Three Chapters on Health and Labor Economics

by

Jacob Thor Watson

(Under the Direction of Joshua L. Kinsler)

Abstract

This dissertation covers three separate topics in the fields of health and labor economics. The first chapter estimates the causal effect of retirement and spousal retirement on health and health-related behaviors by exploiting variation in statutory retirement policies in a fuzzy regression discontinuity framework. Using data from the first three waves of the China Health and Retirement Longitudinal Study, I find that own retirement has adverse effects on subjective health status and mental health for both genders, but its effects on health behaviors are mixed. I also find that women produce beneficial retirement spillovers on husbands' health, but do not experience any spillover effects themselves. Lastly, I provide estimates of the retirement effects under a counterfactual policy of raising the official retirement ages.

In the second chapter, I use Texas's constitutional amendment in 1997 that expanded the scope of home equity loans as a source of exogenous variation to estimate the effects of relaxing credit constraints on small businesses. Using standard panel data methods and restricted-use microdata from the U.S. Census Bureau, I find that the Texas amendment increased the use of home equity finance by small businesses, increased new business and job creation and reduced firm exit and job loss. The effects are larger and significant for businesses with fewer than ten employees.

In the third chapter, I examine how leisure-based physical activity affects earnings and wages. Using data from the American Time Use Survey and an instrumental variable approach, I find that in the short run, physical activity does not enhance labor productivity; instead, work time and physical activity are substitutes. A one-hour increase in average weekly physical activities decreases earnings by 1–2% within a location, but has no effect on wages. Meanwhile, I find that in the long run, physical activity does enhance labor productivity. A one-hour increase in average weekly physical activities increases average earnings and wages by 6–7% for a location.

Index words: Credit Constraints, Health Economics, Labor Economics, Physical Activity, Retirement, Small Businesses

Three Chapters on Health and Labor Economics

by

Jacob Thor Watson

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2021

Three Chapters on Health and Labor Economics

by

Jacob Thor Watson

Major Professor: Joshua L. Kinsler

Committee: Ian M. Schmutte
Meghan M. Skira

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
August 2021

Dedication

The first chapter of this dissertation is dedicated to my life partner, Siyu Pan. I could not have finished this work without your support and willingness to answer my esoteric institutional questions. Thank you.

The second chapter of this dissertation is dedicated to the memory of my friend, Tony Steiner, a true gentleman and scholar, who died, unexpectedly, in 2017. I am grateful to be able to honor him with work that overlaps with his interest in entrepreneurship. I simply wish we had been given the time and opportunity to discuss my research before his passing; I think he would have enjoyed the conversation and I would be wiser for it.

The third chapter of this dissertation is dedicated to my fur children Bootsie, Motoko, and Salví for their warmth and support while completing this dissertation.

Acknowledgments

First and foremost, I would like to acknowledge the help and guidance of my dissertation chair, Josh Kinsler. I would also like to thank my committee members, Ian Schmutte and Meghan Skira. Moreover, I would like to thank the following individuals for their helpful comments and support over the years: Melissa Banzhaf, Chris Cornwell, Roozbeh Hosseini, Julie Hotchkiss, Svetlana Pashchenko, and Vincent Pohl. Lastly, I am thankful to my fellow graduate students for their time and support while completing the doctoral program: Lauren Blumberg, Josh Frederick, Kim Groover, Liesl Himmelberger, Meghan Esson, Fan Liang, John Pasquinelli, Rodrigo Saurin, Keshob Sharma, Tim Smith, Scott Spitze, Polin Wang, Tommy Wiesen, Jiyeon Yun, and many others. I acknowledge financial support from the James C. Bonbright Center for the Study of Regulation in the completion of the second chapter of this dissertation. I also acknowledge the U.S. Census Bureau for providing access to the data for the second chapter of this dissertation. Any opinions and conclusions expressed herein are my own and do not necessarily represent the views of the U.S. Census Bureau. Research for the second chapter of this dissertation was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1085. All results have been reviewed to ensure that no confidential information is disclosed.

Contents

List of Figures

List of Tables

Chapter 1

Retirement Blues and Boons: Health and Spillover Effects of Retirement

1.1 Introduction

A global pattern of declining fertility rates and increasing life expectancy poses a serious challenge for policymakers. One concern is that an aging population is associated with increases in chronic diseases and disabilities, which may inflate national health insurance costs or strain public-assistance programs. At the same time, rapid population aging has provoked collective concern over the financial sustainability of public pension programs, leading to gradual increases in the official retirement age for public pension benefits in the United States and Europe. However, such policies were largely implemented without any consideration of spillover effects on the population's health. Although a better understanding of the effect of retirement on health is required to account for the costs and benefits of such policies, the direction and magnitude of the health effects of retirement remain unclear.

This paper estimates the causal own health and spousal spillover effects of retirement using data from the first three waves of the Chinese Health and Retirement Study (CHARLS). I extend the analysis to investigate potential channels through which retirement affects both own health and spousal health by considering changes in health-related behaviors and time use, as well as heterogeneity in the effects across gender and family characteristics. To account for the endogeneity of retirement, I use a fuzzy regression discontinuity (FRD) design that exploits statutory retirement policies in China. These policies induce discontinuous increases in the probability of retirement at specific ages, which can be used as instrumental variables for retirement status. The results show that own retirement has a significant and negative effect on self-reported health and mental health. Analysis of effect heterogeneity, health-related behaviors, and time use data suggests three mechanisms through which retirement affects health: (1) a decrease in moderate and vigorous physical activity; (2) an increase in sedentary activities; (3) an increase in informal childcare for grandchildren. The results also show a wife's retirement has beneficial spillover effects on her husband's mental health that is partly driven by the provision of childcare by the retiring wife.

The primary contributions of this paper are threefold: (1) causal evidence of retirement effects on own health in a context without changes in other entitlement programs; (2) causal evidence of spillover effects of retirement on spouse's health; and (3) extensive evidence for mechanisms that are consistent with the observed adverse health effects of own retirement and beneficial spillover effects found in the paper.

There is a sizable literature that estimates the health effects of retirement using data from Asia, Australia, North America, and Europe (e.g., Coe and Zamarro, 2011; Behncke, 2012; Latif, 2013; Insler, 2014; Eibich, 2015; Motegi et al., 2016; Apouey, et al., 2017; Heller-Sahlgren, 2017; Mazzonna and Peracchi, 2017; Muller and Shaikh, 2018; Yi and Che, 2018). These studies find mixed results, but generally imply retirement is beneficial to physical health and has a negative effect or no effect on mental health. Although one explanation for the inconsistencies found in the literature is weak external validity, a meta-analysis by Nishimura et al. (2018) finds differences in the surveyed countries only play a small role in explaining the mixed results. To help reconcile inconsistencies in the literature, it is necessary to understand the underlying health mechanisms and intra-household externalities of retirement. Only studying the direct health effects of retirement has been insufficient.

I contribute to two strands of literature: (1) the effects of retirement on own health, health-related behaviors, and time use more generally; and (2) the spillover effects of retirement on the health and time use of spouses. Regarding the former strand, this paper complements and extends the work of Yi and Che (2018), who provide evidence on the *causal* health effects of retirement in China. Using different data, they focus exclusively on the effect of own retirement on self-reported health and recent illness, but most significantly, their paper only considers white-collar males and a very limited number of mechanisms.[1] Ignoring the effects on blue-collar workers and females limits the policy relevance and external validity of their findings. The general lack of studies using Chinese data in the literature is surprising since China provides a unique context with which to examine the health effects of retirement.[2] For instance, health insurance coverage does not change upon retirement, and individuals do not qualify for additional social programs for almost another decade (Li et al., 2015). This is in contrast to the United States where eligibility for Medicare overlaps with the social security pension age, making it difficult to disentangle the effects of retirement on health from those of Medicare coverage. Moreover, by focusing on a single country, I am able to examine retirement effects in one institutional setting, avoiding any potential bias

[1]The mechanisms they consider are playing table tennis, practicing Tai Chi, drinking tea, smoking and drinking. Due to data limitations they are only able to examine changes on the extensive margin in these activities.

[2]Motegi et al. (2016) and Nishimura et al. (2018) both use an instrumental variables approach to examine retirement effects on health in a cross-country analysis using data from the HRS and HRS-sister studies including China. However, they point out their first-stage results for China are insignificant thereby preventing any causal interpretation of the results. This failure of instrument relevance arises due to their inclusion of a large number of observations not eligible for standard pensions, particularly rural farmers. Meanwhile, Xue et al. (2017) use data from the China Health and Nutrition Survey to estimate the effect of retirement on cardiovascular disease (CVD) risk factors. However, they describe their results only as associations between retirement and CVD risk factors since they do not use a causal framework.

from cross-country regressions – a common concern when using data from the Survey of Health, Ageing and Retirement in Europe (SHARE).[3]

With regard to the second strand of literature, this paper helps fill a gap as only Bertoni and Brunello (2017) and Muller and Shaikh (2018) have provided causal evidence of retirement spillover effects on the health of a spouse. This stands in contrast to a large connected literature on health-related intra-household externalities arising from unemployment (Lindo, 2011; Kind and Haisken-DeNew, 2012; Marcus, 2013; Bubonya et al., 2017). Similar to job loss, even though only one person retires, the event sets in motion a chain reaction that affects other household members directly and indirectly. For instance, Muller and Shaikh (2018) use SHARE data to find that retirement negatively affects a spouse's self-assessed health partly in response to an increase in alcohol consumption by the couple. Bertoni and Brunello (2017) use Japanese data to evaluate retirement spillover effects on a spouse's health, finding a husband's retirement has a negative effect on his wife's mental health but they do not consider any underlying mechanisms. This paper presents new evidence in the context of China, and expands on these studies by considering changes in biomarkers and time use in addition to physical activity. Moreover, China provides a setting that is uniquely applicable to the study of intra-household retirement externalities. A well-documented pattern of joint retirement in the United States and Europe over preceding decades makes it difficult to disentangle the extent to which spouses affect health in retirement (Hospido and Zamarro, 2014; Stancanelli and Van Soest, 2016). However, joint retirement is relatively rare in China due to statutory retirement ages differing by gender.[4] As discussed above, the estimates will not reflect potential bias due to coinciding eligibility in social programs by a retiring spouse.

Lastly, I contribute to both strands of literature by applying a modern econometric technique to estimate counterfactual changes in own and spousal retirement effects on health if official retirement ages were increased. To the best of my knowledge, no study in the health-retirement literature has estimated such counterfactuals. From a policy standpoint, knowing the direction and magnitude of these counterfactual retirement effects is equally, if not more important than knowing the effects of retirement under current policies. Recent policy debates over raising official retirement ages often center entirely on expectations of such counterfactuals.

The remainder of this paper is organized as follows. Section 1.2 discusses theoretical mechanisms potentially linking retirement to health; Section 1.3 discusses the institutional setting in China; Section 1.4 describes the data; Section 1.5 outlines the identification strategy; Section 1.6 presents the results alongside robustness checks; Section 1.7 considers a counterfactual retirement policy; and Section 1.8 concludes.

[3] To demonstrate this point, Bingley and Martinello (2013) point out differences in eligibility ages for pension benefits across country and gender are correlated with differences in years of schooling, which affect mental health and health-risk behaviors. Consequently, retirement ages are invalid instruments in many of the aforementioned studies using SHARE data without controlling for schooling. Although a number of studies do control for education, it raises concerns about biased estimates due to unobservables similarly correlated with cross-country eligibility ages. For instance, retirement ages in European countries are correlated with social welfare generosity. It is widely acknowledged that social welfare programs are important determinants of health in Europe as they mediate the extent, and impact, of socio-economic position on health (Bambra, 2006; Eikemo et al., 2008; Nelson and Fritzell, 2014).

[4] The standard retirement age for male and female employees in China is 60 and 50, respectively. Section 1.3 provides further institutional details.

1.2 Theoretical Health Effects of Retirement

1.2.1 Own retirement

From a theoretical standpoint, the net effect of retirement on own health is ambiguous. In the Grossman (1972) model of health demand, an increase in leisure time reduces the opportunity cost of time-consuming health investments. This is supported by a strand of empirical literature that finds retirement increases the amount of time spent exercising (Insler, 2014; Eibich, 2015; Motegi et al., 2016; Muller and Shaikh, 2018) and sleeping (Eibich, 2015; Motegi et al., 2016). At the same time, an exit from the labor market eliminates the incentive to invest in health for market activities. This is in line with empirical studies showing retirement increases the likelihood of engaging in health-damaging behaviors such as excessive drinking (Zins et al., 2011; Ayyagari, 2016; Muller and Shaikh, 2018). On the other hand, there is still an incentive to invest in health for nonmarket activities as health is a direct consumption commodity in the model. Ultimately, the theoretical net effect depends on whether the marginal utility of health decreases or increases after retirement, which is impossible to predict.

Moreover, a number of narratives linking retirement and health outside of a Grossman model predict similarly ambiguous effects. For instance, common perceptions of retirement involve a life of travel and relaxation but also an escape from financial pressure and occupational stress (TIAA, 2017). In this light, retirement is beneficial to mental health; however, retirement requires a dramatic transition in daily routines that can be stressful (Braithwaite and Gibson, 1987). Another common perception of retirement is there is an increase in the amount of time spent socializing with family and friends. A recent body of empirical work finds a positive causal effect of social capital on health that is robust across different contexts and definitions (d'Hombres et al., 2010; Ronconi et al., 2012; Rocco et al., 2014; Ho, 2016; Liu et al., 2016). It is theoretically ambiguous, however, how retirement affects social networks. Although there is likely a decline in social interactions with former coworkers, retirees can use their additional leisure time to form new relationships and invest more in current ones. The empirical research is also ambivalent on how retirement affects social networks. Borsch-Supan and Schuth (2014) find that retirement negatively affects the number of friends and former colleagues in a retiree's social network while Fletcher (2014) and Eibich (2015) find no effect on the number of friends.

1.2.2 Spousal retirement

In regard to spousal retirement, the theoretical spillover effect on health is equally ambiguous. For example, nearly every spouse of the newly retired must cope with a simultaneous reduction in household income and a spouse's increased presence at home (MacBride, 1976; Dave et al., 2008; Bonsang et al., 2012).[5] However, even the net effect of these two simple channels is uncertain. Retirement may directly

[5] A reduction in household income is expected since pension replacement rates are usually less than 100%. The average gross pension replacement rate, defined as gross pension entitlement divided by gross pre-retirement earnings, ranged from 21.70% to 79.50% across OECD countries in 2018. In China, the average gross pension replacement rate was 71.6% in 2018 (OECD iLibrary, 2020).

affect the marginal utility of home production (e.g., cooking, repairs, childcare) and make it more attractive for the retiree, while at the same time reducing household expenditures on consumption goods and services bought in the market (Aguiar and Hurst, 2005; Aguiar and Hurst, 2007). A substitution of home production for private expenditure by the retiree may have positive spillover effects on the spouse's health if this results in a reduction of spousal household production. This increase in the spouse's leisure time could be spent on joint leisure activities such as exercise. On the other hand, the spouse may instead increase labor market hours to compensate for the reduction in household income, thereby increasing occupation-related strain.

The health-related behaviors of the retiree may also have spillover effects on the spouse's health. It is well known in the herd theory literature that the propensity of an individual to behave in a certain way can vary with the characteristics and behavior of surrounding individuals (Banerjee, 1992; Manski, 1993). Following a similar line of reasoning, it is plausible to presume that an individual's health-related behaviors (e.g., smoking, drinking, exercise) directly affect those of their spouse. Muller and Shaikh (2018) find that retirement in Europe significantly increases the frequency of alcohol intake by the retiree and the spouse regardless of the spouse's retirement status. This, in part, explains the negative retirement spillover effect on self-reported health estimated by the authors.

1.3 Chinese Institutional Background

1.3.1 Statutory Retirement

In China, statutory retirement ages are set for all employees in the formal labor force, including government, private firms, state-owned and collectively-owned enterprises. They do not apply, however, to the informal sector which comprises 20% to 37% of urban workers (Park et al., 2013).[6] The retirement legislation defines the standard retirement age for male and female employees to be 60 and 50, respectively.[7] However, there are exceptions for certain occupations specific to each gender. Female civil servants and professionals have their retirement pushed back to 55, and high-ranking male government officials retire at 65. Individuals can always choose to retire earlier than their mandated retirement age, but cannot collect their public pensions until the age of their assigned retirement.[8]

Although the statutory retirement ages in China are considered "mandatory" in design, the policy is not perfectly enforced in practice. It is not uncommon for employers to offer informal contracts to employees who process retirement at their firm, thereby allowing an employee to draw their pension and

[6] To my knowledge there are no recent studies estimating the size of the informal sector in rural China. The International Monetary Fund (IMF) estimates the entire informal sector comprised 12% of the Chinese economy in 2015. However, the IMF cautions that due to its criteria "the results might be capturing the informal economy only partially."

[7] *Methods for the Retirement of Workers* in 1978, and *Principles for Government Employees* in 1993.

[8] A man or woman working in a "high-risk" and/or "health-damaging" occupation has the possibility to retire and receive a pension up to five years early with government approval. To qualify for the early retirement option (45 for females, and 55 for males), an employee must have (i) irreparably damaged their health working a physically intensive job, and (ii) made contributions for at least 15 years while working in the formal sector.

continue working full or part time beyond the statutory retirement age (Li et al., 2015). However, these informal employees are no longer protected by labor laws and wages are renegotiated with the employer.

1.3.2 Public Pension System

Prior to 2015, China's public pension system was composed of four programs targeting different segments of the population. However, only the programs for workers in the formal sector, the Basic Old Age Insurance (BOAI) program and the Public Employee Pension (PEP), provided meaningful retirement benefits.[9] The BOAI was established in 1951, and covers formal workers in for-profit enterprises, including both the public and private sector. In 2015, the PEP was merged with the BOAI, making BOAI the uniform program for all formal sector employees. It is comprised of a compulsory scheme with both defined contributions and benefits, plus an individual account pension. On the benefits side, employees with a contribution history of at least 15 years are entitled to the pension benefits. The target replacement rate is about 60% of local average wages for a worker with 35 years of contributions (Fang and Feng, 2018). The BOAI eligibility age aligns with a worker's statutory retirement age: 50 for most females, 55 for female civil servants and professionals, and 60 for males.

The PEP was established in 1953 for civil servants and employees in non-profit government institutions, such as schools and health facilities. It was the most generous pension system, with a notable feature being that it did not require any contributions from employees. After reaching the eligibility age and processing retirement, a public employee was provided a "retirement wage" with an average replacement rate of 80 – 90% of pre-retirement wages (Fang and Feng, 2018). If a public employee left the public sector to work in the private sector prior to retirement, however, then the retirement wage was forfeited. When the PEP was merged into the BOAI in 2015, the contribution and benefit rules for public employees were changed to those of the BOAI. For public employees who retired before the 2015 reform, their pension benefits are unaffected. For those who entered the public sector after the reform, they are enrolled in the BOAI system. For everyone in the middle, i.e., those enrolled in the PEP who did not retire before the reform, there is a transitional arrangement. These public employees must contribute to the BOAI individual account pension, but will receive a modified retirement wage inversely related to their history of contributions. The transition amounts to incrementally phasing out retirement wages for new retirees, with the intention that contributions to individual accounts will replace the loss in benefits. The PEP eligibility age aligns with a worker's statutory retirement age: 55 for females, 60 for most males, and 65 for high-ranking male government officials.

[9] The other two schemes, the New Rural Resident Pension (NRP) and the Urban Resident Pension (URP) were established in 2009 and 2011, respectively. These pension systems were intended to cover informal workers and the unemployed based on household registration, i.e., urban or rural resident. Both programs relied on voluntary contributions in conjunction with government subsidies, and only provided a replacement rate of about 20% of local per capita income (Fang and Feng, 2018). In 2014, the NRP and URP were merged into a uniform Resident Pension system that no longer differentiated based on residency. Participants with a contribution history of at least 15 years are entitled to receive a basic pension upon reaching the age of 60 years old.

1.4 Data

The data used in this paper come from the first three waves of the China Health and Retirement Longitudinal Study (CHARLS) and the supplemental Life History survey. The CHARLS is one of many sister studies designed to mimic the Health and Retirement Study (HRS) conducted in the United States. On a biennial basis, approximately 20,000 individuals across China participate in the study. It consists of a national baseline sample fielded in 2011-12 and an additional replenishment sample fielded during the second wave in 2013-14. A third wave was conducted in 2015-16 without any replenishment. The study is designed to be representative of the Chinese population over the age of 45, covering 150 urban districts and 450 rural villages. The CHARLS has several features particularly attractive for the analysis at hand. It hosts a rich set of indicators related to individuals' health, employment, family demographics, and living conditions. The same information is also collected for every spouse and community-level information is provided separately.

1.4.1 Retirement

The goal of this paper is to identify own and spousal retirement effects on health, so it is essential to discuss what "retirement" implies in this context. There are three common definitions of retirement in the literature: (1) exiting the paid labor force for a set amount of time, (2) self identifying as retired, or (3) a combination of the former two (Coe and Zamarro, 2011; Insler, 2014; Eibich, 2015; Heller-Sahlgren, 2017). Unlike the HRS or SHARE, respondents in the CHARLS are never asked whether they identify themselves as retired; instead they report whether retirement procedures have been completed at their primary place of employment. As discussed in Section 1.3, completing these procedures does not have a clear implication on an individual's current work status.[10] I define an individual as being "retired" if they are not working for pay and have not actively sought paid work within the past month.[11] It is most likely that retirement affects the health of couples through behavioral adjustments, and that behavioral adjustment only occurs when an individual has exited the labor force. This definition is similar to ones used by Coe and Zamarro (2011) and Heller-Sahlgren (2017), where homemakers, the permanently ill or disabled, and individuals engaged in activities without pay are included in the retirement category. The key difference being that my sample only considers individuals who work, or recently worked, in the formal sector.

1.4.2 Mandatory Retirement Age

The CHARLS does not directly ask respondents their mandated retirement age in the survey, so I use related information to identify individuals subject to the statutory retirement policy and then determine

[10] In the data, I find that 19% of individuals who have completed retirement procedures are still engaged in paid work to some capacity.

[11] As a robustness check, I use an alternative definition of retirement based on a combination of retirement procedures being completed and not working for pay within the past month.

their mandated retirement age. This process involves leveraging survey questions on retirement expectations, pension benefits, and employment history in a step-wise process. The exact details are outlined in the online appendix. There is noise in the assigned mandated retirement ages due to potential misclassification, so I also consider a sensitivity check to the estimation sample in the online appendix.

1.4.3 Outcome variables

I consider a set of health measures and biomarkers from the data. More specifically, I examine the effects of retirement on self-reported health status (SRHS), the CESD-10 depression scale, body mass index (BMI), and blood pressure (systolic/diastolic) to measure changes in the physical and mental health of individuals. Observing both self-reported health measures and biomarkers provides a comprehensive look into the health effects of retirement. I also assess the impact retirement has on a range of health-related behaviors to understand the mechanisms through which retirement can affect health. The most salient are smoking, alcohol consumption, and physical inactivity since these risk behaviors are often associated with several chronic diseases (Linardakis et al., 2015).

Measures of Health

In the CHARLS, respondents are asked to rate their general health status on a five-point scale ranging from 1 "Excellent" to 5 "Poor". This is identical to the self-reported health status question found in the HRS and its sister studies that is commonly used in the health literature (Coe and Lindeboom, 2008; Johnston and Lee, 2009; Eibich, 2015; Muller and Shaikh, 2018). Although SRHS is a subjective measure of health, there is evidence that it captures valuable information that is unobserved by objective health characteristics (Fisher et al., 2005; Ambrasat et al., 2011; Insler, 2014). Moreover, its frequent use within the literature provides a consistent basis upon which to compare results across studies.

With regard to mental health, depression is the most common mental health problem among the elderly and is often correlated with a decline in physical health (Andresen, 1994; Boey, 1999). The CHARLS includes ten questions out of the original 20 used by Radloff (1977) to create the Center for Epidemiological Studies Depression (CESD-20) scale. Respondents are asked to rate how often they have had problems with the following in the past week: irritability, concentration, depression, fatigue, pessimism, fear, sadness, loneliness, motivation. These questions generally overlap with the eight and twelve depression-related questions asked in the HRS and SHARE, respectively. In those two surveys, however, respondents only answer "yes" or "no" to having problems with an issue over the past two weeks (HRS) or month (SHARE). I collapse responses in the data into binary variables that capture the specific problems on the extensive margin, thereby mimicking the HRS and SHARE. This allows for a cleaner interpretation and provides more consistency when comparing my results with the literature. I then construct a variable based on the CESD-10 scale first used by Kohout et al. (1993) to screen for depressive symptoms among older adults in the general American population.[12] This depression scale has been validated by psychiatrists across multiple populations, including elderly Chinese citizens (Irwin et al., 1999; Cheng and Chan, 2005).

In the CHARLS, height and weight are measured for the respondent, unlike the HRS and SHARE where respondents provide self-reported measurements. Using this data, I calculate the body mass index of respondents. The general rule of thumb developed by the World Health Organization is that a BMI of less than 18.5 qualifies as *underweight* and may indicate malnutrition or other health problems, while a BMI equal to or greater than 25 is considered *overweight*, and above 30 is considered *obese*. Obesity is a major risk factor for heart disease and cardiac events.

Unless a medical complication was of concern, every respondent had their blood pressure measured three times during the interview by an accompanying nurse. In line with the medical literature, I use the average of the last two to estimate resting blood pressure - one variable for the systolic reading and another for the diastolic reading (Xue et al., 2017). According to the American Heart Association, a blood pressure between 120/80 and 140/90 mmHg is known as *prehypertension*, which indicates an increased risk of developing high cholesterol, obesity, and diabetes. A reading above 140/90 mmHg is defined as *hypertension* and increases the risk of heart attack, stroke, coronary heart disease, heart failure, and kidney failure.

Health-related Behaviors

I consider two forms of exercise on the intensive and extensive margins: moderate and vigorous physical activity.[13] In the CHARLS, moderate physical activity records the number of days per week an individual reports having engaged in activities that require a moderate level of energy for at least ten consecutive minutes (e.g., carrying light loads or mopping the floor). Vigorous physical activity reports the number of days per week an individual engaged in activities that require a high level of energy for at least ten consecutive minutes (e.g., intensive sports, heavy lifting, or agricultural work). Regular physical activity and exercise improves immune system functions, helps prevent heart disease and diabetes, and lowers the risk of dementia.

Tobacco consumption is known to be a major risk factor for cardiovascular and respiratory diseases. I capture tobacco behavior using two variables; one represents the extensive margin of use as an indicator for whether or not an individual has "smoked" in the past year.[14] The other variable captures the intensive margin; i.e., if a respondent indicated they had smoked in the past year, then they are prompted to estimate how many cigarettes per day they consume. Alcohol consumption is another avoidable risk factor that contributes substantially to a host of diseases. Drinking habits are captured by two indicators; the first is an indicator for whether or not an individual consumed any alcohol in the past year, while the other

[12] A CESD-10 score is simply a count of the number of times a respondent answered "yes" to having one of the depression-related problems. This is similar to the CESD-8 and Euro-D scales commonly used in the literature with the HRS and SHARE, respectively.

[13] Unlike the HRS, neither the CHARLS nor SHARE ask about time spent on "mildly energetic" physical activities such as vacuuming or doing laundry.

[14] This includes chewing tobacco, pipe tobacco, self-rolled tobacco, cigarettes or cigars.

reflects the intensive margin through frequency of alcohol consumption in the past year. Individuals were asked to estimate whether they consumed alcohol on a daily basis ("Daily" to "More than twice a day"), weekly basis ("Once a week" to "4 to 6 days a week"), or monthly basis ("Once a month" to "2 to 3 days a month").

1.4.4 Covariates

I make use of additional variables to investigate heterogeneous effects and test the robustness of the results. These include gender, education, number of own children, regional characteristics, spousal retirement status, and whether a grandchild co-resides in the household. Education is measured by a dummy variable for individuals who completed an education beyond high school or vocational training. Regional characteristics are captured by a dummy variable for whether a province is considered urban in the survey. I also consider effect heterogeneity based on the presence of a co-residing grandchild in Section 1.6.4. A dummy for co-residing grandchild indicates a grandchild resided in the household for over six months in the past year.

1.4.5 Sample criteria

The main estimation sample is comprised of married individuals whose spouses also participated in the survey.[15] Furthermore, each individual must be subject to the statutory retirement policy, as outlined in the online appendix.[16] I use individuals whose age is within a window of ± 5 years from their mandated retirement age.[17] This confines the analysis to individuals within reasonable range of the statutory retirement cutoff, while still providing enough precision to investigate heterogeneous effects. Once these restrictions have been applied, the sample consists of an unbalanced panel of 4,630 person-year observations.[18]

Summary statistics by gender are presented in Table 1.1. Overall, males and females in the estimation sample differ significantly from one another in their health and health-related behaviors. Noticeable examples are in their consumption of alcohol and smoking behavior. Approximately 60% of males in the sample consumed alcohol in the past year and those who did were most likely to do so on a daily basis. Meanwhile, roughly 15% of females consumed alcohol in the past year and those who did were more likely to be monthly drinkers. In a similar fashion, over half the males in the sample smoke compared to only 5% of females. On average, male smokers consume 18 cigarettes per day while female smokers consume 11

[15] Less than 2% of couples in the data have information on only one individual.

[16] This creates an implicit criteria that both members of a couple must have worked in the formal sector at some point. In the data, 39% of couples meet this condition; 24% of couples have one member who works, or has worked, in the formal sector; 33% of couples have both members work, or have worked, only in the informal sector; and 4% of couples have at least one member who has never worked.

[17] I allow the window size to vary as a sensitivity test, with results reported in the online appendix. I also find similar results using two optimal bandwidth selectors outlined in Calonico et al. (2018).

[18] The sample is comprised of 2,024 unique individuals forming 1,102 unique couples.

cigarettes per day. These differences in health outcomes and behaviors highlight the importance of looking into retirement and spillover effect heterogeneity by gender later on in Section 1.6.4.

1.5 Identification Strategy

1.5.1 Endogeneity of retirement

Any valid research design used to evaluate the causal effect of retirement on health must take into account that the former is likely endogenous to the latter. In this context, the literature identifies two sources of endogeneity: omitted variable bias and reverse causality. Omitted variable bias might be induced through differences in unobserved individual characteristics that influence both health and the retirement decision. For example, the birth of a grandchild may increase the probability of retirement, and provision of informal childcare may decrease smoking (creating a downward bias). Reverse causality poses a more serious problem as several studies show that health affects the retirement decision (McGarry, 2004). In order to address the issue of endogeneity, I use a fuzzy regression discontinuity (FRD) design exploiting the discontinuity in retirement rates at the statutory retirement ages.

1.5.2 Fuzzy Regression Discontinuity

The general idea behind a FRD design is to exploit exogenous institutional rules that partially determine treatment. An assignment or forcing variable is used to determine the treatment status of individuals. Those above a known cutoff receive the treatment while those below are not treated. As a result, a discontinuity in the outcome variable at the cutoff can be interpreted as the causal effect of treatment if individuals cannot precisely manipulate the assignment variable around the cutoff, ensuring local random assignment (Lee and Lemieux, 2010).

In the context of China, statutory retirement ages create a strong incentive to retire at specific ages. Individuals cannot draw pension benefits prior to the retirement age and can no longer work under formal labor laws afterward. However, the discontinuity in the probability of being retired is not sharp at the mandated retirement ages due to imperfect compliance; i.e., an individual could retire early without accessing pension benefits or enter the informal sector after retiring from a formal job. This leads to a FRD design where the forcing variable is an individual's (spouse's) *centered* age; that is, an individual's (spouse's) age minus their mandated retirement age (mret).[19] The CHARLS provides data on the month and year of birth for each respondent as well as the month and year of the interview so centered age can be treated as a continuous variable. The estimated effect should be interpreted as the effect on compliers retiring once they are mandated to do so by exceeding the statutory age threshold.

[19] For example, a female public school teacher whose actual age is 52 would have a centered age of -3 since she faces a mandated retirement age of 55 for being a female civil servant.

Estimation of the retirement effects and retirement spillover effects amounts to using the statutory retirement ages as instrumental variables for own and spousal retirement status. I apply Two-Stage Least Squares (2SLS) to estimate equations of the following form:[20][21]

$$\text{Health}_{it} = \beta_0 + \tau_1 R_{it} + \beta_1 C_{it} + \beta_2 C_{it} R_{it} + \tau_2 R^S_{it} + \beta_3 C^S_{it} + \beta_4 C^S_{it} R^S_{it} + \lambda_t + \zeta_p + \epsilon_{it} \tag{1.1}$$

$$\text{R}_{it} = \alpha_0 + \alpha_1 C_{it} + \alpha_2 C^S_{it} + \alpha_3 T_{it} + \alpha_4 C_{it} T_{it} + \alpha_5 T^S_{it} + \alpha_6 C^S_{it} T^S_{it} + \lambda_t + \zeta_p + u_{it} \tag{1.2}$$

$$\text{R}^S_{it} = \gamma_0 + \gamma_1 C_{it} + \gamma_2 C^S_{it} + \gamma_3 T_{it} + \gamma_4 C_{it} T_{it} + \gamma_5 T^S_{it} + \gamma_6 C^S_{it} T^S_{it} + \lambda_t + \zeta_p + v_{it} \tag{1.3}$$

$$C_{it}\text{R}_{it} = \psi_0 + \psi_1 C_{it} + \psi_2 C^S_{it} + \psi_3 T_{it} + \psi_4 C_{it} T_{it} + \psi_5 T^S_{it} + \psi_6 C^S_{it} T^S_{it} + \lambda_t + \zeta_p + \eta_{it} \tag{1.4}$$

$$C^S_{it}\text{R}^S_{it} = \omega_0 + \omega_1 C_{it} + \omega_2 C^S_{it} + \omega_3 T_{it} + \omega_4 C_{it} T_{it} + \omega_5 T^S_{it} + \omega_6 C^S_{it} T^S_{it} + \lambda_t + \zeta_p + \rho_{it} \tag{1.5}$$

where Health$_{it}$ is a health outcome or behavior outlined in Section 1.4.3; R_{it} is an indicator for own retirement; R^S_{it} is the corresponding indicator for spousal retirement; $C_{it} = (age_{it} - mret_i)$ is own centered age at the mandated retirement age; C^S_{it} is the respective spouse's centered age. I instrument for own and spousal retirement using the respective treatment indicators $T_{it} = I_{[C_{it} \geq 0]}$ and $T^S_{it} = I_{[C^S_{it} \geq 0]}$ which capture whether an individual's (spouse's) age passes the mandated retirement age. I include the interaction of the retirement indicators and centered age in Equation (1.1) to allow for the possibility that centered age follows a different trend on each side of the threshold. Since these interaction terms are also endogenous variables, the interaction of the treatment indicators and centered age $C_{it}T_{it}$ and $C^S_{it}T^S_{it}$ are used as additional instruments for identification. Lastly, λ_t and ζ_p are year and province indicator variables, respectively. While ϵ_{it}, u_{it}, v_{it}, η_{it}, and ρ_{it} are idiosyncratic error terms. Standard errors are clustered at the individual level for all results.[22]

In Equation (1.1) the causal effects of own retirement and spousal retirement on $Health_{it}$ are captured by the parameters τ_1 and τ_2. Meanwhile, Equations (1.2) and (1.3) are first-stage equations linking the endogenous variables R_{it} and R^S_{it} to the instruments. The analogous first-stage equations for $C_{it}R_{it}$ and $C^S_{it}R^S_{it}$ are represented by Equations (1.4) and (1.5). As with any IV framework, the estimated treatment effects $\hat{\tau}_1$ and $\hat{\tau}_2$ are interpreted as a local average treatment effects (LATE) on compliers. Specifically,

[20] I also consider a nonlinear specification in Section 1.6.5 as a sensitivity check.

[21] Note, the application of a 2SLS specification here is equivalent to a FRD design with a rectangular kernel.

[22] Clustering the standard errors at the household level does not alter the significance of any estimated effects. Results available upon request.

these estimates are the average treatment effect of retirement for individuals (spouses) entering retirement due to reaching the mandated retirement age.

Discontinuity in Treatment

As an initial check for instrumental relevance, Figure 1.1 shows the share of retirees by centered age in bins of six months. The graphs in this section include local linear fitted lines with a 95% confidence interval on each side of the threshold. Although roughly 25% of individuals retire prior to their mandated retirement age, there is a clear discontinuous jump in retirement after crossing the threshold. Above the statutory retirement threshold, the share of retirees increases abruptly by almost 30 percentage points. This discontinuity reflects the fact that the statutory retirement policy and the corresponding retirement benefits are strong incentives for employees in the formal sector to enter retirement once they pass the statutory retirement threshold. Furthermore, Figure 1.2 shows the share of retirees for men and women separately. Although men are slightly more likely to retire before reaching their mandated retirement ages, the discontinuous jump in retirement is nearly equal for both genders.

Table 1.2 presents the first-stage results for two model specifications: Model (1) which does not include the interaction terms $C_{it}R_{it}$ and $C_{it}^{S}R_{it}^{S}$ in the second-stage equation or use the interaction of the treatment indicators and centered age $C_{it}T_{it}$ and $C_{it}^{S}T_{it}^{S}$ as instruments; and Model (2) which includes the two interaction terms in the second-stage equation and uses the two corresponding instruments. The Kleibergen-Paap *rk* Wald F-statistics are 61.32 and 56.74 for Model (1) and (2), respectively, suggesting that the discontinuities satisfy instrument relevance.[23] As an additional robustness check, the Sanderson-Windmeijer F-statistic for each first-stage is reported in Table 1.2. This F-statistic can be used as a diagnostic for whether a *particular* endogenous regressor is weakly identified. All six SW F-statistics are greater than 100, further supporting the validity of using the discontinuities as instruments.[24]

1.5.3 Validity checks

There are two identifying assumptions required for the FRD design to have a causal interpretation. First, the outcome variable must be a smooth function of the forcing variable. It seems plausible to assume the health-age profile of individuals is smooth considering aging is a gradual process. Second, I must assume individuals cannot precisely control the forcing variable (centered age) near the threshold. On the one hand, this assumption is not threatened by individuals directly manipulating their ages in the CHARLS.

[23] Stock-Yogo critical values for clustered standard errors are unavailable, but critical values for i.i.d. errors can be used as a supplemental check to detect weak identification (Baum et al., 2007). In Model (1), the Stock-Yogo critical value for the weak identification test is 7.03 at the 10% significance level. Unfortunately, Stock and Yogo (2005) only provide critical values for up to three endogenous variables so I do not know the exact critical value for the weak identification test in the case of Model (2). However, as the number of endogeneous variables (n) and the number of instruments (k) increase in the just-identified case (n=k) the critical value always decreases for a given level of significance. Thus, the critical value of 7.03 from the n=k=2 case in Model (1) is an upperbound of the critical value for the weak identification test of Model (2) where n=k=4.

[24] The Stock-Yogo critical value for the weak identification test at the 10% significance level is 19.93 for each first stage of Model (1) and 10.27 for each first stage of Model (2).

The month and year of birth that I use to calculate age is not self-reported but rather from government-issued documents. On the other hand, it is theoretically possible for someone to manipulate their centered age by changing occupations. However, there are a number of reasons why this is unlikely to occur in practice.

The occupations with higher statutory retirement ages (55 for females, and 65 for males) have barriers to entry that would be difficult to surmount later in life. In the case of men, the higher retirement age is for high-ranking government officials. These people start climbing the bureaucratic ladder early in life and only a small handful reach this level (Bo, 2019). In the case of women, the higher retirement age is positively correlated with educational attainment since these are civil servants and professionals (Feng and Zhang, 2018). It's highly unlikely a woman facing the standard retirement age would have the professional qualifications to enter one of these jobs near retirement. The more likely issue would be a man (woman) facing the higher retirement age switching occupations before the age of 60 (50) in order to access his (her) public pension early. However, this would likely involve surrendering most of his (her) pension benefits from the original occupation, alongside a loss in social status and presumably lower wages during her time in the new occupation.[25]

As a validity check, I investigate how many individuals in the data exhibit a change in their mandated retirement ages during the survey. I find less than 2% of observations within a ten-year window of the statutory retirement threshold have any change in the retirement age assigned across survey waves.[26] As an additional validity check, Figure 1.4 tests for the presence of a discontinuity in the density of centered age at the statutory retirement threshold by running kernel local linear regressions of the density separately on both sides of the threshold. Visual inspection suggests no manipulation of the forcing variable since it appears smooth near the threshold. A formal McCrary (2008) test for a discontinuity in the distribution function estimates that the distribution increases by 0.034 percentage points at the threshold, with a standard error of 0.078. Thus, a t-test of the null hypothesis of continuity in the distribution function fails to reject.

Continuity of Baseline Variables

If there is no manipulation of assignment, then all observed and unobserved characteristics should be balanced near the threshold so treatment is "as if randomly assigned." That is, individuals below the threshold represent a valid control group for those just above the threshold and any comparison between the two reveals local causal effects (Lee and Lemieux, 2010). A common validity check in the regression discontinuity design is to see whether observable baseline characteristics are locally balanced near the threshold. If there exists a discontinuity in a predetermined variable, this would cast doubt on the identification strategy.

[25] Most men and women with higher statutory retirement ages are enrolled in the Public Employee Pension (PEP) program. The main pension benefits provided by the PEP are "retirement wages", which are only available to individuals who process retirement procedures while employed in the public sector.

[26] As a sensitivity check, I re-estimate the main results without these individuals. I find no changes in the results (available upon request).

Figure 1.3 shows the share of females, number of living children, proportion of those with a higher education, and share of those living in an urban province in the estimation sample. All four graphs do not show signs of a concerning discontinuity near the threshold, validating local random assignment. Overall, the validity checks support the identification strategy and provide no evidence of violations of the key assumptions. I conduct additional placebo tests in Section 1.6.5.

1.6 Results

I begin by presenting the estimated own and spousal retirement effects for two model specifications. The first model implements the most basic FRD estimator which only uses the treatment indicators for crossing the statutory retirement threshold as instruments for own and spousal retirement. The second model extends the first model by using interactions of own (spousal) retirement with own (spouse's) centered age as additional endogenous variables and interaction terms of the respective treatment indicators with centered age as additional instruments. This is the preferred model, as outlined Section 1.5.2, which allows for the possibility that centered age follows a different trend on each side of the threshold.

1.6.1 Retirement effects on health

Table 1.3 shows the estimated effects of own and spousal retirement on four health outcomes. In both model specifications, I find a positive effect of own retirement on self-reported health status (SRHS), indicating a detrimental effect, that is significant at the 1% level. This result stands in agreement with recent studies that find adverse effects of retirement on health (e.g., Mazzonna and Peracchi, 2017; Shia, 2018) but differs from estimates in another strand of literature that suggest beneficial health effects (e.g., Coe and Zamarro, 2011; Insler, 2014; Eibich, 2015; Blake and Garrouste, 2017; Che and Li, 2018). However, I find no statistically significant effect on health-related biomarkers such as body mass index or blood pressure. This observed difference between SRHS and biomarkers highlights the importance of considering multiple measures of health outcomes. Moreover, there is evidence that subjective measures of health capture valuable information that is unobserved by object health characteristics (Fisher et al., 2005; Ambrasat et al., 2011; Insler, 2014). In addition, there is effect heterogeneity by gender that is masking some of retirement's adverse health effects on BMI and blood pressure in the full sample. I discuss this heterogeneity later in Subsection 1.6.4.

Meanwhile, I find a detrimental effect of own retirement on mental health as measured by the CESD scale that is significant at the 5% level in both model specifications. These results support the growing literature on the negative effects of retirement on cognitive abilities and mental well-being (Mazzonna and Peracchi, 2012; Bertoni and Brunello, 2017; Heller-Sahlgren, 2017; Mazzonna and Peracchi, 2017). Furthermore, I estimate the effect of retirement on the ten components of the CESD scale individually to gain a better understanding of how retirement affects mental health.[27] I find that retirement increases the probability of being more easily bothered and experiencing restless sleep at the 10% level. Specifically,

[27] Results for each of the ten individual components available upon request.

the estimated effects suggest the likelihood of feeling bothered or having restless sleep in the past week increase by 17.2 and 14.6 percentage points, respectively. For comparison, the baseline probabilities in the five years before reaching the mandated retirement age for feeling depressed and restless sleep are 38% and 45%, respectively, so the effects are relatively large.

Regarding spillover effects, I do not find any effect of a spouse's retirement on subjective health measures or biomarkers for either specification. This differs from the beneficial spillovers on self-assessed health found in Muller and Shaikh (2018). I investigate spousal spillovers further in Subsection 1.6.4 and consider effect heterogeneity by gender.

1.6.2 Retirement effects on health-related behaviors

Retirement alters several aspects of daily life that could potentially contribute to the observed health effects. Most notably, it increases leisure time and changes the daily environment. It is likely that behavioral adjustments partly explain the adverse health effects reported in the previous subsection. I investigate this further by estimating the effect of retirement on health-related behaviors in this subsection, and time use in the following subsection.

Table 1.4 presents the estimated effects of own and spousal retirement on moderate and vigorous forms of exercise. I find that own retirement has a negative effect on moderate physical activity along the intensive margin at the 5% level. The estimates suggest there is a 1.21 – 1.37 decrease in the number of days per week spent engaging in moderate exercise, excluding zeros. This effect is large considering the average baseline number of days engaging in moderate physical activities is 5.8 in the five years prior to the mandated retirement age. Moreover, Table 1.4 also shows a large negative effect of retirement on vigorous physical activities along the extensive margin indicating an 18 – 20 percentage point decline in the probability that is significant at the 5% level. This is even more striking when considering on average, only 26% of the sample engages in vigorous physical activities in the five years prior to reaching the statutory retirement threshold. The most likely reason is a decrease in on-the-job physical activities. Overall, these results are consistent with the adverse health effects of retirement on SRHS observed in Subsection 1.6.1.

Unlike my results, previous literature on the health effects of own retirement has found strong positive effects on engaging in physical activity (Eibich, 2015; Kampfen and Maurer, 2016; Motegi et al., 2016; Zhao et al., 2017; Che and Li, 2018; Muller and Shaikh, 2018). One explanation for the different findings lies in the definition of physical activity. All of the previous studies only consider changes in physical activity along the extensive margin, so they would be unable to capture the reduction in moderate exercise along the intensive margin that I find.[28] Moreover, there is substantial heterogeneity in the forms of physical activity (vigorous, moderate, and light) considered by each study and whether on-the-job physical activity is included. For instance, Che and Li (2018) find a positive effect of retirement on engaging in Tai Chi or playing table tennis for Chinese white-collar men. Given the CHARLS only asks about moder-

[28] Prior studies do not consider effects along the intensive margin largely due to data constraints in the most popular health-retirement surveys (e.g., HRS, SHARE, SOEP). These surveys often limit responses on the frequency of physical activity to predetermined categories; e.g., "more than once a week", "once a week", "one to three times a month", and "never".

ate and vigorous forms of physical activity, it is unlikely that I would be able to capture changes in these activities.[29]

Table 1.5 displays the estimated effects of own and spousal retirement on smoking and drinking behaviors. I only find an effect on smoking along the intensive margin that reveals a 2.19 – 2.36 reduction in cigarette consumption at the 10% level. Compared to the average baseline consumption of 18.83 cigarettes, this represents a 11.6 – 12.5% decrease in cigarette consumption. This reduction in smoking is in line with most of the retirement literature, but does not align with the adverse effect on SRHS found in Subsection 1.6.1. In addition, there is substantial effect heterogeneity that I discuss later in Subsection 1.6.4.

Finally, I find no retirement spillover effects on a spouse's health-related behaviors. This result contrasts with the increases in alcohol and cigarette consumption by the retiree's spouse reported in Muller and Shaikh (2018) using SHARE data. In large part, their results stem from couples engaging in shared leisure activities. In the context of China, gender norms have a strong influence on alcohol and smoking behaviors as shown in the summary statistics in Table 1.1. Consequently, it is not surprising that I fail to find a similar pattern of spillover effects on smoking and drinking behaviors.

1.6.3 Retirement effects on time use

I examine a wide range of time use variables in the CHARLS to help explain the adverse effects own retirement has on self-reported physical and mental health observed in Table 1.3. Moreover, it is important to determine how individuals use their extra time in retirement to help direct public policy measures to mitigate these adverse effects. For brevity, I only report significant findings unless otherwise relevant. The most striking change in time use is in the provision of informal childcare for grandchildren. I estimate that own retirement increases the likelihood of providing childcare in the past year by 14 percentage points at the 5% level, nearly double the baseline probability of providing childcare in the five years prior to reaching the mandated retirement age. Moreover, I find that retirement increases childcare on the intensive margin by 3,582 annual hours (149 days) at the 5% level. This constitutes roughly a 30% increase in annual hours spent on childcare compared to the baseline. Although the effect of childcare on health is ambiguous, this substantial time transfer partly explains the lack of additional investment, or even reduction, in health-promoting physical activities.

The literature investigating the effect of childcare on grandparents' health reports mixed results. On the one hand, providing care for a young child can be exhausting both physically and mentally (Blustein et al., 2004). On the other hand, care giving provides a sense of purpose and meaning that may be lost after retiring from work (Rozario et al., 2004). In China, there is a traditional expectation for grandparents to be the provider of daily care during the early years of a child's life (Xu, 2019). This would suggest a much more intensive form of childcare than what is provided by grandparents in the United States or Europe. For instance, Ho (2015) estimates grandparents in the United States provide an average of 5.9 hours of childcare per week using HRS data. Grandparents in the CHARLS report providing an average of 58.9

[29]The description of moderate (vigorous) activities in the CHARLS questionnaire is an activity that "makes you breathe (much) harder than normal" for at least 10 minutes. It is unlikely a respondent in the CHARLS would consider Tai Chi or table tennis a moderate or vigorous form of physical activity since both activities rarely involve sustained heavy breathing.

hours per week, and those in the estimation sample report an average of 57.8 weekly hours. Chen and Liu (2012) use an instrumental variables approach to estimate the health effects of childcare by grandparents using Chinese data. They find that high intensity levels of care (15 weekly hours or more) have an adverse effect on health, while low levels have a protective effect. This finding potentially explains some of the estimated adverse health effects of retirement, considering the estimated retirement effect on childcare along the intensive margin easily satisfies the high intensity definition.

I also find that retirement increases the likelihood of engaging in sedentary activities in the past month (e.g., card games, board games, using the computer) by 16 percentage points at the 5% level. This represents a 37% increase compared to the baseline. Unfortunately, the CHARLS only reports the frequency spent engaging in each activity by three categories (daily, weekly, less often) so interpreting any change on the intensive margin is difficult. It is well acknowledged in the medical literature that a sedentary lifestyle is a health risk, so the increase in sedentary activities on the extensive margin may help explain the adverse retirement effect on SRHS.

To provide additional evidence of child care and sedentary activities as underlying mechanisms, I re-estimate the model for SRHS and CESD with each proposed mechanism included as a control variable.[30] If the adverse health effects of retirement are induced by changes in these variables, I would expect the estimated effect of retirement on SRHS and CESD to decrease in magnitude for each case as the pathway is being shutdown. Furthermore, I would anticipate the potential mechanisms to be positively associated with the outcome variables, suggesting an adverse relationship with health. A similar exercise is used by Eibich (2015) to provide suggestive evidence for mechanisms driving retirement effects on health in the German population. I find the estimated retirement effects on SRHS and CESD decrease by 0.04 – 0.08 and 0.07 – 0.15, respectively, while the signs on sedentary activities, child care, and child care hours are all positive. Furthermore, the significance of the estimated retirement effects slightly decrease and all of the mechanisms are at least significant at the 10% level.

1.6.4 Heterogeneous effects

I examine whether there is heterogeneity in the estimated effects to provide additional insight on the underlying mechanisms. I begin by considering heterogeneity across gender since the two populations face substantially different statutory retirement ages and social norms. Table 1.6 shows there is in fact effect heterogeneity by gender. First, the estimated effects of own retirement are detrimental to subjective measures of health for both genders but noticeably larger for women. Furthermore, there is a significant effect on body mass index and blood pressure for women that was not seen in the pooled sample. Compared to the average baseline woman five years before the statutory retirement threshold, these effects represent a 3.5% increase in BMI and 3.2%/3.7% increase in systolic/diastolic blood pressure. With regard to health-related behaviors, I find the reduction in smoking on the intensive margin observed in Table 1.5 is unsurprisingly driven solely by males, who are thirteen times more likely to smoke compared to females. The retirement effect on engaging in vigorous physical activity is stronger for males than females, which

[30] Results from this exercise are available upon request.

may be due to occupational differences. More surprisingly, the effect on the amount of moderate physical activity seen in the pooled sample appears to be driven by females. Lastly, I find that women produce beneficial spillover effects on husbands' health that nearly offset husbands' own detrimental retirement effects. However, men do not produce any spillover effects on wives.

One potential explanation for this heterogeneity in own retirement effects is the onset of menopause, which is commonly associated with weight gain and increases in blood pressure. The average age of women in the estimation sample at the time of menopause is 53, so its transitional phase (perimenopause) would occur a few years prior - near the general statutory retirement age for women. To check this theory, I re-estimate the model for women with a dummy for menopause as a control variable. I find no change in the retirement effect on SRHS but a small decrease in the magnitude of the coefficients on CESD, BMI, blood pressure, amount of moderate exercise, and probability of vigorous exercise. Moreover, the estimated effect on blood pressure is no longer statistically significant at the 10% level (results are available upon request). This suggests biological differences between men and women partly explain the effect heterogeneity.

Another potential explanation for the effect heterogeneity by gender could be differential changes in time use. In the left half of Table 1.7, I examine effect heterogeneity by gender in the provision of childcare and engaging in sedentary activities. I find little difference between men and women in the likelihood of engaging in sedentary activities within the past month. However, I find own retirement significantly increases the likelihood of providing childcare by women (23 percentage points) but not by men. This difference in the provision of childcare seems intuitive given women tend to retire first in China. That is, provision of childcare by a retiring wife likely induces her working husband to also provide childcare, albeit less intensively. As a result, there is no significant change in childcare along the extensive margin when the husband retires since he is already providing a limited amount prior to retirement. Supporting this hypothesis, I estimate that women produce a positive spillover effect of 10 percentage points on the likelihood of husbands providing childcare but no spillover effect on the intensive margin. Meanwhile, I find that own retirement increases childcare on the intensive margin by both men and women. The estimated effect for men is relatively larger in magnitude, though not significant. This difference in effect size is likely due to a lower baseline amount of childcare hours provided, prior to retirement, by men compared to women.

An increase in childcare may help explain the reduction in cigarette consumption by men and beneficial spillover effects on husbands' mental health. For instance, a reduction in smoking on the intensive margin by men could be due to altruistic grandfathers curtailing their production of second-hand smoke out of concern for the grandchild's health. Meanwhile, the beneficial spillover effect on husbands' mental health could be due to an increase in grandfathers interacting with grandchildren in a low intensive form. For example, commonly cited benefits to the mental health of grandparents providing childcare are an increased purpose for living and optimism about the future (Grindstead et al., 2003). In the right half of Table 1.7, I examine the effect of own and spousal retirement on a subset of the individual components of the CESD scale. I find that women produce a significant increase in husbands' likelihood to report "feeling hopeful about the future" and slight decrease in reporting "everything I did was an effort".

Meanwhile, women experience an increase in restless sleep and "feeling more easily bothered than usual" from own retirement. While these results are consistent with the strand of literature finding beneficial effects of low-intensity childcare on grandparents' mental health, and detrimental effects of high-intensity childcare on grandparents' health, it is still speculative.

Finally, I consider effect heterogeneity by spouse's retirement status.[31] It is possible that I find no spillover effects on health or health-related behaviors in Tables 1.3, 1.4, and 1.5 because increases in joint leisure activities may only occur among couples who are both retired. In other words, retired couples spend more time together on shared leisure activities compared to couples with one retired spouse and one still in the labor force.[32] As Table 1.8 shows, however, I do not find any evidence to support this claim as the estimated spillover effects are not significantly different by spouse's retirement status. Due to the gender differences in statutory retirement ages, one potential concern is that only a few women may be in the subsample of non-retired individuals, and only a few men in the subsample of retired individuals. If that were the case, then the top panel of Table 1.8 would only capture the spillover effect of a wife's retirement on a non-retired husband's health while the bottom panel would capture the spillover effect of a husband's retirement on a retired wife's health. However, the summary statistics reported in Table 1.1 dispel this concern since only 40% of women are retired and 54% of men are non-retired in the estimation sample.

1.6.5 Robustness checks

I check the robustness of the findings in Tables A.1 and A.2 in the online appendix with respect to the window size, inclusion of baseline covariates, and model specification. I also consider a placebo test using observations that should not be affected by the statutory retirement policy.

The bandwidth is an essential part of any regression discontinuity design, since a trade off must be made between bias and precision. I re-estimate the model using a window of 2, 3, and 7 years around the statutory retirement threshold. Results are provided in Table A.1 in the online appendix. Although some of the effects are less significant due to a loss in precision, the estimated coefficients do not deviate much from those reported in Tables 1.3, 1.4, and 1.5. This suggests the results are not systematically biased.

Assuming the identification strategy is valid, the inclusion of baseline covariates as controls in the model should not alter the results. I re-estimate the model with the inclusion of gender, number of kids, and a dummy for higher education with results reported in Table A.2 in the online appendix. Overall, the inclusion of covariates does not influence the estimated effects, suggesting the identification strategy is sound.

Misspecification in the model's functional form could bias the estimated retirement effects, so I re-estimate the model with the inclusion of own (spouse's) squared centered age and corresponding interac-

[31] Ideally, I would consider effect heterogeneity by spouse's retirement status *and* gender. However, a large amount of statistical power is lost from slicing the sample along two dimensions due to very small cell sizes.

[32] Joint retirement, i.e., couples retiring within one year of each other, is relatively uncommon in China. Only 14% of all couples in the data jointly retire compared to around 30% in the United States and Europe over the past two decades based on HRS and SHARE data, respectively.

tion terms with own (spousal) retirement. I also include instruments analogous to the original interaction terms for identification. Results are reported in Table A.2 in the online appendix. In general, this produces results nearly identical to those already presented.

As a placebo test, I estimate the model on the out-of-sample observations that should not be affected by the statutory retirement policy, namely agricultural workers and a small number of people who are self-employed or working for a family business. Although these individuals can choose to enroll in a state pension program, the retirement benefits are relatively small. As a result, the financial incentive to retire at the age of 60, when pension benefits are first available, is weak. Moreover, a change in formal labor rights due to reaching the statutory retirement age should have no effect on the employment of these individuals. I assign placebo mandated retirement ages based on the standard age for each gender (50 for all women, and 60 for all men). If the empirical strategy is valid, there should be no significant own or spousal retirement effects on health. Results in Table A.2 in the online appendix confirm this hypothesis.

1.7 Counterfactual Policy Analysis

From a policy standpoint, it is important to know how the direction and magnitude of the estimated retirement effects would change if official retirement ages were increased. To this end, recent work by Dong and Lewbel (2015) and Cerulli et al. (2017) provides a way to estimate how the local average treatment effect (LATE) from a sharp or fuzzy RD design would change if the RD threshold changed. The basic idea is to identify the derivative of the LATE with respect to the threshold. In the context of this paper, this is analogous to identifying the partial derivative of an estimated retirement effect with respect to the statutory retirement threshold. Dong and Lewbel (2015) call this partial derivative the marginal threshold treatment effect (MTTE). The MTTE can be used to approximate the impact on an estimated retirement effect from a marginal change in the statutory retirement ages similar to how a price elasticity is used to approximate the effect of a marginal change in price.

The first step is to estimate a treatment effect derivative (TED) for a given retirement effect. The TED is essentially the derivative of the LATE with respect to the forcing variable. In other words, the TED is the partial derivative of an estimated retirement effect with respect to centered age. Dong and Lewbel (2015) show that the standard assumptions required by empirical applications of both sharp and fuzzy RD methods are sufficient to consistently estimate a TED. Technical details on how to estimate a TED are provided in the online appendix.

Under an assumption of local policy invariance, the TED is equivalent to the MTTE. Dong and Lewbel (2015) describe local policy invariance as "assuming that the function that describes how the RD LATE varies with the running variable does not itself change when the policy threshold changes infinitesimally. It is essentially a ceteris paribus assumption of the type commonly employed in partial equilibrium analyses." In the context of increasing the statutory retirement ages, local policy invariance implies that, holding an individual's age fixed at her mandatory retirement age, the expected difference in her health outcomes between being retired and not being retired would not change if for all other compliers, their mandatory retirement age changed a small amount (e.g., a month or two). This assumption seems plausible in the current context since it would require strong health spillover effects among late-in-life peers to produce any substantial general equilibrium effects.[33] The most obvious potential spillover effect would come from a spouse having a marginal change in their mandatory retirement age, but this spillover is controlled for in all of the regressions.

1.7.1 External Validity Test

Regardless of whether the local policy invariance assumption holds, an additional benefit of estimating the TED is that it can be used to asses the stability and, hence, external validity of the main results in Section 1.6.1. Recall that the estimated retirement effects throughout the paper only apply to people who are compliers at the statutory retirement threshold. It is therefore important to examine whether other people near, but not at the statutory retirement threshold, would experience similar retirement effects in sign and magnitude. If this is *not* the case, then there could be concerns about instability in the results. Instability does not mean that the estimated retirement effects are invalid, but rather that they need to be interpreted cautiously.

A TED significantly different from zero and relatively large in magnitude would suggest the estimated retirement effects are sensitive to very small changes in the statutory retirement threshold. Cerulli et al. (2017) provide a way to determine what qualifies as "relatively large" by constructing a measure they call the Relative TED.[34] They propose a Relative TED smaller than one in magnitude, alongside a statistically significant TED, suggests potential instability in the estimated retirement effect. While a Relative TED larger than one in magnitude or statistically insignificant TED, suggests some external validity, since it implies individuals near, but not at, the statutory retirement threshold likely have treatment effects of similar magnitude and sign to those right at the threshold.

Furthermore, Cerulli et al. (2017) introduce a closely-related concept to the TED called the Complier Probability Derivative (CPD), which measures the stability of the complier population in fuzzy RD designs. A relatively large and statistically significant CPD would suggest that the population of compliers changes dramatically with a small change in the statutory retirement threshold, thereby posing another threat to external validity. Analogous to the Relative TED, Cerulli et al. (2017) also construct a Relative CPD.[35] A Relative CPD smaller than one in magnitude, alongside a statistically significant CPD, suggests potential instability in the complier population.

Calculations of the TEDs and CPDs for the main results are reported in Table 1.9. To help test for instability in the estimated retirement effects, Relative TEDs and Relative CPDs are also included in the table. Overall, the calculated TEDs are not statistically significant, and those that are significant have a

[33] Local policy invariance could be violated if there is strong clumping of peers at a specific age. For example, if an individual's friends were all the same age as the individual, increasing the statutory retirement age for everyone else would induce the friends to retire a little later than the individual. A shift in retirement by friends to a later age may cause the individual to perceive her own health in retirement through a different lens. That is, if a retired individual uses friends as a basis of comparison, those friends would now be comprised of mostly non-retired individuals.

[34] The Relative TED is equal to the absolute value of the LATE divided by the product of the bandwidth and the TED.

[35] The Relative CPD is equal to the first-stage discontinuity in treatment probability divided by the product of the bandwidth and the CPD.

Relative TED greater than one in magnitude, except for in one case. Meanwhile, most of the CPDs are statistically significant but the Relative CPDs are all greater than one in magnitude. Overall, the results suggest no issues with instability in the estimated retirement effects and, hence, there is some external validity in the retirement effects.

1.7.2 A Five-year Increase in Statutory Retirement Ages

Over the past decade, the Chinese government repeatedly signaled that it would begin increasing the statutory retirement ages by 2020. In fact, this policy was made a priority in the most recent five-year plan (2016-2020) for the Ministry of Human Resources and Social Security. However, as of 2020, no definitive plan has been ratified by the Chinese government. The most conservative proposal is to increase the statutory retirement ages for everyone by five years, while the most drastic proposal is to change the retirement age to 65 for everyone, which would increase the general retirement age for women by 15 years (Feng et al., 2019).

I estimate the retirement effects under a counterfactual five-year increase to everyone's statutory retirement ages. Given the government's reluctance to change the retirement ages and the public resistance policymakers frequently face on the matter, the policy ultimately implemented will likely be a conservative one. Moreover, estimating the counterfactual retirement effects requires extrapolation of the TED. Unless a stronger policy invariance assumption holds, the extrapolation becomes more noisy the further away the counterfactual retirement threshold is moved from the current threshold. Consequently, counterfactual estimates for the most conservative policy proposal, a five-year increase, are the most feasible to estimate.

Results are reported in Table 1.10. For ease of reference, I also include the retirement effects originally estimated in Section 1.6 under the current statutory retirement policy. I report counterfactual retirement effects under the new policy by gender, since significant effect heterogeneity was found between males and females. I omit estimates of spousal retirement effects for females from the table since no statistically significant spillover effects were originally observed in Section 1.6.4.

Given local policy invariance, I find that later retirement ages would mostly improve the health effects of retirement for men. In particular, the detrimental retirement effect on SRHS is nearly gone. There is also a marginal decrease in smoking along the intensive margin, and a marginal increase in physical activity along the extensive margin compared to the current retirement effects. However, there is a marginal increase in the retirement effect on CESD, indicating that later mandatory retirement ages would make the mental health effects of retirement more harmful to men. With regards to spillover effects, the results indicate that women continue to produce beneficial retirement spillovers on husbands' SRHS and CESD. The spillover effects on SRHS are essentially unchanged from what men currently experience from a wife's retirement, but the beneficial spillover effects on mental health would be significantly larger under the new policy.

For women, I find that later mandatory retirement ages appear to exacerbate the detrimental health effects of retirement. Unlike men, women experience a marginal increase in the retirement effect on SRHS, indicating the retirement effect on self-reported health would be worse under the new policy. Similar to men, there is a marginal increase in the retirement effect on CESD, indicating that later mandatory retirement ages would make the mental health effects of retirement more harmful to both genders. There is also a marginal decrease in moderate exercise along the intensive margin, and slight increase in BMI compared to current retirement effects for women.

1.8 Conclusion

This paper estimates the short-term effects of own and spousal retirement on health and health-related behaviors using a fuzzy regression discontinuity design that exploits statutory retirement policies in China. The results indicate that own retirement has adverse effects on subjective health status and mental health for both genders but its effects are relatively larger for women. I also find that women produce beneficial retirement spillovers on husbands' health but do not experience any spillover effects themselves. A reduction in physical activities, an increase in sedentary activities, and intensive childcare for grandchildren appear to be underlying channels. Lastly, I find that raising the statutory retirement ages for everyone by five years would benefit men, with exception of mental health, but not women.

Overall, the results in this paper indicate policymakers do not face a trade off between ensuring solvency of state pension funds and protecting the health of the elderly. Policy proposals seeking to increase the retirement age not only benefit pension systems, but may also provide an efficient way to delay the adverse health effects of retirement. Moreover, counterfactual policy results suggest that increasing the retirement age could potentially dilute many of the adverse health effects of retirement experienced by men. These findings are timely and relevant for China where efforts to simultaneously formalize the labor market and expand state pension programs in a rapidly aging population are creating concerns over the future sustainability of pension funds.

Finally, policy debates over the costs and benefits of retirement need to account for potential externalities. For instance, a recent trend to equalize male and female retirement ages in China and European countries is largely motivated by efforts to shore up pension funds and allocate human resources more efficiently. Although a definitive plan has not been ratified in China, one proposal is to gradually increase the retirement ages to 65 for both men and women. There is little to no concern on how this would affect joint retirement patterns, which in turn could alter the effect of retirement on couples after adjusting for intra-household externalities. Considering the adverse health effects of retirement on husbands can be negated by the beneficial spillover effects from a wife's retirement, structuring official retirement ages to incentivize joint retirement could help alleviate retirement related health issues experienced by men.

1.9 References

Ambrasat, J., et al., (2011). Comparing the Predictive Power of Subjective and Objective Health Indicators: Changes in Handgrip Strength and Overall Satisfaction with Life as Predictors of Mortality. SOEP Working Paper No. 398.

Aguiar, M., and Hurst, E., (2005). Consumption versus Expenditure. Journal of Political Economy, 113, 919-948.

Aguiar, M., and Hurst, E., (2007). Life-Cycle Prices and Production. American Economic Review, 97, 1533-1559.

Andresen, E., et al., (1994). Screening for depression in well older adults: evaluation of a short form of the CES-D. American Journal of Preventive Medicine, 10, 77-84.

Angrist, J., and Rokkanen, M., (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. Journal of American Statistical Association, 110(512),1331-1344.

Apouey, B., et al., (2017). Retirement and Unexpected Health Shocks. IZA Discussion Paper 11226.

Ayyagari, P., et al., (2016). The Impact of Retirement on Smoking Behavior. Eastern Economic Journal, 42, 270-287.

Bambra, C., (2005). Health Status and the Worlds of Welfare. Social Policy & Society, 5, 53-62.

Baum, C., et al., (2007). Enhanced Routines for Instrumental Variables/GMM Estimation and Testing. Boston College Economics Working Paper, No. 667.

Banerjee, A., (1992). A Simple Model of Herd Behavior. Quarterly Journal of Economics, 107, 797-817.

Bingley, P., and Martinello, A., (2013). Mental retirement and schooling. European Economic Review, 63, 292-298.

Behncke S., (2012). Does retirement trigger ill health?. Health Economics, 21, 282-300.

Belloni, M., et al., (2016). The Effect On Mental Health Of Retiring During The Economic Crisis. Health Economics, 25, 126-140.

Bertoni, M., and Brunello, G., (2017). Pappa Ante Portas: The effect of the husband's retirement on the wife's mental health in Japan. Social Science and Medicine, 175, 135-142.

Blake, H., and Garrouste, C., (2017). Collateral Efects of a Pension Reform in France. PSE Working Papers n2012-25.

Bo, Zhiyue (2019). Chinese Provincial Leaders: Economic Performance and Political Mobility Since 1949. Routledge, 2019.

Boey, K., (1999). Cross-validation of a short form of the CES-D in Chinese elderly. International Journal of Geriatric Psychiatry, 14, 608-617.

Bonsang, E., et al., (2012). Does retirement affect cognitive functioning. Journal of Health Economics, 31, 490-501.

Borsch-supan, A., and Schuth, M., (2014). Early Retirement, Mental Health, and Social Networks. Discoveries in the Economics of Aging, 225-250.

Braithwaite, V., and Gibson, D., (1987). Adjustment to Retirement: What We Know and What We Need to Know. Aging and Society, 7, 1-18.

Bubonya, M., et al., (2017). Job loss and the mental health of spouses and adolescent children. IZA Journal of Labor Economics, 6:6.

Bunnings, C., et al., (2017). Fear Of Unemployment And Its Effect On The Mental Health Of Spouses. Health Economics, 26, 104-117.

Calonico, S., et al., (2018). Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. Working Paper, University of Michigan.

Cerulli, G., et al., (2017) Testing Stability of Regression Discontinuity Models. Regression Discontinuity Designs. Advances in Econometrics, Vol. 38, Emerald Publishing Limited, 317-339.

Charles, K., (2004). Is retirement depressing? : Labor force inactivity and psychological well-being in later life. Research in Labor Economics, 23, 269-299.

Che, Y., and Li, X., (2018). Retirement and Health: Evidence From China. China Economic Review, 49, 84-95.

Chen, F., and Liu, G., (2012). The Health Implications of Grandparents Caring for Grandchildren in China. The Journal of Gerontology, 67(1), 99-112.

Cheng, S., and Chan, A., (2005). Measuring Psychological Well-Being in the Chinese. Personality and Individual Differences, 38, 1307.

Coe, N., and Lindeboom, M., (2008). Does Retirement Kill You? Evidence from Early Retirement Windows. IZA Discussion Paper 3817.

Coe, N., and Zamarro, G., (2011). Retirement effects on health in Europe. Journal of Health Economics, 30, 77-86.

Dave, D., et al., (2008). The Effects of Retirement on Physical and Mental Health Outcomes. Southern Economic Journal, Southern Economic Association, 75(2), 497-523.

Dong, Y., and Lewbel, A., (2015). Identifying The Effect of Changing The Policy Threshold In Regression Discontinuity Models. Review of Economics and Statistics, 97(5), 1081-1092.

d'Hombres, B., et al., (2010). Does social capital determine health? Evidence from eight transition countries. Health Economics, 19, 56-74.

Eibich, P., (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. Journal of Health Economics, 43, 1-12.

Eikemo, T., et al., (2008). Welfare state regimes and differences in self-perceived health in Europe: A multilevel analysis. Social Science & Medicine, 66, 2281-95.

Fang, H., and Feng, J., (2018). The Chinese Pension System. University of Pennsylvania, Pension Research Council Working Paper WP2019-6.

Feng, J., and Zhang, X., (2018). Retirement and Grandchild Care in Urban China. Feminist Economics, 24:2, 240-264.

Feng, Q., et al., (2019). Age of Retirement and Human Capital in an Aging China, 2015-2050. European Journal of Population, 35, 29-62.

Fisher, G., et al., (2005). Documentation of Chronic Disease Measures in the HRS/AHEAD. Survey Research Center, University of Michigan.

Gelman, A., and Imbens, G., (2014). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. No. w20405. National Bureau of Economic Research.

Godard, M., (2016). Gaining weight through retirement? Results from the SHARE survey. Journal of Health Economics, 45, 27-46.

Grossman, M., (1972). On the Concept of Health Capital and the Demand for Health. The Journal of Political Economy, 80, 223-255.

Hernaes, E., et al., (2013). Does retirement age impact mortality?. Journal of Health Economics, 32, 586-598.

Ho, C., (2015). Grandchild care, intergenerational transfers, and grandparents' labor supply. Rev Econ Household, 13, 359-384.

Ho, C., (2016). Better Health With More Friends: The Role of Social Capital in Producing Health. Health Economics, 25, 91-100.

Heller-Sahlgren, G., (2017). Retirement Blues. Journal of Health Economics, 54, 66-78.

Hospido, L., and Zamarro, G., (2014). Retirement Patterns of Couples in Europe. IZA Discussion Papers 7926, Institute for the Study of Labor.

Insler, M., (2014). The Health Consequences of Retirement. Journal of Human Resources, 49(1), 195-233.

Johnston, D., and Lee, W., (2009). Retiring to the good life? The short-term effects of retirement on health. Economic Letters, 103, 8-11.

Kampfen, F., and Maurer, J., (2016). Time to burn (calories)? The impact of retirement on physical activity among mature Americans. Journal of Health Economics, 45, 91-102.

Kind, M., (2012). Unexpected Victims: How Parents' Unemployment Affects Their Children's Life Satisfaction. Melbourne Institute Working Paper Series, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Kohout, F., et al., (1993). Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. Journal of Aging and Health, 5(2), 179-93.

Ku, L., et al., (2012). The health effects of caregiving by grandparents in Taiwan: an instrumental variable estimation. Review of Economics of the Household, 10, 521-540.

Latif, E., (2013). The impact of retirement on mental health in Canada. Journal of Mental Health Policy and Economics, 16(1), 35-46.

Lee, D., and Lemieux, T., (2010). Regression discontinuity designs in economics. Journal of Economic Literature, 48, 281-355.

Li, H., et al., (2015). The Retirement Consumption Puzzle in China. American Economic Review, 105(5), 437-41.

Lindo, J., (2011). Parental job loss and infant health. Journal of Health Economics, 30, 869-879.

Liu, G., et al., (2016). How does social capital matter to the health status of older adults? Evidence form the China Health and Retirement Longitudinal Survey. Economics and Human Biology, 22, 177-189.

MacBride, A., (1976). Retirement as a life crisis: myth or reality?. Canadian Psychiatric Association Journal, 72, 547-556.

Manski, C., (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies, 60, 531-542.

Marcus, J., (2013). The Effect of Unemployment on the Mental Health of Spouses - Evidence from plant closures in Germany. Journal of Health Economics, 32, 546-558.

Mazzonna, F., and Peracchi, F., (2012). Ageing, cognitive abilities and retirement. European Economic Review, 56, 691-710.

Mazzonna, F., and Peracchi, F., (2017). Unhealthy Retirement?. The Journal of Human Resources, 52, 1, 128-151.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics, 142, 2, 698-714.

Motegi, H., et al., (2016). Does Retirement Change Lifestyle Habits?. The Japanese Economic Review, 67, 2, 169-191.

Nishimura, Y., et al., (2018). What Explains the Difference in the Effect of Retirement on Health? Evidence from Global Aging Data. Journal of Economic Surveys, 32, 792-847.

Muller, T., and Shaikh, M., (2018). Your Retirement and My Health Behaviour: Evidence on Retirement Externalities from a Fuzzy Regression Discontinuity Design. Journal of Health Economics, 57, 45-59.

Park, A., et al., (2012). Informal Employment in Urban China: Measurement and Implications. World Bank Working Paper 77737.

Nelson, K., and Fritzell, J., (2014). Welfare states and population health: The role of minimum income benefits for mortality. Social Science & Medicine, 112, 63-71.

Neuman, K., (2008). Quit Your Job and Get Healthier? The Effect of Retirement on Health. Journal of Labor Research, 29, 177-201.

Radloff, L., (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement, 1, 385-401.

Rocco, L., et al., (2014). From Social Capital To Health - And Back. Health Economics, 23, 586-605.

Ronconi, L., et al., (2012). Social capital and self-rated health in Argentina. Health Economics, 21, 201-208.

Sahlgren, G., (2012). Work 'til you drop: Short- and longer-term health effects of retirement in Europe. IFN Working Paper No. 928.

Shai, Ori (2018). Is retirement good for men's health? Evidence using a change in the retirement age in Israel. Journal of Health Economics, 57, 15-30.

Stancanelli, E., and Van Soest, A., (2012). Retirement and Home Production: A Regression Discontinuity Approach. American Economic Review: Papers and Proceedings, 102(3), 600-605.

Stancanelli, E., and Van Soest, A., (2016). Partners' leisure time truly together upon retirement. IZA Journal of Labor Policy 5:12.

Stock, J., and Yogo, M., (2005). Testing for Weak Instruments in Linear IV Regression. In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press, 80-108.

TIAA (2017). Transition to Retirement Survey. Teachers Insurance and Annuity Association and College Retirement Equities Fund, New York, NY.

Xue, B., et al., (2017). The Associations Between Retirement and Cardiovascular Disease Risk Factors in China: A 20-Year Prospective Study. American Journal of Epidemiology, 185(8), 688-696.

Xue, Hongwei (2019). Physical and Mental Health of Chinese Grandparents Caring for Grandchildren and Great-Grandparents. Social Science & Medicine, 229, 106-116.

Zhao, M., et al., (2017). Retiring for better health? Evidence from health investment behaviors in Japan. Japan and the World Economy, 42, 56-63.

Zins, M., et al., (2011). Effect of Retirement on Alcohol Consumption: Longitudinal Evidence from the French Gazel Cohort Study. Public Library of Science One, 6(10), e26531.

1.10 Tables and Figures

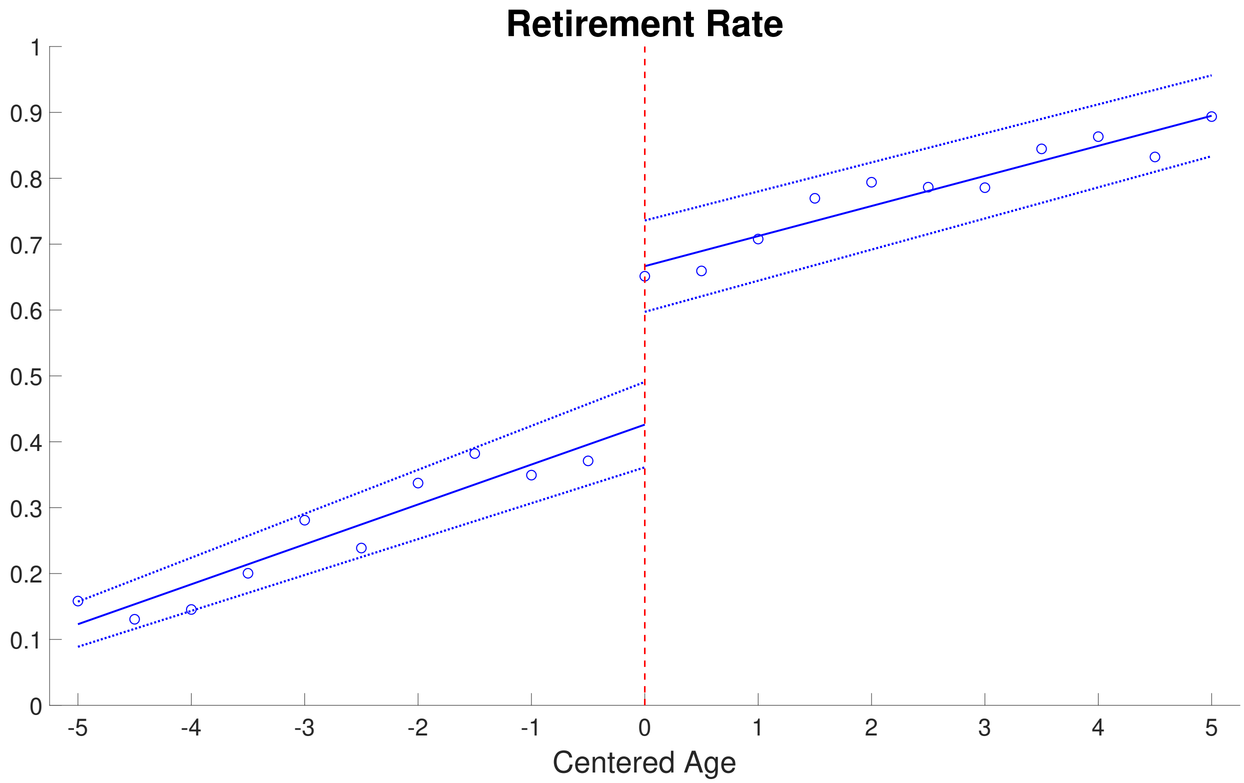


Figure 1.1: Retirement Rate by Centered Age

Note: The figure shows the share of retirees by age centered around the statutory retirement age using bins of six months overlaid with locally fitted linear lines on each side of the threshold. The blue dotted lines represent a 95% confidence interval.

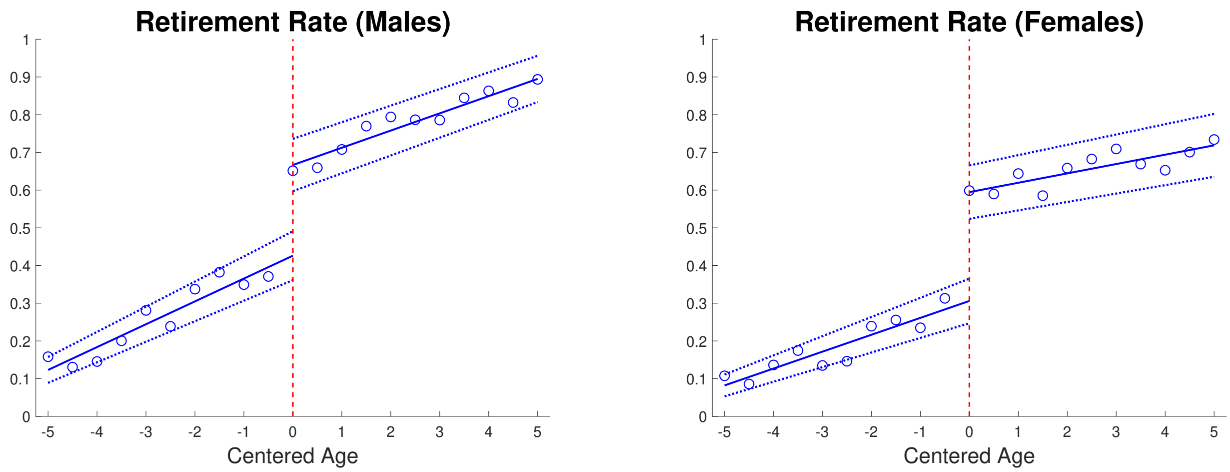


Figure 1.2: Retirement Rate by Gender

Note: The figure shows the share of retirees for each gender by age centered around the statutory retirement age using bins of six months overlaid with locally fitted linear lines on each side of the threshold. The blue dotted lines represent a 95% confidence interval.

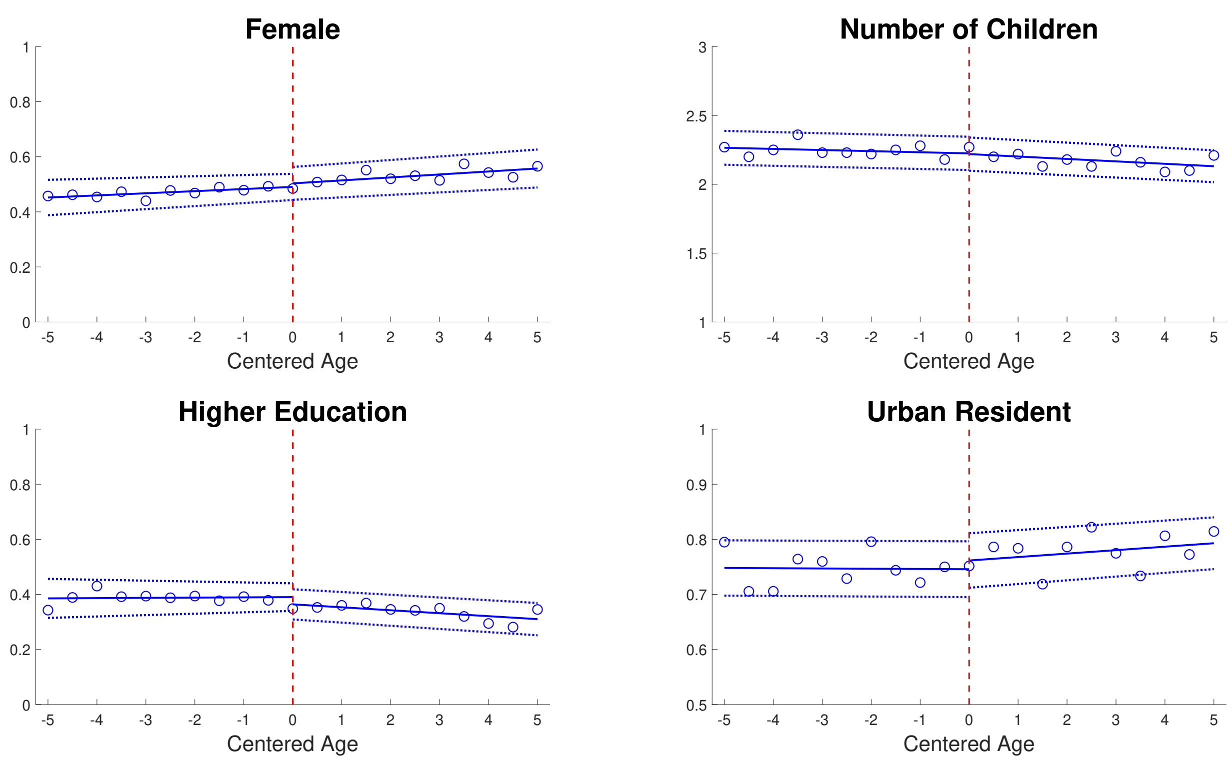


Figure 1.3: Baseline Covariates by Centered Age

Note: Figures of the proportion of females, average number of children, share of those with a higher education, and the proportion of urban residents by age centered around the statutory retirement age using bins of six months overlaid with locally fitted linear lines on each side of the threshold. The blue dotted lines represent a 95% confidence interval.

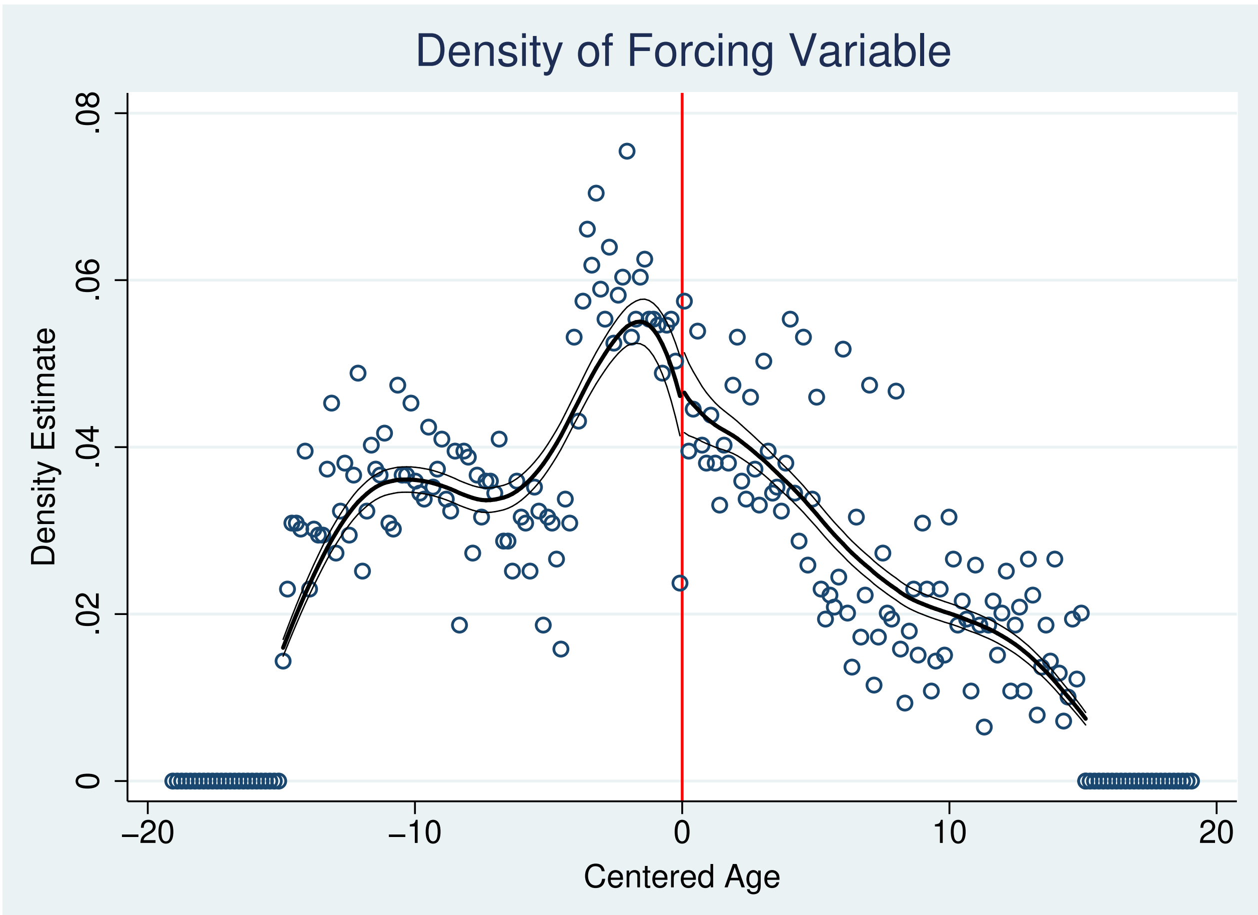


Figure 1.4: McCrary Test

Note: The figure shows the distribution of the forcing variable, centered age, around the statutory retirement threshold using bins of one month. The density function is estimated separately on both sides of the threshold using kernel local linear regressions, overlaid in black with a 95% confidence interval. Visual inspection suggests no manipulation of the forcing variable since it appears smooth near the threshold.

Table 1.1: Summary Statistics

	Male	Female	t-stat	Total	Min	Max	N
Age	56.54 (5.31)	54.40 (5.09)	31.93	55.46 (5.31)	45.08	66.58	4,630
Female	–	–	–	0.50 (0.50)	0	1	4,630
Retired	0.40 (0.49)	0.46 (0.50)	-2.91	0.43 (0.49)	0	1	4,630
Urban Province	0.74 (0.44)	0.74 (0.44)	0	0.74 (0.44)	0	1	4,630
Higher Education	0.38 (0.49)	0.34 (0.49)	1.74	0.35 (0.47)	0	1	3,362
Number of Children	1.95 (1.04)	1.95 (1.04)	0	1.95 (1.04)	0	8	4,630
Co-residing Grandchild	0.25 (0.23)	0.25 (0.23)	0	0.25 (0.23)	0	1	4,630
Household Income	34,270 (25,777)	34,270 (25,777)	0	34,270 (25,777)	600	252,000	1,400
Self Reported Health Status							
"Excellent"	0.02 (0.12)	0.01 (0.11)	0.94	0.01 (0.12)	0	1	4,143
"Very Good"	0.16 (0.37)	0.12 (0.33)	2.17	0.14 (0.35)	0	1	4,143
"Good"	0.19 (0.40)	0.15 (0.35)	2.10	0.17 (0.38)	0	1	4,143
"Fair"	0.50 (0.50)	0.54 (0.50)	-1.82	0.52 (0.50)	0	1	4,143
"Poor"	0.13 (0.34)	0.18 (0.39)	-2.66	0.16 (0.36)	0	1	4,143
CESD	2.99 (2.35)	3.64 (2.60)	-13.51	3.34 (2.51)	0	10	4,279
Systolic blood pressure	128.62 (18.94)	123.90 (19.12)	31.09	126.17 (19.18)	64.5	199	3,304
Diastolic blood pressure	78.45 (12.40)	75.00 (11.75)	28.53	76.66 (12.19)	42	138	3,304
Body mass index	24.50 (3.59)	24.77 (3.85)	-4.02	24.64 (3.73)	15.39	62.89	3,296
Physical Activity							
Moderate PA (ext. margin)	0.53 (0.50)	0.59 (0.49)	-1.82	0.56 (0.50)	0	1	1,836
Days per week of moderate PA	5.79 (1.92)	5.98 (1.88)	-2.25	5.89 (1.90)	1	7	1,075
Vigorous PA (ext. margin)	0.23 (0.38)	0.14 (0.36)	3.16	0.26 (0.41)	0	1	1,836
Days per week of vigorous PA	5.36 (2.17)	5.07 (2.24)	2.13	5.25 (2.20)	1	7	478
Smoke (ext. margin)	0.52 (0.50)	0.04 (0.19)	26.75	0.26 (0.44)	0	1	4,287
Cigarettes per day (no zeros)	19.25 (10.36)	12.43 (8.23)	30.38	19.73 (10.37)	1	70	1,058
Alcohol (ext. margin)	0.60 (0.49)	0.15 (0.36)	23.48	0.38 (0.48)	0	1	4,630
Drinking Frequency							
"Once a month"	0.06 (0.24)	0.21 (0.41)	-4.60	0.08 (0.27)	0	1	1,227
"Two or three times a month"	0.14 (0.34)	0.25 (0.44)	-3.08	0.15 (0.36)	0	1	1,227
"Once a week"	0.10 (0.30)	0.11 (0.31)	-0.31	0.10 (0.30)	0	1	1,227
"Two or three times a week"	0.14 (0.35)	0.17 (0.37)	-0.87	0.15 (0.35)	0	1	1,227
"Four or six times a week"	0.05 (0.22)	0.05 (0.21)	0.02	0.05 (0.22)	0	1	1,227
"Once a day"	0.33 (0.47)	0.19 (0.39)	3.73	0.31 (0.46)	0	1	1,227
"Twice a day"	0.15 (0.36)	0.02 (0.14)	4.05	0.13 (0.34)	0	1	1,227
"More than twice a day"	0.03 (0.17)	0.00 (0.07)	1.51	0.03 (0.16)	0	1	1,227

Note: The table presents summary statistics for the sample constructed from the data using a window of five years around the statutory retirement age. Mean values by gender with standard deviations in parentheses. The third column provides the t-statistic for the equality of means between males and females.

Table 1.2: First-stage Effects

	(1)	(2)
Dependent Variable: Own Retirement		
I(Age > Retirement Age)	0.30***	0.29***
	(0.02)	(0.02)
I(Age > Retirement Age) x Centered Age	–	0.03***
		(0.01)
Sanderson-Windmeijer F-statistic	160.68	250.79
Dependent Variable: Spousal Retirement		
I(Spouse's Age > Retirement Age)	0.31***	0.29***
	(0.04)	(0.02)
I(Spouse's Age > Retirement Age) x Spouse's Centered Age	–	0.03***
		(0.01)
Sanderson-Windmeijer F-statistic	152.50	233.45
Dependent Variable: Own Retirement x Centered Age		
I(Age > Retirement Age)	–	0.00
		(0.06)
I(Age > Retirement Age) x Centered Age	–	0.55***
		(0.04)
Sanderson-Windmeijer F-statistic	–	255.09
Dependent Variable: Spousal Retirement x Spouse's Centered Age		
I(Spouse's Age > Retirement Age)	–	0.04
		(0.06)
I(Spouse's Age > Retirement Age) x Spouse's Centered Age	–	0.51***
		(0.04)
Sanderson-Windmeijer F-statistic	–	256.31
Year Fixed Effects	Yes	Yes
Province Fixed Effects	Yes	Yes
Interaction Term	No	Yes
Kleibergen-Paap Wald *rk* F-statistic	61.32	56.74
Observations	4,630	4,630

Note: The table presents first-stage results using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: Retirement Effects on Physical and Mental Health

Outcome Variable	SRHS		CESD		BMI		Blood Pressure	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Retired	0.26***	0.26***	0.25**	0.28**	0.52	0.51	0.85 / 2.58	0.82 / 2.42
	(0.10)	(0.10)	(0.12)	(0.14)	(0.47)	(0.47)	(2.62) / (1.80)	(2.60) / (1.80)
Spouse Retired	-0.04	-0.04	0.12	0.13	0.05	0.13	1.12 / 2.68	1.17 / 2.75
	(0.10)	(0.10)	(0.33)	(0.33)	(0.43)	(0.43)	(2.88) / (1.84)	(2.87) / (1.83)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	No	Yes	No	Yes	No	Yes	No	Yes
Observations	4,143	4,143	4,279	4,279	3,296	3,296	3,304 / 3,304	3,304 / 3,304

Note: The table shows Fuzzy Regression Discontinuity estimates of the effects of own and spouse's retirement using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level. Outcome variables: SRHS is self-reported health status (1 "Excellent" – 5 "Poor"), CESD is the number of indicators for depression (0–10), BMI is body mass index, Blood Pressure is systolic/diastolic blood pressure reading.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 1.4: Retirement Effects on Physical Activity

Outcome Variable	Moderate PA		Moderate PA (days)		Vigorous PA		Vigorous PA (days)	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Retired	-0.10	-0.11	-1.21**	-1.37**	-0.18**	-0.20**	0.41	1.27
	(0.10)	(0.10)	(0.49)	(0.55)	(0.08)	(0.08)	(1.24)	(1.45)
Spouse Retired	-0.01	-0.00	-0.39	-0.36	0.00	0.02	0.31	0.63
	(0.09)	(0.10)	(0.57)	(0.57)	(0.09)	(0.09)	(0.56)	(0.73)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,836	1,836	1,075	1,075	1,836	1,836	478	478

Note: The table shows Fuzzy Regression Discontinuity estimates of the effects of own and spouse's retirement using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level. Outcome variables: Moderate PA is a dummy for moderate physical activity in the past month, Moderate PA (days) is the number of weekly days spent on moderate physical activity conditional on exercising, Vigorous PA and Vigorous PA (days) are similarly defined.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 1.5: Retirement Effects on Smoking and Drinking

Outcome Variable	Smoke		Cigarettes(no zeros)		Alcohol		Alcohol Frequency	
Specification	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Retired	-0.06	-0.05	-2.36**	-2.19**	-0.07	-0.07	-0.48	-0.45
	(0.04)	(0.04)	(1.20)	(1.11)	(0.05)	(0.05)	(0.43)	(0.44)
Spouse Retired	0.00	0.01	2.27	3.28	0.02	0.02	-0.25	-0.23
	(0.04)	(0.04)	(2.30)	(2.29)	(0.04)	(0.04)	(0.51)	(0.51)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	No	Yes	No	Yes	No	Yes	No	Yes
Observations	4,287	4,287	1,058	1,058	4,630	4,630	1,227	1,227

Note: The table shows Fuzzy Regression Discontinuity estimates of the effects of own and spouse's retirement using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level. Outcome variables: Smoke is a dummy for having smoked in the past year, Cigarettes (no zeros) is the number of daily cigarettes conditional on smoking, Alcohol is a dummy for having consumed alcohol in the past year, and Alcohol Frequency is a categorical measure of weekly alcohol consumption conditional on drinking.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: Retirement Effect Heterogeneity By Gender

Outcome Variable	SRHS	CESD	BMI	Blood Pressure	Moderate PA	Moderate PA (days)	Vigorous PA	Vigorous PA (days)	Smoke	Cigarettes (no zeros)	Alcohol	Alcohol Frequency
Subsample: Males												
Retired	0.19**	0.11**	0.52	-2.78 / -1.89	-0.14	-2.11	-0.29***	0.96	-0.01	-2.32**	-0.10	-0.62
	(0.10)	(0.05)	(0.81)	(5.08) / (3.42)	(0.20)	(1.60)	(0.10)	(2.88)	(0.08)	(1.06)	(0.08)	(0.46)
Spouse Retired	-0.17*	-0.25**	0.16	1.89 / 2.19	-0.04	-0.86	-0.02	0.66	0.04	3.23	-0.02	-0.37
	(0.10)	(0.13)	(0.69)	(3.36) / (2.33)	(0.13)	(0.67)	(0.12)	(0.79)	(0.08)	(2.68)	(0.08)	(0.45)
Observations	2,119	2,088	1,581	1,586 / 1,586	885	500	885	295	1,975	995	2,300	1,088
Subsample: Females												
Retired	0.35***	0.38***	0.86**	3.57* / 3.37*	-0.13	-1.17**	-0.12**	1.66	-0.05	1.80	-0.01	2.13
	(0.14)	(0.15)	(0.41)	(2.27) / (1.81)	(0.11)	(0.55)	(0.06)	(1.67)	(0.03)	(19.22)	(0.06)	(3.34)
Spouse Retired	-0.02	0.08	-0.59	-2.54 / -1.41	0.04	0.83	0.02	0.16	-0.02	-9.59	0.04	3.21
	(0.16)	(0.49)	(0.48)	(5.27) / (3.13)	(0.19)	(1.28)	(0.13)	(2.01)	(0.04)	(19.40)	(0.06)	(2.99)
Observations	2,024	2,191	1,715	1,717 / 1,717	951	575	951	183	2,312	63	2,331	139
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows Fuzzy Regression Discontinuity estimates of the effects of own and spouse's retirement for the subsamples of males and females using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level. Outcome variables: SRHS is self-reported health status (1 "Excellent" – 5 "Poor"), CESD is the number of indicators for depression (0–10), BMI is body mass index, Blood Pressure is systolic/diastolic blood pressure reading, Moderate PA is a dummy for moderate physical activity in the past month, Moderate PA (days) is the number of weekly days spent on moderate physical activity conditional on exercising, Vigorous PA and Vigorous PA (days) are similarly defined, Smoke is a dummy for having smoked in the past year, Cigarettes (no zeros) is the number of daily cigarettes conditional on smoking, Alcohol is a dummy for having consumed alcohol in the past year, and Alcohol Frequency is a categorical measure of weekly alcohol consumption conditional on drinking.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Retirement Effects on Time Use and Selective CESD Components

Outcome Variable	Provide Childcare	Childcare Hours (no zeros)	Sedentary Activities	Hopeful about the future	Everything is an effort	Restless sleep	More easily bothered
Subsample: Males							
Retired	0.06	4,108.49	0.13**	0.09	0.24*	0.05	0.04
	(0.15)	(2,854.41)	(0.06)	(0.13)	(0.14)	(0.13)	(0.13)
Spouse Retired	0.10*	-135.26	-0.02	0.23**	-0.05*	0.14	0.16
	(0.06)	(991.46)	(0.06)	(0.11)	(0.03)	(0.14)	(0.13)
Observations	1,559	918	1,776	2,088	2,088	2,088	2,088
Subsample: Females							
Retired	0.23**	3,123.67**	0.19**	0.12	0.05	0.22**	0.32**
	(0.11)	(1,278.28)	(0.08)	(0.16)	(0.14)	(0.11)	(0.15)
Spouse Retired	0.07	-573.17	-0.09	-0.04	0.05	-0.03	-0.07
	(0.07)	(1,535.49)	(0.07)	(0.13)	(0.11)	(0.13)	(0.14)
Observations	1,826	1,117	1,813	2,191	2,191	2,191	2,191
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows Fuzzy Regression Discontinuity estimates of the effects of own and spouse's retirement for the subsamples of males and females using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level. Outcome variables: Provide Childcare is an indicator for whether a respondent provided care for a grandchild in the past year (conditional on having a grandchild), Childcare Hours is the number of annual hours spent caring for a grandchild conditional on providing care, Sedentary Activities is an indicator for whether time was spent on board games, card games, or using the computer in the past month, Hopeful about the future is an indicator for "I feel hopeful about the future" in the past week, Everything is an effort is an indicator for "I feel like everything I did was an effort" in the past week, Restless sleep is an indicator for "my sleep was restless" in the past week, and More easily bothered is an indicator for "I was bothered by things that do not usually bother me" in the past week.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Retirement Effect Heterogeneity By Retirement Status

Outcome Variable	SRHS	CESD	BMI	Blood Pressure	Moderate PA	Moderate PA (days)	Vigorous PA	Vigorous PA (days)	Smoke	Cigarettes (no zeros)	Alcohol	Alcohol Frequency
Subsample: Non-retired												
Spouse Retired	-0.10	-0.09	0.78	1.82 / 2.51	-0.10	-0.79	0.03	0.11	-0.03	2.61	0.04	-0.57
	(0.15)	(0.41)	(0.60)	(3.54) / (2.30)	(0.12)	(0.66)	(0.12)	(0.75)	(0.07)	(2.46)	(0.07)	(0.52)
Observations	2,508	2,562	1,880	1,887 / 1,887	1,035	683	1,035	319	2,567	715	2,812	873
Subsample: Retired												
Spouse Retired	0.10	0.43	-0.49	1.60 / 3.30	0.10	1.90	0.01	4.64	0.04	2.09	0.00	1.84
	(0.14)	(0.59)	(0.59)	(5.28) / (3.23)	(0.19)	(1.22)	(0.11)	(4.37)	(0.04)	(4.15)	(0.05)	(1.21)
Observations	1,635	1,717	1,416	1,417 / 1,417	801	392	801	159	1,720	343	1,818	354
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table shows Fuzzy Regression Discontinuity estimates of the effects of own and spouse's retirement for the subsamples of retirees and non-retirees using a window of five years around the statutory retirement age. Standard errors are clustered at the individual level. Outcome variables: SRHS is self-reported health status (1 "Excellent" – 5 "Poor"), CESD is the number of indicators for depression (0–10), BMI is body mass index, Blood Pressure is systolic/diastolic blood pressure reading, Moderate PA is a dummy for moderate physical activity in the past month, Moderate PA (days) is the number of weekly days spent on moderate physical activity conditional on exercising, Vigorous PA and Vigorous PA (days) are similarly defined, Smoke is a dummy for having smoked in the past year, Cigarettes (no zeros) is the number of daily cigarettes conditional on smoking, Alcohol is a dummy for having consumed alcohol in the past year, and Alcohol Frequency is a categorical measure of weekly alcohol consumption conditional on drinking.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: Treatment Effect Derivatives (TED) and Complier Probability Derivatives (CPD) of the Retirement Effects

Outcome Variable	SRHS	CESD	BMI	Blood Pressure	Moderate PA	Moderate PA (days)	Vigorous PA	Vigorous PA (days)	Smoke	Cigarettes (no zeros)	Alcohol	Alcohol Frequency
Treatment: Own Retirement												
TED	-0.02**	0.05**	0.09	0.16 / -0.08	-0.02	-0.18	0.03**	0.07	0.00	-0.39	0.02	-0.09
	(0.01)	(0.03)	(0.16)	(0.48) / (0.57)	(0.04)	(0.24)	(0.02)	(0.08)	(0.02)	(0.54)	(0.02)	(0.21)
Relative TED	2.61	1.12	0.85	1.03 / 6.05	1.09	1.52	1.25	3.63	2.51	1.13	0.72	0.99
CPD	0.03***	0.03***	0.04	0.04*** / 0.04***	0.03***	0.05***	0.03***	0.04	0.03***	0.02**	0.03***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01) / (0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Relative CPD	1.91	1.93	1.02	1.01 / 1.01	1.92	1.81	1.92	1.43	1.09	1.77	1.23	1.78
Treatment: Spousal Retirement												
TED	0.01*	-0.05	-0.10	0.55 / -0.09	-0.01	0.06*	-0.02	0.18	-0.03	-0.87	-0.01	-0.02
	(0.01)	(0.07)	(0.27)	(1.09) / (0.65)	(0.04)	(0.24)	(0.04)	(0.22)	(0.03)	(1.16)	(0.02)	(0.13)
Relative TED	0.81	0.52	1.04	0.43 / 6.11	0.12	1.21	0.19	0.68	0.09	0.75	0.46	2.09
CPD	0.03***	0.03***	0.04	0.04*** / 0.04***	0.03***	0.04***	0.04***	0.02	0.03***	0.02**	0.03***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01) / (0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Relative CPD	1.92	1.95	1.03	1.04 / 1.05	1.88	1.98	1.88	2.29	1.13	1.45	1.22	1.13
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,143	4,279	3,296	3,304 / 3,304	1,836	1,075	1,836	478	4,287	1,058	4,630	1,227

Note: The table shows the estimated Treatment Effect Derivatives (TED) and Complier Probability Derivatives (CPD) for the main results in Tables 3, 4, and 5 of Section **??**. These derivatives are estimated following procedures outlined in Dong and Lewbel (2015) and Cerulli et al. (2017). Relative TED (CPD) can be used alongside the TED (CPD) to test the stability and, hence, external validity of the main results: if either TED or CPD is statistically significant, and the magnitude of its Relative measure is less than one, this suggests instability in the estimated retirement effect. Robust standard errors are in parentheses. Outcome variables: SRHS is self-reported health status (1 "Excellent" – 5 "Poor"), CESD is the number of indicators for depression (0–10), BMI is body mass index, Blood Pressure is systolic/diastolic blood pressure reading, Moderate PA is a dummy for moderate physical activity in the past month, Moderate PA (days) is the number of weekly days spent on moderate physical activity conditional on exercising, Vigorous PA and Vigorous PA (days) are similarly defined, Smoke is a dummy for having smoked in the past year, Cigarettes (no zeros) is the number of daily cigarettes conditional on smoking, Alcohol is a dummy for having consumed alcohol in the past year, and Alcohol Frequency is a categorical measure of weekly alcohol consumption conditional on drinking.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.10: Counterfactual Retirement Effects for a Five-year Increase in Statutory Retirement Ages

Outcome Variable	SRHS	CESD	BMI	Blood Pressure	Moderate PA	Moderate PA (days)	Vigorous PA	Vigorous PA (days)	Smoke	Cigarettes (no zeros)	Alcohol	Alcohol Frequency
Sample: Males												
Treatment: Own Retirement												
Current Retirement Policy	0.19**	0.11**	0.52	-2.78 / -1.89	-0.14	-2.11	-0.29***	0.96	-0.01	-2.32**	-0.10	-0.62
	(0.10)	(0.05)	(0.81)	(5.08) / (3.42)	(0.20)	(1.60)	(0.10)	(2.88)	(0.08)	(1.06)	(0.08)	(0.46)
New Policy (5-year Increase)	0.03*	0.32**	0.44	-2.58 / -1.19	-0.06	-3.06	-0.15**	1.49	-0.03	-4.07*	-0.12	-1.04
	(0.02)	(0.15)	(0.73)	(4.87) / (2.93)	(0.13)	(2.24)	(0.07)	(2.94)	(0.08)	(2.111)	(0.08)	(0.50)
Observations	2,119	2,088	1,581	1,586 / 1,586	885	500	885	295	1,975	995	2,300	1,088
Treatment: Spousal Retirement												
Current Retirement Policy	-0.17*	-0.25**	0.16	1.89 / 2.19	-0.04	-0.86	-0.02	0.66	0.04	3.23	-0.02	-0.37
	(0.10)	(0.13)	(0.69)	(3.36) / (2.33)	(0.13)	(0.67)	(0.12)	(0.79)	(0.08)	(2.68)	(0.08)	(0.45)
New Policy (5-year Increase)	-0.19*	-0.56**	-0.14	4.02 / 2.39	-0.11	-0.55	-0.14	1.46	-0.03	-0.92	-0.08	-0.41
	(0.10)	(0.27)	(0.77)	(3.88) / (2.44)	(0.15)	(0.62)	(0.14)	(1.03)	(0.08)	(1.46)	(0.09)	(0.48)
Observations	2,119	2,088	1,581	1,586 / 1,586	885	500	885	295	1,975	995	2,300	1,088
Subsample: Females												
Treatment: Own Retirement												
Current Retirement Policy	0.35**	0.38***	0.86**	3.57* / 3.37*	-0.13	-1.17**	-0.12**	1.66	-0.05	1.80	-0.01	2.13
	(0.14)	(0.15)	(0.41)	(2.27) / (1.81)	(0.11)	(0.55)	(0.06)	(1.67)	(0.03)	(19.22)	(0.06)	(3.34)
New Policy (5-year Increase)	0.41**	0.71***	1.21*	4.62 / 3.52	-0.28	-1.87*	-0.05*	1.91	-0.01	3.32	0.10	3.67
	(0.16)	(0.26)	(0.64)	(2.88) / (2.19)	(0.18)	(0.96)	(0.03)	(1.75)	(0.03)	(20.90)	(0.07)	(3.89)
Observations	2,024	2,191	1,715	1,717 / 1,717	951	575	951	183	2,312	63	2,331	139
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interaction Term	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Current Retirement Policy refers to estimated retirement effects under the current statutory retirement ages. These estimates are identical to the results reported in Table 6. New Policy refers to estimated retirement effects under a counterfactual policy where statutory retirement ages are increased by five years for everyone. These are estimated by applying procedures outlined in Dong and Lewbel (2015). Robust standard errors are in parentheses. Standard errors for the retirement effects under the New Policy are computed using the delta method. Outcome variables: SRHS is self-reported health status (1 "Excellent" – 5 "Poor"), CESD is the number of indicators for depression (0–10), BMI is body mass index, Blood Pressure is systolic/diastolic blood pressure reading, Moderate PA is a dummy for moderate physical activity in the past month, Moderate PA (days) is the number of weekly days spent on moderate physical activity conditional on exercising, Vigorous PA and Vigorous PA (days) are similarly defined, Smoke is a dummy for having smoked in the past year, Cigarettes (no zeros) is the number of daily cigarettes conditional on smoking, Alcohol is a dummy for having consumed alcohol in the past year, and Alcohol Frequency is a categorical measure of weekly alcohol consumption conditional on drinking.
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Chapter 2

Home Equity Lending, Credit Constraints and Small Business in the US[1]

2.1 Introduction

This paper provides evidence on the effects of credit constraints on small business activity in the US, using as a natural experiment the timing of a constitutional amendment in Texas in 1997 that relaxed severe restrictions on home equity loans. Since the work of Abdallah and Lastrapes (2012), Texas's amendment has been exploited as a source of exogenous variation in credit constraints – and interpreted as a credit supply shock – to study the role of financial market imperfections in a wide variety of contexts.[2] To examine the role of credit constraints for small business, we use restricted access business data from the Census Bureau to compare small business outcomes – new business creation, exits of existing firms, job creation and destruction, and job reallocation – before and after the amendment in Texas. We rely on difference-in-differences methods to isolate causal effects of Texas's credit supply shock on these outcomes.

It is well understood that small businesses and entrepreneurs play an important role in the US economy. According to the Census Bureau's Business Dynamics Statistics (which is derived from the Longitudinal Business Database), businesses with fewer than ten employees accounted for about three-quarters of all firms that hire workers, employed one-eighth of all workers and were responsible for one-quarter of net job growth annually on average over the period 1992-1997. The typical start-up company is small, employing four to six workers in its first year.

Unlike large firms and corporations, small firms and would-be entrepreneurs are likely to face binding constraints on borrowing to finance business activity (Schmalz et al., 2017). For example, the National Small Business Association (NSBA) year-end and mid-year reports over the past decade claim that around 30% of respondents to the NSBA's routine survey of entrepreneurs cannot receive 'necessary' funding

[1]This chapter is joint work with William D. Lastrapes and Ian M. Schmutte.

[2]See, for example, Zevelev (2020), Kumar (2019), Kumar (2018), and Stolper (2015).

for their business. And the 2012 NSBA Small Business Access to Capital Survey reported that 45% of small business owners who were denied a loan could not obtain a loan because of a low credit score or insufficient collateral. Debt finance, collateralized by housing equity, helps relax these constraints and is the most common form of external finance for individuals running their own business or starting a new one (Robb and Robinson, 2014). Studying the ability of housing to ease borrowing constraints for small business adds to our knowledge of financial market imperfections and has important implications for theory and policy. At the same time, we gain a better understanding of how housing markets spillover into the wider economy.

Before 1998, households, small business owners and entrepreneurs in Texas were unable to use equity in their homes to support borrowing for consumption or to finance business ventures. Until then, the Texas State Constitution prohibited mortgage lending for all but a very limited set of expenditures, such as the original purchase of the house and home improvements. But citizens in the state amended their constitution in 1997 to relax these restrictions, thereby providing a new source of funding for private sector activity. At the time of this amendment, no other state in the US restricted home equity loans for general purposes, including for use as collateralizing small business loans, as strictly as Texas.

We interpret this political event in Texas as an exogenous relaxation of credit constraints for the state's entrepreneurs and small businesses, and one with the potential to greatly increase the supply of credit. Contemporary accounts estimate home equity in the state during the 1990's at up to $200 billion, implying a range of collateralized lending of $4 billion to $10 billion annually (Abdallah and Lastrapes, 2012, p. 118). Whether the timing of this event is exogenous is a matter of judgment, since there was no random assignment of 'treatment' across agents; however, Abdallah and Lastrapes (2012), p.100, through a careful reading of the circumstances of the amendment's enactment, provide supporting evidence that the amendment was an incidental source of variation in credit availability rather than a response to the demand for credit. As such, the event qualifies as a natural experiment and allows us to disentangle the effects of changes in credit supply from those of credit demand using a difference-in-differences empirical strategy. Abdallah and Lastrapes (2012), using similar methods but different data, estimate that this surge in credit availability led to an increase in state- and county-wide retail sales in Texas of 2 to 7 percent, but do not study the potential impact on business activity.

The primary source for the small business data used in our main analysis is the Longitudinal Business Database (LBD) compiled by the US Census Bureau. We use these data to construct a balanced panel of 159 Texas and non-Texas counties from 1992 to 2003 – a period spanning the constitutional amendment date – for annual rates of business creation and destruction, and job creation and destruction. This data set is a rich source of information covering the universe of registered business establishments in the US. We also use the Bureau's Survey of Business Owners (SBO) in 1992 and 2007 to examine a prior question, which aids the interpretation of the main results: whether the Texas amendment increased the incidence of home equity loans by small business owners to start, acquire or expand their business. Both the LBD and SBO are restricted-use microdata requiring special permission from the Census Bureau to ensure confidentiality. We rely on these data as well as other public-use data to construct control variables in the main analysis, and to help sort out the potential economic mechanisms driving the reduced form results.

The Census Bureau collects and compiles other data, specifically County Business Patterns (CBP) and Business Dynamics Statistics (BDS), that are relevant to our study and freely available for public use. Our preliminary analysis with these data shows only weak effects of the Texas amendment on small business growth, entry and exit rates, and job creation and destruction rates; we report and briefly discuss these results in the online appendix. However, these data lack the richness of the firm-level restricted sources, and have known measurement issues. For example, outliers in the BDS data in Economic Census years of our sample (1992, 1997 and 2002), caused by problems arising from reconciling the timing of entry and exit among multi-establishment firms (Jarmin and Miranda, 2002), can bias inference. The restricted-access data that we use in the main analysis distinguish between single- and multi-unit firms, which allows us to control for this problem by focusing on single-unit firms. Most importantly, restricting our analysis to single-unit firms also focuses attention on firms most likely to be affected by the law – smaller businesses that are typical users of home equity financing. Using the restricted-access microdata is thus essential for a full exploration into the small business effects of the change in Texas home equity law.

We describe the microdata and the empirical modeling strategy for the primary aims of the paper in section 2.2, and report results using these restricted data in section 2.3. In the first sub-section, we provide the first evidence that the Texas amendment directly affected the use of home equity as a source of funding for small business activities. Using the SBO data, we find that 5.1 percent of business owners outside Texas reported using home equity in 1992, but that just 0.3 percent did in Texas. By 2007, Texas business owners were using home equity at the same rate as the rest of the country: 6.6 versus 6.8 percent. Clearly, Texas law prior to the amendment was effective in shutting down the use of home equity to finance business investment. This finding holds up in our more formal analysis, which controls for fixed effects and other control variables. The findings are important, not just for our study, but for any analysis that assumes first-order effects on borrowing behavior to motivate the timing of the Texas amendment as an instrument for relaxing credit constraints.

Our main findings are that small business dynamics and employment were significantly affected by the amendment, with business and job creation rising and destruction falling. These beneficial effects were larger for small firms, which is consistent with the notion that small businesses tend to be more credit constrained larger firms. These findings could be due either to a supply-side channel, where credit availability reduces finance costs for small businesses, or a demand-side channel, where credit availability enhances small business through enhanced purchasing power of households in buying their products, as suggested by the findings of Abdallah and Lastrapes (2012). Using sub-samples to identify sensitivity to treatment, we conclude that the evidence presented in section 2.3 leans more heavily toward supply-side factors as driving our reduced form results.

The literature on credit constraints and small business is vast.[3] The extant work closest to ours is Kerr et al. (2019), which has similar aims, relies on data from the SBO and LBD (although not exclusively) and in the first part of their paper exploits the Texas credit natural experiment as a source of exogenous variation. Our work strongly complements theirs, answering some questions from the data that theirs

[3] Examples include Nykvist (2008), Johansson (2000), Hurst and Lusardi (2004), Evans and Jovanovic (1989), Jensen et al. (2014), and Lelarge et al. (2010).

does not. That paper focuses on the role of housing prices as a driver of the amendment's effect, but most significantly it ignores the effects of the amendment on business exit and job destruction rates, focusing exclusively on birth and creation rates like most of the literature in this area. Ignoring exit and destruction downplays local equilibrium effects and forecloses any inference regarding *net* effects of credit availability on small businesses. As noted above we show in our microdata analysis that the Texas amendment actually did increase home equity lending for small business, a result not documented by Kerr et al. (2019). We provide supporting evidence for the effectiveness of the natural experiment by showing equal pre-treatment trends. And finally, we use triple difference-in-differences models to present robust evidence for whether supply or demand channels dominate. Kerr et al. (2019) speculate on the underlying economic mechanisms driving the results, but provide little evidence on these mechanisms using the LBD, although they explore these issues with another data set.

2.2 Data

The Census Bureau's Longitudinal Business Database (LBD) tracks establishments and firms that have at least one paid employee. The LBD is based on an edited version of the Census Business Register, which the Bureau uses as a sampling frame for all economic surveys. Using information from other sources, the LBD tracks establishments and firms consistently over time in a way that addresses changes in ownership. The longitudinal consistency makes it possible to track when businesses are born and when they die. The LBD also includes information about payroll and employment, as well as industry and location. From these data we construct annual, county-level aggregates of the main variables of interest.[4]

In this paper we limit our sample to single-unit firms, as do, for example, Jarmin and Krizan (2010). A single-unit firm owns or operates just one establishment, which is defined as a single physical location of business. Hereafter and unless otherwise noted, we use the term 'firm' to mean such single-unit entities. As we note above, one reason we focus on single-unit firms is to ensure consistency of the timing of entry and exit of the observations in the sample. The other reason is that we do not expect home equity financing to be a common source of funding for multi-unit firms and so limit our sample to businesses for which the change in policy would matter at the margin. While we potentially sacrifice some external validity with this sample, we gain precision and reduce the chances that existing effects of relaxed credit constraints are masked by firms for which credit constraints do not bind.

The LBD records an individual firm's birth year as the first year it appears in the Bureau's records. Total births in county i during year t, b_{it}, is the number of firms newly formed in that county during year t, computed by adding up all firms with birth year t in county location i. Likewise, total firm deaths in county i year t, d_{it}, is the sum of firms in that county for which the last year the Bureau observes the firm's existence from these sources is year t. We convert these measures to rates, following standard practice of the Census Bureau,[5] by normalizing these flows on the two-year average of total number of firms for

[4] For additional details see census.gov/programs-surveys/ces/data/restricted-use-data/longitudinal-business-database.html and Jarmin and Miranda (2002).

[5] www.census.gov/programs-surveys/bds/documentation/methodology.html and Davis et al. (1996).

county i to obtain entry and exit rates, respectively:

$$ER_{it} = \frac{b_{it}}{1/2(n_{it} + n_{i,t-1})}$$
$$XR_{it} = \frac{d_{it}}{1/2(n_{it} + n_{i,t-1})}$$

where n_{it} is the total number of firms in the sample in county i, year t.

With respect to the labor market, we directly observe from the LBD the number of paid employees on payroll in March for each firm and each year. To measure job creation and destruction at the county level, both flow concepts as with firm births and deaths, we compute year-to-year changes in the employment measure for each firm, then aggregate. The number of new jobs created in county i and year t, jc_{it}, is the sum of all employment increases from year $t-1$ to t for expanding firms in that county, including new jobs from firm births. Jobs destroyed, jd_{it}, is the sum of all job decreases coming from contracting firms in the county-year, including losses from firm deaths. To be specific,

$$jc_{it} = \sum_{h=1}^{n_{it}} (E_{iht} - E_{ih,t-1}), \text{ for } E_{iht} - E_{ih,t-1} > 0$$
$$jd_{it} = \sum_{h=1}^{n_{it}} |E_{iht} - E_{ih,t-1}|, \text{ for } E_{iht} - E_{ih,t-1} < 0$$

where E_{iht} is the level of employment in firm h in county i in (March of) year t . We again convert to relative magnitudes using a similar base as firm births and deaths:

$$JCR_{it} = \frac{jc_{it}}{1/2(E_{it} + E_{i,t-1})}$$
$$JDR_{it} = \frac{jd_{it}}{1/2(E_{it} + E_{i,t-1})}$$

where $E_{it} = \sum_{h=1}^{n_{it}} E_{iht}$. We also compute a county's excess reallocation rate as

$$ERR_{it} = JCR_{it} + JDR_{it} - |JCR_{it} - JDR_{it}|.$$

The excess reallocation rate measures job reallocation exceeding the amount needed to support net employment growth.

Table 2.1 reports sample means and standard deviations of these variables for the time prior to the amendment separately for Texas and non-Texas counties in states bordering Texas and bordering the border states (our second control group described below); disclosure rules from the Census Bureau preclude reporting of post-amendment statistics. There are no major differences in these statistics across the two samples.

The Survey of Business Owners (SBO) is conducted every five years on the same schedule as the Economic Census. Prior to 1997, the survey was called the Characteristics of Business Owners (CBO). Both surveys target the universe of businesses that filed tax forms reporting business income, with or without employees. The survey "provides basic economic, demographic, and sociological data on the characteristics of minority, women, and non-minority male business owners and their business activities" (Census Bureau, 1992). However, the exact questions and sample change from year to year. We are primarily interested in the response to a single question that was asked in the 1992 CBO and the 2007 SBO: "For the current owners(s), as of December [year], what was the source(s) of capital used to start or acquire this business?" In those years, one of the available options was "Personal/family home equity loan".[6]

Table 2.2 shows sample proportions from the survey for the entire sample of firms (not broken down by treatment and control) for the years 1992 and 2007. There are over one million weighted observations for each year. The first three rows reflect answers to question #66 about the source of finance; the remaining are for demographic features of the data that are used as controls in the analysis of section 2.3.1. The proportion of businesses using home equity loans was generally steady, rising by about 1.7 percentage points from 1992 to 2007. Over the same time period, there was more than a ten percentage point rise in 'no loans' used to start or acquire a business at the expense of other loans.

We also use publicly available data to construct county-level variables for each year, included in the difference-in-difference analysis to control for time-varying factors that influence local business outcomes that differ on average between counties inside and outside Texas. From the Census Bureau's Population Estimates we obtain county-level population estimates by gender, race, and ethnicity. From the Bureau of Labor Local Area Unemployment Statistics we obtain unemployment rate and labor force figures. From the Internal Revenue Services' Statistics of Income we obtain adjusted gross income and wages, county-level housing price indices come from the Federal Housing Finance Agency's Annual House Price Indexes, and median housing prices by county are from the Federal Financial Institutions Examination Council's Mortgage Lending Assessment Data Files. For additional details about the data, see the online appendix.

In the analysis below, we consider individual firms in Texas to be in the treatment group, while those outside Texas are in the control group. For the main results we select three sets of control groups: 1) counties located in states in the Census's South region (Delaware, Maryland, West Virginia, Virginia, Kentucky, North Carolina, Tennessee, South Carolina, Georgia, Alabama, Mississippi, Florida, Louisiana, Arkansas and Oklahoma), 2) counties located in a border state of Texas or a border state of a Texas border state, and

[6]The relevant question is #66 in Figure 2.1, which is a replica of p. 7 of the survey in 2007. We compare responses to this question to question 14c from the 1992 CBO form 1, which has slightly different wording but is essentially the same. Comparability over time in the underlying population is addressed in technical documentation provided by the Census Bureau (Census Bureau, 1992, 2006) and summarized in Appendix A of (Fairlie and Robb (2008). Between 1992 and 2002 Census eliminated businesses with receipts under $1000 from the sample, and added C corporations. The sectoral scope of the survey also expanded between 1992 and 2002 to include information, FIRE, real estate, and health-care. For our microdata analysis we use comparable samples adjusted by reweighting. Fairlie (2010) and Fairlie and Robb (2008) demonstrate the feasibility of using the restricted CBO and SBO to study changes over time in characteristics of small businesses. Question #70 in the 2007 form has no analogue in the 1992 survey.

3) counties located in a border state of Texas plus Colorado plus Kansas.[7] Generally, none of our results are sensitive to the control group used, so we report results primarily for the second group. Since we use public-access data to construct control variables, we restrict our sample to counties for which we have complete data over the sample period; this restricts both the treated and untreated samples to counties with a population of at least 30,000 and at least 100 annual housing transactions for each year during the sample period.[8] In the end, we include 47 counties from Texas, and 112 counties in the non-Texas control group. The map in figure 2.2 shows both the control and treatment counties used in our empirical work.

2.3 Main results

We ask three empirical questions regarding the credit supply shock in Texas: 1) did the Texas amendment increase the incidence of home equity loans for business? 2) did the Texas amendment affect business outcomes and do such effects depend on firm size? and 3) were these effects the results of supply-side or demand-side factors?

2.3.1 Did the Texas amendment increase home equity loans?

As we note in the introduction, the survey data from the SBO imply that 0.3% of Texas business owners used a home equity loan to start or acquire their business in 1992, compared to 5.1% for non-Texas owners, strongly suggesting that the law was binding for Texas entrepreneurs. In 2007, home-equity usage caught up in Texas: 6.6% of businesses in the state relied on a home-equity loan – a large rise from 1992 – while 6.8% used such loans outside of Texas, a much smaller increase. The fact that Texas 'looks like' other states after the amendment is solid *prima facie* evidence that the amendment significantly relaxed borrowing constraints in Texas, but we provide more formal evidence from a difference-in-differences specification in the context of a linear probability model:

$$z_{ht} = \alpha_0 + \alpha_1 \phi_t + \delta \tau_h \phi_t + \beta X_{ht} + \theta_s + \epsilon_{ht} \tag{2.1}$$

for $h = 1, \cdots, N$ and $t = 1992, 2007$. N is the total number of firms in the survey, $\tau_h = 1$ if firm h resides in Texas and 0 otherwise, $\phi_t = 1$ for $t = 2007$ and 0 for 1992, θ_s is a state fixed effect, and the dependent variable $z_{ht} = 1$ if firm h reports the use of home equity loans and 0 otherwise. The control variables in X_{ht} include primary business-owner demographics (age, sex, race, education) and firm characteristics (single owned, franchise, exporter, non-employment history, industry sector).[9] Inference relies

[7] We could not limit the third control group to only Texas border states because of disclosure restrictions. Results for the first and third control groups are reported in the online appendix.

[8] The Census Bureau's Population Estimates are only available for counties with a population of at least 30,000 and the Federal Housing Finance Agency's Annual House Price Indexes are only available for counties with at least 100 annual housing transactions.

[9] In separate regressions we considered more detailed industry group controls rather than industry sectors; our results were robust. We are bound by more stringent disclosure-avoidance rules and cannot release specific results from the industry group control regressions.

on cluster-robust standard errors, which account for arbitrary spatial correlation across counties within their particular state.[10]

Table 2.3 shows that the initial inference holds up, even after controlling for demographic factors and state fixed effects.[11] All else the same, the proportion of Texas business owners using home equity loans rose by almost 7 percentage points relative to non-Texas businesses, although the use of such loans remained lower on average in Texas. As we might expect, the greater incidence of home-equity loans reduced the use of other loans. The survey data do not allow us to determine whether the dollar value of new home-equity loans simply replaced the lost value of other loans. Yet we can safely infer that, even if home equity loans replaced other loans one-for-one in value, the former are presumably less expensive alternatives which could have important benefits for credit-constrained firms.

Ferman and Pinto (2019) have shown that cluster-robust standard errors of the type we use above do not perform well in models with only one treated group and two time periods as in our case using the SBO survey data for 1992 and 2007. We thus rely on the method proposed by Ferman and Pinto (2019) to ensure that our results are robust to the potential size distortions. Their method is an extension of the cluster residual bootstrap with the null hypothesis imposed, with residuals corrected for heteroskedasticity. Table 2.4 reports two sets of results for each model: a standard cluster residual bootstrap and a separate bootstrap with the adjustment (both with 100,000 draws). The null hypothesis is that the true point estimate for δ is equal to zero, while Table 2.4 reports p-values derived from bootstrap-corrected inference for the coefficients reported in the top panel of Table 2.3. The p-values in Table 2.4 are all, but for one case, less than 5%, which supports the robustness of our findings.

In sum, we document a robust fact that the Texas constitutional amendment, by freeing up the use of home equity as collateral, indeed relaxed binding constraints on borrowing by small businesses in Texas. This finding helps make sense of our results in the next sub-section, but also generally supports other research that relies on the amendment as a natural experiment for analyzing credit supply shocks.

2.3.2 Did the Texas amendment affect small business dynamics?

We examine this question in two steps, now relying on the annual data from the LBD over the years 1992 to 2003. First, we estimate the broader effect of the amendment on all small businesses in Texas using a standard difference-in-differences design with two-way fixed effects. We then split the sample by firm size, estimating a triple difference-in-differences model by adding a size indicator and interaction terms. The motivation to consider differential effects on large and small firms is that the latter are more likely to face binding credit constraints than the former.

[10] Our standard errors will underestimate sampling uncertainty to the extent that there is spatial correlation across counties in different states, but will overestimate such uncertainty if spatial clustering occurs away from state borders (Kelly, 2020; Cameron and Miller, 2015). A glance at Figure 2.2 suggests that our data appear to be consistent with the latter; thus, our reported test statistics are likely to be conservative and to understate tests for statistical significance. In any case, the coefficients are consistent and the magnitude of the estimates are unaffected by potential cross-state spatial correlation. We cannot release these results owing to disclosure concerns.

[11] Although not shown in the table, the results are robust to the inclusion of county-level fixed effects.

The difference-in-differences model is

$$y_{it} = \alpha_0 + \alpha_1\tau_i + \alpha_2\phi_t + \delta\tau_i\phi_t + \beta X_{ht} + \epsilon_{it} \tag{2.2}$$

for $i = 1, \cdots, 159$ counties (47 in Texas and 112 outside of Texas) and t = 1992 to 2003, a total of 1,908 county/year observations. We estimate the model alternately for dependent variables $y = ER, XR$, JCR, JDR, EER, where we construct these variables using observations from all businesses in the sample. τ_i is set to one for Texas counties and ϕ_t is set to one post-1997; so δ, the treatment effect, measures the difference between how Texas outcomes change after the amendment to how non-Texas counties change. X_{it} is a set of geographic and demographic controls; in particular, county-level proportion of males, race proportions, unemployment rate, population growth rate and wage growth rate.[12]

Panel A of Table 2.5 shows the estimation results from this model for two cases, without control variables ($\beta = 0$) to the left and with controls to the right. Focusing on the latter, we find that the amendment has a statistically significant positive effect on firm entry and negative effect on firm exit, both findings consistent with a loosening of binding credit constraints.[13] The estimates imply that after the amendment the rate of new business creation in Texas counties was almost 18 basis points higher than in the control counties; based off the pre-amendment entry-rate sample mean of around 5% (Table 2.1), this rise is equal to a 3.5% increase. Although this increase is non-trivial, it is only 17% of the entry rate standard deviation of 1.067 in pre-amendment Texas. The coefficient estimate on exit is of similar absolute magnitude, but reflects only a 1.75% increase off the higher mean exit rate. The estimates for the labor market variables, JCR, JDR and ERR are of consistent signs, but are not significantly different from zero. The relatively small effects are consistent with the findings of Kerr et al. (2019). When we add state-specific time trends (Panel B) there are some changes: the effect on entry gets larger in absolute magnitude, while the effect on exit becomes smaller and statistically insignificant. There is now a larger and statistically significant increase in the negative effect on ERR. However, the overall interpretation – that the effects are non-trivial but small – does not change.

Given that small businesses are more likely to face credit constraints than larger ones, we split the sample of business outcomes into rates across firm size. Let $s_j = 1$, where j indexes size category, indicate an entry/exit or creation/destruction rate computed from firms hiring fewer than ten employees – we call these firms 'small'.[14] For example, for $s_j = 1$ ER_{ijt} is the entry rate computed by adding up all births in county i during year t for only small firms so defined. s_j is zero when j does not correspond to a small firm. We then estimate the following triple difference-in-differences specification:

[12] Our results are robust to minor variations on the control variables, such as using the employment-to-population ratio instead of the unemployment rate and lagged values instead of contemporaneous ones.

[13] As in Table 2.3, we compute and report cluster-robust standard errors; however, because in these models we rely on multiple treated units and time periods, inference is not subject to the critique raised by Ferman and Pinto (2019).

[14] The ten-employee threshold to distinguish small versus large firms is common in the literature; for example, see Table 7 in Kerr et al. (2019). The firm size classification used by Census and the BLS goes Size Class 1 (1-4 employees), Size Class 2 (5-9 employees), Size Class 3 (10-19), etc. The average startup has between 4-6 employees (as mentioned in the introduction), so the average new employer would appear in either size class 1 or 2; hence, we combined the two size classes.

$$y_{ijt} = \alpha_0 + \alpha_1\tau_i + \alpha_2 s_j + \alpha_3\phi_t + \alpha_4\tau_i s_j + \alpha_5 s_j\phi_t + \delta_1\tau_i\phi_t + \delta_2\tau_i s_j\phi_t + \beta X_{ijt} + \epsilon_{ijt}. \quad (2.3)$$

We use the same sample as before, but since we have dichotomized it by two size categories the number of county-year observations doubles to over 3,800.[15] Given this specification, δ_1 is the change in non-small Texas firms from before to after the amendment minus the change in non-small, non-Texas firms; for small firms the analogous treatment effect is the sum $\delta_1 + \delta_2$, so δ_2 is the differential treatment effect between small and non-small firms. These estimates are reported in Panel C of Table 2.5 for the model without state-specific time trends, and in Panel D with such effects.

The inference is clear – the amendment's effects are driven primarily through its impact on smaller firms as we define them here. The positive response of the entry rate almost doubles for smaller firms to 33 basis points ($\delta_1 + \delta_2$), while the response for non-small firms is essentially zero. The smaller firm response is over 30% of the standard deviation of post-amendment Texas entry rates. Instead of declining by 17.6 basis points for all firms as implied by model (2.2), the small business exit rate declines by almost 58 basis points, which is more than half of the average variation. The treatment effect for small businesses job creation rate is 77 basis points, an increase of 5.5% off the full sample mean job creation rate of 14%, and 16.5% of average variation. The standard errors indicate that these treatment effects are statistically significantly different for small firms compared to non-small firms.[16] As can be seen, the addition of state-specific time trends has no effect on the results.

The excess reallocation rate rises by almost two percentage points (200 basis points) for small firms only, while it falls about half a percentage point among larger firms (falls by three-quarters of a percentage point when state-specific time trends are included). The excess reallocation rate is an indicator of business dynamism, as it measures the movement of economic activity from contracting to expanding firms. This finding suggests that credit constraints may have been hindering productive entrepreneurs from entering or expanding their businesses prior to the law change.

One possible concern about the results in Table 2.5 is that they reflect shifts in size bins rather than overall increases or decreases in entry or exit. As a quick check, we re-estimated the basic model in equation (2.2) on the subsample of newly established firms in each year (to rule out reallocations across size classes) with dependent variables set alternately to the number of entrants, average size classification and average number of employees. For the model with all controls, including county and year fixed effects, Texas counties after the amendment on average had 18 additional newly established firms, while average employment among newly established firms declined by 1.6 employees, both estimates statistically significant. This suggests that our findings are not due solely to reallocations of business across size classes.

Figures 2.3 and 2.4 compare trends before and after the amendment by replacing the pre-, post-amendment dummy and its interactions (e.g., post$\times$Texas and post$\times$small and post$\times$Texas$\times$small) with a set of T-1 year dummies and their respective interactions, excluding 1997 as the reference year. The figures show

[15]Note that we are not splitting the sample in the usual sense of relaxing a restriction that equates the effects of large and small firms. We recompute the dependent variables for the different size categories, thereby increasing the sample size.

[16]Census disclosure-avoidance rules prevent us from releasing summary statistics by our size cuts.

the results for the model without state-specific time trends. In the post period, the estimates can be interpreted as year-specific treatment effects; but now relative to 1997 outcomes instead of the average across the pre-period. The similarity of the pre-treatment trends supports our use of the Texas Amendment as a natural experiment, as well as the results in Kerr et al. (2019).

2.3.3 Were supply or demand factors at play?

We now consider whether the reduced-form effects identified in the previous section are primarily due to supply or demand factors. On the supply side, the relaxation of borrowing constraints improves access to credit markets for businesses generally unable to borrow in external capital markets. This improvement reduces the cost of finance and production, thereby enhancing entry and job creation and limiting exit and job destruction. On the demand side, the Texas amendment relaxed constraints not only for businesses, but for consumers as well. Increased overall demand by consumers for small business products and service might induce those businesses to expand or proliferate. Both of these channels could operate in the data.

First, note that the differential effects reported above for small and large firms are mostly consistent with the supply-side channel. If demand-side effects dominate, then increased spending should be proportionately disbursed across the output of both small and large firms, and we should see no differential effect. If the supply channel dominates, small firms are more likely to be sensitive to the increase in credit availability than larger firms with lower-cost finance, so we should expect to see the differential that we report.

However, because some evidence in the literature suggests that small firms appear to be more responsive to many types of shocks (Fort et al., 2013; Adelino et al., 2017), inference drawn from our firm size results should be supplemented with additional evidence. We look along two other dimensions to gauge the intensity of the treatment effect in order to tease out which channel is dominant. The first distinguishes between businesses that produce goods in 'tradable' versus 'non-tradable' industries. If relaxing home equity loan restrictions primarily affects overall demand, then activity of businesses selling goods in local markets that are not easily traded in other localities – like restaurants – will increase more than for goods tradable outside of Texas – like manufactured goods or software publishers. Supply-side effects dominate if treatment effects do not differ across these groups. The second dimension looks at locational variation in property values. Businesses in areas with high housing prices, all else the same, have greater scope for using home equity to finance their ventures than those in low-value areas. While high-value areas will also exhibit stronger demand-side effects, much of the higher spending is likely to be disbursed out of the area. Thus, businesses in high-value areas will be more sensitive to the treatment effect than those in low-value areas if supply factors are prominent.

As in the previous sub-section, we estimate triple difference-in-differences models of the form of equation (2.3), where s_j is now an indicator binary variable for either tradable or non-tradable sector, or alternately for high or low value property. In the first regression, s_j is set to one when the firms used to compute entry, exit, job creation, and job destruction rates are in industries that produce goods that are easily tradable across state borders.[17] For supply factors to matter more, δ_2 (the differential treatment effect for firms selling tradeables) should be statistically insignificant; for demand factors to matter more, δ_2 should be negative for entry and job creation rates, and positive for exit and job destruction rates. In the second regression, s_j is set to one for firms in high property value areas.[18] If supply factors are dominant, δ_2 (the differential treatment effect for high-property-value areas) should be positive for entry and creation and negative for exit and destruction, while for demand it should be statistically zero.

From Table 2.6 we see that the results from this exercise are mostly consistent with the size regressions, and support the conclusion that supply-side effects dominate. In the top panel, only the δ_2 estimate for entry rates is consistent with the demand channel and statistically significant – it reveals that the treatment effect is smaller for firms selling tradable goods, and is indeed less than half the estimate for nontradeables. Nonetheless, this finding is not robust across control groups. The middle panel of the table shows the results for our third non-Texas control group, whereas the top panel uses the baseline second control group. The middle panel reveals that the estimate of δ_2 using this alternative control group has a large standard error and is not statistically significant.

The third panel reports the results for which s_j is the binary variable measuring high versus low property value areas. This panel supports the dominance of supply-side factors since the estimates of δ_2 are consistent with greater small business expansionary effects and smaller contractionary effects for high-value areas. Note that the effects for job destruction and excess reallocation rate are particularly large. We can also tie the greater small business effects more directly to the use of home equity loans by small firms in high-value areas: in Table 2.7 the 2007 SBO data show a positive and statistically significant correlation between the use of home equity loans for business (as measured by a zero-one dummy variable) and county-level median home prices and with a county-level housing price index. This finding lends support to our claim that supply-side factors dominate.

These results also allow us to more directly compare our results to those of Kerr et al. (2019). In the section of their paper that deals with the Texas amendment, they estimate a variant of our triple diff-in-diff model in equation (2.3), but find only small effects of house prices on the treatment effect as implied by estimates of their model's parameter analogous to our δ_2. We can only speculate as to the reason for the different results, but it may be due to different control groups and their use of house price indices rather than our binary high-low price variable. Whereas the dummy variable approach ignores some variation in house prices, it better allows for potential non-linear effects. They also work with core-based statistical areas instead of counties, which limits their samples to urban and sub-urban areas which, all else the same, lowers the precision of their estimates.

[17] We use NAICS codes to classify each firm as belong either to tradable, non-tradeable, or other sector following (Mian and Sufi 2012, Appendix Table 1): "we define a 4-digit NAICS industry as tradable if it has imports plus exports equal to at least \$10,000 per worker, or if total exports plus imports for the NAICS 4-digit industry exceeds \$500M. Non-tradable industries are defined as the retail sector and restaurants."

[18] The property value classification is based on the median of county home prices for our Texas counties in 1997. This is approximately \$115,000 in constant 2019 dollars. If a county's median home prices are above \$75,000 in 1997 (whether in Texas or not) then we consider it a high property value county throughout the analysis. In Texas, this procedures splits Texas counties evenly between low and high classification, whereas almost 28% of non-Texas counties are are classified as low.

2.4 Conclusion

We find that the Texas amendment had statistically significant, yet overall modest, effects on small business activity, including job creation and destruction. However, the effect is larger for small firms, those with fewer than ten employees. Our results support a growing body of evidence that the activity of small and young businesses is hampered by limitations on the availability of credit, particularly housing collateral. We complement much of the recent research on this topic, which has often used changes in house prices or housing supply elasticities to act as instruments for collateral availability. Changes in the value of house prices are generally endogenous to business conditions facing entrepreneurs. The Texas amendment, by contrast, resulted in an exogenous shift in the availability of housing collateral and credit for entrepreneurs.

By adding to the ensemble of evidence that collateral constraints matter for small business activity, we contribute to external validity. While we cannot extrapolate our results beyond the specific case we study, our results support a class of models that suggests restrictions on collateral inhibit entrepreneurs from growing their businesses. In adding to this evidence, it is particularly meaningful that we are able to distinguish between supply and demand channels. If the Texas amendment only affected small businesses indirectly, through increased consumption demand, we could not say much about the direct importance of credit availability for small business outcomes.

Availability of liquidity has been a key policy concern during the COVID-19 pandemic of 2020 and 2021. The literature our results support suggests that credit constraints may matter a lot for the viability of small and young businesses. We find that increased availability of credit to small businesses increases their likelihood of survival. It is less clear that expanded credit resulted in fewer job losses for the Texas businesses we study. Perhaps most importantly, we find that expanded credit increased business dynamism, as measured through the excess reallocation rate. This finding suggests that some small businesses that are otherwise productive are prevented from expanding owing to credit constraints. If so, many of the small businesses getting hammered by the pandemic and lacking access to credit could otherwise be productive, implying that the destruction of business wrought by the pandemic is not creative destruction. This line of thinking supports policies that subsidize credit to small businesses, subject to the means of financing such subsidies, to help them stay afloat until business conditions return to normal.

2.5 References

Abdallah, C., and Lastrapes, W., (2012). Home Equity Lending and Retail Spending: Evidence from a Natural Experiment in Texas, American Economic Journal: Macroeconomics, 4, 94–125.

Adelino, M., et al., (2017). Firm Age, Investment Opportunities, and Job Creation, Journal of Finance.

Cameron, A., and Miller, D., (2015). A Practitioner's Guide to Cluster-Robust Inference, Journal of Human Resources, 50, 317–72.

Census Bureau, U.C. (1992). The 1992 Economic Census: Characteristics of Businesses Owners, Company Statistics Series CBO92-1, U.S. Census Bureau. ——— (2006). Characteristics of Businesses: 2002, Company Statistics Series SB02-00CS-CB, U.S. Census Bureau.

Davis, S., et al., (1996). Job Creation and Destruction, MIT Press.

Evans, D., and Jovanovic, B., (1989). An Estimated Model of Entrepreneurial Choice under Liquidity Constraints, Journal of Political Economy, 97, 808–27.

Fairlie, R., and Robb, A., (2008). Race and entrepreneurial success: Black-, Asian-, and white-owned businesses in the United States, MIT Press.

Fairlie, Robert (2013). Entrepreneurship, Economic Conditions and the Great Recession, Journal of Economics and Management Strategy.

Ferman, B., and Pinto, C., (2019). Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity, Review of Economics and Statistics, 101, 452–467.

Fort, T., et al., (2013). How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size, IMF Economic Review, 61, 520–559.

Fort, T., and Klimek, S., (2018). The Effects of Industry Classification Changes on US Employment Composition, Working Paper 18-28.

Hurst, E., and Lusardi, A., (2004). Liquidity Constraints, Household Wealth, and Entrepreneurship, Journal of Political Economy, 112, 319–347.

Jarmin, R., and Krizan, C., (2010). Past Experience and Future Success: New Evidence on Owner Characteristics and Firm Performance, Working Paper 10-24, Center for Economic Studies, U.S. Census Bureau.

Jarmin, R., and Miranda, J., (2002). The Longitudinal Business Database, Working Paper 02-17, Center for Economic Studies, U.S. Census Bureau.

Jensen, T., et al., (2014). Housing Collateral, Credit Constraints and Entrepreneurship - Evidence from a Mortgage Reform, NBER Working Papers 20583, National Bureau of Economic Research, Inc.

Johansson, Edvard (2000). Self-Employment and Liquidity Constraints: Evidence from Finland, Scandinavian Journal of Economics, 102, 123–34.

Kelly, Morgan (2020). Understanding Persistence, CEPR Discussion Paper No. DP15246.

Kerr, S., et al., (2019). House Prices, Home Equity and Entrepreneurship: Evidence from U.S. Census Micro Data, Harvard Business School Working Papers 15-069, Harvard Business School.

Kumar, Anil (2018). Do Restrictions on Home Equity Extraction Contribute to Lower Mortgage Defaults? Evidence from a Policy Discontinuity at the Texas Border, American Economic Journal: Economic Policy, 10, 268–297.

Kumar, A. and Liang, C., (2019). Credit constraints and GDP growth: Evidence from a natural experiment, Economics Letters, 181, 190 – 194.

Lelarge, C., et al., (2010). Entrepreneurship and Credit Constraints: Evidence from a French Loan Guarantee Program, in International Differences in Entrepreneurship, National Bureau of Economic Research, Inc, NBER Chapters, 243–273.

Mian, A., and Sufi, A., (2012). What explains high unemployment? The aggregate demand channel, NBER Working Paper 17830, National Bureau of Economic Research, Inc.

Nykvist, Jenny (2008). Entrepreneurship and Liquidity Constraints: Evidence from Sweden, Scandinavian Journal of Economics, 110, 23–43.

Robb, A., and Robinson, D., (2014). The Capital Structure Decisions of New Firms, Review of Financial Studies, 27, 153–179.

Schmalz, M., et al., (2017). Housing Collateral and Entrepreneurship, Journal of Finance, 72, 99–132.

Stolper, Harold (2015). Home Equity Credit and College Access: Evidence from Texas Home Lending Laws, Columbia University working paper.

Zevelev, Albert (2020). Does Collateral Value Affect Asset Prices? Evidence from a Natural Experiment in Texas, Available from SSRN 2815609.

2.6 Tables and Figures

Figure 2.1: Page 7 of the 2007 Survey of Business Owners

Business

65 In what year was this business originally established?

- ☐ Before 1980
- ☐ 1980 – 1989
- ☐ 1990 – 1999
- ☐ 2000 – 2002
- ☐ 2003
- ☐ 2004
- ☐ 2005
- ☐ 2006
- ☐ 2007
- ☐ Don't know

66 **A.** For the owner(s) as of December 31, 2007, what was the source(s) of capital used to start or acquire this business?
Mark X all that apply.

- ☐ Personal/family savings of owner(s)
- ☐ Personal/family assets other than savings of owner(s)
- ☐ Personal/family home equity loan
- ☐ Personal/business credit card(s)
- ☐ Business loan from federal, state, or local government
- ☐ Government-guaranteed business loan from a bank or financial institution
- ☐ Business loan from a bank or financial institution
- ☐ Business loan/investment from family/friend(s)
- ☐ Investment by venture capitalist(s) *(An early-stage investment in exchange for ownership equity by an individual, outside group, or business not directly involved in the overall operation and management of the business.)*
- ☐ Grants
- ☐ Other source(s) of capital
- ☐ Don't know
- ☐ None needed – *Go to* **67**

B. For the owner(s) as of December 31, 2007, what was the total amount of capital used to start or acquire this business? *(Capital includes savings, other assets, and borrowed funds of owner(s)).*

- ☐ Less than $5,000
- ☐ $5,000 – $9,999
- ☐ $10,000 – $24,999
- ☐ $25,000 – $49,999
- ☐ $50,000 – $99,999
- ☐ $100,000 – $249,999
- ☐ $250,000 – $999,999
- ☐ $1,000,000 or more
- ☐ Don't know

67 In 2007, did this business operate primarily from somebody's home?

- ☐ Yes
- ☐ No

68 In 2007, did this business operate as a franchise?

- ☐ Yes
- ☐ No

69 In 2007, did a franchiser own more than 50% of this business?

- ☐ Yes
- ☐ No

70 In 2007, were any of the following sources used to finance expansion or capital improvement(s) for this business?
Mark X all that apply.

- ☐ Personal/family savings of owner(s)
- ☐ Personal/family assets other than savings of owner(s)
- ☐ Personal/family home equity loan
- ☐ Personal/business credit card(s)
- ☐ Business loan from federal, state, or local government
- ☐ Government-guaranteed business loan from a bank or financial institution
- ☐ Business loan from a bank or financial institution
- ☐ Business loan/investment from family/friend(s)
- ☐ Investment by venture capitalist(s) *(An early-stage investment in exchange for ownership equity by an individual, outside group, or business not directly involved in the overall operation and management of the business.)*
- ☐ Business profits and/or assets
- ☐ Grants
- ☐ Other source(s) of capital
- ☐ Don't know
- ☐ Did not have access to capital
- ☐ Did not expand or make capital improvement(s)

71 In 2007, which of the following types of customers accounted for 10% or more of this business's total sales of goods and/or services?
Mark X all that apply.

- ☐ Federal government
- ☐ State and local government, including school districts, transportation authorities, etc.
- ☐ Other businesses and/or organizations, including distributors of your product(s)
- ☐ Individuals

72 In 2007, what percent of this business's total sales of goods and/or services consisted of exports outside the United States?

- ☐ None
- ☐ Less than 1%
- ☐ 1% - 4%
- ☐ 5% - 9%
- ☐ 10% - 19%
- ☐ 20% - 49%
- ☐ 50% - 99%
- ☐ 100%
- ☐ Don't know

73 In 2007, did this business establish operations outside the United States?

- ☐ Yes
- ☐ No

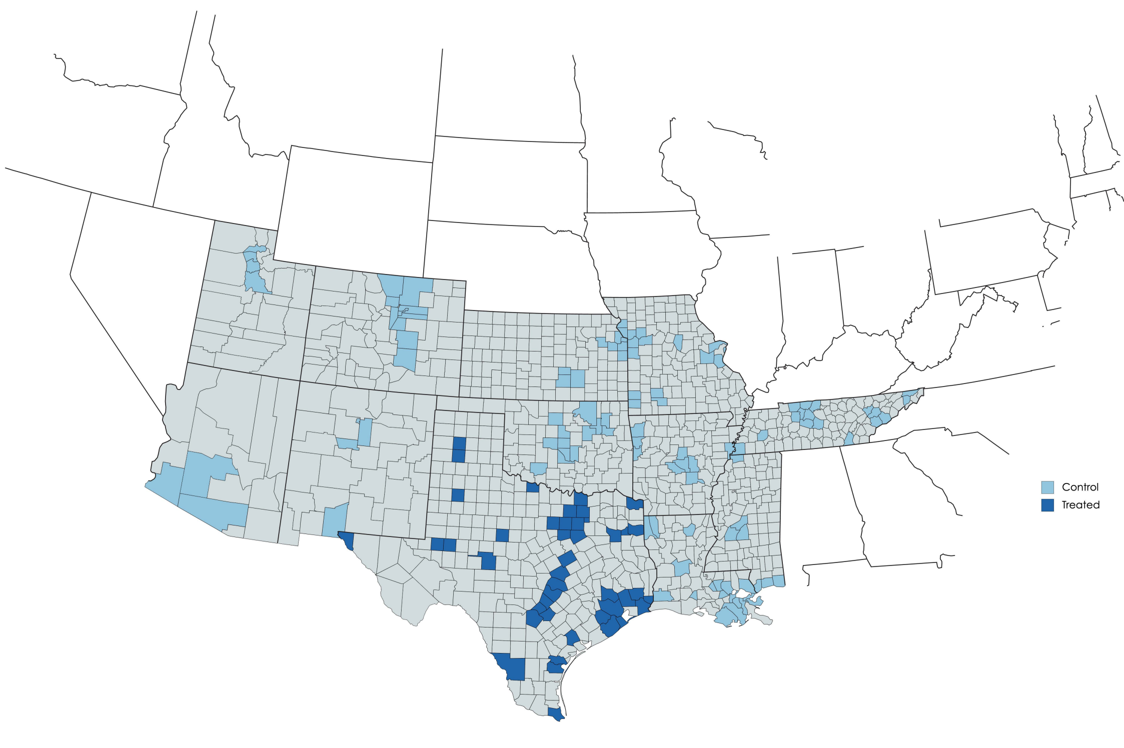
74 In 2007, did this business outsource or transfer any business function and/or service to a company outside the United States?

- ☐ Yes
- ☐ No

Please turn to the next page to continue.

FORM SBO-1 (01/02/2008)

Figure 2.2: Treatment and Control Counties



Note: The non-Texas counties are from the second control group: border states of Texas plus border states of a border state

Figure 2.3: Entry and Exit Rates by Firm Size

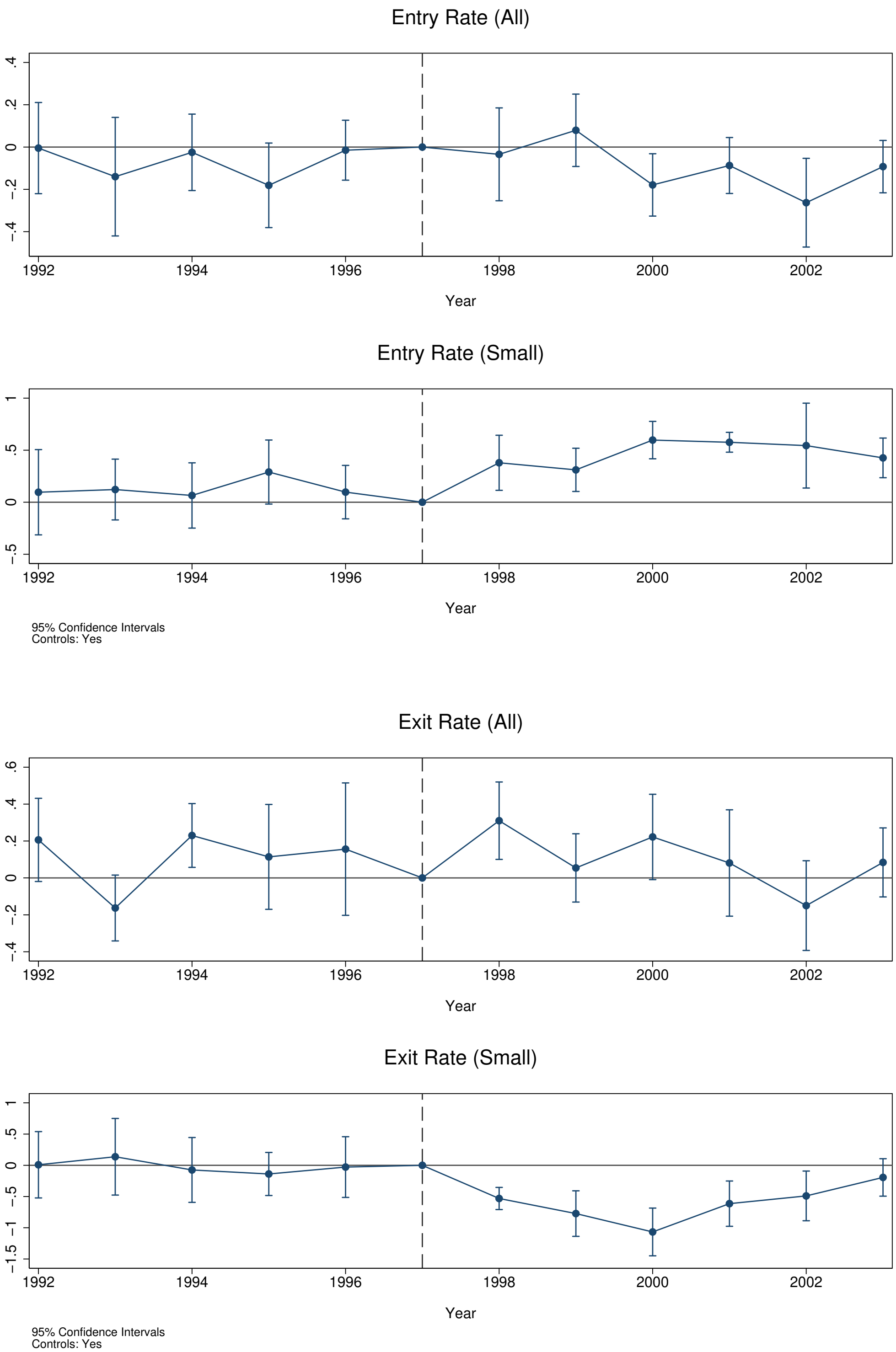


Figure 2.4: Job Destruction and Creation Rates by Firm Size

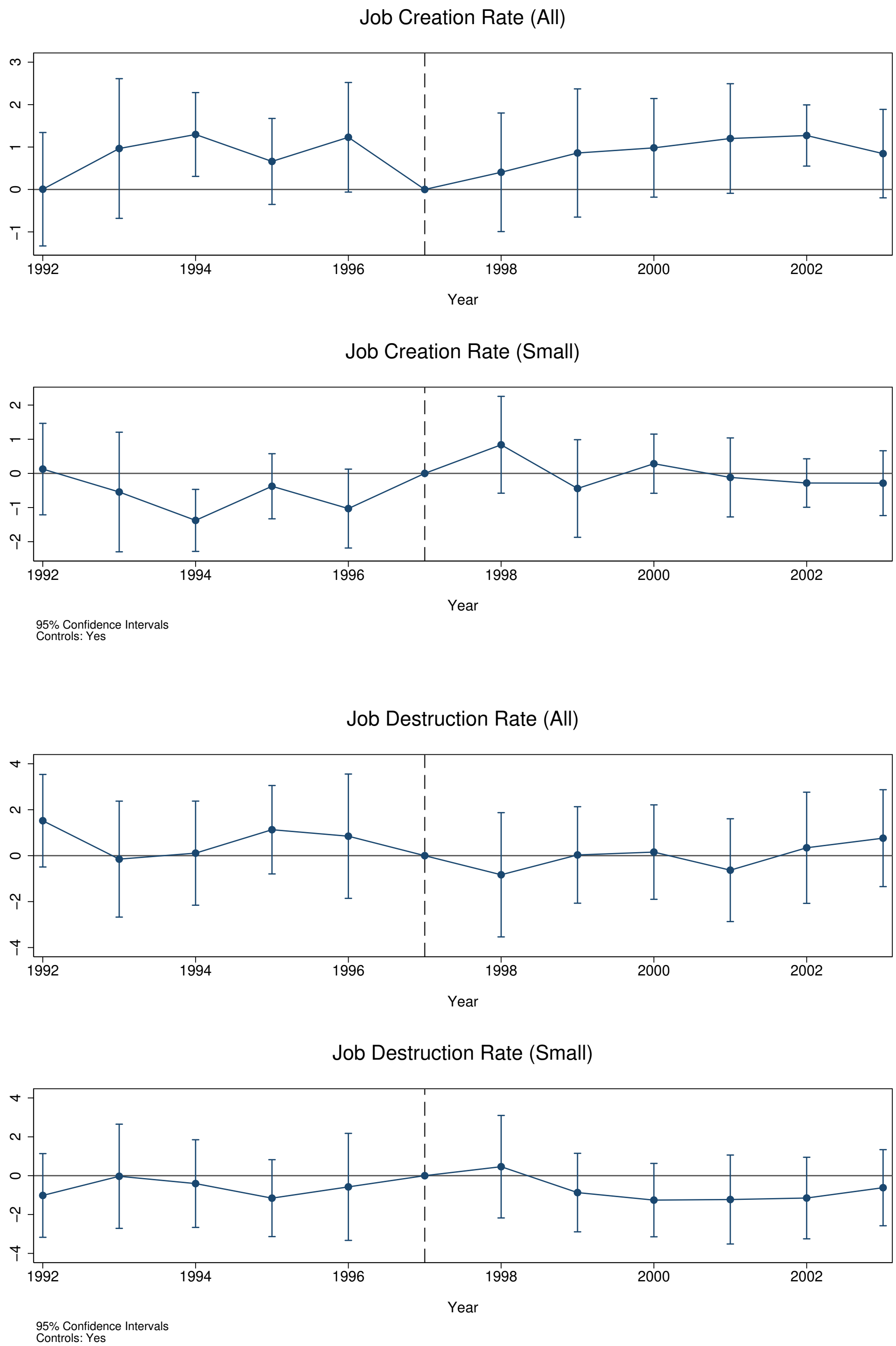


Table 2.1: Sample Statistics: LBD Constructed Variables, 1992-1997

	Non-Texas		Texas	
Variable	mean	st. dev.	mean	st. dev.
Entry rate	4.877	1.075	5.182	1.067
Exit rate	9.851	1.253	10.34	1.126
Job creation rate	14.56	5.951	13.99	4.678
Job destruction rate	13.25	6.619	12.65	5.471
Job reallocation rate	23.68	5.974	23.35	4.433

Note: There are 672 county-year observations for the non-Texas sample and 282 for the Texas sample. Because of confidentiality restrictions, we report sample statistics only for the limited time sample. Non-Texas states are those in our baseline control group of border and border-to-border states.

Table 2.2: Sample Statistics from the Survey of Business Owners

	1992			2007		
Variable	All	Texas	Non-Texas	All	Texas	Non-Texas
Home Loan	5.04	—	—	6.77	—	—
Other Loan	43.71	—	—	31.32	—	—
No Loan	51.25	42.29	52.05	61.91	61.52	61.91
Age $<$ 35	10.91	6.27	11.32	1.82	1.79	1.82
35 $<$ Age $<$ 54	60.93	63.04	60.74	48.64	46.58	48.78
Age $>$ 54	28.16	30.70	27.93	49.55	51.64	49.40
Asian	4.69	4.40	4.72	5.23	5.58	5.22
Black	2.19	2.39	2.17	1.34	1.42	1.33
Hispanic	3.43	12.72	2.61	0.20	0.23	0.20
White	88.75	78.13	89.70	92.30	91.67	92.33
Male	79.97	75.15	80.40	79.33	77.15	79.46
Some HS	9.34	14.79	8.86	3.55	3.68	3.55
HS Grad	28.48	22.89	28.98	18.79	14.92	19.05
Some Col	25.18	22.50	25.42	25.88	23.86	26.02
Col Grad	13.24	13.99	13.18	25.54	28.04	25.35
Post Col	23.75	25.83	23.57	26.24	29.49	26.04
Franchise	3.84	2.49	3.96	3.06	4.42	2.97
Exporter	0.61	0.42	0.63	2.73	3.01	2.71
N.E. History	58.66	60.33	58.51	5.11	5.88	5.07
Midwest	16.65	0.00	18.13	21.89	0.00	23.24
Northeast	25.34	0.00	27.59	24.21	0.00	25.69
South	32.41	100	26.41	32.87	100	28.77
West	25.59	0.00	27.87	21.03	0.00	22.31
AFFM	3.75	6.15	3.53	3.33	4.22	3.28
Construction	14.06	12.16	14.23	13.68	11.04	13.84
Manufacturing	3.87	1.38	4.10	6.07	5.33	6.11
TCEGS	3.01	3.24	2.99	3.37	3.39	3.37
Wholesale Trade	3.75	6.50	3.51	7.18	7.80	7.14
Retail Trade	23.60	25.38	23.44	15.99	13.49	16.14
FIRE	5.52	7.07	5.38	8.30	8.88	8.26
Services	42.44	38.11	42.82	42.08	45.85	41.85
Tradable	3.76	4.29	3.71	6.43	6.69	6.42
Non-tradable	26.08	29.83	25.74	20.84	18.44	20.99
Ambiguous	70.16	65.88	70.54	72.72	74.87	72.59
N (Weighted Count)	1,172,000	95,500	1,076,000	1,130,000	65,000	1,066,000

Note: A dash indicates suppression due to disclosure-avoidance rules set by the US Census Bureau. N is rounded for confidentiality. Weighted statistics for single-unit, employer firms in the 1992 Characteristics of Business Owners Survey and 2007 Survey of Business Owners. The data are linked to the LBD to obtain information on industry SIC/NAICS codes. We also check for potential links with the ILBD to ascertain whether a business had a history of being a non-employer.

Table 2.3: Effects of Texas Amendment on Loan Use

Home-equity loans	Model 1	Model 2	Model 3
δ (Texas*post)	0.0627***	0.0626***	0.0681***
	(0.0145)	(0.0136)	(0.0174)
α_2 (Texas)	-0.0514***	-0.0273***	-0.0298**
	(0.0067)	(0.0081)	(0.0110)
α_1 (Post)	0.0162*	0.0188**	0.0167**
	(0.0065)	(0.0060)	(0.0064)
Other loans	Model 1	Model 2	Model 3
δ (Texas*post)	-0.0603*	-0.0723**	-0.0953***
	(0.0243)	(0.0233)	(0.0228)
α_2 (Texas)	0.149*	0.0684	0.0746
	(0.0610)	(0.0793)	(0.0820)
α_1 (Post)	-0.115***	-0.108***	-0.0961***
	(0.0201)	(0.0168)	(0.0203)
State FE	No	Yes	Yes
Controls	No	No	Yes
N (Weighted Count)	2,302,000	2,302,000	2,302,000

Note: Cluster-robust standard errors at the state level in parentheses. Weighted regression results for single-unit firms. Controls include owner demographics (age, sex, race, education) and firm characteristics (single owned, franchise, exporter, non-employment history, industry sector).
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 2.4: Effect of Texas Amendment on Loan Use – FP Robust Check

Home equity loans	Model 1	Model 2	Model 3
P-value w/o adjustment	0.03658	0.03717	0.02538
P-value w/ FP adjustment	0.04545	0.02293	0.00274
N (Simulations)	100000	100000	100000
Other loans	Model 1	Model 2	Model 3
P-value w/o adjustment	0.1015	0.0385	0.0181
P-value w/ FP adjustment	0.0013	0	0
N (Simulations)	100000	100000	100000

Note: Reported p-values are from cluster residual bootstrapping simulations with and without a heteroskedasticity correction applied to the residuals outlined by Ferman and Pinto (2019).

Table 2.5: Effects of Texas amendment on Business Outcomes

	ER	XR	JCR	JDR	ERR	ER	XR	JCR	JDR	ERR
Panel A: All firms (eq. 2.2); $N = 1,908$ county-year observations										
δ (Texas*Post)	0.205***	-0.107	0.349	-0.563**	-0.108	0.179***	-0.176**	0.15	-0.502	-0.248
	(0.0791)	(0.0874)	(0.3310)	(0.2460)	(0.2230)	(0.0656)	(0.0862)	(0.2770)	(0.3490)	(0.2090)
Panel B: All firms (eq. 2.2) including state-time trends; $N = 1,908$ county-year observations										
δ (Texas*Post)	0.275***	-0.0778	-0.121	-0.786	-0.692***	0.260***	-0.0827	-0.176	-0.812	-0.784***
	(0.0429)	(0.1280)	(0.6600)	(0.7100)	(0.1920)	(0.0393)	(0.1290)	(0.6590)	(0.7690)	(0.1910)
Panel C: By size (eq. 2.3); $N = 3,816$ county-year observations										
δ_1 (Texas*Post)	-0.0025	0.0741	0.312	-0.595	-0.377*	-0.030	0.0151	0.235	-0.624*	-0.462***
	(0.0286)	(0.0559)	(0.2810)	(0.3100)	(0.1550)	(0.0395)	(0.0896)	(0.2810)	(0.3580)	(0.1600)
δ_2 (Texas*post*small)	0.359***	-0.597***	0.527**	-0.256	1.994***	0.360***	-0.597***	0.532**	-0.249	1.992***
	(0.0896)	(0.1420)	(0.2500)	(0.4270)	(0.2720)	(0.0900)	(0.1430)	(0.2500)	(0.4260)	(0.2720)
Panel D: By size (eq. 2.3) including state-time trends; $N = 3,816$ county-year observations										
δ_1 (Texas*Post)	0.035	0.0838	0.245	-0.826*	-0.689***	0.0186	0.0819	-0.195	-0.839	-0.761***
	(0.0890)	(0.1690)	(0.4270)	(0.5690)	(0.2250)	(0.0808)	(0.1730)	(0.4170)	(0.5990)	(0.2990)
δ_2 (Texas*post*small)	0.359***	-0.597***	0.527**	-0.256	1.994***	0.359***	-0.598***	0.531**	-0.249	1.991***
	(0.0898)	(0.1420)	(0.2500)	(0.4270)	(0.2720)	(0.0901)	(0.1430)	(0.2500)	(0.4270)	(0.2610)
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

Note: ER = entry rate, XR = exit rate, JCR = job creation rate, JDR = job destruction rate, ERR = excess reallocation rate. All regressions include county and year fixed effects. Cluster-robust standard errors at the state level in parentheses. Regression results for single-unit firms. Controls include county-level demographics and labor market conditions.
$* p < 0.10$, $** p < 0.05$, $*** p < 0.01$

Table 2.6: Effects of Texas Amendment on Business Outcomes: Supply or Demand Channels

Tradable-1	ER	XR	JCR	JDR	ERR	ER	XR	JCR	JDR	ERR
δ_1	0.319***	0.183	0.573**	0.273	1.269***	0.332***	0.089	0.713***	0.249	1.490***
	(0.0772)	(0.1870)	(0.1930)	(0.2050)	(0.2880)	(0.0678)	(0.2160)	(0.1050)	(0.2960)	(0.2840)
δ_2	-0.150*	0.444	0.256	0.230	-0.719	-0.155*	0.455	0.260	0.233	-0.708
	(0.0786)	(0.3350)	(0.4620)	(0.4880)	(0.6010)	(0.0773)	(0.3330)	(0.4620)	(0.4890)	(0.5980)
N (County-year)	3816	3816	3816	3816	3816	3816	3816	3816	3816	3816
Tradable-2	ER	XR	JCR	JDR	ERR	ER	XR	JCR	JDR	ERR
δ_1	0.362**	0.238	0.48	-0.12	1.157*	0.288*	0.181	0.459	-0.024	1.245*
	(0.0896)	(0.2040)	(0.2080)	(0.2640)	(0.3740)	(0.1030)	(0.2980)	(0.2370)	(0.3590)	(0.4770)
δ_2	-0.13	-0.851	-0.0135	-0.471	-1.296	-0.139	-0.833	-0.0083	-0.464	-1.276
	(0.1450)	(0.5360)	(0.3460)	(0.5500)	(0.8340)	(0.1410)	(0.5360)	(0.3470)	(0.5520)	(0.8300)
N (County-year)	2640	2640	2640	2640	2640	2640	2640	2640	2640	2640
Hi-value	ER	XR	JCR	JDR	ERR	ER	XR	JCR	JDR	ERR
δ_1	-0.0849	-0.0656	-0.00577	-0.181	0.0358	-0.101	-0.117*	-0.22	-0.00815	-0.0786
	(0.1160)	(0.0764)	(0.3980)	(0.6350)	(0.5780)	(0.0953)	(0.0696)	(0.4060)	(0.7210)	(0.5140)
δ_2	0.177	-0.150	0.798	-2.543**	-2.330**	0.136*	-0.21**	0.733	-2.757**	-2.414**
	(0.0836)	(0.0892)	(0.4150)	(0.6660)	(0.6270)	(0.0723)	(0.0997)	(0.4890)	(0.7760)	(0.6070)
N (County-year)	1908	1908	1908	1908	1908	1908	1908	1908	1908	1908
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

Note: ER = entry rate, XR = exit rate, JCR = job creation rate, JDR = job destruction rate, ERR = excess reallocation rate. All regressions include county and year fixed effects. Cluster-robust standard errors at the state level in parentheses. Regression results for single-unit firms. Controls include county-level demographics and labor market conditions.

$* \, p < 0.10$, $** \, p < 0.05$, $*** \, p < 0.01$

Table 2.7: Correlation between Home Loan Use and Housing Prices

Home-equity loans	Model 1	Model 2	Model 3	Model 4
ln(hvalue)	0.0156	0.2920**	–	–
	(0.0163)	(0.1420)		
ln(hpi)	–	–	0.0676**	0.3090**
			(0.0331)	(0.1510)
State FE	No	Yes	No	Yes
Controls	No	Yes	No	Yes
N (Weighted Count)	65,000	65,000	65,000	65,000

Note: Cluster-robust standard errors at the state level in parentheses. Weighted regression results for single-unit firms established within the past year at the time of the 2007 Survey of Business Owners. "hvalue" is county-level median home prices for the past year, while "hpi" is county-level housing price index for the past year. Controls include owner demographics (age, sex, race, education) and firm characteristics (single owned and industry group). N is rounded for confidentiality.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Chapter 3

No Pain, No Gain: The Labor Market Return to Physical Activity

3.1 Introduction

The health benefits associated with leisure-based physical activity are well-known and widely accepted among both policymakers and the public, yet a global trend of insufficient physical activity has persisted over the past two decades.[1] In the United States, almost 50% of the adult population does not meet the recommended Physical Activity Guidelines for aerobic physical activity, while 80% of the adult population does not meet the Physical Activity Guidelines for both aerobic and muscle-strengthening activity (CDC, 2018). Persistence of physical inactivity this high is surprising given the extent to which local and state governments subsidize public parks and recreation facilities in the United States (Walls, 2009). Moreover, there are a number of community outreach programs designed to increase physical activity. For example, in 2007, the American College of Sports Medicine, with endorsement from the the U.S. Office of the Surgeon General, launched an initiative to mobilize physicians and healthcare workers to promote physical activity in clinical care and educate patients on the health benefits of physical activities. However, neither policymakers or outreach programs convey to the public the role that physical activity plays in improving labor market outcomes despite a growing body of evidence.

In this paper, I use data from the American Time Use Survey (ATUS) and an instrumental variable approach to estimate the short and long-run effects of leisure-based physical activity on earnings and wages. More specifically, I use daily sunset time and average sunset time as plausible sources of exogenous variation in time spent on physical activities. By using sunset time as an instrument for physical activity, I make a methodological contribution to the empirical literature on the labor market effects of physical activity since most of these studies rely on using a fixed-effects or matching estimator to address

[1] US Department of Health and Human Services (2018) provides an extensive literature review on the health effects of physical activities. Guthold et al. (2018) provides a meta-analysis on global trends in physical activities.

endogeneity (Lechner, 2009; Rooth, 2011; Hyytinen and Lahtonen, 2013; Lechner and Sari, 2015; Lechner and Downward, 2017). I also complement this body of research by providing further evidence on effect heterogeneity across physical activities and the potential mechanisms through which physical activities affect earnings. Lastly, my paper contributes to the growing body of empirical research on the impact of non-labor time uses on labor market outcomes.[2]

The empirical literature on the labor market returns to physical activity is relatively small but growing. One strand of this literature estimates the effect of participating in sports or athletic programs on future labor market outcomes. Barron et al. (2000) is the first study to take a causal approach in estimating the labor market returns to sports participation. Using data from the National Longitudinal Study of Youth 1979, they find that participation in a high school athletic program increases future wages by 4.2–14.8% for men, roughly 12 years after graduation. Ewing (2007) extends Barron et al. (2000) to include fringe benefits, finding that participation in a high school athletic program also increases fringe benefits by 6% for men. Both Lechner (2009) and Lechner and Downward (2017) find that regular engagement in leisure-based sports increases future earnings. Regarding other labor market outcomes, Stevenson (2010) finds that sports participation during high school slightly increases labor force participation among females. Finally, using a field experiment, Rooth (2011) finds that job applicants who signal sports skills on a resume experience a higher callback rate of about 2 percentage points.

Meanwhile, a second strand of this literature estimates the effect of engaging in any form of physical activity, not just sports and athletic programs, on future earnings. Using administrative data and within variation among brothers, Rooth (2011) finds that physically fit males earn 2–5% more over a 10-to-20-year period. In a similar vein, using administrative data on male twins, Hyytinen and Lahtonen (2013) finds that being physically active increases earnings by 14–17% over a 15-year period. Lastly, Lechner and Sari (2015) finds that engaging in vigorous physical activities increases earnings by 10–20% over an 8-to-12-year period.

The positive labor market effects observed in the empirical literature could result through one or more of the following causal channels. First, the most obvious mechanism is that time spent on physical activity is an investment in health capital. In the Grossman (1972) model, time and goods are invested, via health production functions, to influence next period's level of health capital, which affects how much healthy and unhealthy time individuals experience in the subsequent period. In this context, physical activities could increase labor market attachment by lowering absenteeism due to illness. Furthermore, health capital may directly affect an individual's productivity in the labor market. For example, obesity increases the risk of diabetes, high blood pressure, asthma, and other chronic diseases, which reduce labor productivity (Ruhm, 2007; Liu and Zhu, 2014). A related mechanism through which physical activity can positively affect labor market outcomes is known as the "beauty" effect (Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998). Individuals who are more physically fit may be perceived as beautiful by others, and in some occupations attractive workers can be more productive. For example, restaurant customers may prefer to be served by more attractive servers.

[2]See for example, Hersch and Stratton (2002) and Hersch (2009) on the labor market effects of home production; Gibson and Shrader (2018) and Costa-Font and Fleche (2020) on the effects of sleep; Lu et al. (2020) on the effects of commuting time; Van Houtven et al. (2013) and Schmitz and Westphal (2017) on the effects of informal care.

A third potential mechanism is that engaging in physical activities, particularly team sports, can be an investment in social capital (Seippel, 2006; Lechner and Downward, 2017). By participating in physical activities with others, individuals can develop social and teamwork skills that improve labor market productivity (Aguilera and Bernabe, 2005). Furthermore, engaging in a shared activity provides a way for individuals to enhance their social networks. A recent body of research documents the important role local networks and social interactions play in searching for higher-paying jobs (Bayer et al., 2008; Schmutte, 2015). Lastly, physical activities can serve as a signal to potential employers that an individual is in good health and highly motivated, and will therefore be a productive worker (Rooth, 2011).

Although each of these channels provide an intuitive causal mechanism through which physical activity can affect labor market outcomes, it is equally possible that the positive relationship between earnings and physical activities observed in the literature is attributable to non-causal mechanisms. The fundamental challenge in estimating the labor market returns to physical activity is that individuals endogenously select their level of physical activity. Consequently, estimates of the labor market returns will be biased by differences in unobserved individual characteristics that influence both earnings and the physical activity decision. For example, an individual's underlying health status may directly affect both physical activity levels and earnings (Cutler and Glaeser, 2005). The existing empirical literature relies heavily on fixed-effects or matching estimators to address issues with endogeneity.[3] However, in order to recover causal estimates, matching methods require a conditional independence assumption, which is unlikely to hold in this context since individuals self select into physical activities based on a number of unobservables, e.g., self-discipline, motivation, genetic traits. Meanwhile, a fixed-effects estimator cannot address time-invariant unobserved heterogeneity that is correlated with both physical activity and earnings. For example, physically active people are more conscious about eating a healthy diet and higher income people are able to afford healthier foods (Rao et al., 2013). It is also empirically difficult to control for diet, so it is always an unobservable confounding variable in the physical activity literature. If an individual's diet is time-varying, which is most likely the case, then a fixed-effects estimator will not recover causal estimates.

To account for the endogeneity of physical activity, I use an instrumental variable approach that exploits variation in sunset time. I find that in the short run, physical activity does not enhance labor productivity; instead, it appears work time and physical activity are substitutes. A one-hour increase in average weekly physical activities decreases earnings by 1–2% within a location, but has no effect on wages. However, I find that in the long run, physical activity does enhance labor productivity. A one-hour increase in average weekly physical activities increases average earnings and wages by 6–7% within a location. I also find spending time on more health-promoting physical activities doubles the long-run returns to earnings, providing evidence supporting the health channel. Meanwhile, I find evidence at odds with the social capital channel. Spending time alone on physical activity has a slightly stronger effect on earnings

[3]Two studies use an instrumental variable approach. Barron et al. (2000) use high school size, library books per student, faculty-student ratio, county mean family income, proportion of families headed by women, proportion of families who lived in the same county/city/house for the past five years as instruments for participation in high school athletics. Stevenson (2010) uses variation in pre–Title IX levels of high school athletic participation among males as an instrument for female athletic participation in high school.

than my main long-run estimates, while the effect of spending time with others on physical activity is statistically indistinguishable from zero.

The remainder of this paper is organized as follows. Section 3.2 describes the data; Section 3.3 outlines the identification strategy; Section 3.4 presents short and long-run results alongside robustness checks; and Section 3.5 concludes.

3.2 Data

The data used in this paper come from the American Time Use Survey (ATUS) and the Current Population Survey (CPS). The ATUS, which is conducted by the U.S. Census Bureau for the Bureau of Labor Statistics (BLS), measures the amount of time individuals spend on daily activities by randomly interviewing people and constructing one-day time diaries, spanning 4 a.m. on the previous day to 4 a.m. on the day of interview. The first set of interviews began in January 2003, and additional interviews have been conducted continuously every month to the present day. On average, approximately 12,000 individuals are interviewed annually.

The ATUS draws from a sample of households that have recently completed the final interview of the CPS, a monthly labor force survey in the United States. The sample of households are selected to ensure that estimates are nationally representative. One individual age 15 or over is randomly chosen from each sampled household. This person is interviewed by telephone about their activities on the day before the interview. By employing a short recall period and requiring all activities to total to 24 hours, the ATUS minimizes common biases associated with time diaries (Hamermesh et al., 2005). For each activity, the interviewer records the respondent's description, time spent on the activity and end time. The activities are later coded into one of over 400 detailed categories, along with supplemental information, allowing me to observe, for example, whether a workout session involved jogging alone or playing racquetball with a friend. Unfortunately, each respondent participates in the ATUS a single time, so it is impossible to construct an individual-level panel using the data.[4]

It is possible to link the ATUS with the CPS using unique individual-level identifiers, thereby providing a set of rich demographic and labor market variables for each respondent in the ATUS. Of particular importance for the analysis at hand, the CPS provides information on weekly earnings, wages, and geographic location. I discuss each one individually in further detail below.

3.2.1 Location and Sunset Time

The ATUS does not provide location information for respondents, but this information can be obtained by linking the data to the CPS. The majority of individuals in the CPS have location at the county level; however, this information is suppressed for anyone residing in a county with a population under 100,000. When county-level information is suppressed, the CPS either provides location information on the Cen-

[4] A more detailed description of the ATUS is provided by Bureau of Labor Statistics (2015).

sus Core-based Statistical Area (CBSA) or state.[5] Using this information, I geocode individuals' locations into geographical coordinates based on county or CBSA centroid when possible. For the remaining individuals, I use population-weighted state centroids provided by the Census Bureau to geocode individuals' locations. Lastly, using the geocoded data and the day of interview, I calculate sunset time for each individual using solar algorithms from Meeus (1991) and adapted by Gibson and Shrader (2018). I also compute annual average sunset time by averaging over all observations for a given year in a location.

3.2.2 Labor Market Information

I consider two measures of labor productivity: "usual weekly earnings" and "hourly earnings at main job". The weekly earnings variable is from the final wave of the CPS unless an individual changed jobs or employment status in the time between the final CPS wave and the ATUS interview.[6] Earnings information is provided for all respondents who report positive labor income and are not self employed. Additional labor market information, such as hourly wage rate and occupation codes, are similarly derived from the final CPS wave or the ATUS depending on whether a respondent changed jobs in the interim. Hourly earnings information is only provided for workers who receive an hourly wage, which comprises roughly half of the main estimation sample.

3.2.3 Physical Activity

For my measure of physical activity, I use cumulative time spent on any form of leisure-based physical activity from the ATUS time diary.[7] I convert the original time duration measured in daily minutes to weekly hours in order to better align with my measure of earnings, which is reported as usual weekly earnings. As a robustness check, I consider different forms of physical activities in Section 3.4.3. Excluding less common forms of physical activities, such as rodeo competitions and fencing, changes the point estimates very little, but does improves statistical significance and increases precision of the estimated first stage.

3.2.4 Sample Criteria

In this paper, I use the first seventeen years of data from the ATUS, spanning 2003 to 2019. The main estimation sample is comprised of prime-age, full-time workers who report receiving some weekly earnings from either a primary or secondary job. Furthermore, night-shift workers are excluded from the sample since these individuals' physical activity is unlikely to be affected by my instrument, sunset time.[8] The

[5] I observe 44% of individuals at the county level, 32% at the CBSA level, and 24% at the state level.

[6] Usual weekly earnings are updated during the ATUS interview for approximately one third of employed respondents. As a result, for the majority of respondents, there is a timing mismatch between earnings information and the time diaries because the CPS interview predates the ATUS interview by three months on average.

[7] A list of all physical activities in the ATUS is provided in Table 3.9.

[8] Night-shift workers comprise less than 5% of the sample.

final sample size is 87,168 unique individuals spread across 589 locations. Summary statistics for the main estimation sample are provided in Table 3.8.

3.3 Identification Strategy

3.3.1 Endogeneity of Physical Activity

In order to evaluate the causal effect of leisure-based physical activity on earnings, a valid research design must account for the endogeneity of individuals' decisions regarding physical activity. It is likely that several sources of omitted variable bias exist in a naive regression of earnings on physical activity. For example, an individual's underlying health status may directly affect both physical activity levels and earnings (Cutler and Glaeser, 2005). By controlling for health status it is possible to limit such confounding, but this is not ideal since physical activity itself affects health. Directly controlling for health status would shutdown one of the most likely channels through which physical activity affects earnings. Other potential omitted variables are sociability and motivation. More social-oriented people may engage in physical activities with others, such as team sports, and perform better in the labor market due to better networking ability. In a similar vein, more motivated individuals may stick with an exercise routine longer, while also performing better at work due to their higher levels of motivation. Empirically, it is difficult to control for sociability and motivation. To address potential issues of endogeneity, I use an instrumental variable approach that exploits sunset time as a plausible source of exogenous variation in physical activity.

3.3.2 Two-Stage Least Squares Estimation

The key idea behind an instrumental variable approach is to find a variable that is correlated with the causal variable of interest, physical activity in this case, but uncorrelated with other determinants of the dependent variable, earnings. The most common instrumental variable estimator is Two-Stage Least Squares (2SLS), which involves first regressing the variable of interest, physical activity, on the instrumental variable, and then regressing earnings on the predicted values of physical activity, based on the coefficients estimated from the first-stage regression. Intuitively, this approach solves the omitted variables problem by only using variation in physical activity that is uncorrelated with the omitted variables to estimate the relationship between physical activity and earnings (Angrist and Krueger, 2001).

3.3.3 Short-run Effects

When estimating the short-run effects of physical activity on earnings, I apply 2SLS to estimate equations of the following form:

$$\ln(W_{ijt}) = \tau_{SR}\, PA_{ijt} + \mathbf{X}'_{it}\beta + \gamma_j + \epsilon_{ijt} \tag{3.1}$$

$$PA_{ijt} = \alpha_{SR}\,\text{sunset}_{jt} + \mathbf{X}'_{it}\delta + \lambda_j + u_{ijt} \tag{3.2}$$

where W_{ijt} is weekly earnings or hourly wage rate observed for individual i in location j at time t; PA_{ijt} is leisure-based physical activity measured in weekly hours; sunset$_{jt}$ is daily sunset time measured in hours (zero is midnight); $\mathbf{X}_{it}$ is a set of individual-level controls (age, age squared, race, sex, occupation, day of the week, year, an indicator for holiday); γ_j and λ_j are location fixed effects; ϵ_{ijt} and u_{ijt} are random error terms. The location fixed effects are at the smallest geographic level observed for an individual (county, CBSA, or state).[9] Likewise, the standard errors are clustered at the smallest geographic level observed for an individual. As shown in Section 3.4.2, the results are robust to clustering at the state level.

The first-stage equation (3.2) links the endogenous variable PA_{it} to the instrument, sunset$_{jt}$; while, the parameter τ_{SR} from the second-stage equation (3.1) captures the causal effect of physical activity on earnings. If daily sunset time is a valid instrument for physical activity, then Equations (3.1) and (3.2) can be used to recover causal estimates of the effect of physical activity on earnings. Since the instrument varies on a daily basis, it will identify short-run variation in physical activity (Frazis and Stewart, 2012).

3.3.4 Long-run Effects

When estimating the long-run effects of physical activity on earnings, I collapse the estimation sample by location to form a cross section of 589 locations.[10] I then apply 2SLS to estimate equations of the following form:

$$\ln(W_j) = \tau_{LR}\,PA_j + \mathbf{X}'_j\Lambda + \eta_j \tag{3.3}$$

$$PA_j = \alpha_{LR}\,\text{sunset}_j + \mathbf{X}'_j\zeta + v_j \tag{3.4}$$

where W_j is average weekly earnings or hourly wage rate for location j; PA_j is average leisure-based physical activity measured in weekly hours; sunset$_j$ is annual average sunset time measured in hours (zero is midnight); $\mathbf{X}_j$ is a set of location-specific controls (mean age, mean squared age, proportion of females, race shares, occupation shares, population density, coastal indicator, coastal distance, latitude); η_j and v_j are random error terms. Given the limited sample size, I use heteroskedastic-robust standard errors instead of cluster-robust standard errors.[11] As shown in Section 3.4.2, however, the results are robust to clustering at the state level.

Similar to the short-run equations, the first-stage equation (3.4) links the endogenous variable PA_j to the instrument, sunset$_j$; while, the parameter τ_{LR} from the second-stage equation (3.3) captures the causal

[9] See my discussion in Section 3.2.1 for more details on geocoding the data.

[10] Following Solon et al. (2015), I use weights for each location based on counts of the underlying ATUS observations to correct for heteroskedasticity.

[11] Unlike the short-run analysis, it is not possible to cluster at the smallest geographic level observed for an individual since each cluster would only have a single observation.

effect of physical activity on earnings. Since average sunset time is fixed for a given location, it identifies long-run variation in physical activity by exploiting spatial differences (Frazis and Stewart, 2012).

As with any instrumental variable approach, the estimated treatment effects $\hat{\tau}_{SR}$ and $\hat{\tau}_{LR}$ are interpreted as local average treatment effects (LATE) of the complier population. However, since identification relies on location-level variation, these estimates potentially capture productivity spillovers across workers within a location. As a result, the estimated treatment effects capture the effect of increasing average physical activity in a location, not the effect of increasing individual-level physical activity.

3.3.5 Relevance of the Instrument

As an initial check for instrumental relevance, Figure 3.1 shows the proportion of full-time workers engaging in leisure-based physical activities by time of day. Although participation in physical activities begins to decline prior to sunset, the rate of decline quickly accelerates after sunset. This pattern suggests that sunset time is relevant to decisions on when to engage in physical activity; particularly for those who do so in the evening, which is the most popular time of day. The instrumental relevance of sunset time comes from biological and institutional features. First, the human biological clock naturally follows sun patterns (Roenneberg et al., 2007). As a result, a setting sun provides an external cue that induces the human body to begin preparing for sleep by producing more melatonin. Second, many public areas, such as parks and beaches, have curfew ordinances that begin at sunset. Since outdoors is the most common place for people to engage in physical activities, these curfews make it more difficult to do so after the sun sets (Dunton et al., 2008). Absent external influences, changes in sunset time may simply shift when individuals start and stop a physical activity but leave total duration of physical activities unchanged. However, full-time workers face coordination constraints in the morning due to work and school schedules (Hamermesh et al., 2008). As a result, an earlier sunset time, should reduce the total amount of time available for people to engage in physical activities.

To test this hypothesis, Table 3.1 presents regression results of the effect of daily sunset time on the start time, stop time, and duration of leisure-based physical activities among full-time workers. The empirical results support my claim, an earlier sunset time results in earlier start and stop times of physical activities, but the effect is stronger for stop times. The cumulative effect is that the sun setting one hour earlier reduces the duration of physical activities by 4.6 minutes, on average. The mean duration for a physical activity is roughly 70 minutes, so this constitutes a 6% reduction. More importantly, there is almost six hours of variation in daily sunset time across the continental United States throughout the year, meaning physical activity duration can vary by almost 28 minutes due to sunset time.

Lastly, I provide estimates of the first-stage effect of daily sunset time on cumulative weekly time spent on leisure-based physical activities for my main estimation sample in the first column of Table 3.3. In line with Table 3.9, the effect of sunset time on cumulative physical activity time is modest but statistically significant. A one-hour increase in daily sunset time results in an additional 18 minutes per week of physical activity. The accompanying Cragg-Donald F-statistic is 319.524, further supporting the validity of using sunset time as an instrument and suggesting there are no issues with weak instrument bias.[12]

3.3.6 Threats to Internal Validity

Identification of the short-run effects requires that daily sunset time within a location not covary with unobserved determinants of earnings. One potential threat to this assumption is time spent sleeping since there is a biological relationship between sleep patterns and daylight (Roenneberg et al., 1998; Hubert et al., 2007). Moreover, Gibson and Shrader (2018) find that sleep has a positive effect on measures of labor productivity. Together this implies that the instrument is likely correlated with the the error term in Equation (3.1). I investigate this issue further in Section 3.4.1, but find my results are robust to potential confounding from sleep. A more difficult problem to address is that variation in daily sunset time is more or less perfectly correlated with daily sunrise time and total daylight within a location. Consequently, I cannot control for the effects of either when estimating Equations (3.2) and (3.1). It may be the case, for example, that total daylight has a positive effect on both earnings and physical activity via mental health. A number of studies find that daylight has a positive effect on mental health, so the short-run effects of physical activity on earnings would be biased upward if total daylight time is indeed a confounder (Lambert et al., 2002; Berk et al., 2008; Beute and de Kort, 2018).

Another potential threat to instrumental validity is seasonal variations in the labor market since sunset time follows a seasonal pattern over the calendar year. However, abrupt changes in sunset time during the spring and fall caused by daylight savings time severe the link between short-run variation in sunset and seasonal features, such as the Christmas shopping season (Gibson and Shrader, 2018). A more difficult identification issue is the timing mismatch between observed earnings and time uses from the ATUS since the majority of respondents' earnings information is from their last CPS interview, not at the time of the ATUS interview. Even for those respondents who provide information on their earnings at the time of the ATUS interview, there is a lag between when changes in productivity are reflected in wages. As a result, my short-run estimates of the effect of physical activity on earnings will be biased toward zero. More specifically, the magnitude of the short-run estimates should be between zero and one-quarter of the true effects.[13]

Meanwhile, identification of the long-run effects requires variation in average sunset time across locations be uncorrelated with other labor market factors. Within a time zone, average sunset time is a linear function of longitude, and therefore correlated with distance from the Eastern and Western coastlines. Accordingly, I include controls for coastal distance and whether a location is a coastal county in all model specifications. As a robustness check, I also control for longitude and find little change in the long-run results. In the continental United States, all locations experience roughly the same amount of average daylight over the year, so unlike in my short-run analysis, this does not present a source of potential omitted

[12] The respective Stock-Yogo critical value for the weak identification test based on 2SLS size is 16.38 at the 10% significance level (Stock and Yogo, 2005).

[13] Following Barattieri et al. (2014), Gibson and Shrader (2018) derive predictions of the attenuation bias resulting from the timing mismatch in the ATUS.

variable bias. It is possible that average sunset time changes the time of day people work, and if workers are more productive at particular times then the instrument would be correlated with the error term in Equation (3.3). However, using data from the ATUS, Hamermesh et al. (2008) find that work schedules are not affected by sunrise or sunset times.

Lastly, there is always the possibility of a spurious correlation between average sunset time and unobserved location-specific determinants of earnings. For example, Gibson and Shrader (2018) find a statistically significant relationship between average sunset time and population density; hence, I control for population density in all model specifications. Since additional unobserved determinants of earnings may exist, I consider a number of robustness checks in Section 3.4.2.

3.4 Results

3.4.1 Main Results

I begin by considering the short-run effects of physical activity. Table 3.2 shows the estimated short-run effects of leisure-based physical activity on weekly earnings and hourly wages. The first model specification does not control for time spent on sleep, while the second specification does control for sleep time.[14] In both specifications, I find a small negative effect of physical activity on weekly earnings. More specifically, a one-hour increase in average weekly time spent on physical activity within a location results in a 1.1–1.4% decrease in weekly earnings. It is theoretically unlikely that leisure-based physical activity negatively affects work productivity, so this decrease in weekly earnings likely reflects a reduction in work hours. To test this hypothesis, the third and fourth columns of Table 3.2 show the estimated short-run effects of physical activity on hourly wages for respondents who report one.[15] Although the point estimates are negative, their estimated magnitudes are roughly half that of the earnings estimates and not statistically significant. If the negative short-run effect of physical activity on earnings was driven by a reduction in productivity, and not a change in work hours, then I would expect to find evidence of a comparable effect on hourly wages.

Following a method by Allen and Rehbeck (2020), I can estimate time use complementarity to assess whether time spent on physical activity is indeed being substituted for work time in the short run. In practical terms, their approach involves regressing mutually exclusive time use categories (e.g., leisure-based physical activity, work, sleep, and residual leisure time) individually on a time shifter, sunset time in this case, and then using the ratio of two regression coefficients to determine complementarity. A positive ratio implies that the two time uses are complements, while a negative ratio implies they are substitutes. Table 3.3 shows the effect of daily sunset time on four mutually exclusive categories of time use.[16] The ratio

[14] As I discuss in Section 3.3.6, time spent on sleep is likely a confounding variable. Directly controlling for sleep, however, forces all changes in physical activity to come out of work or leisure time. Consequently, the estimates would be biased if other time uses respond to sunset time and covary with unobserved determinants of earnings.

[15] The sample size is reduced by half, but the first-stage effect remains strong, suggesting no issues with weak instrument bias.

[16] Note, the first column is the respective first-stage effect for the model specification in the first column of Table 3.2.

of interest is negative (0.3106/-0.0585), suggesting that physical activity and work time are substitutes in the short run. At the same time, physical activity and sleep are substitutes (0.3106/-0.4932). If additional physical activity comes at the expense of sleep, which Gibson and Shrader (2018) find to be productivity enhancing, then that could explain the observed reduction in earnings absent any change in work hours. However, the second model specification of Table 3.2 explicitly controls for sleep, thereby shutting down this potential channel, and still estimates a negative earnings effect of physical activity.

The literature on physical activity and earnings provides empirical evidence on two primary channels through which physical activity can affect earnings: (1) an investment in health capital; and (2) an investment in social capital (Lechner, 2009; Rooth, 2011; Lechner and Sari, 2015; Sari and Lechner, 2015; Lechner and Downward, 2017). Both of these channels likely require a longer time horizon to be reflected in labor market outcomes than provided by my short-run analysis. In fact, the majority of empirical studies that find an effect of physical activity, or athletic participation, on earnings consider a time window greater than ten years (Barron et al., 2000; Ewing, 2007; Rooth, 2011; Hyytinen and Lahtonen, 2013; Lechner and Sari, 2015; Lechner and Downward, 2017). Moreover, as I discuss in Section 3.3.6, attenuation bias from the timing mismatch between earnings and physical activity mutes any productivity gains in the short run.

I now consider the long-run effects of physical activity. Table 3.5 shows the estimated long-run effects of leisure-based physical activity on weekly earnings and hourly wages. Analogous to the short-run results, the first model specification does not control for average time spent on sleep in a location, while the second specification does control for average sleep time. In both specifications, I find a positive effect of physical activity on average weekly earnings. A one-hour increase in average physical activity increases average earnings by 6.1–6.7% for a location. Although these estimates may appear large, other empirical studies on the long-run effect of physical activity, or athletic participation, on earnings range from 4–20% (Barron et al., 2000; Ewing, 2007; Lechner, 2009; Rooth, 2011; Hyytinen and Lahtonen, 2013; Lechner and Sari, 2015; Lechner and Downward, 2017). Furthermore, mean physical activity is roughly 2 hours per week across locations, so a one-hour increase in average physical activity for a location would involve a substantial 50% increase in time spent on physical activities.

An increase in average weekly earnings could reflect an increase in average hourly wages, average weekly work hours, or both. The third and fourth columns of Table 3.5 show that a one-hour increase in average physical activity increases the average hourly wage rate by 7.1–7.5% for a location, suggesting most of the effect on average earnings is due to an increase in productivity, not an increase in work hours. A bit of caution is warranted when discussing the estimated long-run effects since the first-stage F-statistics are smaller than the generally accepted rule of thumb, 10 (Stock and Yogo, 2005). The estimates may suffer from weak instrument bias. That said, I consider a robustness check in Section 3.4.3 that returns similar long-run estimates but has a much stronger first-stage effect.

3.4.2 Robustness Checks

Identification of the short-run effects comes from seasonal variation in sunset time within a location, so unobserved seasonal variables may pose a threat to internal validity. Table 3.4 reports a number of robustness checks that I run to assess some of these concerns. First, the inclusion of quarter fixed effects slightly increases both the magnitude and standard error of the estimated short-run effect, but it remains statistically significant. Within a calendar quarter there is less variation in daily sunset time to use for identification, so it is unsurprising that the estimate is less precise. Second, controlling for daily average temperature more than doubles the magnitude and standard error of the estimated short-run effect, but the effect is still statistically significant at the 10% level. Third, excluding federal holidays and the traditional holiday season (Thanksgiving through New Year's Day) has almost no effect on the magnitude or precision of the estimated short-run effect. The short-run results are also robust to higher-order clustering of the standard errors or the inclusion of additional demographic variables (education, marital status, and number of children). Lastly, the results are robust to the exclusion of observations located in East or West coast states, so dense metropolitans (e.g., New York, Los Angeles, Washington D.C.) are not driving the results.

Meanwhile, identification of the long-run effects requires variation in average sunset time be uncorrelated with other labor market factors across locations. Table 3.6 presents results from multiple robustness checks to help alleviate concerns over a spurious correlation between average sunset time and unobserved location-specific determinants of earnings. First, the inclusion of regional fixed effects slightly increases the magnitude of the long-run point estimate, but it does not alter the statistical significance. Second, directly controlling for longitude, in addition to coastal distance, increases the magnitude of the long-run effect by less than a percentage point, while leaving the statistical significance unchanged. Third, excluding observations located in high-wage metropolitans (Boston, Chicago, Los Angeles, New York City, and San Francisco) reduces the magnitude of the point estimate but increases precision, suggesting that the results are not being driven by densely populated urban areas.[17] The long-run results are also robust to clustering the standard errors at the state level or the inclusion of additional location-level demographic variables (education shares, proportion married, average number of children).

3.4.3 Effect Heterogeneity by Type of Physical Activity

My main results take an agnostic view on what constitutes physical activity by including time spent on all "Sports, Exercise, and Recreation" categories listed in the ATUS. However, a few of the activities the ATUS considers physical activities are eccentric, such as equestrian sports and rodeo competitions, or involve limited amounts of physical movement, such as hunting or vehicle racing. When policymakers or researchers discuss the importance of increasing population levels of physical activity, they are referring to more traditional exercise routines with health-promoting benefits, not vehicle racing. If my results are being driven by these uncommon forms of physical activity, then any implied policy implications or comparisons with the related literature may be incorrect. As a robustness check, I re-estimate the long-run effect of physical activity on earnings using a subset of physical activities that I consider "traditional" forms of exercise.[18] Results are reported in the first column of Table 3.7. I find that the effect of tradi-

[17] As with my set of short-run robustness checks, I also re-estimate the long-run effect while excluding observations located in East or West coast states. The point estimate remains positive, but the sample size is too small to provide precise inference.

[18] A list of the "traditional" physical activities is provided in Table 3.9.

tional physical activities on earnings is slightly larger in magnitude compared to my main results. More importantly, precision increases and the first-stage F-statistic more than doubles. It is unlikely decisions on when to go vehicle racing or participate in a rodeo competition are affected by sunset time, so it is unsurprising that the first-stage effect of average sunset on physical activity is stronger when excluding such activities. This stronger first-stage effect helps alleviate concerns over weak instrument bias mentioned in Section 3.4.1.

By considering effect heterogeneity by type of physical activity, I can also shed some light on the potential mechanisms through which physical activity affects earnings. The long-run earnings effect of traditional physical activities may be larger than my main results because, on average, these activities are more health promoting. To better test this hypothesis, I consider a more objective list of health-promoting physical activities. First, I obtain measures on the average metabolic equivalents (MET) of each physical activity in the ATUS from Tudor-Locke et al. (2009). An activity's MET is the average amount of energy expended per unit of time relative to sitting still (resting metabolic rate). For example, running has an average MET of 7.5, so the average person exerts 7.5 times more energy running than sitting still. Using this information, I separate all physical activities in the ATUS into High (Low) MET PA based on whether the activity's MET is above (below) the median MET.[19]

The second and third columns of Table 3.7 show the long-run effect of High and Low MET physical activity on earnings, respectively. I find that the effect of High MET physical activities on earnings is nearly double the size of my main results, and statistically significant at the 5% level. Engaging in a High MET physical activity should be more health promoting than spending an equivalent amount of time on an average physical activity, so the much larger effect size provides evidence in support of the health channel. Moreover, the estimated effect of Low MET physical activities on earnings is statistically indistinguishable from zero. The difference between the High and Low MET estimates may, however, be solely driven by the fact that the first-stage effect for Low MET is weak so caution is warranted.[20]

Another potential mechanism through which physical activity can affect earnings is social capital accumulation (Barron et al., 2000; Seippel, 2006; Lechner and Downward, 2017). This strand of literature suggests that engaging in physical activities with others may help individuals develop social or teamwork skills that are valued in the labor market. To test this hypothesis, I consider effect heterogeneity in physical activity on earnings by the presence of other people while engaging in a physical activity. I use supplemental information from the ATUS on whether an activity is done alone or with others (e.g., friends, neighbors, family members) to categorize physical activities into two categories: "Solo PA" and "Non-solo PA". Columns four and five of Table 3.7 report results for Solo physical activities and Non-solo physical activities, respectively. I find that the effect of Solo PA on earnings is larger in magnitude compared to my main results, and statistically significant at the 1% level. Moreover, the first-stage F-statistic nearly doubles, similar to my traditional physical activity results. In contrast, the effect of Non-solo PA on earnings is statistically indistinguishable from zero. The difference between Solo PA and Non-solo PA may be partly driven by the fact that individuals engage in High MET activities more often alone (34% of the time) than

[19] Table 3.9 categorizes each physical activity in the ATUS by either High MET PA or Low MET PA.

[20] The exclusion restriction for Low MET physical activity is likely violated since the instrument is relevant for High physical activities.

with others (26% of the time), assuming the health channel is at play. However, the low first-stage effect for Non-solo PA limits the amount of inference that can be drawn from such comparisons.[21]

3.5 Conclusion

In this paper, I examine how leisure-based physical activity affects earnings and wages. I find that in the short run, physical activity does not enhance labor productivity; instead, it appears work time and physical activity are substitutes. A one-hour increase in average weekly physical activities decreases earnings by 1–2% within a location, but has no effect on wages. However, I find that in the long run, physical activity does enhance labor productivity. A one-hour increase in average weekly physical activities increases average earnings and wages by 6–7% within a location. Back-of-the-envelope calculations of the annual income effect implied by my estimated long-run effect help put things into perspective. If average weekly physical activity increased by one hour within a location, then average annual income would increase by $3,340, assuming no change in work hours. As I discuss in Section 3.3.5, however, physical activity and work are substitutes so it is likely that an increase in physical activity would reduce work time. Assuming half of the additional hour spent on physical activity per week comes out of work time, the estimated annual income effect would be $2,760. If the entire additional hour of physical activity came out of work time, the estimated annual income effect would be $2,180.

My results support the small but growing body of literature on the labor market returns to physical activity, and the importance of non-labor time use on labor market outcomes. By using a new instrument for physical activity, I extend recent empirical research on the topic, which mostly relies on using fixed-effects or matching estimators to address endogeneity. I also complement this body of research by providing additional evidence on effect heterogeneity by type of physical activity and on the potential mechanisms through which physical activity affects labor productivity. My results provide further support for the health channel as the most likely mechanism behind the labor market returns to physical activity, though future research on the *causal* health effects of physical activity is warranted. An extensive body of medical research finds a positive association between physical activity and health outcomes, but there is a shortage of causal evidence.[22] Understanding the first-order effects of physical activity on health is necessary to gain a better understanding of how physical activity affects labor market outcomes.

[21] The exclusion restriction for Non-solo physical activity is likely violated since the instrument is relevant for Solo physical activity.

[22] To my knowledge, only Lechner (2009), Lechner and Sari (2015), and Sari and Lechner (2015) take a *causal* approach toward estimating the health effects of physical activity.

3.6 References

Aguilera, V., Bernabe, M., (2005). The Impact of Social Capital on the Earnings of Puerto Rican Migrants. The Sociological Quarterly, 46, 569–592.

Allen, R., and Rehbeck, J., (2020). Hicksian Complementarity and Perturbed Utility Models. Economic Theory Bulletin, 8, 245-261.

Angrist, J., and Krueger, A., (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. Journal of Economic Perspectives, 15(4), 69-85.

Barattieri, A., et al., (2014). Some Evidence on The Importance of Sticky Wages. American Economic Journal: Macroeconomics, 6(1), 70-101.

Barron, J., et al., (2000). The Effects of High School Athletic Participation on Education and Labor Market Outcomes. The Review of Economics and Statistics, 82, 409–421.

Bayer, P., et al., (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. Journal of Political Economy, 116, 1150–1194.

Berk, M., et al., (2008). Small Shifts in Diurnal Rhythm Are Associated with an Increase in Suicide: The Effect of Daylight Saving. Sleep and Biological Rhythms, 6, 22–25.

Beute, F., and de Kort, Y., (2018). The Natural Context of Wellbeing: Ecological Momentary Assessment of the Influence of Nature and Daylight on Affect and Stress for Individuals with Depression Levels Varying from None to Clinical. Health & Policy, 49, 7-18.

Biddle, J., and Hamermesh, D., (1998). Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre. Journal of Labor Economics, 16(1), 172–201.

Bureau of Labor Statistics, (2015). American Time Use Survey User's Guide.

Center for Disease Control, (2018). Early Release of Selected Estimates from the 2018 National Health Interview Survey.

Costa-Font, J., and Fleche, S., (2020). Child Sleep and Mother Labour Market Outcomes. Journal of Health Economics, 69, 102258.

Cutler, D., and Glaeser, E., (2005). What Explains Differences in Smoking, Drinking, and Other Health-related Behaviors? American Economic Review, 95, 238-242.

Dunton, G., et al., (2008). Social and Physical Environments of Sports and Exercise Reported Among Adults in the American Time Use Survey. Preventative Medicine, 47, 519-524.

Ewing, Bradley (2007). The Labor Market Effects of High School Athletic Participation: Evidence From Wage and Fringe Benefit Differentials. Journal of Sports Economics, 8(3), 255-265.

Frazis, H., and Stewart, J., (2012). How to Think about Time-Use Data: What Inferences Can We Make about Long- and Short-Run Time Use from Time Diaries?. Annals of Economics and Statistics, 105–106, 231–245.

Gibson, M., and Shrader, J., (2018). Time Use and Productivity: The Returns to Sleep. Review of Economics and Statistics, 100(5), 783–798.

Grossman, M., (1972). On the Concept of Health Capital and the Demand for Health. The Journal of Political Economy, 80, 223-255.

Guthold, R., et al., (2018). Worldwide Trends in Insufficient Physical Activity from 2001 to 2016: A Pooled Analysis of 358 Population-based Surveys with 1.9 Million Participants. Lancet Global Health, 6, 1077-1086.

Hamermesh, D., et al., (2005). The American Time-Use Survey. Journal of Economic Perspectives, 19, 221–232.

Hamermesh, D., et al., (2008). Cues for Timing and Coordination: Latitude, Letterman, and Longitude. Journal of Labor Economics, 26, 223–246.

Hamermesh, D., and Biddle, J., (1994). Beauty and the Labor Market. American Economic Review, 84(5), 1174–1194.

Hersch, Joni (2009). Home Production and Wages: Evidence from the American Time Use Survey. Review of Economics of the Household 7, 159–178.

Hersch, J., and Stratton, L., (2002). Housework and Wages. Journal of Human Resources, 37(1), 217-229.

Hubert, M., et al., (1998). Seasonal and Diurnal Patterns of Human Illumination under Natural Conditions. Chronobiology International 15 (1998), 59–70.

Hyytinen, A., and Lahtonen, J., (2013). The Effect of Physical Activity on Long-term Income. Social Science and Medicine, 96, 129-137.

Lambert, G., et al., (2002). Effect of Sunlight and Season on Serotonin Turnover in the Brain. Lancet, 360, 1840–1842.

Lechner, Michael (2009). Long-run Labour Market and Health Effects of Individual Sports Activities. Journal of Health Economics, 28, 839–854.

Lechner, M., and Sari, N., (2015). Labor Market Effects of Sports and Exercise: Evidence from Canadian Panel Data. Labour Economics, 35, 1-15.

Lechner, M., and Downward, P., (2017). Heterogeneous Sports Participation and Labour Market Outcomes in England. Applied Economics, 49:4, 335-348.

Liu, X., and Zhu, C., (2014). Will Knowing Diabetes Affect Labor Income? Evidence from a Natural Experiment. Economics Letters, 124(1), 74-78.

Lu, Y., et al., (2020). How Does Improvement in Commuting Affect Employees? Evidence from a Natural Experiment. SSRN Working Paper 3456299.

Maruyama, S., and Yin, Q., (2012). The opportunity cost of exercise: Do Higher-earning Australians Exercise Longer, Harder, or Both?. Health Policy, 106, 187-194.

Meeus, Jean H., (1991). Astronomical Algorithms. Richmond, VA: Willmann-Bell.

Mullahy, J., and Robert, S., (2010). No Time To Lose: Time Constraints and Physical Activity in the Production of Health. Review of Economics of the Household, 8, 409-432.

Rao et al., (2013). Do Healthier Foods and Diet Patterns Cost More than Less Healthy Options? A Systematic Review and Meta-analysis. British Medical Journal, 3, 004277.

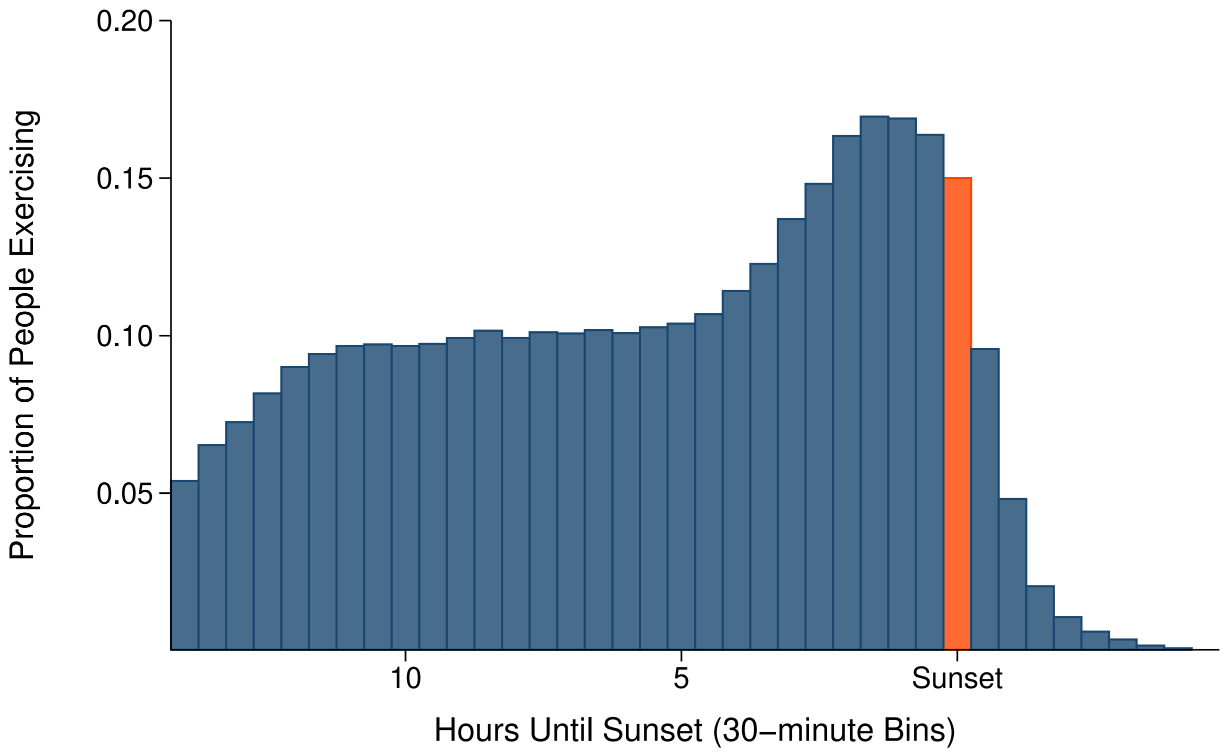
Roenneberg, T., et al., (2007). The Human Circadian Clock Entrains to Sun Time. Current Biology, 17, R44–R45.

Rooth, Dan-Olof (2011). Work Out or Out of Work — The Labor Market Return To Physical Fitness and Leisure Sports Activities. Labour Economics, 18, 399-409.

Ruhm, C., (2007). Current and Future Prelevence of Obesity and Severe Obesity in the United States. Forum for Health Economics & Policy, 10(2), 1–26.
Sari, N., and Lechner, M., (2015). Long-run Health Effects of Sports and Exercise in Canada. Canadian Centre for Health Economics, Working Paper No. 150018.
Schmitz, H., and Westphal, M., (2017). Informal Care and Long-term Labor Market Outcomes. Journal of Health Economics, 56, 1-18.
Schuttoff, U., et al., (2017). Sports Participation and Social Capital Formation During Adolescence. Social Science Quarterly, 99(2), 683-698.
Schmutte, Ian (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. Journal of Labor Economics, 3(1), 1-32.
Seippel, Ornulf (2006). Sport and Social Capital. Acta Sociologica, 49(2), 169-193.
Stevenson, Betsey (2010). Beyond the Classroom: Using Title IX to Measure the Return to High School Sports. The Review of Economics and Statistics, 92(2), 284–301.
Shuval, K., et al., (2017). Income, Physical Activity, Sedentary Behavior, and the 'Weekend Warrior' Among U.S. Adults. Preventive Medicine, 103, 91-97.
Stock, J., and Yogo, M., (2005). Testing for Weak Instruments in Linear IV Regression. In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press, 80-108.
Tudor-Locke, C., et al., (2009). Linking the American Time Use Survey (ATUS) and the Compendium of Physical Activities: Methods and Rationale. Journal of Physical Activity and Health, 6(3), 347-353.
U.S. Department of Health and Human Services, (2018). The 2018 Physical Activity Guidelines Advisory Committee Scientific Report.
Van Houtven, C., et al., (2013). The Effect of Informal Care on Work and Wages. Journal of Health Economics, 32, 240-252.
Walls, Margaret (2009). Parks and Recreation in the United States: Local Park Systems. Resources for the Future.

3.7 Tables and Figures

Figure 3.1: Physical Activity by Time of Day



Source: American Time Use Survey (ATUS), 2003 – 2019. Proportion of full–time workers engaged in a physical activity by centered sunset time. Time diaries weighted by day of the week.

Table 3.1: Effect of Sunset on Physical Activity Routines

	Start Time	Stop Time	Duration
Daily Sunset Time	0.3344***	0.4114***	4.6230***
	(0.0965)	(0.0977)	(0.9766)
Controls	Yes	Yes	Yes
Observations	21,152	21,152	21,152

Note: Cluster-robust standard errors at the location level in parentheses. Regression results for full-time workers from the ATUS, 2003 - 2019. Daily Sunset Time is measured in hours (zero is midnight). Start Time and Stop Time are the respective beginning and end of a recorded leisure-based physical activity measured in hours. Duration is the length of a recorded physical activity in minutes. Controls include individual-level demographics (age, age squared, sex, race, education, marital status, number of children, occupation) and indicators for day of the week, holiday, month, and year.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.2: 2SLS Estimates of the Short-run Effect of Physical Activity on Earnings and Wages

	(1) ln(earnings)	(2) ln(earnings)	(1) ln(wage)	(2) ln(wage)
Physical Activity	-0.0113**	-0.0143***	-0.0056	-0.0077
	(0.0049)	(0.0050)	(0.0055)	(0.0056)
Control for Sleep	No	Yes	No	Yes
Observations	87,168	87,168	45,798	45,798
First-stage F-Stat	319.524	305.928	118.771	114.234
Mean (LHS Variable)	$1018/week	$1018/week	$17.52/hour	$17.52/hour

Note: Cluster-robust standard errors at the location level in parentheses. Regression results for full-time workers from the ATUS, 2003 - 2019. Physical Activity is cumulative time spent on leisure-based physical activities measured in hours per week. Earnings are an individual's reported usual weekly earnings, while wage is their reported hourly wage. All models include controls for individual-level demographics (age, age squared, sex, race, occupation), location fixed effects, indicators for day of the week, holiday, and year.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.3: First-stage Estimates of the Short-run Effect of Sunset on Time Use

	PA	Work	Sleep	Leisure
Daily Sunset Time	0.3106***	-0.0585*	-0.4932***	-0.1013*
	(0.0175)	(0.0347)	(0.0372)	(0.0609)
Controls	Yes	Yes	Yes	Yes
Observations	87,168	87,168	87,168	87,168
Mean (LHS Variable)	1.99 hrs/wk	31.58 hrs/wk	57.48 hrs/wk	73.25 hrs/wk

Note: Cluster-robust standard errors at the location level in parentheses. Regression results for full-time workers from the ATUS, 2003 - 2019. Daily Sunset Time is measured in hours (zero is midnight). PA is cumulative time spent on leisure-based physical activities measured in hours per week. Work and Sleep are similarly defined. Leisure is the remaining time spent on activities, net physical activities, work, and sleep. Controls include individual-level demographics (age, age squared, sex, race, occupation), location fixed effects, indicators for day of the week, holiday, and year.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.4: Robustness Checks of the Short-run 2SLS Estimates

		ln(earnings)			ln(earnings)
Quarter Fixed Effects			*More Demographic Controls*		
	PA	-0.0.0193*		PA	-0.0096**
		(0.0109)			(0.0047)
	Obs	87,168		Obs	87,168
Temperature Control			*State Clustering*		
	PA	-0.0296*		PA	-0.0114**
		(0.0174)			(0.0050)
	Obs	87,071		Obs	87,168
No Holidays			*No East or West Coast*		
	PA	-0.0124**		PA	-0.0167**
		(0.0060)			(0.0070)
	Obs	74,518		Obs	34,718

Note: Cluster-robust standard errors at the location level in parentheses. Regression results for full-time workers from the ATUS, 2003 - 2019. PA is cumulative time spent on leisure-based physical activities measured in hours per week. Unless otherwise stated, all models include controls for individual-level demographics (age, age squared, sex, race, occupation), location fixed effects, indicators for day of the week, holiday, and year. More Demographic Controls includes education, marital status, and number of children. Temperature Control is average daily temperature. No Holidays excludes time diaries from Thanksgiving through New Year's Day or a federal holiday.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.5: 2SLS Estimates of the Long-run Effect of Physical Activity on Earnings and Wages

	(1) ln(earnings)	(2) ln(earnings)	(1) ln(wage)	(2) ln(wage)
Physical Activity	0.0672**	0.0613**	0.0756**	0.0709**
	(0.0271)	(0.0250)	(0.0315)	(0.0297)
Control for Sleep	No	Yes	No	Yes
Observations	589	589	586	586
First-stage F-Stat	5.255	5.299	5.274	5.311
Mean (LHS Variable)	$1023/week	$1023/week	$18.07/hour	$18.07/hour

Note: Heteroskedastic-robust standard errors in parentheses. Weighted regression results for full-time workers from the ATUS, 2003 - 2019, aggregated at the location level. Physical Activity is average cumulative time spent on leisure-based physical activities in a location measured in hours per week. Earnings are average usual weekly earnings for a location, while wages are the average hourly wage for a location. All models include controls for location-level demographics (mean age, mean squared age, proportion of females, race shares, occupation shares, population density) and geography (coastal county indicator, coastal distance, latitude).
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.6: Robustness Checks of the Long-run 2SLS Estimates

		ln(earnings)			ln(earnings)
Regional Fixed Effects			*More Demographic Controls*		
	PA	0.0809**		PA	0.0709*
		(0.0398)			(0.0407)
	Obs	589		Obs	589
Longitude Control			*State Clustering*		
	PA	0.0750**		PA	0.0672**
		(0.0344)			(0.0276)
	Obs	589		Obs	589
No High-wage Cities			*No East or West Coast*		
	PA	0.0491***		PA	0.1029
		(0.0170)			(0.0879)
	Obs	536		Obs	211

Note: Heteroskedastic-robust standard errors in parentheses. Weighted regression results for full-time workers from the ATUS, 2003 - 2019, aggregated to the location level. Physical Activity is average cumulative time spent on leisure-based physical activities in a location measured in hours per week. Unless otherwise stated, all models include controls for location-level demographics (mean age, mean squared age, proportion of females, race shares, occupation shares, population density) and geography (coastal county indicator, coastal distance, latitude). More Demographic Controls includes education shares, proportion married, mean number of children. No High-wage Cities excludes counties in Boston, Chicago, Los Angeles, New York City, and San Francisco metropolitan statistical areas.
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.7: Heterogeneous Long-run Effects of Physical Activity on Earnings by Type of Exercise

	(1) ln(earnings)	(2) ln(earnings)	(3) ln(earnings)	(4) ln(earnings)	(5) ln(earnings)
Traditional PA	0.0789^{***}	–	–	–	–
	(0.0246)				
High MET PA	–	0.1241^{**}	–	–	–
		(0.0503)			
Low MET PA	–	–	0.1470	–	–
			(0.1206)		
Solo PA	–	–	–	0.1002^{***}	–
				(0.0320)	
Non-solo PA	–	–	–	–	0.2045
					(0.2058)
Observations	589	589	589	589	589
First-stage F-Stat	10.420	8.093	1.376	9.183	0.771

Note: Heteroskedastic-robust standard errors in parentheses. Weighted regression results for full-time workers from the ATUS, 2003 - 2019, aggregated to the location level. Traditional PA is average cumulative time spent on traditional leisure-based exercise routines (a complete list is provided in Table 3.9). High (Low) MET PA is average cumulative time spent on leisure-based physical activities with a metabolic equivalent above or below the median metabolic equivalent. Solo PA and Non-solo PA is average cumulative time spent on leisure-based physical activities alone or with others, respectively. All models include controls for location-level demographics (mean age, mean squared age, proportion of females, race shares, occupation shares, population density) and geography (coastal county indicator, coastal distance, latitude).
$*p < 0.10, **p < 0.05, ***p < 0.01$

Table 3.8: Sample Statistics from the American Time Use Survey

Variable	Mean	Std Dev	Observations
Weekly Earnings	1018.61	(652.97)	87,168
Hourly Wages	17.52	(9.94)	45,798
Sunset (hours)	18.77	(1.34)	87,168
Baseline PA (hrs/week)	1.99	(6.53)	87,168
Traditional PA (hrs/week)	1.08	(3.97)	87,168
High MET PA (hrs/week)	0.48	(2.74)	87,168
Low MET PA (hrs/week)	1.51	(5.89)	87,168
Solo PA (hrs/week)	0.79	(3.67)	87,168
Non-solo PA (hrs/week)	1.20	(5.57)	87,168
Work (hrs/week)	31.58	(31.38)	87,168
Sleep (hrs/week)	57.48	(15.46)	87,168
Leisure (hrs/week)	73.25	(27.45)	87,168
Age (years)	41.87	(11.02)	87,168
Number of Children	0.95	(1.11)	87,168
Indicator Variables			
Female	0.48	(0.50)	87,168
High School or Less	0.30	(0.46)	87,168
Some College	0.28	(0.45)	87,168
College or More	0.42	(0.49)	87,168
White	0.81	(0.39)	87,168
Black	0.13	(0.33)	87,168
Asian	0.04	(0.20)	87,168
Other Race	0.02	(0.14)	87,168

Note: Summary statistics for full-time workers from the ATUS, 2003 - 2019.

Table 3.9: List of All Physical Activities in the ATUS

Physical Activity	Traditional	High MET	Low MET
Aerobics	✓	✓	
Baseball	✓		✓
Basketball	✓	✓	
Biking	✓	✓	
Billiards			✓
Boating			✓
Bowling			✓
Climbing, Spelunking, Caving	✓	✓	
Dancing			✓
Equestrian Sports			✓
Fencing		✓	
Fishing			✓
Football	✓	✓	
Golfing			✓
Gymnastics	✓		✓
Hiking	✓	✓	
Hockey	✓	✓	
Hunting			✓
Martial Arts	✓	✓	
Racquet Sports	✓	✓	
Rodeo Competition		✓	
Rollerblading		✓	
Rugby	✓	✓	
Running	✓	✓	
Skiing, Ice Skating, Snowboarding	✓	✓	
Soccer	✓	✓	
Softball	✓		✓
Cardiovascular Equipment	✓	✓	
Vehicle Racing			✓
Volleyball	✓	✓	
Walking	✓		✓
Water Sports	✓		✓
Weightlifting	✓		✓
Working Out	✓		✓
Wrestling		✓	
Yoga			✓
Playing Sports	✓		✓

Note: The list of traditional physical activities is subjective, but roughly corresponds to what an average individual would consider common forms of leisure-based exercise. High (Low) MET physical activities have a metabolic equivalent measure above (below) the median metabolic equivalent of the listed leisure-based physical activities.