective: Introduction to the Special Issue. *Research on Aging*, 30(2):127–136.

Maasoumi, E. (1986). The Measurement and Decomposition of Multidimensional Inequality. *Econometrica*, pages 991–997.

Maasoumi, E. and Nickelsburg, G. (1988). Multivariate Measures of Well-Being and an Analysis of Inequality in the Michigan Data. *Journal of Business & Economic Statistics*, 6(3):327–334.

Maasoumi, E. and Nickesburg, G. (1983). Measuring Joint Inequality in Incomes, Wealth and Schooling and a Decomposition. *Department of Economics Working Paper, 83-4, Indiana University, Bloomington*.

Manduca, R. (2018). Income Inequality and the Persistence of Racial Economic Disparities. *Sociological Science*, 5:182–205.

Manton, K. G., P. C. H. . J. K. W. (1987). Health Differentials between Blacks and Whites: Recent Trends in Mortality and Morbidity. *The Milbank Quarterly*, 65:129–199.

Margo, R. A. (2016). Obama, Katrina, and the Persistence of Racial Inequality. *The Journal of Economic History*, 76(2):301–341.

Martin, L. G. and Soldo, B. J. (1997). *Racial and Ethnic Differences in the Health of Older Americans*. National Academies Press.

Mead, M. N. (2008). Benefits of sunlight: a bright spot for human health.

Miller, R. and Bairoliya, N. (2022). Health, Longevity, and Welfare Inequality of Older Americans. *(Forthcoming) The Review of Economics and Statistics*.

Miller, R., Chin, S., and Sedai, A. K. (2022). The welfare cost of late-life depression. *Journal of Economic Behavior & Organization*, 204:15–36.

Minkler, M., Fuller-Thomson, E., and Guralnik, J. M. (2006). Gradient of Disability across the Socioeconomic Spectrum in the United States. *New England Journal of Medicine*, 355(7):695–703.

Mullahy, J. (2016). Estimation of multivariate probit models via bivariate probit. *Stata Journal*, 16(1):37–51.

Murphy, K. M. and Topel, R. H. (2006). The Value of Health and Longevity. *Journal of Political Economy*, 114(5):871–904.

Nandwani, P. and Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1):81.

Nguyen, D. Q., Vu, T., and Nguyen, A. T. (2020). BERTweet: A pre-trained language model for English Tweets. *arXiv preprint arXiv:2005.10200*.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Oishi, S., Kesebir, S., and Diener, E. (2011). Income inequality and happiness. *Psychological Science*, 22(9):1095–1100.

Olson, C. A. (1981). An analysis of wage differentials received by workers on dangerous jobs. *Journal of Human Resources*, pages 167–185.

Ozawa, M. N. and Tseng, H.-y. (2000). Differences in Net Worth between Elderly Black People and Elderly White People. *Social Work Research*, 24(2):96–108.

Pager, D. and Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, 34:181–209.

Patton, G. C., Sawyer, S. M., Santelli, J. S., Ross, D. A., Afifi, R., Allen, N. B., Arora, M., Azzopardi, P., Baldwin, W., Bonell, C., et al. (2016). Our Future: A Lancet Commission on Adolescent Health and Wellbeing. *The Lancet*, 387(10036):2423–2478.

Pew Research Center (2018). Income Inequality in the U.S. is Rising Most Rapidly Among Asians. https://www.pewresearch.org/social-trends/wp-content/uploads/sites/3/2018/07/Pew_Research_Center_Inequality-Report_FINAL.pdf.

Pollack, C. E., Chideya, S., Cubbin, C., Williams, B., Dekker, M., and Braveman, P. (2007). Should Health Studies Measure Wealth?: A Systematic Review. *American Journal of Preventive Medicine*, 33(3):250–264.

Raghavan, S., Vassy, J. L., Ho, Y.-L., Song, R. J., Gagnon, D. R., Cho, K., Wilson, P. W., and Phillips, L. S. (2019). Diabetes Mellitus–Related All-Cause and Cardiovascular Mortality in a National Cohort of Adults. *Journal of the American Heart Association*, 8(4):e011295.

Rapsomaniki, E., Timmis, A., George, J., Pujades-Rodriguez, M., Shah, A. D., Denaxas, S., White, I. R., Caulfield, M. J., Deanfield, J. E., Smeeth, L., et al. (2014). Blood Pressure and Incidence of Twelve Cardiovascular Diseases: Lifetime Risks, Healthy Life-Years Lost, and Age-Specific Associations in 1· 25 Million People. *The Lancet*, 383(9932):1899–1911.

Rohde, N. and Guest, R. (2013). Multidimensional Racial Inequality in the United States. *Social Indicators Research*, 114(2):591–605.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2020). IPUMS USA: Version 10.0 [dataset]. *Minneapolis, MN: IPUMS*, 10:D010.

Ryen, L. and Svensson, M. (2015). The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. *Health Economics*, 24(10):1289–1301.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2):93.

Schwarz, N. and Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications.

Sexton, A. L. and Beatty, T. K. (2014). Behavioral responses to daylight savings time. *Journal of Economic Behavior & Organization*, 107:290–307.

Shah, J. and Gurbani, S. (2019). Association of vitamin D deficiency and mood disorders: A systematic review. *Vitam. D Defic*.

Shapiro, T. and Kenty-Drane, J. (2005). The Racial Wealth Gap. In *African Americans in the US Economy*, volume 175, page 177. Rowman & Littlefield Lanham.

Shea, D. G., Miles, T., and Hayward, M. (1996). The Health-Wealth Connection: Racial Differences. *The Gerontologist*, 36(3):342–349.

Smith, A. C. (2016). Spring Forward at Your Own Risk: Daylight Saving Time and Fatal Vehicle Crashes. *American Economic Journal: Applied Economics*, 8(2):65–91.

Smith, G. D. and Egger, M. (1993). Socioeconomic Differentials in Wealth and Health. *British Medical Journal*, 307(6912):1085.

Smith, J. P. et al. (1997). Wealth Inequality Among Older Americans. *Journals of Gerontology Series B*, 52:74–81.

Smith, R. S. (1979). Compensating Wage Differentials and Public Policy: A Review. *ILR Review*, 32(3):339–352.

Spence, M. (1978). Job market signaling. In *Uncertainty in Economics*, pages 281–306. Elsevier.

Steptoe, A., Deaton, A., and Stone, A. A. (2015). Subjective Wellbeing, Health, and Ageing. *The Lancet*, 385(9968):640–648.

Strife, S. and Downey, L. (2009). Childhood Development and Access to Nature: A New Direction for Environmental Inequality Research. *Organization & Environment*, 22(1):99–122.

Takasu, N. N., Hashimoto, S., Yamanaka, Y., Tanahashi, Y., Yamazaki, A., Honma, S., and Honma, K.-i. (2006). Repeated exposures to daytime bright light increase nocturnal melatonin rise and maintain circadian phase in young subjects under fixed sleep schedule. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 291(6):R1799–R1807.

Tobaldini, E., Costantino, G., Solbiati, M., Cogliati, C., Kara, T., Nobili, L., and Montano, N. (2017). Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. *Neuroscience & Biobehavioral Reviews*, 74:321–329.

Toro, W., Tigre, R., and Sampaio, B. (2015). Daylight Saving Time and incidence of myocardial infarction: evidence from a regression discontinuity design. *Economics Letters*, 136:1–4.

Valdez, P., Ramirez, C., Garcia, A., and Garcia, E. (1997). Adjustment of sleep to daylight saving time during weekdays and weekends. *Chronobiology International*, 14:170.

Van de Vliert, E. and Rentfrow, P. J. (2021). Who is more prone to depression at higher latitudes? Islanders or mainlanders? *Current Research in Ecological and Social Psychology*, 2:100012.

Veleva, B. I., van Bezooijen, R. L., Chel, V. G., Numans, M. E., and Caljouw, M. A. (2018). Effect of ultraviolet light on mood, depressive disorders and well-being. *Photodermatology, Photoimmunology & Photomedicine*, 34(5):288–297.

Vespa, J., Armstrong, D. M., and Medina, L. (2018). *Demographic Turning Points for the United States: Population Projections for 2020 to 2060*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Villanueva, E. (2007). Estimating Compensating Wage Differentials Using Voluntary Job Changes: Evidence from Germany. *ILR Review*, 60(4):544–561.

Wallis, D. E., Penckofer, S., and Sizemore, G. W. (2008). The "sunshine deficit" and cardiovascular disease. *Circulation*, 118(14):1476–1485.

Wang, J., Fan, Y., Palacios, J., Chai, Y., Guetta-Jeanrenaud, N., Obradovich, N., Zhou, C., and Zheng, S. (2022). Global evidence of expressed sentiment alterations during the COVID-19 pandemic. *Nature Human Behaviour*, 6(3):349–358.

Williams, D., Wilson, C., Binstock, R., and George, L. (2001). Race, Ethnicity, and Aging. In *Handbook of Aging and Social Sciences, 5th Edition*, pages 160–178. Academic Press, San Diego, CA.

Yang, C. and Srinivasan, P. (2016). Life satisfaction and the pursuit of happiness on Twitter. *PloS One*, 11(3):e0150881.

Zheng, S., Wang, J., Sun, C., Zhang, X., and Kahn, M. E. (2019). Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behaviour*, 3(3):237–243.

Appendix A

Missing Data, Statistical Procedure, and Supplementary Tables and Figures

A.1 Multiple Imputation of Consumption and Other Missing Data

We follow the procedure of Miller and Bairoliya (2022) to impute consumption and other missing HRS data. This procedure uses the bootstrapping approach for cross-sectional time-series data proposed by Honaker and King (2010) and implemented through the Amelia II software program (Honaker et al., 2011). We use the program to produced twelve complete datasets without missing data and all analyses are conducted on each dataset then combined into a single estimate. We follow Miller and Bairoliya (2022) in selecting the following variables for the imputation model: number of household members, age, aged squared, cubed root of total wealth, log household income, and dummy indicators for cohort, labor force status, gender, race/ethnicity, education, marital status, census division, 1980 census occupation code for longest reported tenure, self-reported health, ADLs, and eight doctor diagnosed health conditions. Additionally, our model accounted for retirement, hours worked, and an alternate measure of consumption that included health spending. To account for the time-series nature of the data, we included lags and leads of consumption, wealth, income, and hours worked in our imputation model. Interested readers can consult the appendix in Miller and Bairoliya (2022) for further details and diagnostic tests that indicate the procedure's effectiveness in imputing missing data in the HRS dataset.

A.2 Forecasting Model

In this section we detail our estimation and simulation procedures, which also closely follow those used by Miller and Bairoliya (2022).

A.2.1 Higher Order Lags

In order to avoid autocorrelation within the structural error terms of the model, it may be necessary to consider additional outcome lags. An extension of the VAR(1) model to higher orders is straightforward, as seen with the following VAR(2) version of our model:

$$AY_{it} = BY_{it-1} + DY_{it-2} + CX_{it} + \epsilon_{it},$$

with the block matrix form of DY_{it-2} given by:

$$\left[\begin{array}{c|cccc} D_{11} & D_{12} & D_{13} & D_{14} & D_{15} \\ \hline D_{21} & d_{22} & d_{23} & d_{24} & d_{25} \\ D_{31} & d_{32} & d_{33} & d_{34} & d_{35} \\ D_{41} & d_{42} & d_{43} & d_{44} & d_{45} \\ D_{51} & d_{52} & d_{53} & d_{54} & d_{55} \end{array}\right] \left[\begin{array}{c} M_{it-2} \\ \hline s_{it-2} \\ r_{it-2} \\ c_{it-2} \\ w_{it-2} \end{array}\right].$$

For example, the coefficient vector D_{51} in this model allows the second lag of the morbidity state vector to have a direct effect on current wealth. We can also shut down any specific lag by setting the appropriate coefficient to zero. For example, excluding the second lag of self-rated health on wealth simply implies setting $d_{52} = 0$.

A.2.2 Estimation

The forecasting model was estimated using a pooled sample of individuals born before 1960 who were aged fifty or older at the time of the survey. The sample included 40,708 unique individuals and a total of 238,091 individual-year observations. Table A.1 presents descriptive statistics for each cohort in the HRS.

Table A.1: Estimation Sample Descriptive Statistics by Cohort

	AHEAD	CODA	EHRS	LHRS	WB	BB	MBB	LBB
Individuals	7,651	4,137	5,255	5,138	3,529	4,610	5,131	4,200
Observations	36,679	27,946	45,283	46,623	28,290	24,761	18,761	5,506
Age (mean)	81.76	75.23	67.64	62.74	60.46	58.46	55.34	52.68
Hypertension (%)	54.76	57.39	53.57	50.73	49.62	49.75	47.65	45.34
Diabetes (%)	15.48	18.85	19.45	18.16	18.33	20.18	19.54	19.62
Cancer (%)	16.94	18.02	14.20	11.19	10.67	8.64	7.72	6.93
Lung disease (%)	9.48	10.19	9.57	8.52	7.20	7.01	7.67	7.59
Heart disease (%)	35.41	31.11	23.18	19.25	16.72	14.79	12.56	10.45
Stroke (%)	15.34	12.26	7.49	6.03	5.73	5.02	4.42	4.32
Psyche problem (%)	11.89	11.59	11.17	12.93	16.94	19.26	19.62	20.08
Arthritis (%)	56.06	60.25	57.59	52.70	51.81	46.35	40.01	33.01
Difficulty with ADLs (%)	40.44	28.87	23.95	21.72	21.97	21.36	19.71	14.94
Self-rated health (%)								
Poor	14.17	10.30	9.24	7.73	6.48	7.56	7.22	7.47
Fair	25.77	21.71	19.39	18.79	16.78	19.56	21.54	22.96
Good	30.91	32.31	31.61	30.99	30.61	30.16	31.16	30.95
Very good	21.38	26.41	28.11	28.95	32.23	30.46	29.36	26.92
Excellent	7.77	9.27	11.66	13.54	13.90	12.26	10.72	11.69
Retired (%)	95.46	91.47	77.30	64.98	59.33	50.21	43.15	34.91
Annual consumption ($1000s, mean)	22.48	24.99	25.02	26.25	26.81	23.39	19.98	18.18
Male (%)	37.48	46.86	45.03	45.27	37.45	42.39	42.40	39.56
Education (%)								
<HS	41.57	32.07	31.17	28.31	21.27	19.97	22.16	22.96
HS	29.71	31.79	32.93	33.13	31.11	24.74	25.19	24.30
Some College	16.45	17.86	18.58	20.45	24.40	28.46	29.71	28.75
College	12.27	18.28	17.32	18.11	23.23	26.82	22.93	23.99
Race (%)								
White	80.96	83.40	75.53	73.02	76.20	61.70	53.15	48.42
Black	12.93	9.74	16.45	16.09	15.07	21.65	26.64	28.41
Other	6.11	6.85	8.02	10.89	8.73	16.65	20.20	23.17

Notes: Children of the Depression denoted by CODA, War Babies by WB, early Baby Boomers by BB, and mid Baby Boomers by MBB. Consumption is reported in real 2010 dollars. Source: HRS.

We estimate each block in our model separately as there is no simultaneity across blocks. As shown in the methods section of the chapter, the consumption and wealth blocks only consist of one equation which follows a standard linear dynamic panel data model with lagged dependent variables and fixed effects at the individual level. We estimate these equations with OLS, but to avoid the Nickell (1981) bias that OLS can generate for this kind of model, we use the Everaert and Pozzi (2007) bootstrap-based method.[22] By including a single period lag of retirement and health on consumption, and two lags of consumption

[22]We implement the bootstrap with De Vos et al. (2015) Stata routine $xtbcfe$.

on itself, we ensure that shocks are not serially correlated in the consumption equation. Similar lags are included in the wealth equation. We also use a VAR(2) system in the retirement, health, and survival equations to maintain consistency with the consumption and wealth models. The self-rated health equation is estimated independently of other VAR blocks via maximum likelihood, while the retirement and mortality equations are estimated independently using standard probit regressions.[23]

Finally, we estimate the morbidity block, which we model as a multivariate probit with correlated shocks. To estimate this model, we use a chain of bivariate probit estimators suggested by Mullahy (2016) because of the large number of outcomes and observations in the HRS. While this approach allows for consistent estimation via maximum likelihood, a potential issue arises due to the absorbing nature of morbidity states. This means, for example, diagnosed hypertension at time t perfectly predicts hypertension at time $t+1$ and we have quasi-complete separation. In a univariate probit model, we could condition on not being diagnosed with the morbidity at time t to solve this issue, but in the bivariate probit this is not possible. Thus, we constrain the infinite coefficients to large values in the bivariate probit, but this does not affect the likelihood or estimates of remaining (non-infinite) coefficients.

The full set of estimation results are shown in Tables A.3-A.5.

A.2.3 Simulations

We used the estimated panel VAR model to construct the expected remaining lifetime utility for a subset of sixty-year-olds from the HRS. Analyses are limited to the EHRS, LHRS, War Babies, and early Baby Boomers cohorts as simulations require data at age fifty-eight and sixty as "initial" conditions. The HRS provides respondent-level analysis weights for each wave, designed to create representative cohort samples of the non-

[23]There is no incidental parameters or initial conditions problem in these models as there is no permanent unobserved heterogeneity or serial correlation. The standard (ordered) probit estimator is consistent and provides asymptotically valid test statistics and standard errors.

institutionalized US population. We used base year weights corresponding to when the cohort was approximately age sixty to analyze the welfare distribution. Specifically, we followed Miller and Bairoliya (2022) and used the 1996 analysis weights for the EHRS, 2000 for the LHRS, 2006 for War Babies, and 2008 for Baby Boomers. As any missing data was imputed among respondents, no individuals were removed from the simulation due to missing item response. However, individuals were removed if they were not surveyed at ages 58-59 and 60-61. For example, any EHRS cohort member interviewed at age 60 in 1996 but missing from the previous survey round would be excluded from the simulation sample but included in the 2000 nationally representative sample. Table A.2 provides a comparison of time invariant characteristics between the weighted representative sample and the sample used in our simulations after dropping these missing cases. The simulation sample was slightly more female, educated, and white in comparison to the representative sample, but the differences were minor and generally consistent across all cohorts.

Table A.2: Representative and Simulation Sample Comparison

	EHRS		LHRS		WB		BB	
	Rep	Sim	Rep	Sim	Rep	Sim	Rep	Sim
	0	1	2	3	4	5	6	7
Individuals	3,096	3,029	3,816	3,541	2,628	2,506	2,911	2,640
Male (%)	47.16	46.33	46.71	46.54	47.93	47.93	48.26	47.56
Education (%)								
<HS	29.27	29.11	25.41	25.48	18.49	18.18	14.52	14.50
HS	33.82	33.97	32.34	32.61	30.65	30.52	25.07	25.17
Some College	19.27	19.27	21.56	21.36	24.26	24.37	29.19	28.95
College	17.64	17.64	20.69	20.54	26.60	26.93	31.23	31.37
Race (%)								
White	83.36	83.90	81.52	81.90	82.48	83.06	79.82	80.25
Black	10.55	10.42	10.16	10.10	9.64	9.14	11.07	10.85
Other	6.09	5.68	8.32	7.99	7.88	7.80	9.12	8.89

Notes: War Babies denoted by WB and Baby Boomers by BB. EHRS cohort inclues those under age 60 in 1992. "Rep" indicates representative sample based on HRS respondent analysis weights. "Sim" indicates simulation sample weighted by the same analysis weights.

Using age sixty data as initial $(t = 0)$ conditions[24], we simulate the remaining life outcomes for each individual (i) as follows:

1. Survival shock u_{i1} is drawn and survival to time $t = 1$ (age 62) is determined according to the mortality equation. If individual survives, move to step two.

2. Morbidity shock vector e_{i1} is drawn from a standard multivariate normal distribution with estimated covariance matrix Σ (see Table A.5). This shock vector along with the model outlined in the methods section is used to compute simulated age 62 morbidity vector M_{i1}.

3. Given age 62 morbidities (M_{i1}), general health shock $\epsilon_{2,i1}$ is drawn and age 62 self-rated health (s_{i1}) is computed.

4. Given age 62 self-rated health (s_{i1}) and morbidities (M_{i1}), retirement shock $\epsilon_{3,i1}$ is drawn to determine age 62 retirement (r_{i1}).

5. Given age 62 retirement, self-rated health, and morbidities (r_{i1}, s_{i1}, M_{i1}), consumption shock $\epsilon_{4,i1}$ is drawn to determine age 62 consumption (c_{i1}).[25]

6. Given all other age 62 outcomes $(c_{i1}, r_{i1}, s_{i1}, M_{i1})$, wealth shock $\epsilon_{5,i1}$ is drawn to determine age 62 wealth (w_{i1}).

7. Steps 1-6 are repeated for $t = 2, 3, \ldots$ until death or $t = 30$ (age 120).

8. Steps 1-7 are repeated 5,000 times for each individual.

Figures A.1-A.4 show a comparison between the average simulated life-cycle profiles and those constructed from available data by race/ethnicity for the EHRS cohort. The

[24] Initial conditions also include unobserved endowments $\hat{\pi}$ estimated using the prediction method of De Vos et al. (2015).

[25] ϵ_4 is drawn from the normal distribution with mean zero and standard deviation determined to match the empirical error distribution of each cohort. Specifically, standard deviations used for EHRS, LHRS, WB, and BB cohorts are 0.49, 0.48, 0.48, and 0.40. Clustering by cohort provides a slightly better fit to the data.

simulations closely match the available aggregated data, indicating that our life-cycle dynamics model is a reasonable approximation of the underlying data generating processes. Note that the data and simulations are the same at age 60 by construction. However, the simulations match the data quite well even up to 24 years later, when the EHRS cohort reaches age 84.

To further demonstrate the accuracy of our model, we compare consumption and health utility means and standard deviations of the data with simulated life-cycle profiles for each birth cohort in Figures A.5-A.6. The simulations match the data well across birth cohorts, further highlighting the advantages of using the VAR approach to forecast joint dynamics accurately.

A.2.4 Figures and Tables

Table A.3: Model Estimates for ADLs, Self-Rated Health, Retirement, Consumption, and Mortality

	ADLs		Self-rated health		Retirement		Consumption		Mortality	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Hyper			-0.271	0.014	0.078	0.036	-0.001	0.014	0.102	0.026
Diab			-0.254	0.018	0.064	0.046	-0.003	0.014	0.100	0.033
Cancer			-0.678	0.019	0.191	0.051	0.028	0.018	0.655	0.026
Lung			-0.464	0.023	0.142	0.069	-0.005	0.020	0.409	0.032
Heart			-0.485	0.016	0.127	0.045	-0.007	0.016	0.193	0.024
Stroke			-0.480	0.022	0.464	0.074	-0.070	0.024	0.243	0.029
Psych			-0.428	0.021	0.392	0.058	-0.051	0.019	0.239	0.029
Arthritis			-0.225	0.014	0.036	0.034	0.018	0.014	-0.023	0.024
ADL			-0.664	0.013	0.368	0.039	-0.060	0.017	0.334	0.019
Health 2					-0.578	0.046	0.049	0.022	-0.332	0.017
Health 3					-0.725	0.047	0.062	0.023	-0.544	0.019
Health 4					-0.741	0.049	0.086	0.023	-0.662	0.022
Health 5 (best)					-0.725	0.054	0.114	0.027	-0.652	0.032
Lag Hyper	0.037	0.031	0.147	0.019	-0.032	0.049	-0.005	0.012	-0.046	0.026
Lag Diab	0.080	0.039	0.084	0.025	-0.010	0.067	-0.003	0.015	0.070	0.034
Lag Cancer	0.028	0.043	0.524	0.028	-0.121	0.079	-0.007	0.015	-0.447	0.028
Lag Lung	0.163	0.049	0.198	0.033	0.007	0.104	-0.010	0.025	-0.120	0.034
Lag Heart	0.075	0.033	0.277	0.022	-0.162	0.068	0.003	0.016	-0.038	0.025
Lag Stroke	0.372	0.046	0.348	0.032	-0.253	0.123	-0.003	0.019	-0.053	0.032
Lag Psych	0.360	0.043	0.241	0.030	-0.141	0.087	0.025	0.018	-0.149	0.032
Lag Arthritis	0.210	0.026	0.115	0.018	0.051	0.045	-0.009	0.012	-0.080	0.023
Lag ADL			0.326	0.018	-0.170	0.057	0.005	0.014	-0.117	0.020
Lag Health 2	-0.229	0.030	0.621	0.014	0.012	0.061	0.012	0.012	-0.039	0.018
Lag Health 3	-0.470	0.031	1.120	0.015	-0.029	0.062	0.013	0.015	-0.078	0.020
Lag Health 4	-0.641	0.033	1.650	0.016	-0.064	0.064	0.014	0.015	-0.115	0.023
Lag Health 5	-0.722	0.040	2.272	0.018	-0.076	0.067	0.017	0.015	-0.148	0.031
Time	-0.049	0.007	0.019	0.003	-0.014	0.010	0.004	0.010	-0.014	0.005
2008+	0.030	0.025	0.011	0.012	-0.023	0.033	-0.044	0.009	0.045	0.021
CODA	0.102	0.031	0.021	0.016	0.092	0.078			-0.010	0.024
Early HRS	0.132	0.044	0.014	0.022	0.091	0.092			-0.047	0.033
Late HRS	0.138	0.057	0.004	0.028	0.014	0.106			-0.064	0.043
War Babies	0.168	0.071	-0.017	0.034	0.064	0.123			-0.116	0.055
Boomers	0.267	0.086	-0.087	0.042	0.040	0.145			-0.143	0.067
Mid Boomers	0.332	0.102	-0.135	0.050	-0.012	0.165			-0.199	0.083
Late Boomers	0.334	0.150	-0.147	0.070	-0.046	0.202			-0.088	0.161
Black	0.098	0.019	-0.066	0.009	0.054	0.023			0.042	0.016
Other race	0.062	0.024	-0.129	0.011	0.010	0.030			-0.169	0.022
Female	-0.011	0.014	0.036	0.007	0.118	0.017			-0.219	0.013
HS grad	-0.083	0.016	0.076	0.008	-0.026	0.023			0.027	0.014
Some college	-0.040	0.019	0.113	0.009	-0.046	0.025			0.009	0.017
College grad	-0.091	0.023	0.188	0.010	-0.052	0.028			-0.007	0.020
Retired							-0.041	0.011	0.202	0.032
Lag Retired	0.104	0.028	-0.026	0.013			-0.024	0.011	-0.022	0.028
Lag2 Retired	0.001	0.026	-0.015	0.012						
Lag Con							0.170	0.005		
Lag2 Con							0.081	0.005		
Constant	-0.927	0.077			-0.849	0.184			-1.763	0.249

Notes: Dependent variable across columns. Multivariate probit results reported for ADLs as dependent outcome. Standard (ordered) probit results reported for self-rated health, mortality, and retirement as dependant outcomes. Linear dynamic panel estimates reported for consumption as outcome. All regressions also include dummies for age. Regressions for ADLs, self-rated health, mortality, and retirement also include dummies for occupation and census division. Regressions for ADLs and self-rated health also includes second lag for all health outcomes.

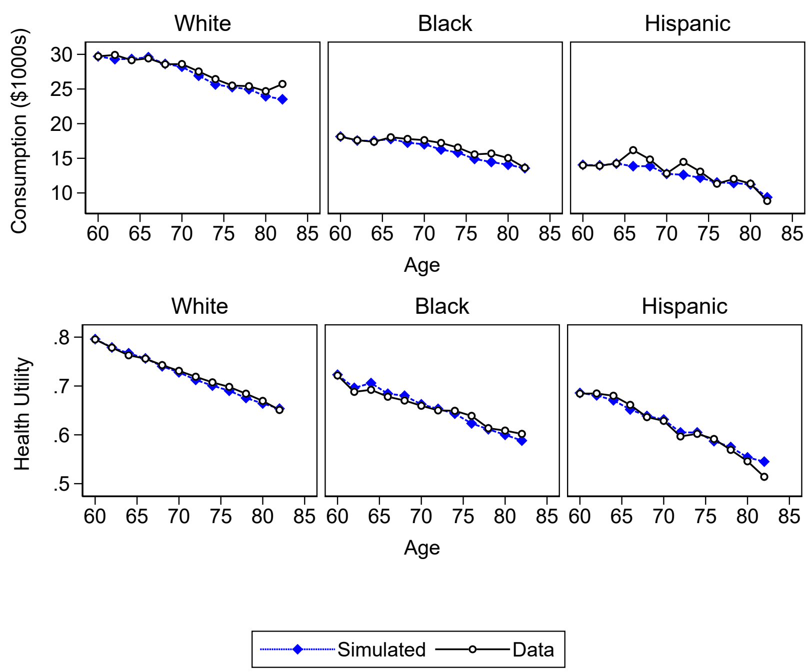
Table A.4: Model Estimates for Morbidities

	Hypertension		Diabetes		Cancer		Lung disease		Heart disease		Stroke		Psych		Arthritis	
Variable	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	SE	SE
Lag Hyper			0.251	0.034	-0.027	0.038	0.087	0.040	0.125	0.033	0.106	0.040	0.143	0.037	0.071	0.033
Lag Diab	0.248	0.050			0.066	0.046	0.052	0.050	0.052	0.043	0.039	0.052	0.027	0.050	0.062	0.044
Lag Cancer	-0.037	0.051	0.015	0.053			0.039	0.058	-0.083	0.049	-0.030	0.058	-0.043	0.059	0.055	0.050
Lag Lung	0.110	0.058	0.093	0.056	0.084	0.058			0.263	0.051	0.018	0.062	0.100	0.062	0.155	0.065
Lag Heart	0.098	0.043	0.085	0.040	0.007	0.041	0.204	0.040			0.172	0.040	0.064	0.041	0.080	0.042
Lag Stroke	0.082	0.066	-0.046	0.063	-0.013	0.057	-0.020	0.062	0.069	0.053			0.256	0.051	-0.040	0.062
Lag Psych	0.053	0.055	0.070	0.051	-0.055	0.058	0.087	0.056	0.083	0.047	0.154	0.053			0.270	0.054
Lag Arthritis	0.085	0.029	-0.006	0.033	-0.012	0.034	0.139	0.036	0.068	0.030	-0.002	0.036	0.117	0.035		
Lag ADL	0.057	0.033	0.029	0.034	0.017	0.034	0.080	0.036	0.075	0.030	0.180	0.032	0.209	0.032	0.167	0.036
Lag Health 2	0.011	0.035	0.001	0.032	-0.045	0.034	-0.069	0.032	-0.101	0.030	-0.122	0.032	-0.207	0.031	-0.070	0.037
Lag Health 3	0.004	0.036	-0.009	0.034	-0.075	0.036	-0.147	0.035	-0.164	0.031	-0.210	0.034	-0.296	0.033	-0.108	0.038
Lag Health 4	-0.036	0.038	-0.083	0.037	-0.106	0.038	-0.305	0.040	-0.250	0.034	-0.240	0.038	-0.388	0.038	-0.152	0.040
Lag Health 5 (best)	-0.119	0.042	-0.215	0.045	-0.139	0.045	-0.442	0.055	-0.291	0.040	-0.362	0.050	-0.476	0.050	-0.245	0.044
Lag2 Hyper			0.039	0.033	0.056	0.038	-0.091	0.040	0.049	0.032	0.036	0.040	-0.087	0.037	0.031	0.033
Lag2 Diab	-0.086	0.053			-0.068	0.049	-0.122	0.053	0.105	0.045	0.087	0.054	-0.008	0.052	-0.056	0.047
Lag2 Cancer	0.018	0.055	-0.011	0.057			0.043	0.061	0.075	0.053	-0.005	0.061	0.061	0.063	-0.014	0.054
Lag2 Lung	-0.167	0.064	-0.073	0.061	0.044	0.063			-0.124	0.056	0.036	0.067	0.019	0.067	-0.077	0.072
Lag2 Heart	-0.049	0.046	0.018	0.042	0.014	0.043	-0.091	0.042			-0.026	0.041	-0.056	0.043	-0.019	0.044
Lag2 Stroke	-0.025	0.075	0.073	0.069	0.002	0.063	0.038	0.068	0.053	0.059			-0.165	0.058	0.026	0.069
Lag2 Psych	-0.019	0.058	-0.063	0.055	0.056	0.061	0.051	0.059	-0.019	0.050	-0.041	0.056			-0.131	0.058
Lag2 Arthre	-0.040	0.030	0.000	0.033	0.049	0.034	-0.044	0.035	0.024	0.029	-0.021	0.036	-0.020	0.034		
Lag2 ADL	-0.060	0.035	0.037	0.036	-0.019	0.036	0.003	0.037	0.005	0.031	-0.093	0.034	-0.061	0.034	-0.055	0.041
Lag2 Health 2	-0.018	0.037	-0.075	0.033	-0.041	0.036	-0.089	0.034	0.008	0.032	-0.055	0.034	-0.069	0.033	0.054	0.041
Lag2 Health 3	-0.019	0.037	-0.070	0.035	-0.009	0.037	-0.144	0.036	-0.015	0.034	-0.065	0.037	-0.115	0.035	0.068	0.042
Lag2 Health 4	-0.033	0.039	-0.113	0.038	-0.002	0.040	-0.197	0.040	-0.043	0.036	-0.053	0.040	-0.203	0.039	0.032	0.043
Lag2 Health 5	-0.060	0.042	-0.143	0.044	0.001	0.045	-0.292	0.052	-0.089	0.041	-0.096	0.048	-0.264	0.049	-0.036	0.047
Time	0.039	0.007	0.029	0.008	0.005	0.008	0.013	0.009	-0.003	0.007	-0.023	0.008	0.006	0.008	-0.033	0.007
2008+	-0.061	0.028	-0.063	0.030	0.015	0.030	-0.015	0.034	-0.047	0.026	0.013	0.032	-0.108	0.033	0.021	0.028
CODA	-0.031	0.038	-0.036	0.042	-0.019	0.039	0.016	0.044	-0.015	0.035	0.017	0.039	0.075	0.042	-0.097	0.038
Early HRS	-0.084	0.053	-0.072	0.058	-0.075	0.054	-0.036	0.062	0.013	0.048	0.004	0.054	0.073	0.058	-0.097	0.053
Late HRS	-0.083	0.066	-0.082	0.073	-0.094	0.069	0.020	0.079	0.042	0.062	0.014	0.069	0.113	0.074	0.004	0.067
War Babies	-0.093	0.082	-0.038	0.090	-0.072	0.086	0.014	0.099	0.059	0.076	0.087	0.087	0.236	0.090	0.121	0.082
Boomers	-0.175	0.100	-0.027	0.110	-0.119	0.106	0.005	0.121	0.103	0.093	0.079	0.106	0.326	0.110	0.166	0.100
Mid Boomers	-0.307	0.118	-0.013	0.129	-0.090	0.127	0.069	0.144	0.126	0.112	0.145	0.128	0.313	0.130	0.200	0.118
Late Boomers	-0.309	0.159	0.117	0.167	-0.208	0.204	0.002	0.220	0.130	0.167	0.479	0.202	0.038	0.200	0.245	0.156
Black	0.197	0.022	0.118	0.021	-0.049	0.022	-0.185	0.026	-0.157	0.020	0.036	0.024	-0.199	0.025	-0.020	0.021
Other race	0.100	0.025	0.260	0.025	-0.200	0.032	-0.282	0.035	-0.218	0.027	-0.117	0.033	-0.017	0.030	-0.105	0.026
Female	0.015	0.015	-0.108	0.017	-0.196	0.017	-0.039	0.020	-0.173	0.015	-0.064	0.019	0.128	0.019	0.161	0.015
HS grad	-0.030	0.019	-0.056	0.020	-0.013	0.020	-0.108	0.022	0.005	0.018	0.035	0.022	-0.065	0.022	-0.046	0.020
Some college	-0.066	0.021	-0.061	0.023	0.026	0.023	-0.110	0.026	0.021	0.021	0.044	0.026	0.000	0.025	-0.010	0.022
College grad	-0.106	0.024	-0.120	0.027	0.033	0.027	-0.226	0.032	-0.053	0.024	0.033	0.030	-0.040	0.030	-0.044	0.024
Lag Retired	-0.036	0.030	0.050	0.031	0.030	0.034	0.057	0.041	0.005	0.031	0.048	0.043	0.085	0.038	0.003	0.029
Lag2 Retired	0.021	0.029	-0.047	0.030	-0.004	0.032	0.004	0.038	0.004	0.030	0.033	0.039	-0.033	0.036	-0.022	0.028
Constant	-1.549	0.086	-2.027	0.093	-1.908	0.093	-2.015	0.107	-1.726	0.084	-2.506	0.112	-1.866	0.097	-1.311	0.087

Notes: Multivarite probit results with dependent variable across columns. Regressions also include dummies for age, occupation, and census division.

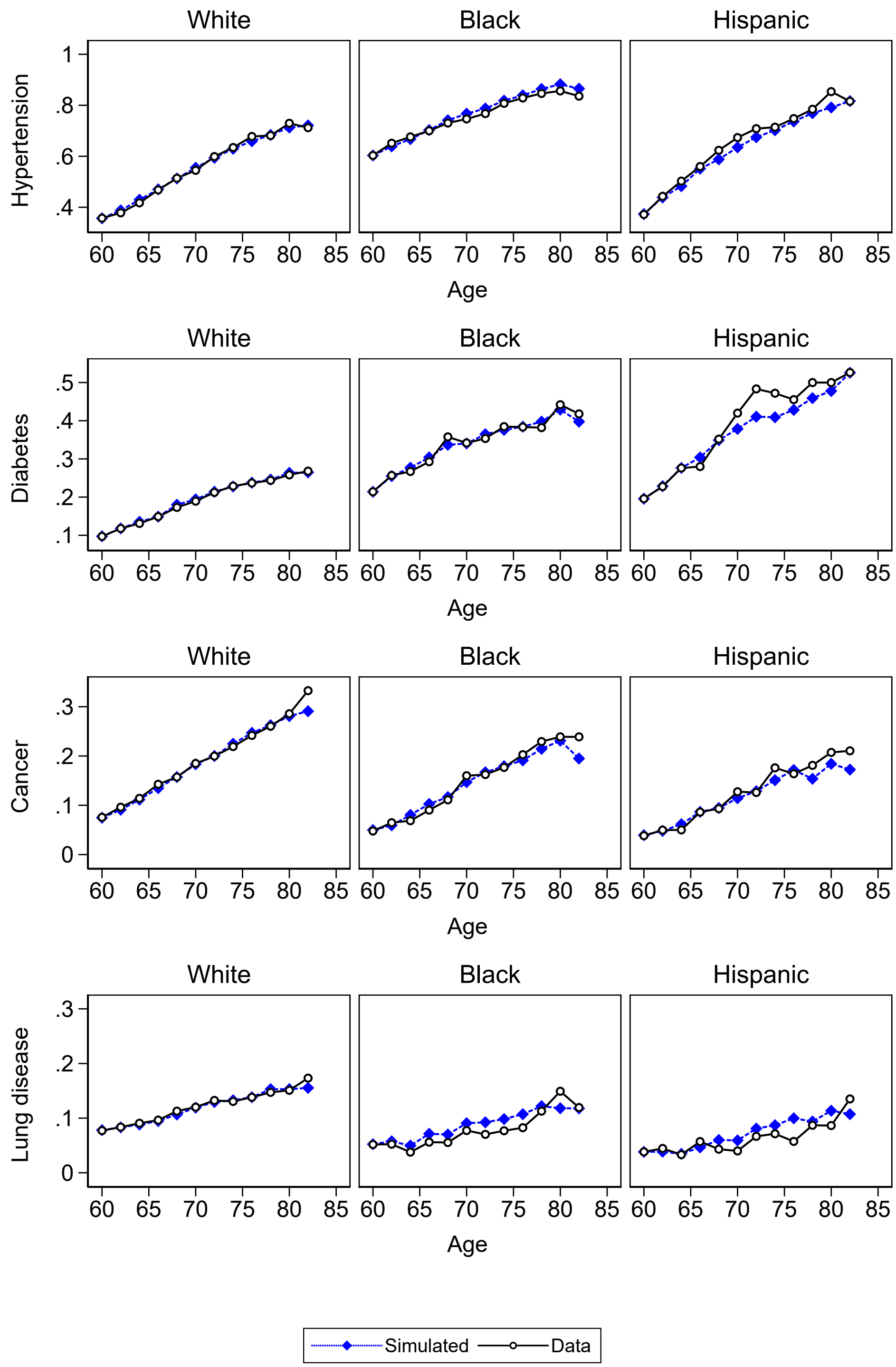
Table A.5: Morbidity Shock Covariance Matrix (Σ)

	Hyper	Diabetes	Cancer	Lung	Heart	Stroke	Psych	Arthritis	ADLs
Hyper	1.00	0.27	0.05	0.09	0.28	0.29	0.14	0.09	0.11
Diabetes	0.27	1.00	0.06	0.06	0.10	0.14	0.08	0.04	0.07
Cancer	0.05	0.06	1.00	0.13	0.04	0.06	0.12	0.06	0.13
Lung	0.09	0.06	0.13	1.00	0.23	0.11	0.18	0.08	0.19
Heart	0.28	0.10	0.04	0.23	1.00	0.28	0.16	0.09	0.14
Stroke	0.29	0.14	0.06	0.11	0.28	1.00	0.21	0.11	0.40
Psych	0.14	0.08	0.12	0.18	0.16	0.21	1.00	0.16	0.29
Arthritis	0.09	0.04	0.06	0.08	0.09	0.11	0.16	1.00	0.26
ADLs	0.11	0.07	0.13	0.19	0.14	0.40	0.29	0.26	1.00



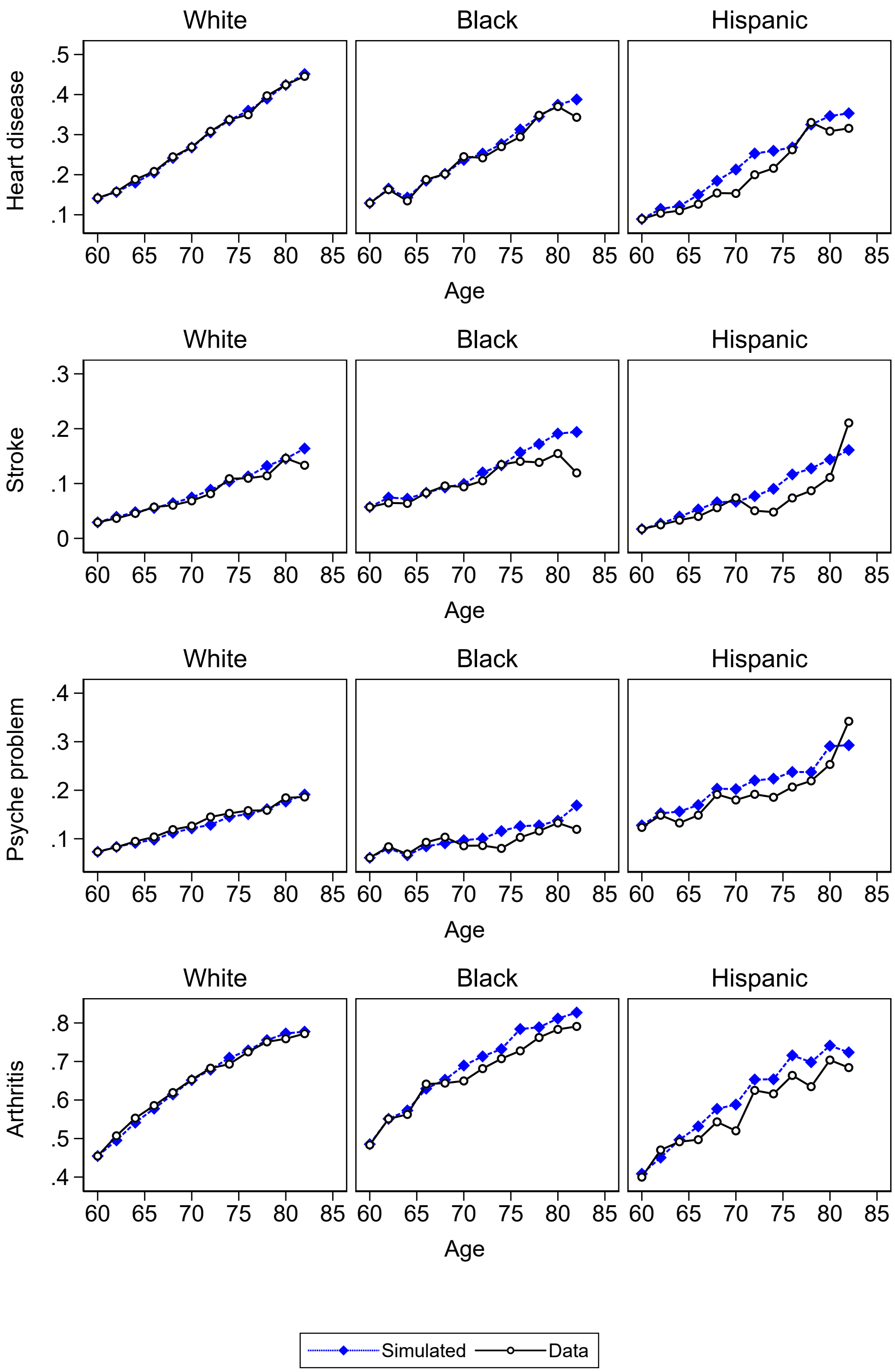
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure A.1: Mean of Life-Cycle Consumption and Health Utility Profiles by Race/Ethnicity



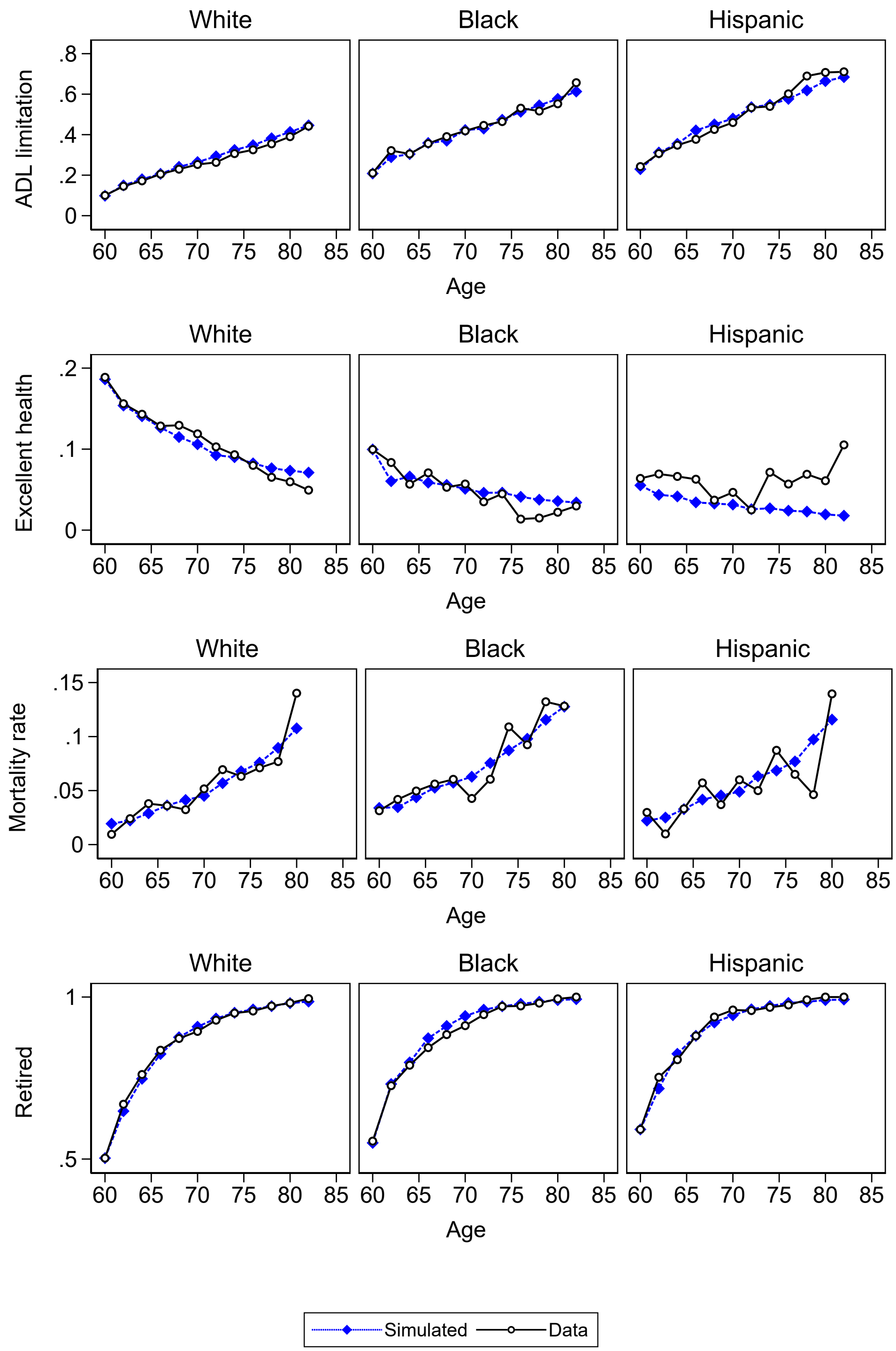
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure A.2: Mean of Life-Cycle Morbidity Profiles by Race/Ethnicity



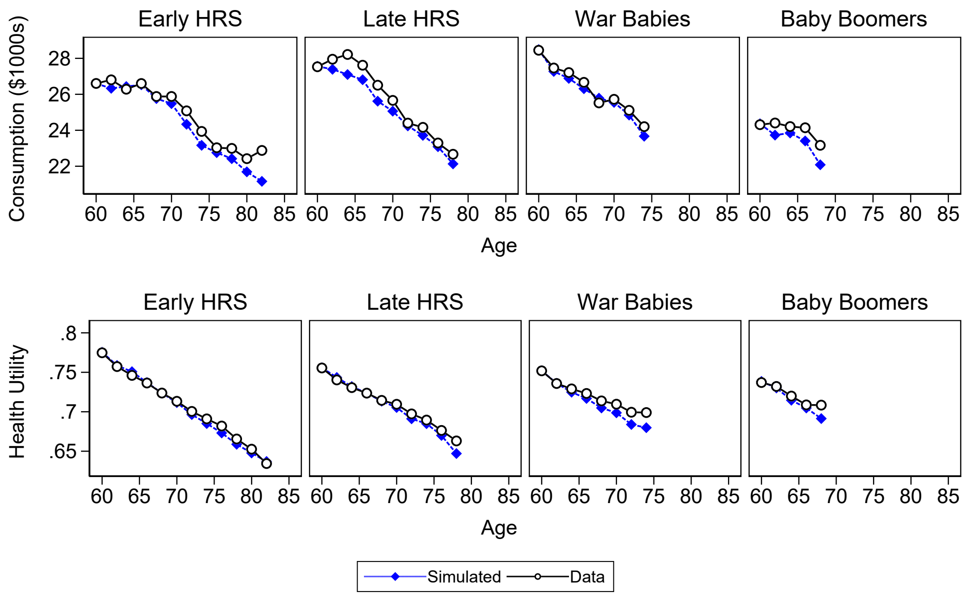
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure A.3: Mean of Life-Cycle Morbidity Profiles by Race/Ethnicity



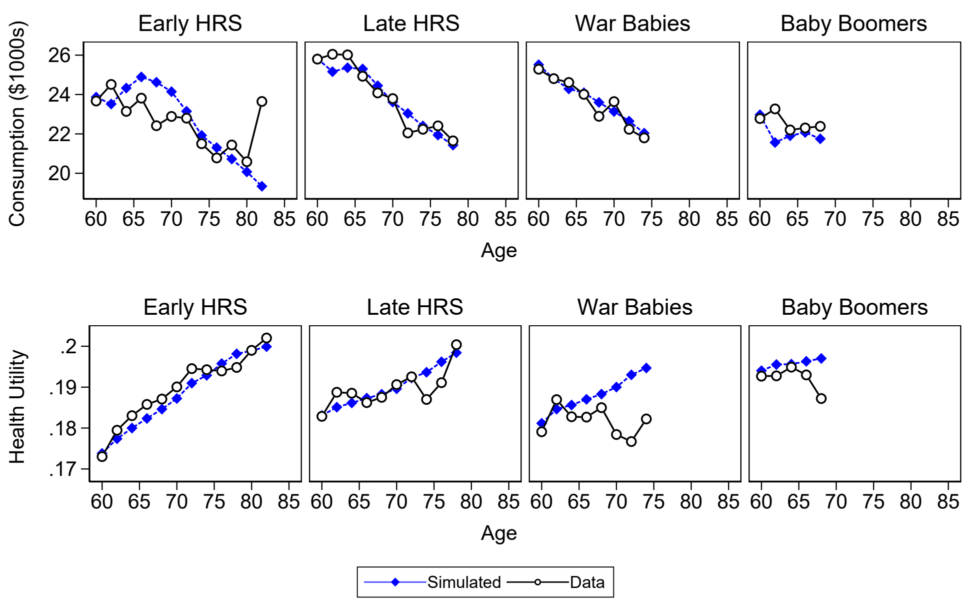
Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in the EHRS cohort by two-year age interval. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure A.4: Mean of Life-Cycle Health, Mortality, and Retirement Profiles by Race/Ethnicity



Notes: "Data" plots mean of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of expected simulated outcome for each observation in the data (i.e. the expected outcome for each person-year observation in the data).

Figure A.5: Mean of Life-Cycle Consumption and Health Utility Profiles by Cohort



Notes: "Data" plots standard deviation of all available data (inclusive of imputed missing values) in HRS by two-year age interval and cohort. "Simulated" plots mean of standard deviations of simulated outcome (i.e. the mean of standard deviations calculated for each of the 5,000 simulation runs).

Figure A.6: Standard Deviation of Consumption and Health Utility Life-Cycle Profiles by Cohort

A.3 Health Utility Weights

We obtain our health utility weights ω from the Health Utilities Index Mark 3 (HUI3) instrument, which was administered to around 1,200 participants in the HRS in 2000. The HUI3 instrument was designed to produce cardinal utility scores on the standard utility scale of 0 (death) to 1 (best health) and has been widely used in studies on health utilities (Furlong et al., 1998; Feeny et al., 2002; Horsman et al., 2003). We use the HUI multi-attribute utility score (*hui3ou*) for our analysis.

The HUI3 was conceptualized such that $u(h_i) = HUI3_i \times u(h_{best})$ for individual i and general utility function $u(.)$, where h_{best} refers to the best possible health state. For example, a year in the best health state is equivalent in utility to two years with $HUI3 = 0.5$. For our model, we adopt the approach of

Table A.6: Estimated Health Utility Weights (γ)

Measure	Weight	SE
Self-rated health		
Fair	0.229	0.026
Good	0.314	0.026
Very good	0.406	0.027
Excellent	0.421	0.031
Hypertension	0.003	0.012
Diabetes	-0.003	0.018
Cancer	0.009	0.017
Lung disease	-0.027	0.022
Heart disease	-0.031	0.015
Stroke	-0.077	0.022
Psych problem	-0.070	0.020
Arthritis	-0.062	0.013
Diff with ADL	-0.158	0.017
Constant	0.516	0.028

Notes: Results from regression of adjusted HUI3 score on self-rated health and morbidities. SE denotes standard error. $R^2 = 0.17$. N = 760.

While this approach is consistent with the interview instructions of the survey, other researchers have questioned whether respondents are fully capable of conceptualizing changes in health states without also considering changes in other aspects of life (Feeny et al., 2018). For instance, respondents may have considered changes in consumption and leisure along with changes in health. In such cases, the appropriate representation of the HUI3 instrument would be as follows:

$$\gamma h \left[\bar{u} + log\left(c\right) + \nu\left(l\right)\right] = HUI3 \times h_{best}\left[\bar{u} + log\left(c_{best}\right) + \nu\left(l_{best}\right)\right].$$

Rearranging terms and setting $h_{best} = 1$ yields:

$$\gamma h = HUI3 \frac{\bar{u} + log\left(c_{best}\right) + \nu\left(l_{best}\right)}{\bar{u} + log\left(c\right) + \nu\left(l\right)}. \quad \text{(A.1)}$$

However,

Table A.7: Estimated Alternate Health Utility Weights (γ)

Measure	Weight	SE
Self-rated health		
Fair	0.263	0.040
Good	0.328	0.039
Very good	0.413	0.041
Excellent	0.401	0.046
Hypertension	-0.002	0.018
Diabetes	0.012	0.025
Cancer	0.004	0.024
Lung disease	-0.036	0.031
Heart disease	-0.047	0.022
Stroke	-0.048	0.031
Psych problem	-0.061	0.029
Arthritis	-0.057	0.020
Diff with ADL	-0.132	0.024
Constant	0.507	0.041

Notes: Results from regression of adjusted HUI3 score on self-rated health and morbidities. SE denotes standard error. $R^2 = 0.17$. N = 760.

A.4 Additional Welfare Results

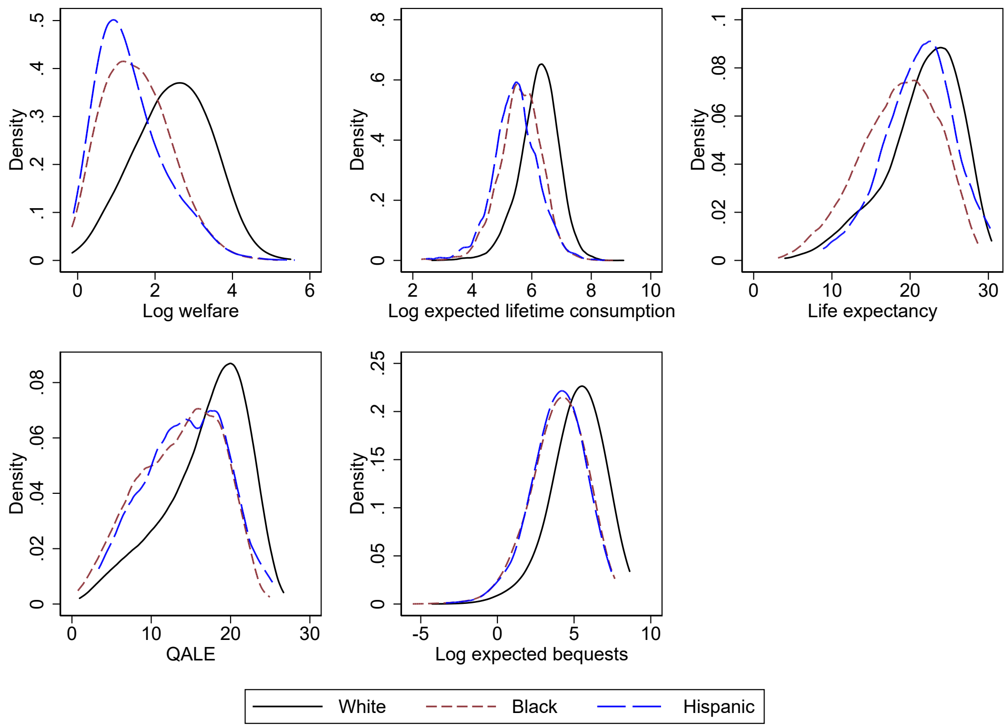


Figure A.7: Distribution of Welfare, Consumption, and Life Expectancy by Race/Ethnicity

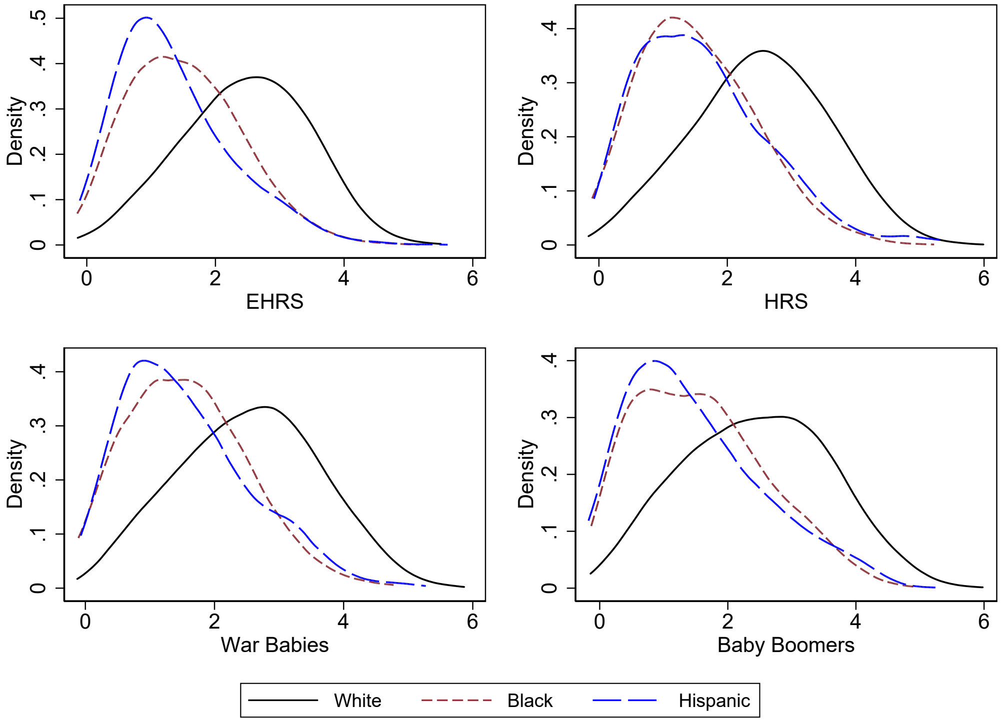


Figure A.8: Distribution of Log Welfare by Race/Ethnicity and Cohort

Appendix B

Construction of Expressed Sentiment from Twitter Data, Challenges in Collecting Tweets with Geolocation and Timestamps, and Supplementary Tables and Figures

B.1 Accessing Twitter's Academic Research API

In this section, we outline the process we used to acquire the Twitter data that were utilized in this chapter. To obtain these data, we utilized Twitter Academic Research API, which grants access to a portion of Twitter's historical public Twitter posts when users make requests. Twitter's license agreement for this API prohibits sharing the raw Twitter data publicly.

To access Twitter's data, the first step is to sign up for an API key through Twitter's Developer Program (https://developer.twitter.com/en/docs/twitter-api). This procedure will give us a set of keys, including API key, API secret key, Access token, and Access token secret, that allow us to access data. Nevertheless, it is important to note that users must agree to Twitter's terms and conditions before using the data. We strongly recommend that users carefully read the terms, particularly because the location information of Twitter's account owners is sensitive.

Next, we wrote a Python script that allowed us to request all historical Twitter data that were geolocated and timestamped within the area given by Figure B.1.[26]

[26] A sample version of this script can be obtained here: https://github.com/schinlfc/fetch_tweet_api.

Figure B.1: Tweet Collection Map

B.2 Challenges in Collecting Tweets with Geolocation and Timestamps

B.2.1 Geolocated Tweets

The collection of tweets with geolocation presented several challenges. Each geolocated tweet from Twitter is attached with a *"place_id"* that needs to be translated into a state. To tackle this problem, we explored four approaches:

1. **Centroid:** The first approach involved using the centroid field of each place and mapping it back into a state using a state shapefile. However, this approach created an issue where some places referred to a broader geographic area, but their centroid would always be in a specific location, creating artificial certainty and leading to errors. For instance, this news story [27] illustrates the problems that arise from this approach. Thus, we abandoned this approach.

2. **Bounding Box:** The second approach involved taking the bounding box of the entire place and matching it against the state shapefile. However, this approach had difficulty with places like "South Carolina, USA". The issue is that if we draw a

[27] https://splinternews.com/how-an-internet-mapping-glitch-turned-a-random-kansas-f-1793856052.

rectangular bounding box around South Carolina and check which states it intersects, it intersects with many states besides South Carolina. To address this problem, we introduced an 80% rule: if 80% of the intersection is in a single state, then the point is resolved to that state. However, this approach did not solve the South Carolina case.

3. **Name:** Our third approach was to use the name of the location. With only 50 states, we were able to hard code a mapping of "South Carolina, USA" to "SC". For cities, the name includes the state, so we wrote a regex (regular expression) to extract the state name from these locations.

4. **Parent Place:** Twitter places have a *"contained_within"* attribute, which we tried to use to match poorly behaved places. However, in cases where a place was difficult to match, its parent place was even harder to match.

In the end, we decided on the following solution: (1) attempt a place name match, (2) if that fails, try a geospatial match, and (3) consult an override file, giving precedence to the override if it matches over the other two rules. Here are some edge cases:

1. "Washington Heights, Manhattan": The place name rule fails because the name does not include a state, and the 80% rule fails because it is too close to the border. Therefore, we hard-coded this exception.

2. "Lincoln Tunnel": This is a tunnel between New York and New Jersey, jointly managed by both states. It does not make sense to assign it to either state. Therefore, we dropped these.

3. "Thomas A Edison Middle School": This place has a bounding box that is only a single point, causing the geospatial match to fail. To solve this, we buffered the bounding box by 0.0001 degrees, approximately 10 meters, in cases where the bounding box had zero area.

4. "Kansas City, MO": Due to the proximity of the city to the Kansas border, the geospatial match struggles with this city. Issues like these led us to prioritize the place name match and fall back to the geospatial match for cities.

B.2.2 Timestamped Tweets

We faced a challenge with the date and time of tweets from Twitter. These are presented in Coordinated Universal Time (UTC), but we need them to be in the local time zone of the user for our analysis. The reason for this is that we want to identify the time of day when the user tweeted. There are two possible solutions to this problem: (1) obtain the time zone from the user information on Twitter, or (2) use the latitude and longitude coordinates and a shapefile of time zones to determine the time zone of the tweets. However, we found out that Twitter had removed the timezone information from their API in 2018, forcing us to consider alternative approaches.

We wrote a Python script to identify the timezone of a location. To our surprise, 98% of locations that were successfully geocoded to a state also corresponded to a single timezone, meaning that we did not require a complex approach. However, we still needed to account for the remaining locations, for which we developed three possible solutions:

1. We took an average of different time zones if a place spans more than one time zone. For instance, if a place is in UTC-6 for 40% and in UTC-5 for 60%, we could compromise and say it is UTC-5:24.

2. We determined which time zones are synonyms for our purposes. For example, if a place is detected as being in both Wisconsin and the upper peninsula of Michigan, but both regions have the same time zone, then this is not an issue.

3. We used Open Street Maps' Nominatim service to look up place names for a location, which would allow us to replace a bounding-box lookup with a more specific polygon lookup.

B.3 Measuring Expressed Sentiment

Twitter is a platform for posting microblog messages, which are called tweets. These messages are limited to 140 characters and may include text, URLs, pictures, usernames, and emoticons. However, they may also contain misspellings and other irrelevant information. To make the data more suitable for mining and feature extraction, we carried out a series of preprocessing steps to remove irrelevant information and eliminate duplicate tweets and retweets. Each of which was processed to extract its main message using Python's Natural Language Toolkit (NLTK).

To preprocess the data, we first used a regular expression (Regex) in Python to detect and discard special characters, such as URLs ("http://url"), retweet (RT), user mention (@), and unwanted punctuation. We kept hashtags (#) as part of the tweet because they often explain the subject of the tweet and contain useful information related to the topic. However, we removed the "#" symbol.

Next, we used various functions of NLTK to convert the tweets to lowercase, remove stop words (i.e., words that do not express any meaning, such as is, a, the, he, them, etc.), tokenize the tweets into individual words or tokens. Once these preprocessing steps were complete, the dataset was ready for sentiment classification. By cleaning the data and removing irrelevant information, we improved the accuracy of the results.

To ensure that the responses only capture changes in general sentiment due to transitions out of and into DST, we removed tweets that contain any DST-related terms.

B.3.1 VADER

We applied the VADER Sentiment Analyzer to the dataset to categorize the sentiments expressed in the tweets. VADER is a widely used rule-based sentiment analysis tool and lexicon that is particularly useful for analyzing social media content (Hutto and Gilbert, 2014). To classify our dataset, we first developed a sentiment intensity analyzer and then used the polarity scores method to determine the sentiment. We then used the VADER

Sentiment Analyzer to classify the preprocessed tweets as positive, negative, neutral, or compound. The compound value is an essential metric for measuring the sentiment of a given tweet. In our proposed method, we used threshold values to categorize tweets as positive, negative, or neutral. The following are the typical threshold values we used in this study: For positive sentiment, we assigned a score of 1 if the compound value was greater than 0.001. For neutral sentiment, we assigned a score of 0 if the compound value was between -0.001 and 0.001. For negative sentiment, we assigned a score of -1 if the compound value was less than -0.001.

B.3.2 AFINN

AFINN determines the sentiment of English words by using a word list or dictionary. To calculate the sentiment of a given text, AFINN calculates the average score of all the words in that text. The AFINN-165 dictionary, created by experts, is used to map words to emotional states. This dictionary consists of $3,382$ words, each with a sentiment score ranging from -5 to 5. A negative emotional state is represented by -5 and a positive emotional state is represented by 5. Examples of word lists used to build the AFINN sentiment score are included in Table B.1.

Table B.1: AFINN Word-Score Examples

Negative		Neutral		Positive	
abuse	-3	achievable	1	adores	3
agonising	-3	affronted	-1	amusements	3
animosity	-2	anticipation	1	appreciate	2
fraud	-4	free	1	fun	4
niggas	-5	oxymoron	-1	superb	5

Notes: The raw scores are displayed, while the standardized scores are utilized in the analysis. A complete list of 3,382 total word-score mappings is available here: https://github.com/fnielsen/afinn/blob/master/afinn/data/AFINN-en-165.txt.

B.3.3 BERTweet

BERTweet is a unique large-scale language model that was specifically pre-trained for English tweets, making it the first of its kind. The pre-training process utilized the RoBERTa procedure, as detailed in Nguyen et al. (2020). Several experiments have demonstrated that BERTweet surpasses strong baseline models, including RoBERTa base and XLM-R base, on different natural language processing tasks (Bansal et al., 2022). To classify tweet sentiment as either positive, neutral, or negative, BERTweet employs a fine-tuning approach. In fine-tuning, the pre-trained language model is trained on a specific downstream task, in this case, sentiment analysis.

The fine-tuning process for sentiment analysis using BERTweet involves utilizing a vast dataset of tweets labeled as positive, neutral, or negative. The dataset is divided into training, validation, and test sets. The training set is utilized to optimize the model's parameters, the validation set is used to adjust the model's hyperparameters, and the test set is used to evaluate the model's performance. During the fine-tuning process, the tweet's input text is fed to BERTweet, and it generates a vector representation of the tweet. This vector is then passed through a classification layer that maps it to one of the three possible sentiment classes: positive, neutral, or negative. The classification layer employs a softmax function to produce a probability distribution over the three classes, and the sentiment label with the highest probability is chosen as the tweet's final sentiment label.

In the context of sentiment analysis with BERTweet, the softmax function is used to produce a probability distribution over the three possible sentiment classes: positive, neutral, and negative. Given a vector of scores, the softmax function takes the exponentials of each score and normalizes them so that they sum to 1. These values signify the likelihood that each class is the correct label for the input tweet. For example, if BERTweet produces a vector of scores of $[3.1, -0.5, 1.7]$ for a particular tweet, the softmax function would first calculate the exponentials of each score:

$$exp(3.1) = 22.20$$

$$exp(-0.5) = 0.61$$

$$exp(1.7) = 5.43$$

Then these values are normalized by dividing them by their sum:

$$(22.20/(22.20 + 0.61 + 5.43)) = 0.787$$

$$(0.61/(22.20 + 0.61 + 5.43)) = 0.018$$

$$(5.43/(22.20 + 0.61 + 5.43)) = 0.195$$

The resulting values of 0.787, 0.018, and 0.195 represent the probabilities that the tweet is positive, neutral, or negative, respectively. It is worth noting that these probabilities sum to 1.

B.4 Supplementary Tables and Figures

This section will present our findings for the AFINN and VADER sentiment scores, as well as revisiting the previous section on the differential returns to sunlight. We will estimate alternative specifications by incorporating different sets of fixed effects for both AFINN and VADER. Additionally, we will revisit our short-run and long-run estimates, and estimate alternative specifications by incorporating different sets of fixed effects for BERT. Finally, we will provide the parallel trend between morning and evening sentiments obtained from our predicted model.

B.4.1 Short-Run Estimates: AFINN and VADER

Table B.2: The Short-Run Effects of Fall Back and Spring Forward on AFINN

Variables	Fall Back	Spring Forward
POST (α_1)	0.010	-0.029***
	(0.008)	(0.005)
EVENING (α_2)	-0.049***	-0.045***
	(0.008)	(0.005)
POST $\times$ EVENING (α_3)	-0.018**	0.020***
	(0.008)	(0.006)
$\alpha_1 + \alpha_3$	-0.008**	-0.009***
	(0.004)	(0.003)
Constant	-0.164	-0.199**
	(0.105)	(0.089)
Observations	383,653	527,567
R-squared	0.003	0.003

Notes: Dependent variable: Standardized AFINN. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10.

Table B.3: The Short-Run Effects of Fall Back and Spring Forward on VADER

Variables	Fall Back	Spring Forward
POST (α_1)	0.027***	-0.033***
	(0.007)	(0.005)
EVENING (α_2)	-0.016**	-0.020***
	(0.008)	(0.005)
POST $\times$ EVENING (α_3)	-0.033***	0.023***
	(0.008)	(0.006)
$\alpha_1 + \alpha_3$	-0.006	-0.009***
	(0.004)	(0.003)
Constant	-0.145	-0.106
	(0.105)	(0.089)
Observations	383,653	527,567
R-squared	0.002	0.002

Notes: Dependent variable: Standardized VADER. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10.

B.4.2 Long-Run Estimates: AFINN and VADER

Table B.4: The Long-Run Effects of Fall Back and Spring Forward on AFINN

Variables	Fall Back	Spring Forward
POST (α_1)	0.028***	-0.026***
	(0.007)	(0.006)
EVENING (α_2)	-0.032***	-0.053***
	(0.008)	(0.005)
POST $\times$ EVENING (α_3)	-0.022***	0.024***
	(0.009)	(0.007)
$\alpha_1 + \alpha_3$	0.006	-0.001
	(0.004)	(0.003)
Constant	-0.121	-0.178**
	(0.109)	(0.085)
Observations	353,394	516,819
R-squared	0.003	0.003

Notes: Dependent variable: Standardized AFINN. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10.

Table B.5: The Long-Run Effects of Fall Back and Spring Forward on VADER

Variables	Fall Back	Spring Forward
POST (α_1)	0.042***	-0.031***
	(0.007)	(0.006)
EVENING (α_2)	0.011	-0.029***
	(0.008)	(0.005)
POST $\times$ EVENING (α_3)	-0.033***	0.035***
	(0.009)	(0.007)
$\alpha_1 + \alpha_3$	0.009**	0.004
	(0.004)	(0.003)
Constant	-0.131	-0.185**
	(0.110)	(0.085)
Observations	353,394	516,819
R-squared	0.002	0.002

Notes: Dependent variable: Standardized VADER. Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Standard errors are in parenthesis. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

B.4.3 Additional Robustness

Table B.6: Sensitivity of the Short-Run Effects of Fall Back on BERT

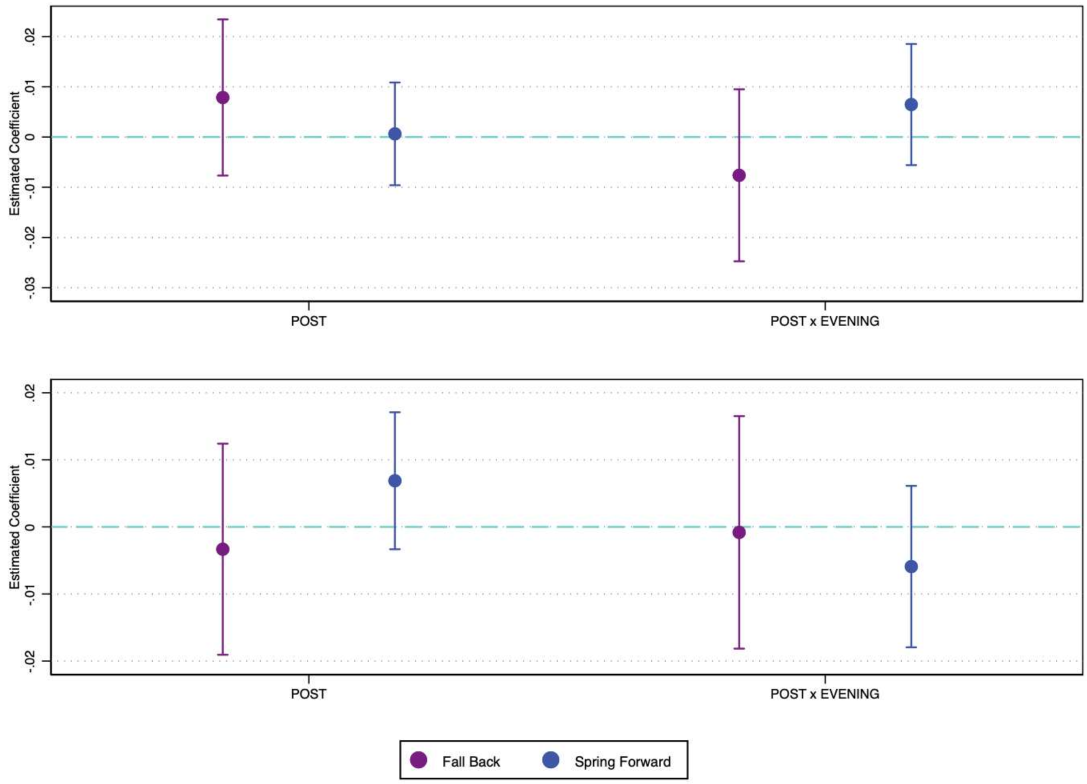
Variables	Model I	Model II	Model III	Model IV
POST	0.004	0.004	0.007	0.007
	(0.007)	(0.007)	(0.007)	(0.007)
EVENING	0.007	0.005	0.006	-0.001
	(0.006)	(0.006)	(0.006)	(0.008)
POST $\times$ EVENING	-0.033***	-0.030***	-0.031***	-0.029***
	(0.008)	(0.008)	(0.008)	(0.008)
Constant	0.058***	0.049***	0.046***	0.055
	(0.009)	(0.009)	(0.009)	(0.105)
Observations	383,653	383,653	383,653	383,653
R-squared	0.003	0.003	0.003	0.003
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Weekend FE	No	Yes	Yes	Yes
Holiday FE	No	No	Yes	Yes
Coordinates	No	No	No	Yes
Volumes	No	No	No	Yes

Notes: Dependent variable: Standardized BERT. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10. Model IV is our main specification.

Table B.7: Sensitivity of the Long-Run Effects of Fall Back on BERT

Variables	Model I	Model II	Model III	Model IV
POST	0.037***	0.039***	0.038***	0.045***
	(0.007)	(0.007)	(0.007)	(0.007)
EVENING	0.006	0.005	0.004	0.036***
	(0.006)	(0.006)	(0.006)	(0.008)
POST $\times$ EVENING	-0.030***	-0.031***	-0.031***	-0.044***
	(0.008)	(0.008)	(0.008)	(0.009)
Constant	0.018**	0.009	0.011	0.242**
	(0.008)	(0.008)	(0.008)	(0.110)
Observations	353,394	353,394	353,394	353,394
R-squared	0.002	0.002	0.002	0.002
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Weekend FE	No	Yes	Yes	Yes
Holiday FE	No	No	Yes	Yes
Coordinates	No	No	No	Yes
Volumes	No	No	No	Yes

Notes: Dependent variable: Standardized BERT. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10. Model IV is our main specification.



Notes: Dependent variables are standardized AFINN (top panel) and VADER (bottom panel). Fall Back regression specification include weekend, holiday, year, and state fixed effects. Additional independent variables in regression: coordinates (latitude, longitude) and tweet volumes. Spring Forward regression specification includes all fixed effects and additional independent variables, except holiday fixed effects. Spikes indicate 95% confidence intervals.

Figure B.2: The Effect of Pseudo-DSTs on AFINN and VADER

Table B.8: Sensitivity of Differential Returns to Sunlight–AFINN

Variables	Model I	Model II	Model III	Model IV	Model V
BRIGHT (β_3)	0.027***	0.028***	0.028***	0.028***	0.028***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
EVENING (β_4)	-0.023***	-0.021***	-0.021***	-0.032***	-0.031***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
BRIGHT $\times$ EVENING (β_5)	-0.033***	-0.031***	-0.031***	-0.013*	-0.014*
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Constant	-0.035***	-0.169**	-0.168**	-0.154**	-0.179**
	(0.006)	(0.073)	(0.073)	(0.073)	(0.073)
Observations	728,885	728,885	728,885	728,885	728,885
R-squared	0.003	0.003	0.003	0.003	0.003
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Weekend FE	No	Yes	Yes	Yes	Yes
Holiday FE	No	No	Yes	Yes	Yes
Season	No	No	No	Yes	No
Day of Year	No	No	No	No	Yes
Coordinates	No	Yes	Yes	Yes	Yes
Volumes	No	Yes	Yes	Yes	Yes

Notes: The dependent variable in our model is the standardized AFINN. As the seasonal indicator and day of the year variables are correlated, we included them separately in the model. Standard errors are in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.10. Model IV is our main specification.

Table B.9: Sensitivity of Differential Returns to Sunlight–VADER

Variables	Model I	Model II	Model III	Model IV	Model V
BRIGHT (β_3)	0.036***	0.039***	0.039***	0.039***	0.039***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
EVENING (β_4)	0.006	0.018***	0.018***	0.012**	0.012**
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
BRIGHT $\times$ EVENING (β_5)	-0.042***	-0.041***	-0.041***	-0.031***	-0.031***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Constant	-0.044***	-0.196***	-0.195***	-0.187**	-0.202***
	(0.006)	(0.073)	(0.073)	(0.073)	(0.073)
Observations	728,885	728,885	728,885	728,885	728,885
R-squared	0.002	0.002	0.002	0.002	0.002
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Weekend FE	No	Yes	Yes	Yes	Yes
Holiday FE	No	No	Yes	Yes	Yes
Season	No	No	No	Yes	No
Day of Year	No	No	No	No	Yes
Coordinates	No	Yes	Yes	Yes	Yes
Volumes	No	Yes	Yes	Yes	Yes

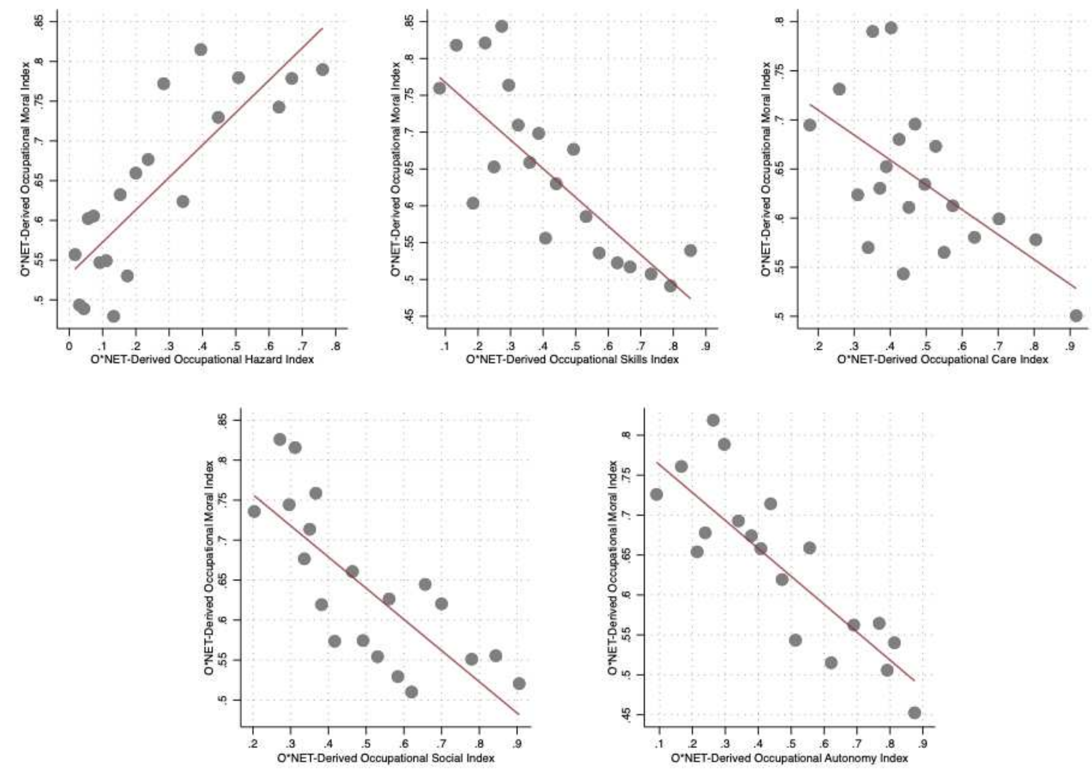
Notes: The dependent variable in our model is the standardized VADER. As the seasonal indicator and day of the year variables are correlated, we included them separately in the model. Standard errors are in parenthesis. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Model IV is our main specification.

Appendix C

Supplementary Tables and Figures

C.1 Correlations

This sections shows the correlations between the occupational moral index and other indices derived from the O*NET data.



Notes: O*NET-derived occupational indices are scaled from zero to one using min-max standardization.

Figure C.1: Correlations between Occupational Moral Index and other Indices

C.2 Additional Results and Robustness Checks

This section includes supplementary results and robustness checks. Specifically, we examine real hourly compensation as the outcome variable and present the corresponding findings in Tables C.1. Additionally, we conduct additional robustness checks by exclusively focusing on occupational and industry switchers and report the results in Table C.2. Furthermore, we present our estimates of the quadratic transformation of the O*NET-derived occupational moral index in Table C.3. Lastly, Table C.4 displays the coefficient estimates of other indices.

Table C.1: Parameter Estimates of Occupational Moral Index vis-a-vis Total Hourly Compensation

Variables	(1)	(2)	(3)	(4)
Occupational Moral Index	-11.202***	-3.059***	-3.059***	-3.059***
	(0.210)	(0.228)	(0.228)	(0.228)
Constant	22.196***	4.509***	4.509***	4.509***
	(0.147)	(0.543)	(0.543)	(0.543)
Observations	72,544	72,544	72,544	72,544
R-squared	0.044	0.467	0.467	0.467
Year FE	No	Yes	Yes	Yes
Region FE	No	Yes	Yes	Yes
Urban FE	No	Yes	Yes	Yes
Industry FE	No	Yes	Yes	Yes
Demography Controls	No	Yes	Yes	No
Human Capital and Labor Supply Controls	No	Yes	Yes	Yes
Job Characteristics Controls	No	Yes	Yes	Yes
Person FE	No	No	No	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable is Real Hourly Compensation. Occupational Moral Index is scaled from zero to one using min-max standardization.

Table C.2: Sensitivity of Parameter Estimates of Change in Occupational Moral Index vis-a-vis Change in Natural Log of Total Hourly Compensation

Variables	(1)	(2)	(3)	(4)
Δ Occupational Moral Index	-0.037**	-0.038**	-0.049***	-0.042**
	(0.015)	(0.015)	(0.018)	(0.018)
Constant	-0.009	0.019	0.033	-0.002
	(0.064)	(0.075)	(0.089)	(0.089)
Observations	25,698	24,796	18,694	29,772
R-squared	0.089	0.086	0.092	0.085
All Controls	Yes	Yes	Yes	Yes
Δ Autonomy Index	Yes	No	No	No
Education Restriction	No	Yes	No	No
Age Restriction	No	No	Yes	No
Occupation Cluster	No	No	No	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The dependent variable is Δ ln(real hourly compensation). All controls consist of a set of year, region, urban, and industry dummies, as well as human capital and labor supply variables, and other job characteristics as specified in Table 3.2. Δ Occupational Moral Index is scaled from zero to one using min-max standardization.

Table C.3: Parameter Estimates of Quadratic Occupational Moral Index vis-a-vis Total Hourly Compensation

Variables	Log Hourly Compensation	Hourly Compensation
Occupational Moral Index	-0.306***	-7.480***
	(0.048)	(1.127)
Occupational Moral Index^2	0.215***	4.887***
	(0.038)	(0.874)
Constant	2.206***	6.582***
	(0.032)	(0.602)
Observations	72,544	72,544
R-squared	0.679	0.681
Year FE	Yes	Yes
Region FE	Yes	Yes
Urban FE	Yes	Yes
Industry FE	Yes	Yes
Demography Controls	No	No
Human Capital and Labor Supply Controls	Yes	Yes
Job Characteristics Controls	Yes	Yes
Person FE	Yes	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Occupational Moral Index is scaled from zero to one using min-max standardization.

Table C.4: Parameter Estimates of O*NET-Derived Occupational Indices vis-a-vis Natural Log of Total Hourly Compensation

Variables	Log Hourly Compensation
Moral Index	-0.049***
	(0.013)
Hazard Index	0.061***
	(0.012)
Skills Index	0.269***
	(0.015)
Care Index	-0.099***
	(0.017)
Social Index	0.032*
	(0.018)
Autonomy Index	-0.011
	(0.012)
Constant	2.140***
	(0.031)
Observations	67,823
R-squared	0.680
Year FE	Yes
Region FE	Yes
Urban FE	Yes
Industry FE	Yes
Demography Controls	No
Human Capital and Labor Supply Controls	Yes
Job Characteristics Controls	Yes
Person FE	Yes

Notes: Robust standard errors denoted in parentheses. Significance is represented as follows: *** p<0.01, ** p<0.05, * p<0.1. O*NET-derived occupational indices are scaled from zero to one using min-max standardization.