UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays on Income and Health

Permalink
https://escholarship.org/uc/item/41q0g9z2

Author
Eli, Shari Jane

Publication Date
2011

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Essays on Income and Health

by

Shari Jane Eli

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Barry Eichengreen, Chair
Professor Brad Delong
Professor Robin Einhorn

Fall 2011

Essays on Income and Health

Copyright 2011
by
Shari Jane Eli

Abstract

Essays on Income and Health

by

Shari Jane Eli
Doctor of Philosophy in Economics

University of California, Berkeley

Professor Barry Eichengreen, Chair

In the three chapters of this dissertation, I examine the effects of changes in income on health measures for three different populations. Each of the three populations live in countries with developing economies: Whites in the post-bellum U.S.; black ex-slaves and freedmen upon Emancipation in the U.S., and individuals living in Indian households at the end of the 20th century. In the first and second chapters, I look at the effect of unearned income on morbidity and mortality, while in the third, I investigate how spatial and temporal differences in energy requirements affected food consumption at the end of the 20th century in India. Taken together, the three chapters focus on the effects of income on individual health measures.

In the first essay, I investigate how increases in individual income contributed to improvements in adult health during the late 19th and early 20th century. To disentangle the effect of income as opposed to medical advancements or public health interventions, I use exogenous variation in income from the first wide-scale entitlement program in the United States: the Union Army pensions. Documenting that Republican Congressional candidates boosted veterans pensions in order to secure votes, I exploit exogenous increases in income stemming from Republican corruption to estimate income effects on morbidity and mortality. The effects of income on disease onset are large - an extra $1 of monthly pension income, a 9% average real income increase, lowered the probability of infectious disease onset by 38%. In addition, I find that an extra $1 of monthly income lowers the crude death rate by .008. I find the largest income effects for infectious illnesses, smaller effects for respiratory and digestive illnesses, and no effect for the onset of most endocrine diseases. Results from this chapter help shape our understanding of the U.S. mortality transition and inform todays debates on the health benefits of cash transfers to adults in regions with wide SES gradients in health, as was the case in the U.S. a century ago.

In the second chapter co-authored with Trevon Logan, we investigate the effect of increases in Union Army pension income on the mortality rates of black veterans who served for the side of the North in the Civil War. It is not possible to use the same source of exogenous variation in pension income as in the first chapter since the

majority of black veterans live in the South and thus are unable to vote. Therefore, they are not a constituency of interest to Republican Congressional candidates. To circumvent endogeneity bias, we use propose a new instrument for pension income and an econometric framework to determine black-white differences in the effect of pension income on health. In addition, we investigate the determinants of differential pension rates among black veterans and present new results.

In the third chapter co-authored with Nicholas Li, we explore the effect of declining energy requirements on the demand for calories in India. Despite the large growth in real expenditure and positive calorie-expenditure elasticities, calories consumed per person in India declined between 1983 and 2005. Similarly, rural households are found to be poorer than urban households but consume more calories on average. We test the energy requirements hypothesis of Deaton and Dreze (2009) as an explanation for these missing calories by using time-use data to impute household energy requirements. To show the differential effects of energy requirements and other factors on food quality and expenditures, we use a simple model to provide intuition and motivate our empirical approach. We find that variations in household characteristics such as size, age, education and occupation predict highly correlated fluctuations in caloric intake and requirements. Labor-saving durables also play an important role in caloric intake. Quantitatively, energy requirements can explain most of the missing calories between urban and rural areas. Over time, differing energy requirements explain approximately one-half of the changes in food quality but only a modest share of the total missing calories, which implies that other factors are important. Our results shed light on the importance of considering variations in energy requirements when formulating welfare and poverty measures.

To Vera, Ronny, Samir and Nick

Contents

List of Figures **iv**

List of Tables **v**

1 Wealth is Health: Pensions and Disease Onset in the Gilded Age **1**
1.1 Introduction . . . 1
1.2 Historical Background . . . 4
1.2.1 The Union Army Pension in the Patronage Era . . . 4
1.2.2 Medicine in the Gilded Age . . . 7
1.2.3 Public Health Interventions . . . 10
1.3 Empirical Strategy . . . 10
1.3.1 Part I . . . 11
1.3.2 Part II . . . 16
1.4 Data . . . 18
1.4.1 Union Army Sample . . . 18
1.4.2 Historical Congressional Election Data . . . 19
1.5 Results . . . 20
1.6 Robustness Checks . . . 23
1.6.1 Threats to Identification . . . 23
1.7 Conclusion and Discussion . . . 26

2 Black Mortality in the Post-Bellum Era: Evidence from Union Army Pensions (with Trevon Logan) **41**
2.1 Introduction . . . 41
2.2 Historical Background . . . 43
2.2.1 The war experience of U.S. Colored Troops . . . 43
2.2.2 Race Differentials in Pension Receipt . . . 44
2.3 Empirical Strategy . . . 46
2.3.1 Identification . . . 46
2.3.2 Econometric Framework . . . 48
2.4 Data . . . 50
2.4.1 United States Colored Troops . . . 50

2.4.2 Examining Surgeons dataset . . . 52
2.4.3 City and County-Level data . . . 53
2.5 Results . . . 53
2.6 Conclusion . . . 54

3 In Search of India's Missing Calories: Energy Requirements and Calorie Consumption (with Nicholas Li) 62
3.1 Introduction . . . 62
3.2 Theory . . . 65
3.3 Decomposing the decline in calories . . . 69
3.3.1 Calories per capita . . . 69
3.3.2 Calorie Engel curves . . . 70
3.3.3 Changes in food composition . . . 71
3.4 Caloric intake and requirements in the cross-section . . . 73
3.4.1 Imputation of energy requirements . . . 73
3.4.2 Household composition and life-cycle effects . . . 75
3.4.3 Occupation and education . . . 77
3.4.4 Other factors . . . 78
3.5 Quantitative assessment of energy requirements hypothesis . . . 79
3.5.1 Rural-Urban Gap . . . 80
3.5.2 Over time gap . . . 81
3.6 Measuring poverty and welfare with caloric requirements . . . 82
3.7 Conclusion . . . 84
3.8 Tables and Figures . . . 85

4 Bibliography 108

A Appendices 113
A.1 Age-based Laws . . . 113
A.2 Imputing caloric intake . . . 114
A.2.1 Calorie Intake . . . 114
A.2.2 Data issues . . . 114
A.2.3 Calorie estimates . . . 116
A.3 Imputing caloric requirements . . . 119

List of Figures

1.1 Congressional District Map for Census Year 1870 29
1.2 Congressional District Map for Census Year 1880 29
1.3 Congressional District Map for Census Year 1890 30
1.4 Congressional District Map for Census Year 1900 30
1.5 Number of Exams per Year 31
1.6 District Vote Share and Pensions for 1873-sample 31
1.7 District Vote Share and Pensions for 1893-sample 32
1.8 Frequency of Exams per year for all veterans 32
1.9 Frequency of First Exams per year for those not yet on rolls 33
1.10 Frequency of Exams per year for those already on the rolls 33

2.1 Number of Exams per Year for Black Veterans 56
2.2 Number of First Exams per Year for Black Veterans 56
2.3 Average Monthly Pension for Black Veterans 60

3.1 Fall in price of non-food p_{nf} (or rise in non-food taste/quality/variety p_f) 99
3.2 Fall in energy requirements $\bar{C}$ 100
3.3 Kernel density estimation of log calories per capita 101
3.4 Locally weighted regression of log per capita calories on log per capita real expenditure (two adult, 3 child household) 101
3.5 Locally weighted regressions for two adult, three child households (logs) 102
3.6 Locally weighted regression of budget share on log per capita real expenditure (two adult, 3 child household) 103
3.7 Food group shares by year/sector 104
3.8 Life-cycle calorie consumption and requirements, relative to 19-22 year old adults 105
3.9 Life-cycle coefficients, relative to 19-22 year olds 106
3.10 Effect of average adult years of schooling on calorie and food consumption/requirements 107

List of Tables

1.1 Summary Statistics for Union Army Veterans 27
1.2 Probit Regression - Predicting Veteran's Entry onto Pension Rolls . . 28
1.3 First Stage Results 34
1.4 Hazard Regression Estimates of the Onset of Respiratory and Digestive Conditions 35
1.5 Hazard Regression Estimates of the Onset of Infectious, Cardiovascular, and Endocrine 36
1.6 Hazard Regression Estimates of the Onset of Hernia and Genito-urinary Conditions 37
1.7 Hazard Estimates for Mortality 37
1.8 Regression Discontinuity - Law of 1907 38
1.9 Logit Model of Migration 39
1.10 Mortality by County: 1870 and 1880 40

2.1 Summary Statistics for Black Union Army Veterans 57
2.2 Probit Regression - Predicting Veteran's Entry onto Pension Rolls . . 58
2.3 OLS Estimates of Pension Receipt for Blacks on Rolls by 1893 59
2.4 Hazard Estimates for Mortality for Black and White Veterans on Pension Rolls by 1893 61

3.1 Estimates of mean per capita calorie consumption in India 86
3.2 Ratio of calories per rupee of expenditure, relative to rice and wheat . 87
3.3 Ratios of mean real expenditures and calories per real expenditure relative to 1983 rural sector 88
3.4 Comparison of Time Use Survey and NSS 89
3.5 Contributions to household caloric intake and requirements relative to male adult 90
3.6 OLS - effects of household size on caloric intake and requirements . . 90
3.7 Coefficients on household type, occupation and education dummies . 91
3.8 Coefficients on other NSS household variables 92
3.9 Means of household characteristics by year and sector 93

3.10 Coefficients on urban dummy for caloric intake and requirements, 1998-1999 TUS and 1999-2000 NSS 94
3.11 Minutes per day on various activities, by sector and gender 95
3.12 Coefficients on sector/year dummies for caloric intake 96
3.13 Poverty headcounts 97
3.14 Calorie-based poverty headcounts (by occupation) for Madhya Pradesh 97
3.15 Welfare gains/losses from caloric requirements 98

A.1 Problem foods for calorie imputation 122
A.2 Cooked meals, meals to other households and free meals (per 30 days) 123
A.3 Daily calories per person: different imputations 124

Acknowledgments

I would like to thank the members of my committee - Brad, Robin, and especially my advisor Barry - for intellectually challenging discussions, encouragement, and spot-on advice. I am also thankful to Dora Costa for several suggestions which substantially improved the first chapter of the dissertation. I thank Trevon Logan for his contribution to the second chapter of this dissertation and insightful comments on the first chapter. I thank Martha Olney for her encouragement and research advice. I also benefitted from the comments of Hoyt Bleakley, Joseph Burton, Louis Cain, David Card, Robert Fogel, Suresh Naidu, James Powell, Christina Romer, and Carlos Villareal. I thank Noelle Yetter at the Center for Population Economics for her advice and assistance with the Union Army sample. I gratefully acknowledge the financial support of the NICHD grant T32-HD007275, NIA grant T32-AG000246, the Center for Population Economics, the All-UC Group in Economic History, and the Institute for Business Economics Research at U.C Berkeley. Last but not least, I thank my husband and best friend Nicholas Li for his intellectual and emotional support as well as his fine sense of humor and good cheer throughout the dissertation-writing process.

Chapter 1

Wealth is Health: Pensions and Disease Onset in the Gilded Age

1.1 Introduction

At the end of the nineteenth and early twentieth century, crude death rates from infectious illnesses fell dramatically in the U.S. Between 1900 and 1930, the crude death rate from tuberculosis dropped by 63% and that of influenza and pneumonia fell by 82%.[1] The dramatic decreases in mortality from infectious illnesses were followed by large gains to life expectancy and decreases in death rates for all age groups. Between 1900 and 1925, the infant mortality rate fell by 53%, and the death rate for those aged 45 to 54 fell by 15%.[2]

Much of the recent work on the U.S. mortality transition between 1870 and 1930 attribute the sharp decrease in death rates from infectious illnesses to the emergence of public health interventions in the period. With the acceptance of the germ theory of disease during the 1880s and 1890s, macro public health initiatives to improve sanitation, construct sewer systems, and chlorinate or filter municipal water sources have been credited as the drivers of the large declines in mortality. Cutler and Miller [2005] find that the purification of water in U.S. cities explains half of the mortality reduction in the early twentieth century. Similarly, Ferrie and Troesken [2005] find that one-third to one-half of the 60% mortality reduction in Chicago between 1850 and 1925 can be explained by the city's efforts to purify water. While these findings shed light on the mortality transition experienced in cities, they do not explain significant drops in morbidity and mortality outside of urban areas.

Besides improvements in public health, the unprecedented advances in medical science during the early twentieth century also play a role in reducing morbidity

[1] (*Historical Statistics*, U.S. Census Bureau [1960], p. 26).
[2] (*Historical Statistics*, U.S. Census Bureau [1960], p. 29).

and mortality. While there were few effective medical treatments in the late 1800s,[3] vaccination against a wide range of infectious illnesses including diphtheria, pertussis, tuberculosis, tetanus and yellow fever were discovered shortly after 1920. During the 1930s, sulfa drugs were developed and widely used to treat infection. Jayachandran et al. [2010] find that sulfa drugs accounted for 2% to 4% of the mortality reduction between 1937 and 1943. McKinlay and McKinlay [1977] posit that medical measures account for approximately 3.5% of the decline in mortality from infectious disease between 1900 and the mid-1970s.

Lastly, improvements in socioeconomic status stemming from increases in individual income greatly contributed to the U.S. mortality transition at the end of the last century. Increased income in this period led to a rise in retirement rates (Costa [1995]) and allowed individuals to exit hazardous work environments. Increases in income in this period also allowed individuals to alter their living arrangements and reduce crowding (Costa [1997]), which decreased prevalence rates of contagious illnesses. In addition, rising income levels led to improved nutrition levels. McKeown [1976] argued that improvements in nutrition explain the majority of the mortality decline before 1900 in the industrialized world.[4] Similarly, Fogel [2004] shows that improvements in socioeconomic status over time led to increases in net nutritional status, body size and longevity. However, Fogel emphasizes that increases in socioeconomic status come not just in the form of increased wages but also in the form of medical care to the individual as well as public health interventions. Fogel, therefore, highlights the difficulty in disentangling the effect of income on health as opposed to the effects of public health interventions or advancements in medical science. By estimating the effect of income on health status at a time when effective medical care for the individual is largely absent, and by finding an exogenous source of variation in income which does not operate through public health interventions, I disentangle confounding effects to estimate the impact of income on health.

In this analysis, I investigate the contribution of increases in individual income to the decrease in adult morbidity and mortality by examining the cohort of Union Army soldiers who enlisted in the American Civil War. There are several advantages to using the Union Army dataset to examine my research question. First, the data set includes demographic and socioeconomic information for veterans until death. Second, the data include detailed information regarding a veteran's health that is not self-reported but rather recorded by surgeons employed by the Pension Bureau. Third, Union Army veterans received sizable and varying pension payments from the U.S. government which were documented from the time of a soldier's enlistment until his death. Finally, Union Army veterans are a representative sample of Northern

[3]Variolation against smallpox was one of the few effective treatments available in the U.S. beginning in the early eighteenth century (Cutler et al. [2006], p. 103).

[4]With specific regard to declining rates of tuberculosis, McKeown [1976] asserts that the 80% decrease in the rate of tuberculosis, which transpired before the advent of any effective treatments for the illness, was likely the result of improved nutrition in the population.

white men of their birth cohort.[5]

One drawback of using the Union Army data set in this analysis is that veterans received pensions for war-related illnesses and disabilities before 1890, and for disabilities unrelated to the war experience after 1890.[6] In order to circumvent the problem of reverse causality, the case in which an individual's health could influence his income, I utilize two sources of exogenous variation in pension income: the presence of close elections for Republican Congressional candidates in the individual's district and changes in pension legislation. As I show, there is evidence that Republican Congressional candidates boosted the pension income of veterans in an effort to secure votes in elections. Another contribution of my research is thus to highlight the pervasiveness of corruption in 19th century politics and show how it helped shape the largest entitlement program the country had ever seen.

I first use a proportional hazard model to estimate the effect of monthly pension income on the probability of contracting or dying from a specific illness from 1873 to 1906. I restrict this first part of the analysis to 1906 because that is the final year before the Pension Bureau began administering pensions based on age as opposed to disability. Estimating income semi-elasticities and proportional hazards for a variety of diseases, I find that respiratory, infectious, digestive, and cardiovascular conditions are significantly responsive to income. For example, I find that an extra dollar of monthly pension income, which is equivalent to 4% of a farm laborer's monthly earnings in 1900[7] or a 9% increase in the average veteran's monthly pension income, decreases the likelihood of developing an infectious illness by an average of 38%, reducing the hazard rate by .0015.

For the period after 1906, when the Bureau began dispensing pensions based on the age of veterans, I use a regression discontinuity framework to test whether older cohorts had a differing chance of disease onset relative to younger cohorts. I find strong effects for respiratory illnesses and some cardiovascular illnesses. Besides exploiting the change in pension legislation, a second benefit of using the regression discontinuity framework in the period after 1906 is that I am able to identify effects based on variation between adjacent birth-year cohorts during a period of growth in public health interventions and emergence of more modern medical treatments.

This paper is organized as follows. Section 2 contains an historical background of the Union Army pension system and the political environment which shaped its evolution. Section 2 also contains a brief account of the medical knowledge and advancements during the lifetime of the Union Army cohort. Section 3 provides the econometric framework. Section 4 describes the data and provides summary statistics. Section 5 contains results, and Section 6 contains robustness checks. Section 7 concludes.

[5] See section 4.1.

[6] After 1890, veterans could still receive pensions for war-related illnesses and disabilities.

[7] Preston and Haines [1991], p. 212-20.)

1.2 Historical Background

1.2.1 The Union Army Pension in the Patronage Era

The U.S. political climate at the end of the 19th century was defined by protectionist tariffs, an unruly patronage system and, with the exception of the Cleveland administrations, Republican domination of the White House. Republicans and Democrats quarreled over how to spend the budget surplus which resulted from the high tariffs, and the proper way of reforming the patronage system. In this intensely corrupt political environment, the Union Army pension system, initiated in 1862 to provide assistance to veterans suffering from war-related disabilities, evolved into a large-scale entitlement program. The pension system was the Republican answer for how to spend the surplus and, as benefits increased, served as a way for Republicans to keep Civil War veterans faithful to the party.

Passage of the General Law of 1862

During the Civil War, 2.7 million men enlisted, and 87% of recruits survived the war.[8] These veterans could qualify for public assistance. Under the General Law passed on July 14, 1862, veterans became eligible for a pension if their physical disabilities were shown to result from the war experience. Therefore, the General Law excluded veterans whose disabilities resulted from aging or any incidents unrelated to the war. Disabilities pensioned under the General Law included loss of sight, hearing, limbs or their usage, and illnesses caught while in the camps. In order to determine whether a causal link existed between illnesses/injuries and war experience, veterans were required to undergo medical examinations by physicians, called examining surgeons, hired by the Pension Bureau. Upon completion of each exam, examining surgeons filled out a surgeon's certificate, which was then sent to the Pension Bureau and added to the veteran's application for pensions.

The Pension Office: A Political Machine

During the 1870s and 1880s, many Republican candidates began boosting pensions awarded to veterans in their constituency in an effort to secure the vote of the old soldier. Evidence of this practice is abundant. While all of the correspondence of the Pension Bureau between 1870 and 1914 is suspiciously missing from the documents of the Pension Bureau found at the National Archives, the correspondence (letters and telegrams) of the Commissioner of Pensions can still be found in texts outlining corruption at the Bureau. Perhaps the best surviving collection of evidence regarding Republican corruption at the Pension Bureau comes from a 104-page booklet *Republican Abuses at the Pension Bureau* put out by the Democratic party before the

[8] Dyer [1959]

election of 1888 (National Democratic Committee [1888]). Similarly, *The Campaign Book of the Democratic Party, 1886*, also provides evidence of Republican corruption at the Pension Bureau and more concretely outlines the methods by which Congressional candidates were able to augment the pensions of their constituents (Polkinhorn [1886]). In these two works, Democrats published private letters between the Commissioner of Pensions and Congressional candidates, testimony from pensioners as well as clerks from the Bureau, and data regarding the party affiliations of all clerks hired by the Bureau during the 1870s and 1880s in order to outline the pervasiveness of the corruption. Much of the evidence provided by the Democratic Party in these two texts, that which is included in this research, is drawn directly from the transcriptions of testimony and evidence given at the 48th Congress' session of the Committee on the Payment of Pensions, Bounty and Back Pay.

In order to secure the votes of Union Army veterans, Republican Congressional candidates wrote letters to the Commissioner of the Pension Bureau on behalf of constituents who had previously applied for pensions. For cases in which a veteran was still waiting for his application to be processed by the Bureau, the candidate asked that the Commissioner speedily review the case and send special examiners to the district if there were remaining questions about a veteran's claim. For veterans already receiving pensions and having applied for an increase, candidates requested that the Commissioner either increase the pensions or that he similarly send a special examiner to the district to reassess the case. In their letters, Republican candidates frequently were found expressly stating that they were interested in boosting pensions in order to win the upcoming election.[9]

The Commissioner, who hired mostly Republican clerks[10] at the Bureau, gave priority to the processing of the claims of veterans in the Republican candidate's district, especially during election season. Clerks were not in charge of deciding the amount of pension paid to veterans. Instead, the Medical Reviewers of the Pension Bureau decided the dollar amount.[11] Clerks had the power to stall the application of

[9]Many examples of letters written from Congressmen to the Commissioner of the Pension Bureau can be found in *Republican Abuses at the Pension Bureau* (National Democratic Committee [1888]). The following is one such example. On August 4, 1884, the Republican Congressman from Ohio wrote the following to Commissioner Dudley: "My dear sir: I am just entering upon my canvass for Congress with a first-rate fighting chance. I know you will render me all the aid which you properly can. Mr. H. H. Wallace, special examiner, just sent to Springfield, Illinois, is one of the best 'button-holers.' If he could be adjacent to Butler and Preble counties, he would be of the very greatest assistance to me. Can't you put him just over the line in Indiana? He is thoroughly well acquainted in Butler and Preble and can do me great good, and he will be zealous to do so. He does not know of me writing this letter, and I write it solely in the interest of my canvass and of a Republican House of Representatives." p. 31.

[10]In 1886, for example, there were 1,472 Republican clerks employed by the Pension Bureau as opposed to only 28 Democratic clerks (*The Campaign Book of the Democratic Party, 1886*, Polkinhorn [1886], p. 7).

[11](*Instructions to the Examining Surgeons*, U.S. Pension Bureau [1882], p. 9)

veterans during election seasons to prevent the loss of votes to Democrats.[12] Therefore the Pension Office remained a Republican political machine used to secure the veteran vote throughout the patronage era.

Democratic candidates for Congress were largely unable to boost the pension payments of their veteran constituents by writing letters to the Commissioner because of the partisan nature of the Bureau during this period. Additionally, there is evidence that veterans supporting Democratic candidates were threatened with the loss of their pension or with having their applications stalled indefinitely by special examiners sent from Washington into districts during election season to aid the Republican cause.[13]

The corrupt practices of the Pension Bureau continued through the late 1880s and into the 1890s. During the election of 1888, Civil War veteran and New York commander of the Grand Army of the Republic James Tanner "had, by his own account, 'plastered Indiana with promises' of more generous pensions under the Republicans. 'God help the surplus revenue!' the new commissioner declared, as he set about handing out new and readjusted pensions with gusto." Once Harrison took office, Tanner was appointed Commissioner of the Pension Bureau.[14] Republican congressmen continued to easily augment pensions by specially appointing clerks, who were loyal to the party, to review the red-taped[15] pension applications.

[12]In 1885, during the session of the Committee on the Payment of Pensions, Bounty and Back Pay (H.D. 2683, 48th Congress, 2d Session), several clerks employed by the Pension Office were sworn in to testify about the ways in which pensions were adjudicated. In the testimony of these clerks, each one says that he was informed by the Commissioner not to reject any cases whatsoever before the election of 1884 for "political purposes." One clerk in particular testified that when he asked the assistant-chief (his superior at the Bureau) about rejecting a specific case, he was told "No no no; we are not rejecting any cases now. If you reject that case it will lose us six votes" (*The Campaign book of the Democratic Party, 1886*, Polkinhorn [1886], p. 45-48).

[13]There are a number of examples beginning on page 49 of *The Campaign Book of the Democratic Party, 1886* (Polkinhorn [1886]). The following is one example of testimony from a veteran: "John M. Mattingly, being first sworn, deposes as follow: My name is John M. Mattingly; I have resided at Olney, Ills., for the past four years; I am forty-two years of age, and a laborer by occupation. During the late war I was a private in Company A, Tenth Kentucky Infantry; I am an applicant for pension, and my claim has been pending since the year 1879. During the political contest of 1884, which resulted in the election of Grover Cleveland, my case was in the hands of Special Examiner Epert for the purpose of examination. When he came to me to take my preliminary statement, he asked me my politics; when I told him I was a Democrat, he told me I was on the wrong side to ever get a pension: that I never would get a pension as long as I voted the Democratic ticket, but that if I voted the Republican ticket my claim for pension would be allowed" (p. 50).

[14]Skocpol [1992], p. 128.

[15]Pension applications arrived to the Department of the Interior (and after 1884, to the Pension Bureau) wrapped in red twill tape. The origins of the phrase 'red tape,' describing excessive bureaucratic regulations that slow or halt decision-making, come from the processes of the Union Army Pension System.

The Passage of the Invalid Pensions Act of 1890 and Age-Based Laws of 1907, 1912 & 1918

Later, under the Invalid Pensions Act of June 27, 1890,[16] which awarded pensions to veterans with disabilities from causes other than the war, previously excluded veterans were finally able to claim pensions on the basis of non-war injuries.[17] The passage of the Invalid Pensions Act of 1890 enabled thousands of veterans to begin receiving pensions and drop the federal government from a surplus to a deficit. By 1892, expenditure on veterans compensation and pensions accounted for approximately 40% of the federal budget.[18]

In 1907, the pension system changed once again by granting payouts based on the veteran's age. Under the 1907 Act, any veteran over age 62, who was otherwise eligible for a pension under the 1890 Act, received $12 per month; if over age 65 received $15 per month; and if over age 70 received $20 per month.[19] In the years 1912 and 1918, these payouts were increased once again.[20] These pension amounts were substantial for the time and large enough to allow veterans to retire.[21] In 1900, the average pensioner received $10.02 per month, which was approximate half of the average monthly income of a farm laborer.[22] By 1890, the Union Army pension was comparable to today's Social Security Program in that approximately 42% of income is replaced by the SSA upon retirement.

1.2.2 Medicine in the Gilded Age

Medical Treatment: From the Age of Heroic Medicine to Fleming

In 1800, only two cities in the U.S. had a hospital - New York and Philadelphia.[23] By the close of the Civil War, the Union boasted over 200 hospitals, which could accommodate as many as 137,000 patients.[24] However, this rapid increase in the number of hospitals did not signify advances in medical science which could alter mortality trends. Physicians of the 1870s still relied on the same treatments used in the 1830s, a period in which few useful treatments were known.[25]

---

[16] The Invalid Pensions Act of 1890 was passed under the Harrison administration as an answer to the budget surplus debate.

[17] Glasson [1918], p. 126.

[18] *Historical Statistics of the United States, Colonial Times to 1957*, U.S. Census Bureau [1960], p.718.

[19] Linares [2001], p. 16.

[20] See Appendix A for exact pension amounts paid by the laws of 1907, 1912 and 1918.

[21] Costa [1995].

[22] Preston and Haines [1991], p. 212-20.

[23] Rosenberg [1987], p. 337.

[24] Rosenberg [1987], p. 97.

[25] One famous medical intervention was bloodletting. Bloodletting was thought to "moderate vascular excitement, reduce inflammation, relieve congestion, allay spasm and pain, relax the muscular

Until the beginning of the twentieth century, many Americans found it preferable to remain at home instead of seeking treatment from private physicians or being admitted to a hospital when they fell ill. Since the primary function of hospitals in this period was to provide patients with a warm, clean place to rest and eat well, all of which income could buy, those of the middle and upper classes, unlike the poor, seldom sought care in hospitals during this period.[26]

Public opinion regarding the efficacy of hospital care for the middle and upper classes changed beginning in the 1890s. With the invention of the x-ray in 1895, the wide acceptance of germ theory by the 1890s, and the use of antiseptics, hospitals made great advances in surgery and boasted dramatically lowered death rates from post-operative infections. By the 1920s, the public viewed hospitals as institutions with physicians able to perform surgery effectively as opposed to the nineteenth century view of the hospital as an ineffective and unsanitary shelter for the disabled poor to convalesce. However, even during the 1920s, hospitals could do little to help non-surgical patients such as those suffering from chronic illnesses.[27] The public's reluctance to seek medical attention from physicians or in hospitals continued until the early 20th century when the germ theory of disease gained wide acceptance and when the research on life-saving vaccines began.

Medical Education and Licensing

In 1813, there were only seven medical colleges in the United States[28] and during the early nineteenth century medical education can be described as poor, at best.[29]

---

system, and arrest hemorrhage" (Kaufman [1976], p. 59). There are instances in which bloodletting can be beneficial according to our understanding of modern medicine. In the case of pulmonary edema, which is an accumulation of fluid in the alveoli of the lungs and is frequently due to heart failure, bloodletting decreases blood pressure and helps fluid drain from the alveoli so that the patient can breathe more easily. Today, diuretics are given which achieve the same outcome by draining fluid from the alveoli. Despite a few cases in which bloodletting is beneficial, the majority of the time it did more harm than good.

A second treatment widely used by physicians of the era was the administration of calomel, which is a diuretic and purgative composed of mercury chloride. Physicians gave their patients calomel for cholera as well as a wide variety of other illnesses and with the usual result of the patient's death from mercury poisoning (Kaufman [1976], p. 59-62).

Since the majority of treatments of physicians were ineffectual and frequently seen as worse than the disease itself, many Americans avoided doctors and sought out self-treatment alternatives. In addition, the emergence of homeopaths and eclectics, who were seen by orthodox medical practitioners as rival groups, helped solidify the common beliefs that physicians administered frightening allopathic treatments with little beneficial results. (Kaufman [1976], p. 62.)

[26]Rosenberg [1987], p. 115-116.

[27]Rosenberg [1987], p. 246.

[28]Kaufman [1976], p. 36.

[29]In response to the growing demand for medical education, medical colleges opened in rural areas. However, since professors at medical schools made their income from students' tuition payments, they were incentivized to admit any student wishing to become a doctor regardless of his academic

With the emergence of the Association of American Medical Colleges and the American Medical Association's Council on Medical Education and an improvement of the public's opinion of traditional physicians during the second half of the 19th century, medical education underwent a reformation during which the requirements for the certification of medical colleges and physicians became stringent.[30]

It was during this era of reform that the Pension Office began hiring examining surgeons to diagnose the ailments resulting in the disability of Union Army veterans. All examining surgeons hired by the Pension Office graduated from U.S. medical schools and were required to have practiced medicine for a minimum of five years prior to their hiring.[31] Examining surgeons were therefore subject to the new licensing laws enacted by states and therefore likely to be quite knowledgeable relative to the other physicians of their time.

Examining surgeons were hired by the Pension Bureau to accurately diagnose and report health conditions and disabilities faced by Union Army veterans, which would prevent veterans from performing manual labor.[32] For diseases which were easily diagnosed without the use of imaging tools such as the x-ray, not invented until 1895, examining surgeons in different parts of the U.S. uniformly classified and named the veterans' illnesses.[33]

---

abilities, (Kaufman [1976], p. 52-53) which lowered the quality of instruction and the curriculum. During this period, students were not required to partake in dissections, study anatomy or undertake clinical work (Kaufman [1976], p. 106). In addition, medical colleges possessed the authority to license physicians to practice, as opposed to independent state medical societies, (Kaufman [1976], p. 96) which lowered the quality of physicians with medical degrees even further.

[30]During the 1870s and 1880s, states mandated that each practicing physician be licensed. In order to receive a license, physicians frequently had to prove they had received a diploma from a college that adhered to the standards put forth by the state's medical board. The requirements for the certification of colleges were more rigorous. In the case of Illinois, medical colleges required students to attend a least three years of instruction. During these three years of instruction, students were required to spend two years gaining clinical and hospital experience as well as engage in dissections (Kaufman [1976], p. 132). These new licensing regulations enforced by individual state boards greatly reduced the number of "quacks" masquerading as experienced physicians and resulted in the closure of many medical colleges which were little more than "diploma mills" (Kaufman [1976], p. 131-132).

[31]US Pension Bureau, Report of the Commissioner of Pensions, June 30 1891,19.

[32](*Instructions to Examining Surgeons*, U.S. Pension Bureau [1882], p. 6).

[33]International Classification of Diseases (ICD) is the standard by which health conditions are registered in the United States and worldwide. The first list of such diseases was composed in 1900. Since then there have been 10 revisions to the list. However, the concept of a standardized list of diseases was first introduced, even earlier, by the British epidemiologist William Farr in 1837. In later years, he advocated for a complete list of diseases grouped by anatomical site and suggested that it be adopted internationally. Farr stressed the uniformity of disease classification across countries. A variant ICD of today was available in an early form by the 1850s. Based on the lists composed by William Farr, which were precursors to the ICD list, doctors made uniform diagnoses. Therefore, diagnoses made by examining surgeons employed by the Pension Bureau, would match diagnoses made by today's physicians for illnesses which do not require imaging tools but instead require an analysis of symptoms. A brief history of the International Classification of Disease can be found at

Further insight into the diagnostic skills and treatment ideas of the examining surgeons hired by the Pension Bureau to assess the illnesses of Union Army veterans can be found in their annual journal *Transactions*. Beginning in 1902, examining surgeons gathered in a different U.S. city each year to discuss diagnostic methods and causal factors leading to the diseases they diagnosed in veterans and in patients of their private practices. While many of the causal links they inferred have been proven inaccurate over the last 100 years, their diagnostic approach closely mirrors many of the practices of doctors today.

1.2.3 Public Health Interventions

While public health interventions were largely absent in rural areas, attempts at securing cleaner sources of water, constructing of sewers, and removing of garbage and other waste had begun in many major cities by the middle of the nineteenth century. Whereas in 1870 access to filtered water was rare in urban America,[34] by 1900, 1.86 million urban Americans had access to filtered water.[35] Besides the introduction of clean water supplies and efforts at waste removal, other public health interventions included the use of quarantine, urban sanitation projects, efforts to secure clean milk and new legislation to reduce air pollution.[36]

The introduction of public health improvements during the period of analysis is not a concern for this study unless the timing and location of these improvements is influenced by the election of Republican Congressmen to office. If this were the case, the exclusion restriction – that close elections can only affect the health of veterans through pension income – would be violated. However, the evidence suggests that public health improvements in cities were enacted by local governments and that the timing of specific projects such as water chlorination were largely arbitrary.[37]

1.3 Empirical Strategy

I divide this analysis into two parts: the period before the enactment of age-based laws from 1865 to 1906 and the period from 1907 to 1935. The reason I divide the analysis is because the Pension Bureau began awarding pensions based on age in the years after 1906, and thus there was little need for new exams.[38] Information

www.who.int/classifications/icd/en/HistoryOfICD.pdf.

[34]Haines [2001], p. 13.

[35]ibid.

[36]ibid.

[37]Cutler and Miller [2005], p. 5-6.

[38]Veterans applying for new pensions or pension increases under the Invalid Pensions Act of 1890 were still required to undergo exams.

about disease onset between the years 1907 and 1920[39] comes primarily from death certificates.

1.3.1 Part I

Identification

Prior to 1890, veterans could only apply under the Law of 1862, which granted pensions based on illnesses and injuries related to the war experience. Therefore, a veteran who had a chronic illness between the years 1862 and 1890 was not likely to apply for a new pension or pension increase if he felt that his illness could not be traced to the war experience during a medical exam. Instead, the veteran was likely to only report his non-war-related condition after the passage of the Law of 1890, which granted pensions to veterans facing any debilitating illness.[40] Figure 1.5 shows the total number of exams undergone by veterans applying for new pensions and pension increases between the years 1865 and 1935. Given the dramatic spike in the number of exams shortly after 1890, it is unlikely that over 4,000 veterans suddenly contracted pensionable illnesses between 1890 and 1892. Instead, veterans were most likely reporting illnesses they faced during the 1870s and 1880s, which they did not report because the illnesses were not pensionable under the Law of 1862. Because veterans underreported non-war-related illnesses during the 1870s and 1880s, I divide the sample of veterans into two parts. The first sub-sample consists of those who first applied for a pension under the Law of 1862 by the year 1873. The second sub-sample consists of those who first applied by 1893 under the Law of 1862 or the Law of 1890. By restricting the second sub-sample to those who applied by 1893, I ignore first illnesses reported by the group of veterans who applied under the Law of 1890 (between 1890 and 1892) because these first illnesses were likely to be reported long after the veteran faced initial symptoms of the illness. I then estimate the effect of pension income on subsequent morbidity and mortality conditions. Finally, I restrict both subsamples to those who were admitted onto the pension rolls.[41]

Under the pension acts prior to 1907, veterans could only receive pensions based

[39]Figure 1.5 shows the drop in the number of exams after 1906. Veterans could still apply for pensions under the preceding laws, such as the General Law of 1862 or the Invalid Pensions Act of 1890. However, practically all those eligible to collect pensions under the General Law of 1862 had already applied by 1907. Therefore, new exams were largely unnecessary.

[40]Illnesses thought to be the result of vicious habits, such as syphilis, were not pensionable under the Law of 1890.

[41]Date of death and cause of death information is largely unavailable for veterans who were not on the pension rolls. Including veterans who did not collect pensions in the sample would bias estimates. This is because there is no disease information for veterans who were not on the pension rolls. If veterans who did not collect pensions died before having the opportunity to claim pensions, their inclusion in the sample will bias estimates. Therefore, I restrict the sample to veterans who applied by the baseline year (1873 or 1893) and who were admitted onto the pension rolls by the baseline year.

on health status. Because of this fact, pension income is endogenous to health outcomes. Since veterans receiving larger pensions were likely to be in worse health, and likely to have continually deteriorating health, relative to those who received smaller pensions, it is difficult to identify the true effects of income on health without finding an exogenous source of variation in pension income. In addition, it is likely that estimates of the effect of pension income on health will be biased downward causing it to seem as though pensions increase the chance of sickness in veterans. However, in actuality, it may simply be that poor health status initially has overshadowed the effects of income. While controlling for prior health status partly controls for this effect, it is preferable to use an exogenous source of variation in pension income. I use variation in Republican Congressional vote share.

My measure of close elections for Republican Congressional candidates is an indicator variable equal to 1 if Republican Congressional vote share is between 0.3 and 0.7 inclusive. Republican Congressional candidates and incumbents were more likely to boost pensions in close elections. In Figure 1.6 to Figure 1.7, the graphs of the relationship between median pension income and Republican Congressional vote share are shown for the 1873-sample in 1888 and the 1893-sample in 1900. From the figures, it is clear that median pension income rises significantly if elections are in the close range. The positive relationship between close elections and pension income remains over the span of all three decades between 1870 and 1900.

In order to use the presence of close elections for Republican Congressional candidates in a pensioner's district as an exogenous source of variation in pension income, the presence of close elections must satisfy the following exclusion restriction: Conditional on my other control variables, close elections can only affect health through their effect on pension income. Three additional assumptions are required. First, close elections must have a quantitatively and statistically significant effect on pension income. Second, individual health status cannot affect the presence of close elections in a district. One example of this might be if sick veterans moved to areas of close elections for Republican Congressional candidates.[42]

Third, it cannot be the case that the likelihood to apply for a pension (and then undergo exams) varies between close election districts and non-close election districts. Figure 1.8 shows the frequency of exams per year for all veterans in the sample. With the exception of the year 1890, there is little difference between non-close and close election districts between 1860 and 1907. Figure 1.9 shows the frequency of first exams per year for veterans not yet on the pensions rolls. Prior to 1890, there is little difference in the frequency of exams between veterans in non-close election districts as opposed to those in close election districts. In 1890, veterans in non-close election districts were approximately 20% more likely to undergo first exams to apply for pensions. Veterans who first applied between 1890-1891 would be excluded from the 1873-sample but would be included in the 1893-sample. After 1893, those in non-close

[42]I show that this is not the case in Section 6.1.1.

election districts undergo first exams at a slightly higher frequency than those in close election districts. However, veterans admitted onto the pension rolls after 1893 are not included in either the 1873 or 1893-subsample and are entirely excluded from this analysis. Figure 1.10 shows the frequency of exams per year for those already on the pension rolls. For those in the 1873-sample, the large biannual spikes in the frequency of exams during the 1870s reflect the requirement that veterans on the rolls return for biannual exams.[43] However, there is virtually no difference in the frequency of applications during the 1870s. With the exception of the mid-1880s and early 1890s, there is little difference in the frequency of applications between close and non-close election districts.

Controls in this section include rank while in the army, birth year, an indicator equal to 1 if the veteran lived in one of the hundred most populated cities, average farm land value per acre, prior health status, prisoner-of-war status and battle wounds. Each of these can affect income. High-ranking veterans received higher pensions than their lower-ranked counterparts. It is important to control for birth year because incidence of disease increases with age. County-level average farm land values per acre are used to control for increasing wealth in the county.

Information about disease comes from the surgeon's certificates. The empirical strategy of this research hinges on the validity of the statements testifying to the diseases faced by veterans on these certificates. It does not affect this research if examining surgeons embellished the severity of illnesses faced by veterans because I am testing the effect of pension income on the *first* incidence of the disease. In data cleaning, disease conditions preceded by adjectives such as "severe" or "acute" were ignored as were explicit descriptions. For example, if a surgeon states that a veteran has "severe" asthma and had "prolonged inspiration, shortened expiration, and whistling heard all over the lungs" during an exam, as was the case for veteran Henry Greeley during his first exam in 1888, the dummy variable for asthma would be 1 for the year 1888 until Greeley's death. Alternatively, if the examining surgeon had simply described Greeley's condition as "asthma caught at camp Banks at Georgetown," the dummy variable would still equal 1. Therefore, the asthma variable is simply a dummy equal to one if the surgeon diagnoses the veteran with asthma, using the exact word "asthma," and zero otherwise.

It is, however, imperative that surgeons report the correct disease conditions as opposed to fabricating a story of illness when the veteran is perfectly healthy. If surgeons made honest mistakes when diagnosing patients, this would just create noise in the dependent variable but would not bias coefficient estimates. However, if surgeons diagnosed veterans with fictitious illnesses when in fact they were perfectly healthy, then this would bias income effects downward because otherwise healthy veterans with high pension incomes would look sick.

[43] In figure 1.9, there are no spikes in the frequency of the exams in the 1870s because this figure shows the frequency of *first* exams.

In addition to using specific disease outcomes in the analysis, I also group the illnesses into one of the following six categories: respiratory, digestive, infectious, cardiovascular, endocrine and genito-urinary.[44]

While physicians may have exaggerated the severity of illnesses faced by veterans in an effort to secure higher pensions for veterans, these exaggerations did not lead to increases in pension amounts. According to the *Instructions to Examining Surgeons* a pamphlet circulated to all examining surgeons,[45] physicians were instructed to rate the severity of the illness as opposed to stating opinions as to how much money a veteran deserved based on his illness.[46] Instead, it was the clerks at the pension bureau who evaluated the applications and determined pension amounts. Surgeons' ratings and pension payments are largely uncorrelated. In fact, for many years, it is not possible to link increases in pension payments to exams precisely because pensions were frequently increased as a favor for party loyalty as opposed to just the incidence of an illness. Therefore, veterans had every incentive to report their medical conditions but their pensions depended on a combination of their illnesses and political corruption in their district.

Econometric Framework

In this research, I use a Weibull proportional hazard model to predict the effect of an extra dollar of monthly pension income[47] on the probability of disease onset (or

[44]Respiratory illnesses include abscess, adhesion, allergy, asthma, atelectasis, bronchiectasis, bronchitis, cavity, edema, effusion, emphysema, empyema, fibrosis, hemoptysis, pleuritis, pneumonia, pneumonitis, pneumothorax, tracheitis, and tuberculosis. Digestive illnesses include constipation, diarrhea, dysentery, dyspepsia, dysphagia, enlarged liver, gallstones, gastroenteritis, malassimilation (malabsorption) and nausea. Infectious illnesses include chancroid, cholera, dengue, diphtheria, gonorrhea, hepatitis, influenza, malaria, meningitis, mumps, orchitis, parotiditis, pertussis, rickettsia, rubeola, salmonella, scarlet fever, septicemia, streptococcus, syphilis, tetanus, typomalaria, and variola. Cardiovascular illnesses include arteriosclerosis, cyanosis, dyspnea, enlarged heart, impaired circulation, murmur, cardiac edema, and palpitation. Endocrine illnesses include diabetes, goiter and enlarged spleen. Genito-urinary illnesses include cystitis, enlarged prostate, nephritis, urethral obstruction, and uremia.

[45]*Instructions to Examining Surgeons* were pamphlets put out by the Pension Bureau and frequently updated to reflect changes in the pension law (U.S. Pension Bureau [1882]). Instructions relevant to Civil War veterans are known to have been published in the years 1871, 1877, 1884, 1891, 1895, 1901, 1905, and 1908, and most of these copies can be found at the Regenstein Library at the University of Chicago. There is evidence that other instructions to surgeons may exist between the years 1871 and 1908, however, at the time of this research, additional pamphlets have not been found.

[46]An example can be found in the *Instructions* published in 1884, p. 8: "All disability which cause a lesser degree of disability than that resulting from anchylosis [the stiffening of joints] of the wrist or ankle should be determined by comparison with anchylosis of the wrist or ankle, and stated by fractions of totals, as 1/8, 1/4, 3/8, 1/2, 5/8, 3/4, or 7/8 of total." Therefore, examining surgeons *rated* disability but did not determine the payouts awarded to veterans.

[47]An extra dollar of monthly pension income is approximately a 4% increase in the income of an average farmer who makes $24 a month in 1900.

mortality). The dependent variable in each regression is an indicator equal to 1 if a veteran faced a particular condition in year t (or died from a particular condition) and 0 otherwise.

The model assumes that a veteran would eventually suffer from the health condition if he lived long enough. If a veteran died from a particular condition, the indicator variable for that condition will equal 1 in the year of the veteran's death. The dataset is expanded to a balanced panel spanning the years 1865 to 1906. If a veteran's death year came before 1907, observations after his death year were dropped from the sample. Since the independent variable, monthly pension income, is not exogenous to health status in all years, the presence of close elections in a pensioner's district is used as exogenous sources of income. However, because the model is non-linear, an instrumental variables approach cannot be used.[48] Instead, I use the control function approach in which the monthly pension income and control variables as well as a cubic of the residual from the first stage is included in the hazard regression. The first-stage takes on the following form:

$$pen_{it} = \gamma_0 + \gamma_1 vote_{it} + \Gamma C_{it} + u_{it} \tag{1.1}$$

where pen = monthly pension income, $vote$ = dummy variable equal to 1 if Republican congressional vote share is between 0.3 and 0.7 and otherwise zero, i = individual, t = year, and C = set of demographic and socio-economic controls. Controls include birth year, battle wounds, previous health conditions, military rank, prisoner-of-war status, and county-level farm land value per acre. The associated residual has the following form:

$$\hat{u}_{it} = pen_{it} - \hat{\gamma}_0 - \hat{\gamma}_1 vote_{it} - \hat{\Gamma} C_{it} \tag{1.2}$$

In the second-stage, I choose

$$\theta = \{\beta, \alpha, \phi\}$$

to maximize the likelihood function

$$\sum_{i=1}^{N} \sum_{b=1873,1893}^{1906} \left\{ d_{it} log[e^{k_{it}\beta} \alpha t^{\alpha-1}] + (1 - d_{it})(-t^{\alpha} e^{k_{it}\beta}) \right\} \tag{1.3}$$

where

$$k_{it}\beta = \beta_0 pen_{it} + \sum_{j=1}^{N} \beta_j (C_j)_{it} + \phi_1 \hat{u}_{it} + \phi_2 \hat{u}_{it}^2 + \phi_3 \hat{u}_{it}^3 + \epsilon_{it} \tag{1.4}$$

[48] Since the second-stage is non-linear, E(u‖vote) must equal zero instead of E(u*vote)=0, which is the weaker assumption necessary for two-stage least squares. Combined with the assumption that the first-stage is linear (necessary for the CF approach), the CF approach is more efficient than two-stage least squares. More importantly, two-stage least squares is inconsistent when the model is not linear in the parameters. For a more detailed explanation of the Control Function versus Instrumental Variables Approach, see Imbens and Wooldridge [2007].

The term d_{it} is an indicator variable equal to 1 if the veteran is alive and 0 if dead or if t=1907. Veterans who have died do not contribute to the likelihood after their death although the hazard model treats them as at risk for illnesses that they have not yet had.[49]

Since the model is a proportional hazard model, it is possible to interpret the effects of a change in income on the hazard. A negative coefficient indicates that the time until onset of disease increases as income increases, which is equivalent to a negative effect on the hazard or the probability of disease onset.

For the purpose of hypothesis testing, I cluster the standard errors by district and year in the first stage. This is because the unit of observation is the individual in a particular year. However, the vote variable is the same within a district-year. Therefore, it is perfectly correlated across all individuals within a district-year.

The standard errors of the second stage do not take account of the fact that the residual from the first stage, which enters in the second stage, is estimated with error. Because of the difficulty in making the usual correction for this in a non-linear setting, I bootstrap the standard errors. However, bootstrapping is only necessary if the coefficient on the residual is significantly different from zero in the second stage.[50]

1.3.2 Part II

In the years between 1907 and 1918, the number of exams dropped significantly because the Bureau dispensed pensions based on year of birth.[51] Figure 1.5 shows the number of exams in each year from 1870 to 1930. It is immediately clear that the largest spike occurred in the 1891-1892 period coinciding with the passage of the Invalid Pensions Act of 1890. After the 1907 law, the number of exams fell rapidly and approached zero in the late 1910s. If a pensioner died in this period, the Pension Bureau received information about the veteran's cause of death and contributing causes of death. In data cleaning, disease indicator variables were set equal to one in the year of death if a veteran died of the disease or the disease was a contributing cause of death.

Using a regression discontinuity approach, I attempt to address the effects that changes in medical treatments and public health interventions may have had on different cohorts. Comparing cohorts that were born only one year apart, I assume that changes in medical treatments and public health interventions had the same effect on the health status of adjacent cohorts.

[49] Conditional on the covariates, the true time to failure is independent from the censoring event. Because I code death due to particular disease as a morbidity, this should be valid.

[50] Imbens and Wooldridge [2007], p. 9).

[51] It was still possible to collect pensions under the rules of the Invalid Pensions Act of 1890, however, most veterans would receive more income if they applied under the new age-based laws.

Identification

The details of the Laws of 1907, 1912, and 1918 can be found in Appendix A. I use the law of 1907 in the regression discontinuity analysis as opposed to all three changes in pension legislation because the law of 1907 provided the largest relative increase in pension income to veterans. Since there are few exams after 1907, I identify income effects on cause of death as opposed to the probability of onset of non-fatal diseases.

Econometric Framework

I compare the cohort born in 1844 to the cohort born in 1843. In 1907, the cohort born in 1844 was age 63 and those born in 1843 were age 64. I compare these cohorts because they were necessarily eligible for a pension increase under the Law of 1907. Assuming that an individual born in 1844 lived the same number of years as an individual born in 1843, the individual born in 1844 would receive more years of pension income under the Law of 1907. Therefore, if increased income decreases the likelihood of mortality from diseases of poverty (such as infectious diseases), then the expectation is that the effect of a pension increase on the mortality rates from diseases of poverty would be negative for the 1844 cohort. With this econometric strategy, the assumption is that, in the absence of income differences, mortality rates for the cohort born in 1843 were the same as for the cohort born in 1844.

I restrict the sample to those born in 1843 and 1844, who lived to be at least 63 years old and thus able to take advantage of the law. The first year in the sample is 1907 and the final year is 1935.[52] I use the following regression discontinuity model:

$$disease_i = \beta_0 + \beta_1 cohort1844_i + \beta_2 C_i + \epsilon_i \tag{1.5}$$

where *disease* is an indicator variable equal to 1 if the veteran ever had the disease. $cohort1844$ is an indicator variable equal to 1 if the veteran was born in 1844 and zero otherwise. Therefore, the reference group in the sample is veterans born in 1843. No other cohorts are present in the sample. In this specification, the "treatment" is receiving an extra ($12 x 12 months =) $144 of lifetime wealth before age 64 due to the change in the pension system. I test whether, conditional on surviving to late life, an extra $144 of lifetime wealth delays the onset of particular diseases.[53] Controls included in this specification are a veteran's rank in the army, his war wounds and POW status.

[52]The last veteran in the sample died in 1935.

[53]Note that because of other laws passed later, the total lifetime effect of the treatment is potentially greater than $144, but only for veterans who survive to 1912 and then to 1918.

1.4 Data

1.4.1 Union Army Sample

I study these issues using the Center for Population Economics' Union Army data set, which consists of information about the military experience, post-war health conditions, and socioeconomic status of veterans. The dataset is compiled from three sources: the military, pension, and medical records; surgeons certificates; and census records. The military records contain demographic and socioeconomic information about the soldier at the time of his enlistment. The pension and medical records provide detailed information regarding the veterans experience with the Union Army pension.

Since two-thirds of white[54] men between the ages of 20 and 45 served in the Union Army, the more than 2 million men who served were representative of the general population of white men in the North.[55] Of the men who served in the Union Army, records for 35,570 white men were compiled. Regiments in the Union Army were comprised of 10 companies. The Fogel group selected 331 white companies at random and compiled information for each of the men within a company. The data used in this analysis consist of 303 out of the 331 companies. Therefore, the 35,570 recruits span 303 companies.

In this analysis, I use two subsamples of the 35,570 recruits in the dataset. The first subsample, the 1873-sample, consists of 1,750 recruits who survived to 1873 and were already on the pension rolls in that year. The second subsample, the 1893-sample, consists of 13,296 recruits who survived to 1893 and were on the pension rolls by that year.

Table 1.1 shows summary statistics for both the 1873- and 1893-sample at baseline. For those on the rolls in 1873, 4.4% suffered from respiratory conditions, 6.91% suffered from digestive conditions and 3.43% suffered from a cardiovascular condition in 1873. For those on the rolls in 1893, 7.22% suffered from a respiratory condition, 24.75% suffered from digestive conditions, and 30% suffered from cardiovascular conditions. Veterans of both the 1873- and 1893-sample were likely to be farmers or retired. Few veterans were professionals of type 1 or 2.[56] Veterans from both samples come largely from New York, Pennsylvania and the Midwest. Table 1.2 presents probit regressions in which I use wartime indicators to predict the likelihood that a veteran would be in the 1873- or 1893-sample in the baseline year. Veterans were 20.7% more likely to be on the pension rolls by 1873 (and thus in the 1873-sample)

[54]Black men were not included in this research because they were largely disenfranchised between 1870 and 1930. Therefore, the exogenous source of variation in pension income - Congressional vote share in a pensioner's district - is not applicable.

[55]Costa and Kahn [2008], p. 2-3.

[56]The Professional (1) category includes manufacturers, educators, attorneys, and other professionals. The Professional (2) category includes clerks, merchants and salesmen.

if they were wounded. 76.51% of veterans in the 1873-sample were wounded as compared with only 34.12% in the 1893-sample. From Table 1.2, it is clear that veterans on the pension rolls by 1893 were two years younger than those on the rolls by 1873.

The surgeons certificates are of particular importance in this research because they describe the illnesses of veterans. Surgeons certificates were written by physicians, who were appointed by the Pension Bureau, to assess the illnesses faced by veterans.

Since veterans underwent many physical examinations, either to apply for the first time or to apply for an increase, health information is available over time. One applicant could undergo as many as twenty exams. There are a total of 79,238 exams in the data set, and on average, each veteran in the sample had 4.94 exams.[57] The average exam year was 1892 (the median is 1891), coinciding with the passing of the 1890 Act, which awarded pensions for disabilities.

1.4.2 Historical Congressional Election Data

In order to determine the Republican Congressional vote share of each veteran's district, it was first necessary to determine the district of residence for veterans in each year. Since the Union Army sample contains the state and city of residence as well as the exact street address of veterans from first pension receipt until death, I matched each veteran to his associated Congressional district between 1870 and 1910.[58] Since district boundaries change rapidly during this period, veterans had to be rematched to districts to account for border changes. Even during years in which a veteran remained at the same residence, he could still be counted as being in a new district if boundaries were redrawn.

To account for changes in district lines, I matched each veteran's address to district maps for census years 1870 to 1910.[59] Having matched veterans to their districts in each year, I then collected election returns for each district in the United States between the years 1870 and 1910.[60]

It was necessary to collect this new dataset of Congressional Election returns by district and to create new GIS district maps, as opposed to using the ICPSR's Historical Election Returns Series and GIS county maps from the NHGIS's Historical state and county boundary files (1790-2000),[61] for two reasons. First, while the

[57]The median number of exams was 4.

[58]If a veteran moved to a new location in a given year, he appears in the data as being in the destination district for that year.

[59]District boundaries change after each census year. Therefore, district boundaries from 1970 to 1979 can be determined by looking at the boundaries of the 1970 census. District boundaries for the maps were found in Kenneth Martis' *The Historical Atlas of the United States Congressional Districts: 1789 - 1983* (Martis [1982]), which had never been digitized.

[60]The election returns were compiled from Michael Dubin's book *United States Congressional Elections, 1788 - 1997* (Dubin [1998]).

[61]The National Historical Geographic Information System has historical county boundary files from 1790 to 2000 available at nhgis.org.

ICPSR data set links counties to districts, it is not possible to be certain of a veteran's district simply by determining his county. This is because many counties were split into multiple districts. Therefore, matching a veteran's address to a county, using GIS county maps, and linking the county to an associated district, using the linkages provided in the ICPSR's Historical Election Returns, 1824-1968, is an inaccurate approach.

Second, the ICPSR series contains vote totals by *county* as opposed to district. While counties are matched to districts in this series, it is not possible to compute the Republican vote share for a district by summing Republican votes over the counties in a district (and diving by total votes) because one county could lie in *two* different districts. Therefore, I collected a new series of election returns by district, which are linked to a new set of GIS maps with historical Congressional district boundaries.[62]

County and City-Level Data

In the analysis, I use county-level data from the ICPSR's Historical, Demographic, Economic and Social Data: The United States, 1790-2002. In particular, I use farm values per acre as a proxy for changing wealth levels in the county. Information on the population size of U.S. cities from the U.S. Census Bureau.[63]

1.5 Results

First-stage results from the specification in Part I are presented in Table 1.3.[64] For the 1873-sample, the coefficient estimate on the close-vote variable is 1.28 and significant at the 1% level. This means that veterans in the 1873-sample living in districts with close elections received an extra $1.28 of pension income relative to their counterparts in non-close election districts. Veterans in the 1893-sample living in close election districts received an extra 88 cents of pension income relative to their counterparts in non-close election districts. Using both samples, living in one of the hundred most populated cities did not significantly affect pension income.

The tables of hazard regression estimates of disease onset are separated into anatomical groups and regressions with and without the control function are presented for both the 1873- and 1893-sample. Table 1.4 shows that respiratory conditions are income sensitive using the control function approach to correct for reverse

[62]The digital GIS maps were created by Carlos Villareal with the support of the All-UC Economic History Dissertation Grant 2010. Figures 1.1 to 1.4 show the GIS maps created from census years 1870, 1880, 1890 and 1900.

[63]The U.S. Census Bureau's Population of the 100 Largest Cities and Other Urban Places in the United States: 1790 to 1990 is available at http://www.census.gov/population/www/.

[64]The dependent variable in the second-stage associated with the first-stage presented in the table is an indicator variable =1 if a veteran had a hernia in year t.

causality.[65] Correcting for reverse causality, an extra $1 of pension income decreases the probability of the onset of a respiratory condition by [1-exp(-.207)=] 18.7% for veterans in the 1893-sample, which is equivalent to a reduction in the hazard rate of 1.68/1000. Estimates of the effect of a $1 increase in pension income on the likelihood of contracting a respiratory illness for the 1873-sample are even larger – the semi-elasticity is -.344, equivalent to a 29.1% reduction in the proportional hazard rate. Table 1.4 also shows estimates for digestive conditions. For veterans in the 1873-sample, increased income does not significantly affect the onset of conditions in the digestive category, however, for veterans in the 1893-sample, income has large effects. Interestingly, for both the 1873- and 1893-sample, increased income does not prevent diarrheal conditions once correcting for reverse causality. This may be because diarrheal conditions in this period are primarily caused by contaminated water supplies shared by both the rich and poor. Therefore, public heath interventions to chlorinate water supplies were especially beneficial because, in light of these estimates, the incidence of diarrheal conditions could not be warded off by additional income.

Table 1.5 shows regression estimates for illnesses in either the infectious, cardiovascular or endocrine category. An extra $1 of pension income decreases the likelihood of the onset of an infectious illness by 38% (a decrease of .0015 in the hazard rate) for veterans in the 1893-sample. However, the probability of a veteran in the 1873-sample contracting an infectious illness does not appear to be significantly influenced by increased income. This may be because veterans in the 1873-sample were likely to be in worse health earlier in the lifecycle (and therefore able to receive pensions at an earlier date) relative to veterans in the 1893-sample causing income to have little beneficial effects.

With regard to the onset of cardiovascular conditions, veterans in both the 1873- and 1893-sample appear to have benefitted enormously from increased income. Cardiovascular conditions such as arteriosclerosis and enlarged heart are conditions associated with affluence and seen in higher rates in developed nations. However, individuals who suffered from acute infections, such as rheumatic fever, could sustain valvular damage to the heart. Therefore, inasmuch as income could ward off infectious illnesses, it could also prevent the likelihood of heart disease resulting from acute infections.

Table 1.5 also shows the effects of income on endocrine diseases. For both the 1873- and 1893-sample, increased income has no effect on diseases in the endocrine category taken together. However, for the 1893-sample, there are negative income effects on the probability of developing diabetes whereas there are no effects for the 1873-sample. This may be because veterans in the 1893-sample receive pensions at older ages relative to the 1873-sample, and there may be negative income effects on the probability of the onset of type-II diabetes (which usually appears at older ages)

[65] Estimates on the left-hand side of Table 1.4 show the results when I estimate the hazard model without using the control function approach – use the second-stage of the specification in Part I without the use of the first-stage residual.

operating through better nutrition. Logan [2009] that calorie expenditure elasticities in developing countries today are significantly lower than calorie elasticity estimates for the U.S. in the late 19th century. Therefore, individuals born at the same time as the Union Army cohort were likely to be significantly underfed relative to the poor today. Given the findings of Logan [2009], increased income most likely operated on health through improvements in nutrition to prevent a range of illnesses from diabetes to infectious conditions.

Table 1.7 shows the effects of income on the crude death rate (all-cause mortality) and on cause of death. I find that an extra $1 of pension income lowers the hazard rate of death from any cause by .008. I also find large income effects for respiratory, digestive and cardiovascular illnesses. However, I do not find significant income effects for infectious illnesses. There are two possible reasons for this finding. First, diseases in the respiratory and digestive category were mostly infectious illnesses but are not in the infectious category because symptoms of the illnesses are confined to either the respiratory system or the digestive system. Therefore, the large declines seen in illnesses such as tuberculosis are predicted by my estimates. Second, while income can delay the onset of infectious illness, the results suggest that income cannot prevent the eventual death of a veteran from an infectious illness relative to other illnesses.

Hazard estimates for mortality from endocrine diseases appear to be negatively and significantly affected by increased income. This is plausible because improvements in diet can greatly alter mortality rates of endocrine diseases such as diabetes.

With regard to the regression discontinuity estimates, infectious diseases shown in Table 1.8 do not seem to be responsive to income. Instead respiratory illnesses and cardiovascular conditions are responsive to income. The negative coefficient on respiratory conditions, for example, means that veterans born in 1844, who received more pension income relative to veterans born in 1843, had a lower probability of dying from respiratory infections.

In Table 1.8, I include the results of the same regression discontinuity approach but instead compare the likelihood of the onset of illness for the 1846-cohort to the (omitted 1845-cohort). Since neither those in the 1845-cohort nor those in the 1846-cohort "missed out" on income from the enactment of the Law of 1907 (as was the case for the 1843-cohort relative to the 1844-cohort), there should be no difference in the rates of disease onset between the two cohorts. In Table 1.8, I include the results of the regression discontinuity analysis when I compare the 1845- to 1846-cohort and do not find a difference in the rates of disease onset with the exception of the respiratory category.

1.6 Robustness Checks

1.6.1 Threats to Identification

Migration Model

The results of this research hinge on the assumption that sick veterans do not migrate to districts in which Republican Congressional candidates face close elections in order to secure higher pensions. Of veterans that move to new districts, an average of 10.44%[66] migrate from districts in which Republican vote share was below 30% or above 70% to districts in which Republican Congressional candidates faced close elections. Therefore, veterans who moved chose districts with closer elections. However, to more concretely determine whether *sicker* veterans migrated to more Republican districts, I use a simple Conditional Logit model of Migration. The dependent variable is the probability that a veteran moves to a new district with close elections when the originating district did not have close elections. The independent variable is the incidence of any disease in the year *t-1*. The model has the following form:

$$Prob(move_{it} = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 sick_{i,t-1} + \Gamma C)}} \tag{1.6}$$

where *sick* is a dummy variable equal to 1 if a veteran has any illness in a given year and C are a set of controls. Controls include a quadratic time trend, body mass index, rank in the army, marital status, occupation, retirement status, prisoner-of-war status, and war wounds.[67]

The results presented in Table 1.9 show that if a veteran was sick with any illness in the previous period, he is 8 percent more likely to move to a close-election district, and the result is significant at the 10%-level. However, upon including the set of controls, the estimate is no longer significant. Therefore, while this was a period of great migration in the United States, Union Army veterans, once they fell sick, did not appear to be moving to districts in which Republicans faced close elections.

Farm Land Values

If Republican Congressional candidates were able to raise income and wealth levels in their districts through means other than increasing the pensions of Union Army veterans, then the identification strategy of this research breaks down. One check on the candidates' abilities to affect the health of their constituents through practices, which increased income or wealth, other than boosting pensions is to determine the effect of close elections on farm land values. Farm land values are a particularly good measure of overall economic prosperity in a veteran's district because the majority of veterans lived in rural areas.

[66] The standard deviation is .306

[67] I cluster by year and district.

Data on the average value of farm land per acre stratified by county over the years 1870 to 1930 come from the Michael Haines' ICPSR's Historical, Demographic, Economic and Social Data: The United States, 1790 - 2002, Dataset 2896 Part 106. I match each county to its associated district[68] and Republican Congressional vote share. I use the following specification:

$$logfarm_{dt} = \gamma_0 + \gamma_1 vote_{dt} + \Gamma C_{dt} + \epsilon_{dt} \tag{1.7}$$

where *logfarm* is the log of average farm land value per acre, *vote* is an indicator variable equal to 1 if Republican vote share is between .3 and .7 inclusive, C is a set of controls including year fixed effects, district time trends and district fixed effects, d denotes the district and t denotes time.

There is no significant effect of the closeness of elections for Republican Congressional candidates on farm land values per acre. The coefficient on the close-election indicator variable is -0.006 and is not significant.[69]

Mortality Trends of Veterans and Non-Veterans

To further check whether Republican Congressional candidates were able to enact policies which benefited the health of their constituents besides increasing the pensions of Union Army veterans, I use a separate mortality dataset to see if non-veteran cohorts living in districts with close elections had lower mortality rates relative to those in all other districts.

The mortality data used in this section come from the U.S. Federal Census Mortality Schedules, 1850-1885.[70] The Mortality Schedules contain information on every person who died during the year preceding the census year.[71] The following information appears in the mortality schedules: Name of deceased, gender, age, race, birth place, cause of death, county and state. I restrict the sample to the post-Civil War census years 1870 and 1880 because they coincide with the years included in the main analysis of the paper. The dataset includes the following states: Arizona (Territory), Arkansas, California, Colorado, District of Columbia, Georgia, Idaho (Territory), Illinois, Iowa, Kansas, Kentucky, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Montana (Territory), Nebraska, Nevada, New Hampshire, New Jersey, North Carolina, Pennsylvania, South Carolina, Texas, Utah (Territory), Vermont, Virginia

[68] In cases where one county overlaps two districts, I map the county to the district in which most of its area lies.

[69] There were 163,020 observations, and the t-statistic for Republican Congressional vote share is -1.26.

[70] This data is found on the Ancestry.com website and was compiled by researchers at the Center for Population Economics at the University of Chicago.

[71] For the 1850, 1860, 1870 and 1880 censuses, collection began in June of the census year. Therefore, information on mortality is collected for the twelve months prior to the June collection for each census year.

and Wisconsin.[72]

Since the Mortality Schedule contains only a listing of those who died in each county for a given census year and not the fraction who died out of the total population, it was necessary to match population totals for each county.[73] However, since this dataset only has population totals as opposed to population stratified by age, it was necessary to determine the age structure (each cohorts share of the total population) by county. This was found using 1870 and 1880 IPUMS 1% samples. I then matched counties to district.[74]

I restrict the dataset to white males and form 10-year birth cohorts beginning in 1760 and ending in 1870. I test whether living in a district with close elections affects the overall mortality in the district as well as cause-specific mortality.

Table 1.10 includes the coefficient estimates on the close-election variable,[75] a dummy variable for the year 1880, and interaction terms (close-election multiplied by cohort) for each cohort. Each 10-year birth cohort represents those who were born during the nine years following the cohort-year label.[76] The majority of those who served in the Civil War were born during the 1830s. From Table 1.10, it is clear that all cohorts had higher mortality rates and cause-specific mortality rates in 1880 relative to 1870. This is apparent from the positive and significant coefficient on the 1880 dummy variable. In order to determine whether living in a district with close Congressional elections affects mortality, it is necessary to add the coefficient on the close-election variable with the coefficient on the interaction term of the cohort of interest. For example, relative to the omitted 1860-cohort, living in a district with close elections caused a -0.03% [0.0000+(-.0003)=-.0003] decrease in the mortality rate of the 1830-cohort. However, this result is not significant. When testing the joint significance of the coefficient of the close-election indicator variable with the Interaction term for the 1830-cohort, the F-statistic is 0.24. The estimates on the close-election variable are jointly significant with the 1790-cohort, 1800-cohort and 1810-cohort interaction terms for the Total Mortality, Respiratory and Infectious disease categories. The estimates on the close-election variable are jointly significant with the 1800-cohort interaction term for the Cardiovascular disease category, and the close-election estimate is not jointly significant with any of the cohort interaction

[72]Some states are left out of the analysis because they have not yet been acquired by ancestry.com. These states are Alabama, Connecticut, Delaware, Florida, Indiana, Maryland, Mississippi, Missouri, New Mexico (Territory), New York, North Dakota, Ohio (Seneca County only), Oregon, Rhode Island, South Dakota, West Virginia, Wyoming (Territory).

[73]County-level population was drawn from Michael Haines' Historical, Demographic, Economic, and Social Data: The United States, 1790 - 2002. Dataset: ICPSR 2896. Part 11: 1870 Census (County and State) and Part 15: 1880 Census (County and State)

[74]If a single county lied in two districts, I matched the county to the district in which encompassed the majority of the county's land.

[75]The close election variable is a dummy equal to 1 if Republican Congressional vote share is between 0.3 and 0.7 inclusive.

[76]For example, the 1770-cohort includes all individuals born between 1771 and 1780 inclusive.

terms in the Digestive category.

While effects are significant for older cohorts, they are positive and small. For example, living in a district with close elections decreases the overall mortality rate for the 1800-cohort by .3% [-.0000+(-.0030)=-.0030], which is the largest significant effect for all cohorts across all categories (columns). Therefore, close elections do not affect health outcomes for the Union Army cohort or any individuals born after 1820. For cohorts born before 1820, the effects are slight causing the main results of this research to be biased downward.

1.7 Conclusion and Discussion

In this research, I explore the impact of income on adult health at the end of the 19th and early 20th century. In order to circumvent reverse causality - the scenario in which health influences income - I use two sources of exogenous variation in pension income: the presence of close elections for Republican candidates of a pensioner's Congressional district and changes in pension legislation. I find large effects and conclude that increases in income, in the absence of public health interventions or medical advancements, profoundly influenced the health trajectories of cohorts born in the mid-nineteenth century for the better.

The results of this research shed light on current debates regarding the health benefits of cash transfers to adults living in developing nations in which the socioeconomic gradient in health is large, as was the case in the U.S. a century ago. Entitlement programs providing cash transfers, of particular interest today, are the South African pensions and Mexico's Oportunidades system. Case [2004] uses data on the South African pension system to determine the effect of income on health status, but is unable to make specific assessments regarding the health of recipients because health status is self-reported in the sample. Investigations of income's effect on health using data from the Oportunidades program suffer from the same problem as health status is similarly self-reported (Fernald et al. [2008]). For regions in which individuals suffer from high rates of infectious illnesses, the results of this research suggest that cash transfers have large benefits. In light of the results, it appears that the tradeoffs between providing cash transfers to adults in developing countries as opposed to public health interventions is narrower than previously thought.

The results of this research are also pertinent to discussions of the relationship between income and retirement, as income may act on retirement through health. First, the design of pension systems in countries with wide SES gradients in health must account for positive income effects on longevity. Second, since income affects current health status and expected longevity, individual decisions regarding labor force participation and retirement are ambiguous and would be an interesting avenue for further research.

Table 1.1: Summary Statistics for Union Army Veterans

	Mean (%) in Baseline Year	
	On rolls in 1873	On rolls in 1893
Disease Conditions		
Respiratory	4.40	7.22
Endocrine	.46	5.64
Genito-urinary	.40	3.56
Digestive	6.91	24.75
Cardiovascular	3.43	30.00
Infectious	1.94	5.06
Occupation Type		
Farmer	35.60	36.82
Professional (1)	2.22	3.03
Professional (2)	5.66	6.19
Artisan	14.74	13.03
Service	3.71	4.61
Manual Labor	9.37	11.44
Retired	24.17	22.50
Married	72.06	78.44
State		
Illinois	12.02	8.01
New York	18.57	10.17
Ohio	13.69	14.67
Pennsylvania	10.41	8.77
Michigan	4.72	6.34
Indiana	4.37	5.50
All Other	36.22	46.54
No. of Veterans	1750	13,296

Table 1.2: Probit Regression - Predicting Veteran's Entry onto Pension Rolls

	Date of Entry onto the Pension Rolls			
	1873		1893	
	Coefficient	Mean (%)	Coefficient	Mean (%)
War Injury/Illness	.207***	76.51	.044***	34.12
	(37.00)		(10.16)	
POW	.031***	13.14	.013*	8.76
	(4.16)		(-16.42)	
Private	.007	71.82	.087***	74.70
	(0.18)		(2.46)	
Sergeant	.014	10.29	.036	8.06
	(0.34)		(1.49)	
Lieutenant	.015	2.29	.024	1.84
	(0.34)		(0.87)	
Musician	.003	.97	.042*	1.45
	(0.08)		(1.82)	
Birth Year	-.003***	1836	-.006***	1838
	(-9.88)		(-16.42)	
No. of Veterans on rolls	1750		13296	
Total no. on veterans	16013		14442	

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

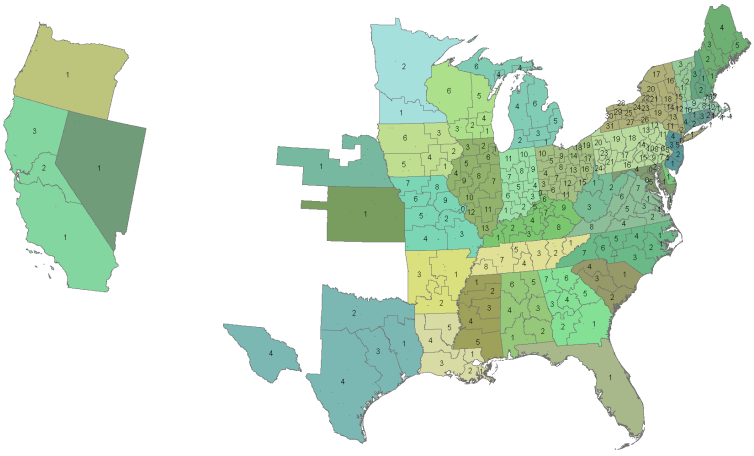


Figure 1.1: Congressional District Map for Census Year 1870

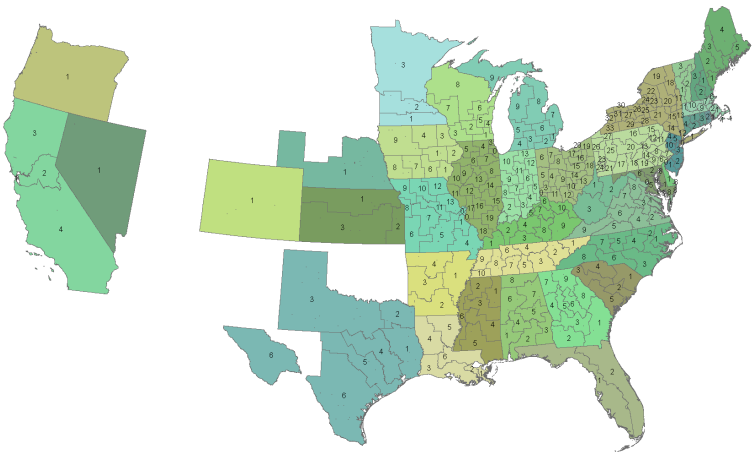


Figure 1.2: Congressional District Map for Census Year 1880

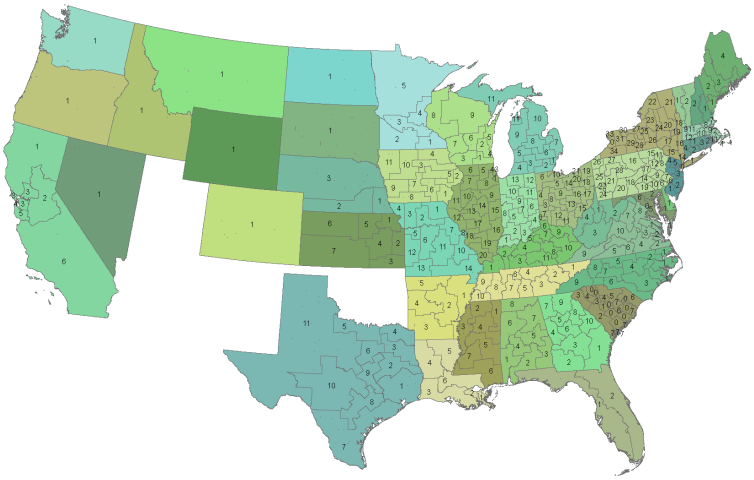


Figure 1.3: Congressional District Map for Census Year 1890

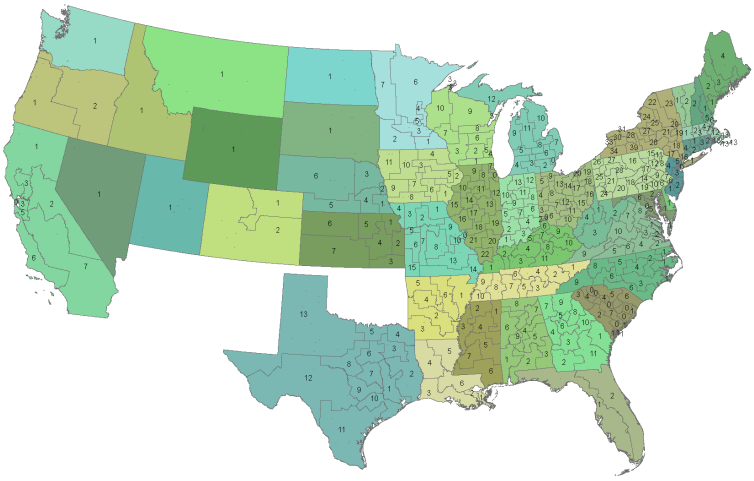


Figure 1.4: Congressional District Map for Census Year 1900

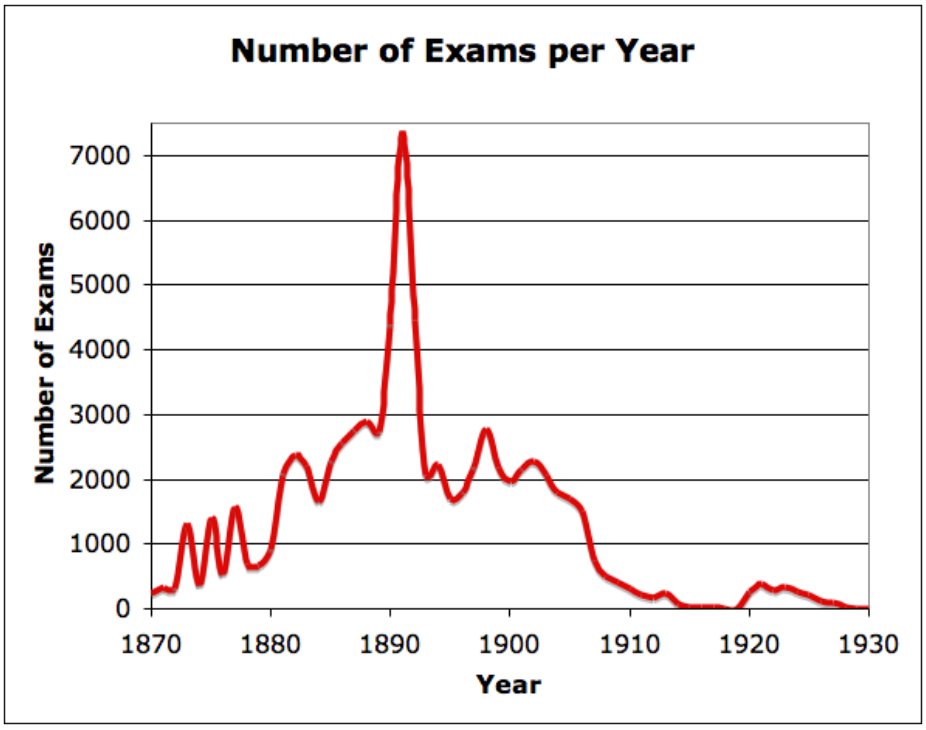


Figure 1.5: Number of Exams per Year

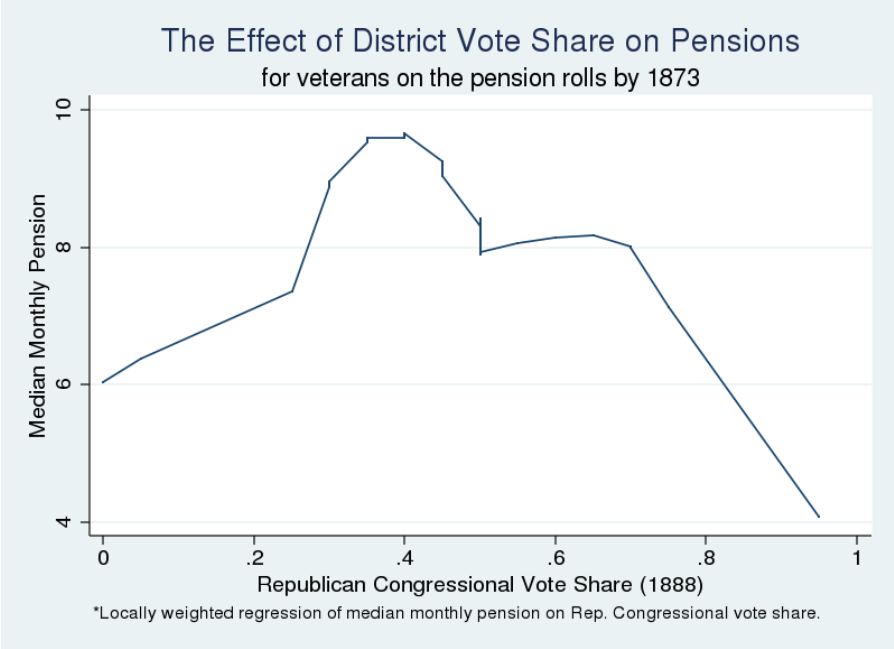


Figure 1.6: District Vote Share and Pensions for 1873-sample

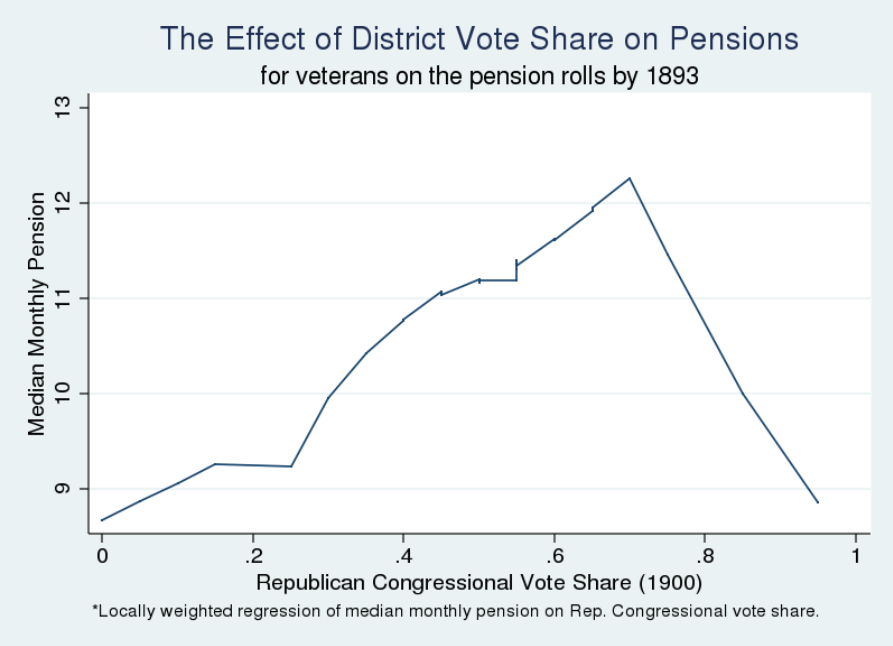


Figure 1.7: District Vote Share and Pensions for 1893-sample

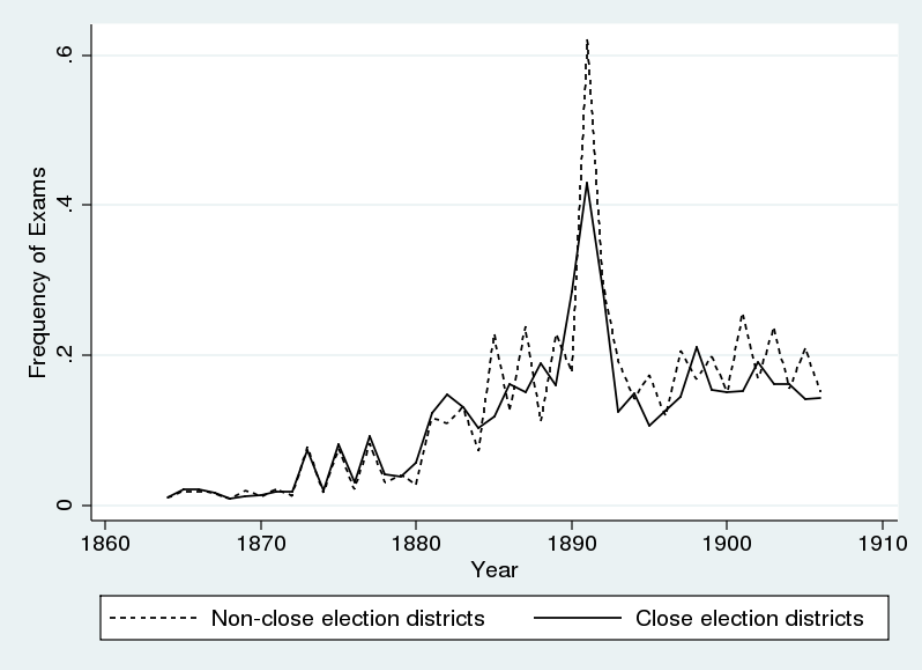


Figure 1.8: Frequency of Exams per year for all veterans

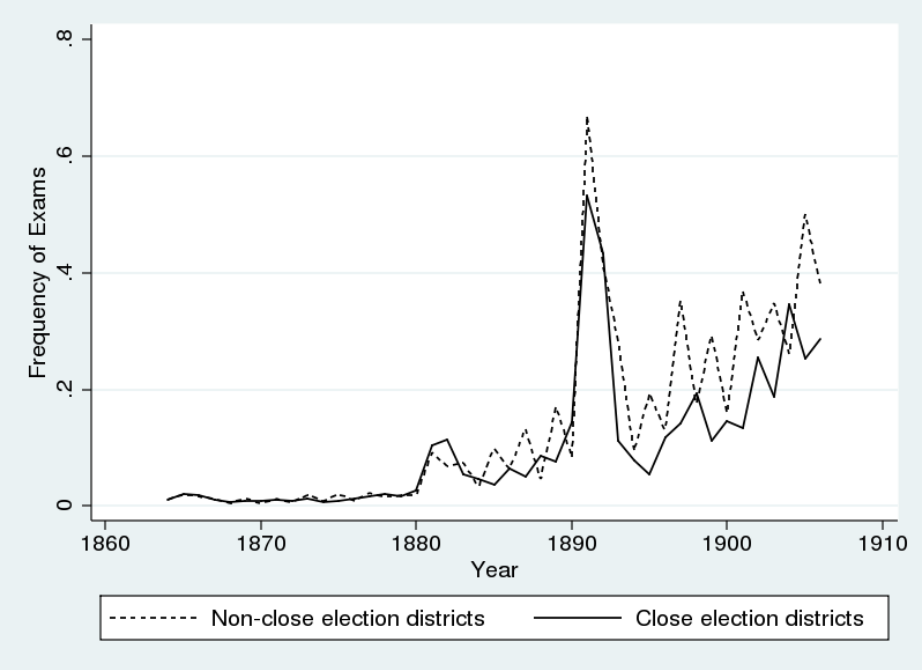


Figure 1.9: Frequency of First Exams per year for those not yet on rolls

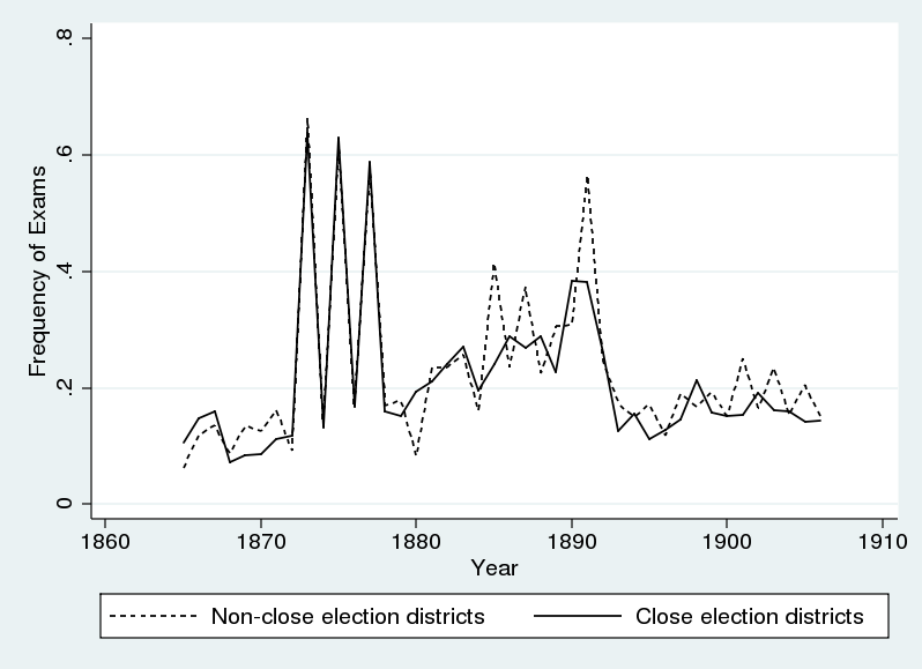


Figure 1.10: Frequency of Exams per year for those already on the rolls

Table 1.3: First Stage Results

	Monthly Pension ($)	
Date of entry onto Pension Rolls:	1873	1893
Vote (=1 if 0.3≤ Republican Vote Share ≤0.7)	1.28***	.88***
	(9.44)	(9.06)
Time Trend	.45***	.24***
	(78.80)	(27.50)
Birth Year	-.05***	-.18***
	(-7.81)	(-44.97)
Top 100 City	-.24	.26
	(-0.24)	(0.53)
Farm Land Value (per acre)	.00***	.00***
	(-5.23)	(-14.18)
Mean pension ($)	11.13	11.92
No. of Veterans	1671	11698

Note: First stage results when second-stage dependent variable of interest is a dummy variable =1 if a veteran had a hernia in the period. t statistics in parentheses. Monthly pension income in 1907 dollars. Controls: Rank in the army, prisoner-of-war status, war wounds and dummies for previous illnesses. * p¡0.05, ** p¡0.01, *** p¡0.001. Clustered by year and district.

Table 1.4: Hazard Regression Estimates of the Onset of Respiratory and Digestive Conditions

	Dependent Variable: Monthly Pension Amount					
	No fix for reverse causality		Correcting for reverse causality			
Date on rolls:	1873	1893	1873	Hazard Rate	1893	Hazard Rate
Respiratory	.003	-.002	-.344***	.006	-.207**	.009
	(.007)	(.004)	(.138)		(.098)	
	[1659]	[11548]	[1659]		[11548]	
Bronchitis	-.007	-.003	.100	.002	-.065	.005
	(.010)	(.005)	(.262)		(.134)	
	[1717]	[12192]	[1717]		[12192]	
Pneumonia	.016	-.003	-.566**	.001	-.210*	.002
	(.010)	(.008)	(.283)		(.119)	
	[1894]	[12855]	[1894]		[12855]	
Digestive	-.018***	-.028***	-.031	.008	-.199**	.015
	(.007)	(.004)	(.150)		(.087)	
	[1603]	[9023]	[1603]		[9023]	
Diarrhea	-.020**	-.017***	-.194	.005	.055	.008
	(.009)	(.005)	(.166)		(.121)	
	[1618]	[9884]	[1618]		[9884]	
Enlarged Liver	-.018***	-.005	-.511***	.006	-.207**	.011
	(.007)	(.004)	(.139)		(.091)	
	[1719]	[10856]	[1719]		[10856]	

Note: Standard errors in parentheses. No. of veterans in square brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Controls: Top 100 city (in population size), time trend, cubic of first-stage residual, birth year, county-level average farm value per acre, rank in the army, prisoner-of-war status, war wounds and previous illnesses. Clustered by year and district.

Table 1.5: Hazard Regression Estimates of the Onset of Infectious, Cardiovascular, and Endocrine

	Dependent Variable: Monthly Pension Amount					
	No correction for reverse causality		Correcting for reverse causality			
Date of Entry onto Pension rolls:	1873	1893	1873	Hazard Rate	1893	Hazard Rate
Infectious Illnesses	-.013	-.024***	-.275	.003	-.478***	.004
	(.010)	(.007)	(.201)		(.125)	
	[1693]	[11829]	[1693]		[11829]	
Cardiovascular	-.009**	-.017***	-.150*	.022	-.285***	.049
	(.004)	(.002)	(.082)		(.050)	
	[1657]	[8152]	[1657]		[8152]	
Arteriosclerosis	-.017	-.001	-.592***	.002	-.505***	.006
	(.011)	(.004)	(.174)		(.106)	
	[1736]	[12880]	[1736]		[12880]	
Dyspnea	-.021***	.002	-.089	.011	-.231***	.034
	(.005)	(.002)	(.116)		(.052)	
	[1715]	[11827]	[1715]		[11827]	
Enlarged Heart	-.003	-.008	-.051	.013	-.272***	.029
	(.004)	(.003)	(.114)		(.059)	
	[1708]	[9946]	[1708]		[9946]	
Endocrine Illnesses	-.003	.005	-.255	.004	-.102	.008
	(.008)	(.003)	(.189)		(.108)	
	[1729]	[11725]	[1729]		[11725]	
Diabetes	.019	.008	-.486	.001	-.449**	.002
	(.012)	(.006)	(.376)		(.194)	
	[1734]	[12884]	[1734]		[12884]	

Note: Standard errors in parentheses. No. of veterans in square brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls: Top 100 city (in population size), time trend, cubic of first-stage residual, birth year, county-level average farm value per acre, rank in the army, prisoner-of-war status, war wounds and previous illnesses.
Clustered by year and district.

Table 1.6: Hazard Regression Estimates of the Onset of Hernia and Genito-urinary Conditions

	Coefficient on Monthly Pension Amount					
	No correction for reverse causality		Correcting for reverse causality			
Date of Entry onto Pension rolls:	1873	1893	1873	Hazard Rate	1893	Hazard Rate
Genito-urinary Illnesses	-.028***	-.022***	.042	.003	-.241***	.009
	(.010)	(.005)	(.196)		(.093)	
	1729	12366	1729		12366	
Cystitis	-.027	-.014**	.036	.001	-.418***	.003
	(.017)	(.007)	(.336)		(.150)	
	1733	12750	1733		12750	
Hernia	-.028***	-.021***	-.410***	.004	-.556***	.005
	(.010)	(.006)	(.147)		(.114)	
	1671	11698	1671		11698	

Note: Standard errors in parentheses. No. of veterans: 17,238. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls: Top 100 city (in population size), time trend, cubic of first-stage residual, birth year, county-level average farm value per acre, rank in the army, prisoner-of-war status, war wounds and previous illnesses. Clustered by year and district.

Table 1.7: Hazard Estimates for Mortality

	Coeff. on Monthly Pension No correction for reverse causality		Coeff. on Monthly Pension Correcting for reverse causality			
Date of Entry on rolls:	1873	1893	1873	Hazard Rate	1893	Hazard Rate
All-cause Mortality	.014***	.010***	-.327***	.018	-.342***	.028
	(.003)	(.001)	(.071)		(.048)	
Respiratory Illnesses	.021***	.014***	-.420***	.003	-.377***	.005
	(.006)	(.003)	(.186)		(.113)	
Infectious Illnesses	.022	.008	-.513	.001	-.264	.002
	(.015)	(.005)	(.390)		(.195)	
Digestive Illnesses	.015**	.0143	-.459**	.002	-.333***	.007
	(.007)	(.002)	(.219)		(.101)	
Cardiovascular Illnesses	.009	.009***	-.281*	.005	-.382***	.013
	(.005)	(.001)	(.155)		(.072)	
Endocrine Illnesses	.019	.019***	-.720**	.001	-.430**	.002
	(.013)	(.003)	(.345)		(.181)	
No. of veterans	1736	13011	1736		13011	

Note: Standard errors in parentheses. No. of vets: 17,238. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls: Top 100 city (in population size), time trend, cubic of first-stage residual, birth year, county-level average farm value per acre, rank in the army, prisoner-of-war status, war wounds and previous illnesses. Clustered by year and district.

Table 1.8: Regression Discontinuity - Law of 1907

	1844-Cohort	1846-Cohort		1844-Cohort	1846-Cohort
Respiratory	-.043***	-.028*	Endocrine	.002	-.003
	(-3.04)	(-1.76)		(-0.30)	(-0.34)
	[1566]	[1221]		[1499]	[1199]
Asthma	-.008*	.004	Diabetes	-.003	.006
	(-1.66)	(0.99)		(-0.87)	(1.25)
	[1825]	[1424]		[1826]	[1427]
Bronchitis	-.013	-.012	Enlarged Spleen	.006	-.007
	(-1.45)	(-1.34)		(1.22)	(-1.39)
	[1691]	[1333]		[1641]	[1240]
Pleuritis	-.002	-.016	Genito-urinary	-.009	.001
	(-1.05)	(-0.53)		(-0.81)	(0.21)
	[1836]	[1273]		[1675]	[1440]
Pneumonia	-.018**	-.013	Cystitis	-.004	.005
	(-2.11)	(-1.25)		(0.49)	(.054)
	[1840]	[1447]		[1794]	[1394]
Tuberculosis	-.007	.004	Enlarged Prostate	-.017	.004
	(-1.48)	(0.70)		(-1.28)	(0.27)
	[1833]	[1445]		[1743]	[1333]
Digestive	.001	-.006	Nephritis	.008	-.014
	(0.08)	(-0.50)		(0.90)	(-1.33)
	[1125]	[851]		[1864]	[1468]
Diarrhea	-.002	.000	Infectious	-.002	-.009
	(-0.40)	(0.03)		(-0.35)	(-1.41)
	[1347]	[1036]		[1649]	[1245]
Dysentery	-.001	-.002	Cholera	.001	-.001
	(-0.64)	(-0.80)		(1.11)	(-0.90)
	[1864]	[1469]		[1882]	[1478]
Enlarged Liver	-.026***	-.003	Cardiovascular	-.029	.021
	(-3.00)	(-0.41)		(-0.88)	(0.59)
	[1335]	[1086]		[698]	[492]
Gastroenteritis	.004*	-.002	Arteriosclerosis	-.009	.008
	(1.55)	(-.039)		(-0.51)	(0.40)
	[1851]	[1433]		[1766]	[1376]
Malassimilation	-.003	-.004	Dyspnea	-.050***	-.001
	(-0.69)	(-0.90)		(-2.58)	(-0.07)
	[1744]	[1334]		[1091]	847
Nausea	.001	.003	Enlarged Heart	-.035**	-.020
	(0.91)	(0.77)		(-1.94)	(-1.07)
	[1848]	[1448]		[1079]	[784]

Note: z statistics in parentheses. No. of veterans in square brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls include the veteran's rank in the army, his war wounds and POW status.

Table 1.9: Logit Model of Migration

Dep. Var.: Prob. of moving to district in which elections are close for Republicans				
Lagged Sickness (any disease)	.080*	-.032	.020	-.065
	(1.70)	(-0.93)	(0.37)	(-1.42)
Time Trend	-.032***	-.011	-.051***	-.013
	(3.56)	(-1.06)	(-2.97)	(-1.28)
Square of Time Trend	.000***	.000*	.001***	.000***
	(4.50)	(1.81)	(4.39)	(3.41)
Only those who move	NO	NO	YES	YES
Controls	NO	YES	NO	YES
$Number of Observations$	651,822			

Controls include BMI, rank, marital status, occupation, retirement status, POW status and war wounds.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.10: Mortality by County: 1870 and 1880

	Total Mortality	Respiratory	Digestive	Infectious	Cardiovascular
Close Election	-.0000	-.0000	.0000	.0001***	-.0001
	(-0.20)	(-0.27)	(0.97)	(3.67)	(-1.14)
Year = 1880	.0025***	.0012***	.0001***	.0002***	.0004***
	(13.98)	(12.78)	(7.37)	(16.40)	(7.54)
Interaction (Cohort =1770)	-.0022	.-0010*	.0000	-.0001*	-.0000
	(-1.21)	(-1.00)	(0.87)	(-1.35)	(-0.10)
Interaction (Cohort = 1780)	-.0025	-.0010	.0001	-.0003	-.0005
	(-1.09)	(-0.81)	(0.24)	(-1.31)	(-0.85)
Interaction (Cohort = 1790)	.0023*	.0018***	.0001	.0001	.0002
	(1.71)	(2.53)	(0.41)	(0.71)	(0.37)
Interaction (Cohort = 1800)	.0030***	.0016***	-.0000	.0001**	.0008***
	(3.05)	(2.43)	(-0.30)	(2.23)	(3.13)
Interaction (Cohort = 1810)	.0013**	.0008***	.0001	.0001	-.0003**
	(2.74)	(2.72)	(0.80)	(1.11)	(2.00)
Interaction (Cohort = 1820)	.0004	.0003	.0000	-.0000	-.0000
	(1.49)	(1.12)	(1.20)	(-0.96)	(-0.14)
Interaction (Cohort = 1830)	-.0003	-.0002	-.0001	-.0001**	-.0000
	(-1.40)	(-0.87)	(-0.20)	(-2.33)	(-0.78)
Interaction (Cohort = 1840)	.0001	-.0000	-.0000	-.0001**	.0001
	(0.46)	(-0.11)	(-0.51)	(-2.04)	(1.21)
Interaction (Cohort = 1850)	-.0003*	-.0003*	.0000	-.0001***	.0001**
	(-1.89)	(-1.86)	(0.67)	(-2.60)	(2.50)
N	12768	12768	12768	12768	12768

t statistics in parentheses. Controls: Cohort dummies for 1770 - 1850. Omitted cohort is 1860.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered by year and district.

Chapter 2

Black Mortality in the Post-Bellum Era: Evidence from Union Army Pensions (with Trevon Logan)

2.1 Introduction

How much of the black-white mortality gap in the late 19th century and early 20th century U.S. can be explained by income differences between the races? In 1900, the black-white differential in life expectancy was approximately 10 years,[1] and city-level differentials in black-white death rates ranged from 13.8 per 1000 in New Orleans to 25.3 per 1000 in Annapolis.[2] While there is still a black-white gap in life-expectancy, the gap has narrowed significantly over the 20th century. According to U.S. Census estimates from 2000, the gap is approximately 5 years.[3]

Given the recent literature on the mortality decline for both the black and white population in this period, it is clear that the two most important determinants of health status were socio-economic status and the introduction of public health interventions.[4] Scholars such as Fogel and McKeown have emphasized the role of increases in standards of living, and in particular improvements in nutrition, on the decline in mortality prior to the introduction of modern medical techniques and drugs such as antibiotics.[5] For those living in cities, public health interventions such as water purification and the construction of sewer systems reduced mortality by as much as 50%[6] However, while the black-white gaps in education and income levels were large

[1] Troesken [2004], p. 2

[2] ibid., p. 6

[3] ibid., p. 2

[4] Eli [2010] contains a discussion of the efficacy of emerging medical treatments, which include the use of sulfa drugs. However, medicine remained largely ineffectual in this period.

[5] Fogel [2004]

[6] Scholars such as Cutler, Miller, Ferrie and Troesken have contributed to this literature. See

in this period, recent research by Troesken has shown that blacks and whites had similar access to clean water in cities.[7] Therefore, racial inequities in access to public goods such as clean water do not explain the black-white mortality gap in the period. After 1908, Troesken shows that equity in access to clean water leads to declining rates of typhoid and other infectious illnesses as well as a *convergence* of black and white mortality rates.[8] However, several question still remain unaddressed given the current literature on public health interventions: 1) What explains the black-white mortality gap for those living in rural areas, which did not see the introduction of sewer systems or water filtration in this period? 2) What explains the black-white mortality gap in cities prior to the large-scale public health interventions of the 1910s?

In this research, we investigate the effect of income increases on black mortality by using evidence from pensions received by black veterans who served as Union Army soldiers during the American Civil War. The benefit of using the cohort of black Union Army veterans is that we have information on pension income, health status and cause of death.[9] Information on the health status of veterans is available because veterans were required to undergo medical exams by qualified surgeons and furnish the Pension Bureau with surgeons' reports in order to qualify for pensions. However, since veterans received pensions based on proof of disability at medical exams, estimates of the effect of pension income on mortality will be biased. Therefore, it will seem as though increases in pension income lead to a higher risk of mortality. Thus, to circumvent endogeneity bias, we propose an exogenous source of variation in pension income: the number of years that the surgeon(s) who examined the veteran had been in practice and working for the Pension Bureau. We hypothesize that younger surgeons hired by the Pension Bureau had differing racial attitudes with regard to the evolution of disease in the black body and were therefore less likely to recommend veterans for pensions or augments in their current pension income. We expect younger surgeons to believe that the blacks were weaker than whites and more likely to succumb to disease ultimately leading to the disappearance of the black race. Younger surgeons were also more likely to be influenced by the contemporary research in which scholars such as Frederick Hoffman posited that increased income would be ineffectual in preventing the onset of disease in blacks.[10]

We construct a proportional hazard model to estimate income effects on the probability of death from any cause and from specific causes. Since data collection for our proposed instrument is currently underway, we instead present coefficient estimates on the effects of pension income and mortality without correcting for reverse

Cutler and Miller [2005] and Ferrie and Troesken [2005].

[7]Troesken [2004]

[8]Troesken [2004], p. 27

[9]This information was kept by the Pension Bureau, which was a sub-division of the federal government's Department of the Interior. These records were digitized by the Center for Population Economics at the University of Chicago.

[10]Hoffman [1896]

causality. Despite the fact that coefficient estimates are biased, they are still informative. We find that the endogeneity bias for blacks is larger than that for whites when estimating the effect of pension income on mortality from any cause. Therefore, conditional on controlling for prior health status, the larger coefficient estimates for blacks relative to whites imply that unobserved health conditions (such as disease severity) have bigger effects on health status for blacks than for whites.

Besides estimating income effects on mortality, we also examine the determinants of pension income and present new evidence on the effect of birthplace on pension income. We find that being born in the Deep South lowered average monthly pension income by almost $2 relative to the pension income received by all other blacks. However, blacks born in other areas of the South or in Border states received average monthly pensions which were not significantly different from those of blacks born in the Midwest, Northeast or the West. These facts suggest that there was an added income penalty for blacks born in the Deep South. Citing narrative evidence,Shaffer [2004] argues that ex-slaves from the South were unlikely to be literate and therefore unable to advocate for themselves in the pension application process. However, this argument does not completely fit with the quantitive evidence we present. Instead, we argue that blacks born in the Deep South were considerably different from those born elsewhere in the South.

This paper is organized as follows. Section 2.2 contains an historical background of the war experience of black veterans of the Union Army and their experiences in trying to secure a pension after the Civil War. Section 2.2 also contains a synopsis of the evolution of scientific racism in the period. Section 2.3 provides the econometric framework and justification of the proposed instrument. Section 2.4 describes the data and provides summary statistics. Section 2.5 contains results, and section 2.6 concludes.

2.2 Historical Background

2.2.1 The war experience of U.S. Colored Troops

The United States Colored Troops were comprised of 186,017 black soldiers, approximately 88,000 of which enlisted in Confederate states.[11] Another 46,000, enlisted in Border states, and of the remaining 46,000, some were likely born into slavery though free by the time of enlistment.[12] Of the 33,000 black soldiers who died during the war, 4,000 died from wounds and the remainder succumbed to disease.[13] The white battle mortality rate was significantly higher than that for blacks[14] owing to

[11] Humphreys [2008], p. 5.

[12] (ibid.)

[13] Humphreys [2008], p. 5-6.

[14] As many as 90,000 (4.5%) of white soldiers died in batter in comparison with 1.8% of blacks. (Humphreys [2008], p. 10).

the fact that blacks were purposely kept out of battle whenever possible. Whites were sent into battle more commonly than blacks because the common belief was that blacks were less "courageous" than whites in combat. Accordingly, troops in black regiments were assigned garrison duty during the war.

While black soldiers were in part recruited for their perceived immunity to disease, they in fact suffered from higher rates of diarrhea, dysentery, pneumonia, tuberculosis, smallpox and malaria. Approximately 9% of white recruits died from disease as opposed to a total of 16.7% of black recruits who succumbed to disease. Therefore, the black mortality rate during the war was higher despite the fact that whites saw more battle. Furthermore, even though blacks suffered higher rates of morbidity and mortality from disease during the war, they had greater difficulty gaining admission to army hospitals than their white counterparts. Even when blacks were able to get care in hospitals, the condition of black hospital wards was hardly therapeutic. Black wards were often unsanitary and lacking in food, clean bedding, clothing and medicine for their patients.[15]

2.2.2 Race Differentials in Pension Receipt

While pension legislation did not differ for white and black veterans, blacks faced difficulty in securing pensions. Between 1861 and 1934, the application success rate for white veterans was 92.6% while the rate for blacks was 75.4%.[16] Furthermore, conditional on receiving a pension during the period between 1879 and 1900, the white/black ratio in pension awards climbed from 1.106 to 1.273.[17] The reasons for the black-white differentials stem from racial discrimination during and after the war.

First, the inability of sick and injured black veterans to gain admittance to hospitals during the war hindered their ability to claim pensions for war-related disabilities after the war. Even though receiving care in a hospital during the war was not necessarily beneficial to a black soldier's health since doctors did not have modern medicines and hospitals remained unsanitary and contagious environments, the inability of black soldiers to seek care in hospitals during the war was detrimental to their ability to claim their pensions after the war. Prior to 1890, veterans could claim a pension under the General Law of 1862 if they faced illnesses or disabilities attributable to the war experience. In order to prove disability, veterans first sent their applications to the Pension Bureau stating their illness or injury. Upon receipt of the application, the bureau assigned each veteran to an examining surgeon[18] in his local area who certified that the veteran did in fact suffer from the disabilities stated on his application. The examining surgeon then completed a Surgeon's Certificate outlining

[15] Humphreys [2008], p. 79.

[16] Shaffer [2004], p. 209.

[17] Wilson [2007].

[18] After 1884, the veteran was likely required to see a board of examining surgeons as the Pension Bureau no longer hired single surgeons to perform exams.

the disabilities and their relationship to the veteran's war experience. The surgeon's certificate was then sent directly from the surgeon's office to the Pension Bureau in Washington, D.C. Once the certificate was received at the bureau, it was compared to the medical records from the war. If a veteran claimed a war-related disability, such as malaria or a gun shot wound, but there was no record of his admission to a war hospital for the illness or injury, he was likely to be denied a pension. Therefore, because black veterans had difficulty gaining access to hospitals during the war, they were unable to secure pensions in as large numbers compared to white veterans after the war.

Second, many black veterans, especially ex-slaves, had difficulty proving that had fought in the Civil War because their post-war names did not match the name given to clerks at enlistment. There were three reasons for name changes: 1) Upon enlisting, slaves often submitted their masters' last names, which were changed once free; 2) Likely to be illiterate, slaves were unable to correct enlistment clerks when their names were misspelled; 3) Many slaves escaped and enlisted under false names to prevent their masters from finding and reclaiming them.[19] Uncertain of a black applicant's identity, pension clerks requested affidavits from members of the applicant's community to prove war service. However affidavits from other community members were often considered unreliable in the case files of black veterans,[20] and pensions were subsequently denied.

The third, and most common, reason for pension denial (or lowered pensions relative to whites) was the inability of black veterans to prove their disabilities. After the passage of the Invalid Pensions Act of 1890, both white and black veterans could claim pensions for disabilities they faced which were unrelated to the war experience. Disabilities deserving of pensions included those which were verifiable during an exam, such as hernias, and those which were diagnosed on the basis of symptoms stated by the veteran, such as chronic diarrhea.[21] Symptom lists from white veterans were trusted by examining surgeons while lists from black veterans were considered suspect. Testimony from black veterans was considered to be "reliable" by examining surgeons if the veteran displayed appropriate behavior typical of middle-class whites.[22] The lack of trust between examining surgeons and many black veterans led surgeons to assign lower disability ratings to blacks as compared to whites. Because the surgeons' certificates for black veterans frequently had lower disability ratings, the medical examiners at the Pension Bureau were more likely to award lower pensions to black veterans.

After the passage of the Act of 1907, the pension bureau began awarding pensions

[19]Shaffer [2004], p. 124.

[20]Shaffer [2004], p. 129.

[21]Wilson [2007] shows that the black/white approval ratio for pensions in which veterans claimed a hernia was 0.896 whereas the rate was 0.404 for diarrhea. For other unverifiable conditions, such as back pain and hearing loss, the black-white approval rates were 0.397 and 0.216 respectively.

[22]Shaffer [2004], p. 130.

to veterans based on their age instead of disability. Under the new act, veterans over the age of 62 were eligible to receive $12 per month, and those over age 65 were eligible for higher amounts.[23] However, even with the passage of the new act, black veterans continued to face difficulty when applying for pensions because many did not know their date of birth.[24] By the Acts of 1912 and 1918, the age-based pension rates were increased yet again. While racial discrimination in the Union Army pension prevented many black veterans from receiving payments, current estimates show that the Pension Bureau dispersed a total of $313 million to black veterans.[25]

2.3 Empirical Strategy

In this analysis, we use the sample of veterans who began collecting pensions by 1893.[26] We follow this sample until 1906, which was the last year before the passage of the age-based laws. After the passage of these laws, the Pension Bureau began dispersing pensions based on age, and thus the ratings of examining surgeons were not a factor in the majority of application decisions. Therefore, the proposed instrument is only valid between 1890 and 1906.

2.3.1 Identification

Prior to 1907, the Pension Bureau awarded payments to veterans based on the degree of their disabilities, which causes pension income to be endogenous with respect to health outcomes. Since an individual's health status determines pension income, we propose an instrument for pension income to accurately estimate the effect of pension income on health status. The instrument we propose is the number of years an examining surgeon or board has been in practice. Surgeons who were more recently hired by the Pension Bureau were more likely to be younger and have graduated from medical school later than their counterparts who had long tenures with the Pension Bureau. Accordingly, newer examining surgeons would be less inclined to give high ratings to disabled black veterans during exams (and suggest that blacks be given higher pensions on surgeons' certificates) because they were likely to espouse the view that increases in income and socioeconomic status were ineffectual in improving the health status of blacks.[27] Without using the instrument, and thus exploiting an

[23] See Appendix A for exact pension amounts paid by the laws of 1907, 1912 and 1918.

[24] Shaffer [2004], p. 129-130.

[25] Shaffer [2004], p. 133.

[26] The sample of veterans collecting pensions prior to 1890 is too small for regression analysis.

[27] During the late nineteenth century, scientific racism gained prominence. In particular, those in the medical community as well as statisticians such as Hoffman (see Hoffman [1896]) began to view the black race as one that would soon perish because of poor heredity. It was believed by many in the medical community that high rates of mortality among blacks could not be helped by increased income.

exogenous source of variation in pension income, the effect of pension payments on mortality will be biased causing it to appear as though increases in pensions cause higher rates of mortality.

Our proposed instrument, the year a surgeon (or group of surgeons) was first hired by the Pension Bureau to administer exams *initialyear*, will be constructed in the following way: 1) if the veteran sees a single surgeon, then *initalyear* is equal to the initial year in which the physician began serving as an pension examining surgeon prior to the date of the veteran's exam; 2) if the veteran sees a board of examining surgeons, then *initialyear* is equal to the initial year that all members of the board have been practicing together.

The Register of Boards of Examining Surgeons, 1862 - 1928, available at Archives I, contains annual rosters of examining surgeons for all years between 1862 and 1928 except for the period between 1887 and 1892 inclusive. The following information can be found in the ledgers: 1) the town, county and state of exam sites; 2) the names of surgeons at each exam site; 3) the name of the congressman or senator who appointed each examining surgeon; 4) the reason for the removal of examining surgeons from their post or the organization of boards is provided. Common reasons for surgeon replacement include death, retirement or Congressional replacement of a previous surgeon. As was common during the patronage era, civil servants, examining surgeons included, were given their posts by newly elected or re-elected politicians as rewards for party loyalty.

In order to use *initialyear* as an instrument for pension income, the use of *initialyear* must satisfy the following exclusion restriction: Conditional on our other control variables, the initial year in which a single surgeon or board began serving for the Bureau cannot impact the health of a veteran through any other channel except through pensions. Several additional assumptions must be made to fully validate the instrument. First, *initialyear* must have a relevant impact on pension income. Second, the health status of black veterans must be independent with respect to the initial year in which the physician began his service for the Bureau. Third, it cannot be the case that black veterans from counties in which examining surgeons have relatively low *initialyear* are more likely to apply for pensions than those from counties with a low *initialyear*, or vice versa.

Controls in this analysis include an indicator equal to 1 if the veteran was born a slave,[28] the veteran's birth year, an indicator equal to 1 if the veteran lived in one of the hundred most populated cities, the veteran's prior health status, prisoner-of-war status and battle wounds. Each of these variables has an impact on pension income. Including a control for slave status at birth is necessary because early-life disease environments, which differ between free and slave states, can affect late-life disease

[28]The veteran is assumed to have been born a slave if his birth state was was a slave state in 1861. These states include Delaware, Maryland, Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Tennessee, Kentucky, Missouri, Arkansas, Mississippi, Louisiana, Oklahoma territory, Texas, the New Mexico Territory, and the Utah Territory.

specific mortality rates. Since the age range of the cohort is wide, controlling for birth year is crucial as older veterans have higher mortality rates. Lastly, we control for whether the veteran lives in one of the hundred most populated cities as a proxy for the disease environment he faced once on the pension rolls.[29]

Details of disease conditions faced during life come from the surgeons' certificates.[30] To control for prior health status, we construct indicator variables equal to 1 if a veteran had a particular illness, which was described on a surgeon's certificate. We group morbidity conditions during life into the following categories: respiratory, digestive, infectious, cardiovascular, endocrine, genitourinary.[31] Underneath the illness descriptions on surgeons' certificates, examining surgeons were required to rate the severity of the disability. While the surgeons' rating was uncorrelated with pension payments for white veterans,[32] the opposite was true for black veterans. Medical reviewers at the Pension Bureau relied heavily on the descriptions and ratings of examining surgeons in the cases of black veterans.[33]

2.3.2 Econometric Framework

Following the framework of Eli [2010], we use a Weibull proportional hazard model to estimate the effect of an extra dollar of monthly pension income[34] on the probability

[29]Since we do not observe the place of residence for veterans prior to their first receipt of pension income, we cannot control for the disease environment veterans faced between the end of the war and their first application date for a pension.

[30]See Eli [2010] Section 3.1.1. for a discussion concerning the accuracy of illnesses described on certificates.

[31]Respiratory illness include abscess, adhesion, allergy, asthma, atelectasis, bronchiectasis, bronchitis, cavity, edema, effusion, emphysema, fibrosis, hemoptysis, pleuritis, pneumonia, pneumonitis, pneumothorax, tracheitis and tuberculosis. Digestive illnesses include constipation, diarrhea, dysentery, dyspepsia, dysphagia, enlarged liver, gallstones, gastro-enteritis, malassimilation (malabsorption) and nausea. Infectious illnesses include chancroid, cholera, dengue, diphtheria, gonorrhoea, hepatitis, influenza, malaria, meningitis, mumps, orchitis, parotiditis, pertussis, rickettsia, rubeola, salmonella, scarlet fever, septicaemia, streptococcus, syphilis, tetanus, typhomalaria, and variola. Cardiovascular illnesses include arteriosclerosis, cyanosis, dyspnea, enlarged heart, impaired circulation, murmur, cardiac edema, and palpitation. Endocrine illnesses include diabetes, goiter and enlarged spleen. Genito-urinary illnesses include cystitis, enlarged prostate, nephritis, urethral obstruction, and uremia.

[32]The most important predictors of pension payment for white veterans was the type of disability and the congressional district of the veteran. See Eli [2010] for further explanation.

[33]Shaffer [2004] presents compelling evidence that the medical reviewers at the Pension Bureau held black applications for pensions to a higher standard than those of whites. For example, Shaffer shows that blacks were twice as likely as whites to have their pension application sent to "special examiners" for re-investigation. (p. 129)

[34]Wage rates for blacks in the U.S. do not exist prior to 1940. However, the average monthly income for a farmer (black or white) was $24 in 1900. Therefore, an extra dollar of pension income was equivalent to 4% of monthly income. Pensions likely replaced a higher percentage of monthly income for blacks than whites.

of death by cause. The dependent variable in the regressions is an indicator equal to 1 if a veteran died from a specific cause in year t.

We create a balanced panel, which spans from 1893 to 1906.[35] If a veteran died prior to 1907, all observations after his death are not included in the sample. Since pension income is a function of health status for all of the years contained in the sample, we propose using the number of years that a single surgeon or board of examining surgeons remains in his post as an instrument for pensions. In this section, we outline the way in which we will estimate income semi-elasticities using the instrument. Since we use a non-linear proportional hazard model, we cannot use a two-stage least squares approach to estimate the coefficients.[36] Therefore, we propose the use of the control function approach in which we include pension income, control variables and a residual from the first stage in the hazard regression. The functional form of the first stage is represented in the following way:

$$pen_{it} = \gamma_0 + \gamma_1 initialyear_{it} + \Gamma C_{it} + u_{it} \tag{2.1}$$

where pen = monthly pension income, $initialyear$ = the initial year a single surgeon or board of surgeons began practicing, i = individual, t = year, and C = set of demographic and socio-economic controls. Controls include birth year, battle wounds, previous health conditions, prisoner-of-war status, slave status at birth and population size of the veteran's current city of residence. The associated residual has the following form:

$$\hat{u}_{it} = pen_{it} - \hat{\gamma}_0 - \hat{\gamma}_1 initialyear_{it} - \hat{\Gamma} C_{it} \tag{2.2}$$

In the second-stage, we choose

$$\theta = \{\beta, \alpha, \phi\}$$

to maximize the likelihood function

$$\sum_{i=1}^{N} \sum_{b=1893}^{1906} \left\{ d_{it} log[e^{k_{it}\beta} \alpha t^{\alpha-1}] + (1 - d_{it})(-t^{\alpha} e^{k_{it}\beta}) \right\} \tag{2.3}$$

where

$$k_{it}\beta = \beta_0 pen_{it} + \sum_{j=1}^{N} \beta_j (C_j)_{it} + \phi_1 \hat{u}_{it} + \epsilon_{it} \tag{2.4}$$

We censor on death, d_{it}, which is an indicator equal to 1 if the veteran is alive in year t or if t=1907. Therefore, veterans do not contribute to the likelihood of death

[35] Wherever missing, we impute pension income by setting pensions equal to rates awarded in the last application. In the same way, whenever disease conditions are missing for year t, we assume that a veteran suffers from the same disease conditions claimed in a prior exam if a veteran has not undergone a new exam in year t.

[36] We follow Eli [2010]. For an in-depth discussion of the control function approach, see Imbens and Wooldridge [2007].

from a particular disease after their death. However, a key assumption of the model is that veterans still face a risk of dying from a particular illness even after dying from a different illness.[37] Said differently, the risk of dying from disease A is interrupted by death from disease B.

Since we estimate a proportional hazard model, coefficient are semi-elasticities representing the effects of varying income on the hazard rate. If the coefficient on pension income is negative, then the probability of death from a particular cause decreases as income increases.

2.4 Data

2.4.1 United States Colored Troops

We study these issues using the United States Colored Troops sample compiled by the Center for Population Economics. Information in the dataset comes from three sources: the military, pension and medical records; surgeons' certificates; and census records. The military, pension and medical records contain socioeconomic and demographic data. These records are of particular importance because they contain information regarding the recruit's town and state of enlistment as well as his place of birth, which taken together are used to determine the slave status of the recruit.[38] The surgeons' certificates provide individual-level descriptions composed by examining surgeons of a black veteran's health status at the time of each exam. Lastly, the census records contain all individual-level data on black recruits collected by enumerators in the census years from 1850 and 1930 inclusive.

Unlike the white Union Army veterans, it is unclear whether black veterans were representative of the average black male in the U.S. after the war. While 1 in 3 white men was rejected from enlisting in the Union Army during the war,[39] the sample of veterans contained in the Union Army sample is representative of the sample of white men found in the 1900 Census. For example, the mortality rate of white veterans in 1900 are similar to those of white men in the 1900 Census.[40] However, no other individual-level data for blacks in the late 19th century exists and so analysis of the mortality of black veterans at this juncture provide the first insights into factors causing high black mortality and the black-white health gap in the period.

Of the soldiers who fought for the Union Army, the Fogel group collected records

[37]The model assumes that the true time to failure (death from a particular illness) is not dependent on the censoring event (death from any illness).

[38]Black veterans who enlisted in slave states are counted as having been slaves at enlistment in the dataset. Efforts by the Fogel group to link black veterans, for whom a master's name was listed, to the Slaves Schedules were unsuccessful because of a lack of necessary identifiers to be certain that accurate linkages have been made.

[39]The statistic for black men: 1 in 4 men was rejected (Humphreys [2008], p. 9).

[40]Costa and Kahn [2008], p. 2-3.

for 6,187 veterans spanning 52 companies of the USCT. In 1873, only 167 of black veterans who survived the war collected pensions. In 1893, this number grew to 1,035 veterans (approximately 16.5% of all recruits). In this analysis, I use only the subsample of veterans who were on the pension rolls by 1893.[41]

Despite being subject to the same laws, whites were able to secure higher pensions than blacks. Of veterans who survived to 1893, black veterans received an average monthly pension of $9.64 while whites received $11.92. Black veterans were also less likely to undergo exams. The average black veteran underwent approximately 4 exams while the average white veteran had 5 exams.

Table 2.1 contains summary statistics on cause of death, state of residence and occupation for black veterans, who were on the pension rolls by 1893. We divide the sample of black veterans into two subsamples: those born free and those born into slavery. Despite the fact that approximately three-quarters of black recruits were slaves at enlistment, just over half of them collected pensions in 1893. This finding coincides with the arguments of Shaffer [2004], who uses narrative evidence to show that ex-slaves faced great difficulties in trying to secure a pension relative to free blacks. Of those who were already on the pension rolls in 1893, the majority lived in Maryland and Kentucky. The largest ex-slaves populations of USCT veterans lived in Tennessee (13.49%) and Missouri (9.63%). Therefore, unlike white veterans, most of whom lived in Ohio, New York and Illinois in 1893, black veterans lived in the South and former Border states. In addition, most ex-slave veterans remained in the South after the war. Blacks, who were slaves at birth, were more likely to die from digestive illnesses than those born free. For the remaining causes of death, which include respiratory, cardiovascular and infectious diseases, the rates for ex-slaves and born-free blacks appear to be similar. Lastly, similar to white veterans, the majority of black veterans, both ex-slave and born-free, were farmers in 1900.[42] In addition to being farmers, many black veterans were day laborers and farm laborers.

Table 2.2 contains estimates from probit regressions in which we use wartime and demographic indicators to predict entry onto the pension rolls in both 1873 and 1893 for the sample of black veterans. We find that veterans with wartime injuries were 17% more likely to be on the pensions in 1873 and 7% more likely in 1893. However, only .62% of black veterans in 1893 had wartime injuries as opposed to 34.12% of whites, which quantitatively emphasizes that black saw fewer battles during the war

[41]Sample selection bias restricts me from including veterans who did not collect pensions by the baseline years 1873 or 1893. For an explanation of the selection bias, see Eli [2010]. As in Eli [2010], dividing the analysis into two subsamples – those who collected pensions in 1873 and those who collected in 1893 – would overcome issues of selection bias. However, since only 167 black veterans collected pension in 1873, the sample size is not large enough for regression analysis.

[42]Veterans were asked to provide occupational information during exams. However, on the surgeons' certificate, occupations fields are left black for 94% of black veterans. However, the Fogel group has matched each individual in the sample of black veterans to census data. Therefore, we take occupational information as reported by black veterans in the Census of 1900.

than whites. As expected, POW status and evidence of wartime illnesses[43] increased the likelihood of entry onto the pension rolls by 1893. However, the most important predictor of entry onto the pension rolls by 1893 was whether or not a black veteran was born in the Deep South. Black veterans born in the Deep South were 33% less likely to be on the pension rolls by 1893 which quantitatively confirms assertions made by Shaffer [2004] concerning the frequent inability of ex-slaves to claim pensions.

Figure 2.1 shows the number of exams per year for all black veterans. While many veterans did undergo exams during the 1880s, the overwhelming majority of veterans underwent exams during the early 1890s. The spike in the graph near 1891 coincides with the passage of the Act of 1890, which allowed veterans to claim pensions for disabilities unrelated to the war experience. Figure 2.2 shows the number of first exams per year for black veterans. The spike in this graph near 1890 shows that the majority of veterans did not apply for pensions prior to the passage of the Act of 1890.

2.4.2 Examining Surgeons dataset

In order to determine the number of years a board of surgeons or single surgeon had been practicing prior to a particular veteran's exam, it was first necessary to collect the Pension Bureau's rosters of examining surgeons for each year between 1866 and 1907.[44] Annual rosters contain the name of surgeon(s), the location of the exam (town, county and state) and the reason for surgeon removal (death, resignation, or Congressional removal). Combining the information for all years, we compute the number of years that each examining surgeon held his post.[45] We then match the surgeon(s) to veterans for each exam by linking the surgeon(s) address from the rosters to the address found on each surgeons' certificate in the USCT dataset.[46]

[43]Pensionable wartime illnesses include smallpox, hepatitis, tuberculosis, scurvy, measles, malaria, typhoid, cholera, diarrhea, fever and pneumonia.

[44]The surgeons' roster for 1866 can be found in *Report of the Secretary of the Interior, 1866* on page 434. The roster for 1872 can be found in *Roster of Examining Surgeons, 1872* on Google Books. The roster for 1875 can be found in *Roster of Examining Surgeons, 1875* found at the National Library of Medicine. Rosters for 1880 - 82, 1885, 1889, 1890, 1894, 1898, and 1902 are found on microfilm at Northwestern University Library. The roster for 1906 can be found in the *American Medical Directory, 1906* on pages 12 25 from Google Books. Since annual rosters are missing for some of the years, we have begun digitizing *The Register of Boards of Examining Surgeons, 1862 - 1928* found at Archives I, which will be completed in early 2012.

[45]Collecting this information from the *Register of Boards of Examining Surgeons, 1862 - 1928* available at Archives I would be ideal since this record contains the exact start and end dates for each surgeons' term. The annual rosters published by the Pension Bureau that we use in this analysis contain the starting and ending year (but not the month and day) of each surgeons' term.

[46]It is not possible to match veterans to surgeons by linking the surgeons' names from the surgeons' certificate to names in the roster because the Fogel group did not collect the surgeons' names from the surgeons' certificates. For this reason, we create linkages using addresses.

2.4.3 City and County-Level data

To determine the population size of each veteran's town of residence, we use population estimates from the U.S. Census Bureau.[47] Exact street addresses for veterans come from the surgeons' certificates, and the town of residence for each veteran is linked to Census population estimates for the associated year. Since the Census Bureau collected population counts decennially, we link population counts to the nearest census year.[48]

2.5 Results

In Table 2.3, we present coefficient estimates of key determinants of monthly pension income, the year of first application and the year of first pension receipt. On average, older veterans receive higher pension payments, apply earlier and receive their first pensions earlier than younger veterans. In particular, being one year older increases monthly pension income by an average of 10 cents. Due to legislated increases in pension amounts, pension payments increase over time in the period between 1893 and 1907. Accordingly, the coefficient on the time trend is positive and significant for the regressions in which monthly pension income and first application year are the dependent variables. As expected, war injuries are significant predictors in all three regressions. Relative to a veteran who did not sustain an injury during the war, those who did received an extra $1.29 in monthly pension income on average. In addition, they first applied for a pension 5.7 years earlier than those who were not injured.

Table 2.3 also shows coefficient estimates for birthplace indicators. Relative to black veterans born in the West, those born in the Deep South received an average of $1.57 less monthly pension income. Given that all blacks in the sample received an average of $9.64 (almost $2 less than whites), veterans born in the Deep South received an average of $8.07. Therefore, even when successful at gaining entry onto the pension rolls, those born in the Deep South received nearly $3 less than whites on average. The pension differential between these two groups is equivalent to 12.5% of a white farmer's income in 1900. Additionally, those born in the Deep South first applied over two years after those born in the West and were admitted onto the pension rolls over three years after those born in the West. As outlined in the historical background section of this paper, those born as slaves often could not prove their identities after the war and were more likely to be illiterate and thus less able to advocate for themselves during the pension application process.[49] The effect of slave

[47]The U.S. Census Bureau's Population of 100 Largest Cities and Other Urban Places in the United States: 1790 to 1990 is available at http://www.census.gov/population/www.

[48]For years ending in the number 5, we link to the following census year. For example, population estimates for 1895 are taken from the estimates reported in 1900.

[49]Shaffer [2004]

status at birth on pension receipt however is stark. Figure 2.3 shows the average monthly pension income per year for veterans already on the pension rolls in a given year. In every year after 1870, those born in the Deep South received lower pensions than those born elsewhere (even compared to those born in other parts of the South or in Border States).

To determine the effect of pension income on cause-specific mortality and all-cause mortality from any cause, we estimate a hazard model of mortality and compare coefficient estimates in Table 2.4. While the coefficient estimates presented are biased due to the fact that we did not instrument for pension income, the results are still informative. First, hazard rates for respiratory and infectious illnesses in the white and black veteran population are identical. In addition, blacks have lower rates of death from digestive and cardiovascular illnesses. However, the hazard rate of death from any cause for the black veteran population is 50% higher than that of the white population. These facts taken together suggest that the specific causes of death from cardiovascular and digestive illnesses may be underreported for black veterans.

Table 2.4 shows that the estimated effects of monthly pension income on the hazard rate of death from any cause is positive for both blacks and whites. Since there is no correction for reverse causality, these estimates are biased. However, a comparison of the black-white differentials in coefficient estimates is revealing. The black-white differential of the coefficient estimate in the hazard regression for mortality from any cause is positive (.006) suggesting that the coefficient estimates for blacks are even more biased than for whites. Said differently, the unobserved severity of health conditions was a stronger predictor of pensions for blacks than whites.

Coefficient estimates on monthly pension income for cause-specific mortality are also presented in Table 2.4. Coefficient estimates for respiratory illnesses are similar for blacks and whites, however the estimate is not significant for the black sample. Interestingly, when the dependent variable is death from cardiovascular illness, the coefficient estimate is positive and significant. The black-white differential in coefficient estimates is positive and large (.029) suggesting that the endogeneity bias is larger for the black sample than the white sample.

2.6 Conclusion

In this research, we set out to investigate the effect of pension income on the black-white gap in mortality rates at the end of the 19th and early 20th century. Because pension income depends partly on health status, pension income is endogenous with respect to mortality. We show empirically that coefficients on monthly pension income are more biased in hazard regressions of the black sample relative to those of the white sample. Taken together, these estimates suggest that on average, black pensioners were sicker than white veterans.

The estimates from our hazard regression also call attention to the need for an

exogenous source of variation in pension income. The instrument used in Eli [2010] – the Republican Congressional vote share in a veteran's district – cannot be used in any analysis using the sample of black veterans because the majority of black veterans lived in the South and were thus unable to vote. Therefore, we propose a new instrument which is applicable to the black sample of veterans: the initial year in which a single surgeon or board of examining surgeons began practicing as an examining surgeon for the Pension Bureau. Single surgeons or boards hired later in the period were more likely to be younger and to have graduated from medical school after surgeons who began were appointed as examining surgeons earlier in the period. Therefore, we argue that racial attitudes towards the black body evolved so that, controlling for the year of the exam, surgeons or boards appointed later in the period were less likely to recommend veterans for pensions or pension increases.

While Wilson [2007] shows the difficulty black veterans faced when trying to claim pensions for the first time, we document the black-white differential in pension income as well as the origins in the income differential between blacks who were born in the Deep South as opposed to those who were born elsewhere. We show that being born in the Deep South was the greatest hindrance to securing a pension equal to that of blacks born in other states in the South. Shaffer [2004] argues that ex-slaves (identified as blacks born anywhere in the South) were frequently unable to prove that they did in fact serve in the war because of name changes after the war or misspellings made my clerks during enlistment. However, we find that blacks born in the South, but outside of the Deep South, did not face any additional difficulty in securing a pension equal to that of blacks from the Midwest, Northeast, Border States or the West. Therefore, Shaffer's arguments do not fully explain pension differentials.

The question remains then whether pension income had a greater effect on mortality for blacks than whites. Given that black veterans were sicker than white veterans on average, as shown in Costa et al. [2007], it may be that income had little effect on mortality rates. However, Costa [2010] shows that increases in pension income had greater effects on retirement for blacks than whites, which suggests that the effect of income on mortality risk may also be greater for blacks than whites. Costa et al. [2007] shows that black veterans were at a higher risk of mortality from cardiovascular conditions relative to whites due to higher rates of acute infections in childhood. Given this evidence, it is still unclear whether increased income could help black veterans overcome early-life penalties or whether these early-life health shocks were permanently scarring (and therefore causing income to be ineffectual). Furthermore, if income is effectual, which diseases are most income sensitive for blacks veterans and how do they compare to findings in Eli [2010] for white veterans? In future research, we plan to use our proposed instrument to answer the questions.

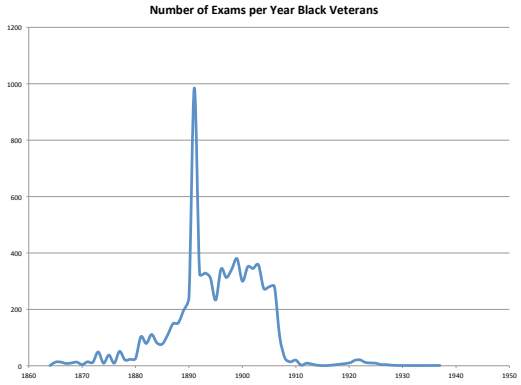


Figure 2.1: Number of Exams per Year for Black Veterans

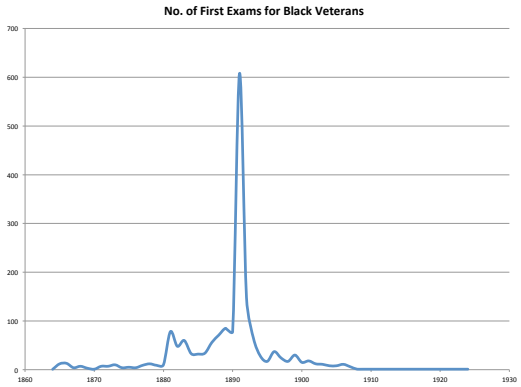


Figure 2.2: Number of First Exams per Year for Black Veterans

Table 2.1: Summary Statistics for Black Union Army Veterans

	Mean (%) in 1893 for veterans on pension rolls	
	All	Slave at Birth
Causes of Death		
Respiratory	9.30	9.40
Digestive	2.55	1.80
Cardiovascular	11.36	11.03
Infectious	2.84	2.71
State of Residence in 1893		
Alabama	4.44	8.09
Arkansas	4.44	6.94
District of Columbia	5.27	5.59
Kansas	4.75	5.39
Kentucky	11.16	2.89
Maryland	15.91	3.28
Missouri	7.44	9.63
Ohio	5.86	4.62
Tennessee	8.16	13.49
Virginia	4.96	8.48
All Other	27.61	31.6
Occupation in 1900		
Barber	1.15	2.62
Carpenter	1.84	1.46
Day Laborer	16.92	10.50
Farm Laborer	5.98	1.75
Farmer	28.31	46.94
All Other	45.20	36.73
No. of Veterans	1021	553

Table 2.2: Probit Regression - Predicting Veteran's Entry onto Pension Rolls

	Date of Entry onto the Pension Rolls			
	1873		1893	
	Coefficient	Mean (%)	Coefficient	Mean (%)
War Injury	.170***	17.20	.073**	0.62
	(10.83)		(2.09)	
War Illness	.016*	59.81	.081***	59.85
	(1.91)		(2.98)	
POW	.055	3.74	.225***	3.87
	(1.63)		(3.68)	
Birth Year	-.001**	1838	-.016***	1839
	(-2.09)		(-8.43)	
Born in Northeast	.026	3.05	-.045	7.30
	(1.09)		(-0.59)	
Born in Midwest	-.010	7.53	-.051	7.68
	(-0.47)		(-.059)	
Born in Border	.011	64.05	.010	64.30
	(0.59)		(0.14)	
Born in Deep South	-.042***	19.46	-.330***	20.53
	(-2.64)		(-4.24)	
Born in South	-.016	29.13	-.108	28.00
	(.355)		(-1.40)	
Fair Skinned	.027*	9.47	-.040	9.49
	(1.90)		(-0.91)	
No. of Veterans on rolls	167		1035	
Total no. on veterans	1953		1549	

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
The omitted birth category is the West.

Table 2.3: OLS Estimates of Pension Receipt for Blacks on Rolls by 1893

	Mo. Pension ($)	1st Application Year	Year of 1st Receipt
Birth Year	-.102***	.033***	.035***
	(-14.17)	(3.90)	(5.04)
Time Trend	.153***	.027*	.017
	(10.87)	(1.66)	(1.25)
War Injury	1.29***	-5.685***	-5.168***
	(8.20)	(-31.22)	(-34.64)
War Illness	.145	-.193	-.060
	(0.97)	(-1.11)	(-0.42)
Birthplace:			
Deep South	-1.571***	2.263***	2.764***
	(-4.70)	(5.88)	(8.78)
South	-.319	1.070***	1.706***
	(-1.04)	(3.02)	(5.87)
Mean	$9.64	1888	1888

Note: No. of veterans: 903. Restricted to observations in years after 1892 and before 1907, Controls include birthplace, POW status and size of enlistment city. Standard errors in (). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
The reference birthplace is the West.

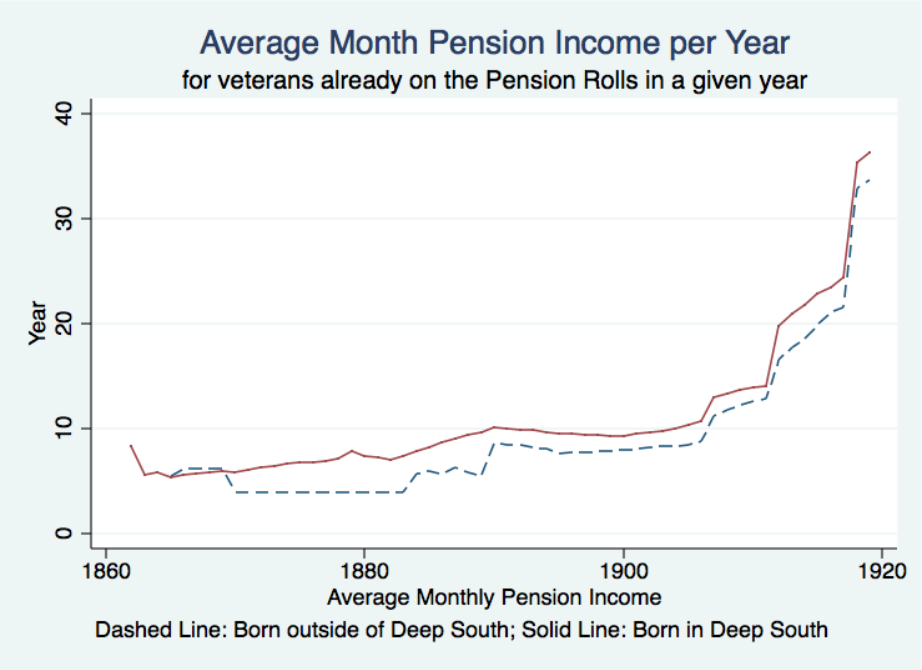


Figure 2.3: Average Monthly Pension for Black Veterans

Table 2.4: Hazard Estimates for Mortality for Black and White Veterans on Pension Rolls by 1893

	Coeff. on Monthly Pension		Hazard Rate	
Sample:	Blacks	Whites	Blacks	Whites
All-cause Mortality	.016**	.010***	.042	.028
	(.020)	(.001)		
Respiratory Illnesses	.015	.014***	.005	.005
	(0.48)	(.003)		
Infectious Illnesses	.045***	.008	.002	.002
	(.004)	(.005)		
Digestive Illnesses	-.011	.0143	.001	.007
	(.869)	(.002)		
Cardiovascular Illnesses	.038***	.009***	.004	.013
	(.003)	(.001)		
No. of veterans	903	13011	903	13011

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Controls: Top 100 city (in population size), time trend, birth year and place, rank in the army, prisoner-of-war status, war wounds and previous illnesses.

Chapter 3

In Search of India's Missing Calories: Energy Requirements and Calorie Consumption (with Nicholas Li)

3.1 Introduction

Deaton and Dreze [2009] use India's National Sample Survey to document a large (10%) decline in mean calorie consumption per capita over the 1983 to 2005 period. This is especially puzzling because the same data indicate significant growth of real expenditure per capita during the period (30%) and imply a positive cross-sectional calorie-expenditure elasticity (0.3), so we would have expected a significant increase in per capita calorie consumption (9%). The same pattern of higher real expenditures and equal or lower calorie consumption is observed between urban and rural households. Understanding the source of these "missing" calories is critical for our understanding of poverty and well-being in India. Poverty measures based on caloric norms would indicate a dramatic increase in poverty rates and greater poverty in urban than rural areas, contrary to conventional wisdom.

Deaton and Dreze [2009] propose the hypothesis that declining household energy requirements may explain the decline in caloric intake. Energy requirements may fall due to substitution of machine or animal power at work, structural change towards less energy-intensive occupations, infrastructure improvements, labor-saving appliances, and improvements in the health environment that increase nutritional absorption. This hypothesis is consistent with evidence of slow but steady progress from anthropometric measures (child heights and weights) and what we call an increase in food quality - the substitution away from staples that are cheap sources of calories, like grains, towards foods that are more expensive per calorie, such as dairy

products or processed foods. Rather than impoverishment, caloric intake differences driven by lower energy requirements would imply greater welfare and lower poverty.

We contribute to this debate by analyzing the effect of variable energy requirements on food consumption patterns in India. We first use a simple model to motivate a decomposition of the calorie-expenditure relationship into a food quality and a food/non-food component, emphasizing that food quality may be a better indicator of variable energy requirements when food is separable in the household utility function and many difficult to measure forces affect the food share budget share. We then use time-use data to directly impute caloric requirements, allowing us to explore the covariation of intake and requirements with respect to household characteristics like age, size, education, and occupation. We ask whether these characteristics are sufficient to eliminate the unexplained calories across sectors and over time. Finally, we provide a preliminary assessment of the effects of variable caloric requirements on poverty and welfare in India.

The decomposition of the calorie-expenditure Engel curve into food quality and food/non-food implies that the urban-rural missing calories are largely explained by differences in food quality, with urban households consuming foods that are expensive per calorie. The missing calories between 1983-1993 can also be explained by a transition to more expensive calories holding real food expenditures constant. By contrast, the 1994-2005 period is characterized by relatively stable food quality but a large decline in food expenditures for most levels of expenditures (e.g. a downward shift in food Engel curves).

Our main innovation is to use time-use data for six Indian states to impute caloric requirements at the individual and household level. While there are limitations to our imputation procedure, our measure of caloric requirements has significant advantages to those provided by the India Council of Medical Research (ICMR), allowing us to examine the impact of variables like household size, age, hours worked, and home production that are not captured by the standard caloric requirement classification. We are also able to use identically defined household characteristics to compare caloric intake and requirements across time-use and consumption surveys. Our results indicate that demographic composition, age, education, occupation all have significant impacts on caloric intake and requirements, with education and occupation being especially important for explaining differences over time and between rural and urban areas. Ownership of labor-saving durables also has a significant impact on caloric intake conditional on total expenditures or food expenditures.

Quantitatively, when we control for these household characteristics we explain as much as 61% of the missing urban calories, rising to 100% when we look only at the food quality component. The missing calories over time are less affected by these household characteristics, declining by only 21% in rural areas and 15% in urban areas, rising to 42% and 60% respectively for the food quality component. Our results suggest that while proxies for energy requirements can explain most of the missing calories due to food quality, we may need other explanations for the general

downward drift in food Engel curves observed in India and other countries. Changes in relative prices, the availability of new and better quality non-food goods, or other factors may play an important role.

We conclude with a preliminary assessment of the implications of variable caloric requirements for measurement of poverty and welfare. Poverty rates based on caloric inadequacy tend to be much higher than official rates, and variable caloric requirements leads to significant differences in poverty rates across states and across occupations within states. While our model is not very realistic, it implies a simple Engel equivalence scale that suggests that urban areas are up to 5% better off (expenditure-equivalent) due to lower caloric requirements, an effect largely driven by occupation as sedentary workers have up to 5% greater welfare than primary sector workers. This effect is large relative to the measured urban-rural average real expenditure gap of about 30%. The gains over time that we attribute to caloric requirements are smaller at around 2%-3% (relative to growth of real expenditures of over 30%).

Though our primary contribution is to explore the source and implications of India's missing calories, we make two additional contributions that extend beyond the Indian context. We provide evidence that household economies of scale in energy requirements may explain the Barten puzzle discussed in Deaton and Paxson [1998].[1] We also find evidence that caloric requirements may play an important role in life-cycle consumption patterns for poor countries.

Unlike the findings of Aguair and Hurst [2005] for the United States and Hicks [2010] for Mexico, we observe large declines in both caloric requirements and intake for older Indian households, with a limited role for consumption smoothing through greater home-production and shopping intensity.

Our paper is also related to the macroeconomics and history literature that uses food consumption patterns to measure welfare and price index bias, including Costa [2001], Hamilton [2001], and Almas [2008]. Assuming relative prices are stable (or can be accurately measured along with their elasticities), shifts in food Engel curves can be used to compute the true cost-of-living index and the bias in official price indexes. If caloric requirements shift food Engel curves, they should be considered another source of "bias" detected by these methods, one that will be particularly important when comparing countries or periods with very different levels of economic development. [2]

[1]The Barten puzzle is can be briefly stated as follows. When there are household economies of scale, larger households are better off holding per capita expenditures constant. Provided private and public goods are not too substitutable, the income effects imply that the share of household expenditures on private goods like food should rise. Data from poor and rich countries shows the opposite, with larger households spending a smaller share of budgets on food, and the effect is larger in developing countries (where substitution effects are likely to be lower). Our findings indicate that, at least for food, the results may be driven by economies of scale in caloric requirements, which are declining in household size at constant per capita expenditures.

[2]Li [2010] uses a different approach to measure welfare gains from greater food variety that can be interpreted as gains from food quality. Those gains are distinct from the ones discussed in this

In section 3.2 we present the model that motivates our empirical analysis. Section 3.3 examines the robustness of the decline in calories and its sources. Section 3.4 links energy intake and requirements to household characteristics in the cross-section. Section 3.5 quantitatively assesses the energy requirements hypothesis as an explanation for the missing calories. Section 3.6 considers implications for poverty and welfare, and section 3.7 concludes.

3.2 Theory

We consider a simple model in which households directly trade-off utility from food quality, food quantity (calories) and non-food, with potentially variable energy requirements. The model assumes additive separability for tractability and is thus not very realistic, but captures the intuition of more systematic approaches to measuring demand for food characteristics rather than individual food quantities or budget shares (see Bouis [1995] for a detailed Food Characteristic Demand System). Our empirical decompositions only require the assumption of weak separability between food and non-food to ensure that a decomposition of calorie Engel curves into a calorie - food expenditure and a food expenditure - total expenditure component is not only interesting, but also potentially informative about a household optimization problem.[3]

Suppose the consumer food subutility problem is given by:

$$\max_{Q,C} \left(C - \bar{C}\right)^{\alpha} (Q)^{1-\alpha} \;\; s.t.\; p_c C + Q \leq X_f \tag{3.1}$$

where Q is food quality, $\bar{C}$ is the minimum calorie requirement [4], C is calories con-

paper as they occur due to within-group substitution and are only partly related to caloric intake. Li [2010] suggests that the size of variety welfare gains that could plausibly be related to caloric requirements are about 1% to 2% in food expenditure equivalent for both the urban-rural gap and the change over time, while the total gains range from over 2% for urban versus rural households to 10% over time. The welfare gains we estimate here operate only through the effect of our proxies for caloric requirements on the food share, and much of the change in food quality that we observe occurs due to across-group substitutions.

[3]Although we later use the model for a preliminary welfare analysis in the spirit of the Engel equivalence scale literature, the results should not be taken too seriously. Even with a more flexible functional form the use of food share to calculate an expenditure-equivalent welfare gain would be subject to the usual criticisms about non-identification of equivalence scales (Lewbel and Pendakur [2007]) and unobserved relative-price effects (which implies to all work in the spirit of Costa [2001],Hamilton [2001], and Almas [2008]).

[4]We model $\bar{C}$ as exogenous, but a simple extension of the model allows caloric requirements to be an input into the expenditure-generating function. This has several implications, most notably, occupations that are more dependent on caloric inputs to generate expenditures will have steeper sloping (and higher) calorie-total expenditure Engel curves. They will also have lower welfare inequality relative to expenditure inequality due to the positive expenditure - caloric requirement correlation. We leave exploration of this aspect of the model for future work, but note that it implies that the

sumed, p_c is price of calorie quantity relative to quality (the price of food quality is normalized to one) and X_f is food expenditure. This functional form is restrictive but the intuition would carry through with more general preferences provided households were maximizing a food utility function that is increasing in both calories and another characteristic (like quality), and for which the relative demand for calories rises with energy requirements.[5]

Solving this problem yields a **calorie-food expenditure Engel curve (CF)**:

$$\textbf{CF: } C = \alpha \frac{X_f}{p_c} + (1-\alpha)\bar{C} \tag{3.2}$$

that is increasing in food expenditures and minimum energy requirements and decreasing in the price of calories relative to food quality. While Q may be unobserved, we have a proxy through the expression for food expenditures per calorie:

$$\frac{X_f}{C+\bar{C}} = \frac{1}{\frac{\alpha}{p_c} + (1-\alpha)\frac{\bar{C}}{X_f}} \tag{3.3}$$

which is increasing in food expenditures but *decreasing* in energy requirements. The model unrealistically predicts a unitary elasticity of calories to food expenditure at high expenditure levels but over a lower range of expenditures could generate an elasticity closer to 0.4-0.6, the empirical range in the NSS data.[6]

Total utility from food is given by

$$U_f = \frac{\alpha^\alpha (1-\alpha)^{1-\alpha}}{p_c^\alpha}\left[X_f - \frac{1-\alpha}{\alpha}\bar{C}p_c\right] \tag{3.4}$$

which is increasing in food expenditures and decreasing in minimum energy requirements. A higher price for calories lowers food utility by making it more expensive to meet minimum requirements and consume additional calories.

measure of welfare advocated by Logan [2009] may be flawed. Logan [2009] suggests that the slope of calorie Engel curves is an intuitive measure of hunger or welfare and potentially superior to using budget share or total calorie consumption, as it takes into account the marginal propensity to consume on a basic necessity that should be falling in the standard of living. However, if the generation of expenditures is more calorie intensive in some areas and periods, this can result in steeper calorie Engel curve slopes regardless of the actual level of welfare of the population, though it will still be the case that *conditional on expenditures* a locally steeper calorie Engel curve slope implies lower welfare.

[5]See Li [2010] for a model in which households could trade-off variety and calories, with average quality effects (measured as expenditures per calorie) arising due to the higher cost per calorie of non-staple varieties.

[6]That the elasticity of calories to food expenditure does not decline over most of the expenditure range in the NSS data may seem surprising. However, one must keep in mind that India has non-trivial obesity rates (as high as 30% in some states) despite even the richest (95th percentile) households having incomes that are low on a global scale (below USD 9,000).

We model demand between food and non-food with a CES function of their sub-utilities:

$$U = \left(U_f^{\frac{\sigma-1}{\sigma}} + U_{nf}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{3.5}$$

with budget constraint $X_f + Q_{nf}(U_{nf})p_{nf} \leq Y$. The price of non-food relative to food is p_{nf} and total expenditure is Y. Let $U_{nf} = \gamma_{nf} Q_{nf}$ where γ_{nf} represents some combination of exogenous shifters that affect preference for non-food, including preferences, quality, and variety. Substituting this and equation 3.4 we have the consumer problem:

$$\max_{X_f, Q_{nf}} \left(\left[\frac{\alpha^\alpha (1-\alpha)^{1-\alpha}}{p_c^\alpha} \left[X_f - \frac{1-\alpha}{\alpha} \bar{C} p_c \right] \right]^{\frac{\sigma-1}{\sigma}} + \left[\gamma_{nf} Q_{nf} \right]^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{3.6}$$

subject to $X_f + Q_{nf} p_{nf} \leq Y$.

Denoting $\gamma_f = \frac{\alpha^\alpha (1-\alpha)^{1-\alpha}}{p_c^\alpha}$ and $\bar{C}^* = p_c(\frac{1-\alpha}{\alpha})\bar{C}$, the **food expenditure - total expenditure Engel curve (FE)** is:

$$\textbf{FE: } X_f = \frac{Y + (\frac{p_{nf}\gamma_f}{\gamma_{nf}})^{1-\sigma} \bar{C}^*}{1 + (\frac{p_{nf}\gamma_f}{\gamma_{nf}})^{1-\sigma}} \tag{3.7}$$

with non-food demand given by $Q_{nf} = \frac{Y - X_f}{p_{nf}}$. If food and non-food are substitutes ($\sigma > 1$) then food expenditure increase in the price of non-food p_{nf} and decrease in the taste shifter for non-food γ_{nf}. Note that the calorie-total expenditure Engel curve (CE) can be derived by substituting equation 3.7 into equation 3.2.

Figure 3.1 provides a graphical interpretation of the effect of a fall in the relatively price of non-food (or alternatively an increase in taste, quality, or variety for non-food relative to food). The CE and FE curves both shift down, but there is no effect on the CE or food quality curves. However, the decline in food expenditures implies that there is movement along the CE and quality curves towards the origin, with both calories and quality (food expenditures per calorie) declining. Supposing the price of cell phones falls, the value of cell-phones rises because of network economies, or the returns to education rise, household will spend less of their budgets on food at any level of expenditure, and this downward adjustment in food leads to a decrease on both a (calorie) quantity and quality margin.

Figure 3.2 depicts a fall in energy requirements. While the CE and FE both shift down, in this case there is also a shift down in the CF curve and a shift up in the food quality curve. The decline in food expenditures reinforces the downward shift in the CE curve, causing movement along the new CF curve (CF_2) towards the origin and making calorie consumption fall further. However, the decline in food expenditures is more than offset by the shift up in the food quality curve, resulting in an increase in food quality at any level of food expenditure. Thus the characteristic

that distinguishes energy requirements from other shifts is that they affect the CF curve and predict increases in food quality, because the downward adjustment in calories is larger that the downward adjustment in food expenditures.

The other interesting comparative static in the system is a change in the relative price of calories relative to quality (p_c). The effect of an increase in p_c is given by:

$$\frac{\partial X_f}{\partial p_c} = -[Y - \bar{C}^*]\frac{1}{(1+Z)^2}\frac{\partial Z}{\partial p_c} + \frac{Z}{1+Z}\frac{\partial \bar{C}^*}{p_c} \tag{3.8}$$

where $Z = (\frac{p_{nf}\gamma_f}{\gamma_{nf}})^{1-\sigma}$. Because $\frac{\partial \bar{C}^*}{\partial p_c} > 0$ and $\frac{\partial Z}{\partial p_c} > 0$ the sign of the expression above depends on the levels of Y and $\bar{C}$. For households with high expenditures relative to energy requirements the first effect dominates and food expenditures fall when the relative price of calories rises.

The model implies that an evaluation of welfare requires measurement of both total expenditures and minimum energy requirements, as indirect utility is given by

$$V = \Omega[Y - \bar{C}^*] \tag{3.9}$$

with $\Omega = \frac{1}{1+Z}\left[\gamma_f^{\frac{\sigma-1}{\sigma}} + \left(\frac{\gamma_{nf}Z}{p_{nf}}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$. Rather than measuring $\bar{C}^*$ directly, we can rearrange equation 3.7 to get $\bar{C}^* = \frac{X_f(1+Z)-Y}{Z}$ and substitute this in to get indirect utility

$$V = \frac{1+Z}{Z}Y[S_{nf}] \tag{3.10}$$

where S_{nf} is the budget share of non-food. We can then use the budget share of non-food to compute Engel equivalence scales in this model. Holding other model parameters constant, households with a lower food share (higher non-food share) due to lower energy requirements have higher utility.

We summarize the implications of our model:

1. The calorie-expenditure Engel curve (CE) can be decomposed into two pieces, the calorie-food expenditure Engel curve (CF) and the food expenditure - total expenditure Engel curve (FE).

2. Lower caloric requirements manifest themselves through a downward shift in both the CF and the FE, raising food quality.

3. Lower relative prices or greater taste/variety/quality of non-food causes a downward shift in the FE with no effect on the CF, lowering food quality..

4. Welfare comparisons that do not adequately account for minimum energy requirements are incomplete, but a simple Engel equivalence scale using differences non-food budget shares orthogonal to real expenditures can be used to analyze the effect of calorie requirements on welfare.

3.3 Decomposing the decline in calories

3.3.1 Calories per capita

The National Sample Survey (NSS) provides the most comprehensive measure of food quantities and expenditures by Indian households. Consumption is measured at the household level using a 30-day recall period.[7] While there are other nutritional surveys in India that cover individual villages or smaller regions, the only other dataset with a wide scope comes from the National Nutritional Monitoring Bureau (NNMB) which covers several states and records calorie data using direct weighing of food over a 24-hour period.

Deaton and Dreze [2009] document the decline in caloric intake in India between 1983 and 2005 in the NSS data. Table 3.1 presents their estimates of per capita calorie consumption over this period. We also present independent estimates calculated by other authors with the same data - surprisingly the different studies disagree on both the direction and magnitude of calorie changes.[8] While Deaton and Dreze [2009] find a large decline in rural areas and modest decline in urban areas, Chatterjee et al. [2007] find a decrease in rural areas and an increase in urban areas, while Kumar and Dey [2007] find an increase in both areas. Both Kumar and Dey [2007] and Chatterjee et al. [2007] find that in recent years urban India has higher per capita consumption of calories than rural India. Below we also report calorie intake from the NNMB as calculated by Deaton and Dreze [2009]. These data, presented at the bottom of table 3.1 show a dramatic decline in calories that is over the double the size of the decline for comparable states in the NSS based on the calculations of Deaton and Drèze.

In the data appendix we discuss our construction of calorie estimates their sensitivity to various imputation assumptions. The three main issues are (1) treatment of food with missing or imprecise quantity data (whose caloric conversions per quantity may be certain), (2) composite or processed food items with unknown calorie conversion factors (even though quantity may be precise), and (3) meals to/from others that bias the numerator or denominator of a household calories per capita. Our preferred calorie estimates, presented at the bottom of table 3.1, use direct calorie conversion whenever possible and make adjustments for meals to/from others. The "group" estimates impute calories for items missing quantities or conversion factors with an aggregate expenditure weighted average of calories per rupee at the group level. The "all food" estimate uses household average calories per rupee divided by two. The measures agree quite closely though the imputation method is quite different. Our

[7] The notable exception is the 55th (1999-2000) survey round which used a 30-day and a 7-day recall period. Critics observed that using a shorter additional recall period biases upward consumption measures over the 30-day period, leading to overestimation of the decline in poverty. See Deaton and Dreze [2002] or Deaton and Kozel [2005] for discussion. When pooling multiple rounds our results are not sensitive to excluding the 55th round entirely.

[8] As none of the studies make explicit the details of data-cleaning and calorie imputation we cannot pinpoint the reason for the divergent estimates.

estimates are close to those of Deaton and Drèze, showing a much larger decline in rural than urban areas and consequent convergence of rural and urban calories per capita.

Figure 3.3 presents kernel density estimates of log per capita calorie consumption using our preferred "group" estimates. In 1983 the urban calorie distribution is shifted left of the rural one, but over time the have compressed and grown similar. The large decline in mean calories per capita in rural areas is thus driven by a right-tail that shrinks much more than the left tail, while for urban households there is a smaller decline as the compression is symmetric.

3.3.2 Calorie Engel curves

Figure 3.4 non-parametrically plots log calories against log real expenditures for different periods and sectors.[9] We restrict the sample to households with an adult (over age 15) male and female with three children. The calorie Engel curves (CE) are upward sloping throughout with a slight decline at the top (the linear estimated slope is around 0.3). Given the rise of real expenditures over 1983-2005, the only way calories could have decreased is with the large shift down over time shown in figure 3.4. For any given year the urban CE lies below the rural CE, explaining how the on average richer urban households consume less calories on average. The downward shift in the curves over time has been greater at the upper end of the expenditure distribution, consistent with the inward shift of the per capita calorie distribution upper tail in figure 3.3.

The model from section 3.2 suggests a decomposition of the CE into a food-total expenditure (FE) and a calorie-food expenditure (CF) Engel curve. Figure 3.5(a) presents the FE curve. While the urban curves are slightly below the rural curves for 1983 and 1993, there is virtually no shift for rural or urban households over the period. There is a large downward shift in the 1994-2005 period, especially at the top of the expenditure distribution. As the CE falls throughout, a downward shift in the CF curve explains virtually all of the decline in calories between 1983-1993. Figure 3.5(b) documents the downward shift over the 1983-1993 period and stability over the 1994-2005 period. The difference between rural and urban areas remains large throughout the entire 1983-2005 period, so rural households consume more calories per rupee of food expenditure. Because the food expenditures are adjusted for rural/urban prices this is not the result of higher average urban prices and must be due to a compositional effect. Our decomposition highlights the potential for multiple explanations of the decline in calories as the source of the shift differs over the 1983-1993 and 1994-2005 periods.

[9]Our total expenditure measure excludes taxes, water charges, and rent to be comparable across all rounds. We use survey-based Tornqvist price indexes (base rural 50th round) to deflate expenditures, with median unit values as prices. This covers 58-83% of aggregate expenditures depending on the sector and survey year. We also trim the 1% tails of the calorie and expenditure distribution.

Before examining food composition we briefly consider the composition of total expenditures and reasons for the downward shift of FE curves over 1994-2005. Figure **??** presents plots of budget share - log real expenditure Engel curves for different categories - food, clothing, fuel and light, intoxicants, medical care, education, entertainment services, transport services, other services, other nondurables and durables. Many of the Engel curves are highly non-linear and differ greatly across sectors, but as the shares sum to one at any level of real expenditure some must shift upward if food shifts downward. For rich households (with three children), education spending has the largest increase accounting for up to half the decline in food share. Durables are also important for these households. Other services, transport services, and fuel and light increase for all households. Poor households have larger increases in other nondurables and clothing.

The surveys provide unit values for clothing, fuel and light, and intoxicants so we can examine relative prices for these goods with the caveat that quality effects on unit values are large for clothing and intoxicants. Over the 1983-2005 period food prices rose 396% (423%) in rural (urban) areas, compared to 459% (381%) for the combined clothing, fuel and light and intoxicants price index. Note that although the relative prices of the food and non-food composites moved in opposite directions in rural and urban areas, the food Engel curves shift down for both. Looking at the individual components, clothing rose by 518% (382%), intoxicants by 559% (731%) and fuel and light by 342% (276%). The rising share of fuel and light may be related to lower relative prices.[10]

Downward drift in food Engel curves over time has been observed in the United States (Costa [2001],Hamilton [2001]) and in the United Kingdom and Spain[11] Some studies interpret this drift as CPI bias which upwardly biases the price indexes used to deflate nominal expenditures. Disaggregating by group suggests that it is difficult to rule out other factors like prices, variety, quality and tastes, to which we add caloric requirements. Group Engel curves do not shift in parallel or follow a systematic pattern based on expenditure elasticities, implying that substitution effects are important and Engel curve shifts may be problematic for estimating bias in income.

3.3.3 Changes in food composition

Mean real food expenditure changed little over the 1983-2005 period. Table 3.3 documents real food expenditure growth of 6.7% (5.7%) in rural areas and 9.1% (7.3%)

[10]While relative prices (including relative returns to education) may explain many of these shifts, the introduction and diffusion of consumer goods may also play a role (Li [2010]). While the number of distinct food and clothing items per household rises, the increase for durable and nondurable goods (personal care and effects, toilet articles, and sundry) is twice as large. Consumption growth in these categories may be due to in part to an increase in the range of goods available due to advances in advertising, infrastructure and retailing.

[11]Our calculations.

in urban areas depending on whether we use a base-period weighted or Tornqvist price index. The table also presents the differences in calories per real food expenditure over time and across sectors (normalized by 1983 rural sector) that are consistent with figure 3.5(b). Urban households consume significantly less calories per real food expenditure and there is a decline over time in both sectors concentrated in the earlier 1983-1993 period. The size of the decline is similar to figure 3.5(b) but here we pool all households regardless of size and make calculations based on the average household.

Figure 3.7(a) presents a pie chart for ten different food groupings that emphasize the importance of grains in food expenditures, especially for rural households, and their declining importance over time. Figure 3.7(b) shows that grains are even more important as a share of total calories. Linking these two figures together are the differences in calories per rupee of the different food groups, presented in table 3.2. We normalize the price per calorie using rice and wheat, the two most important foods in India, which together make up between 22-40% of food expenditures and 46%-56% of total calories across the years and sectors in our sample.

Three different factors could affect total calories per real rupee of food expenditure - (1)changes in the food budget share of different food groups, (2)changes in inter-group relative prices, and (3)changes in composition within food groups (including the effects of intra-group relative prices). This suggests a simple (non-additive) decomposition. We can write mean calories as:

$$cal = \frac{X_f}{P_f} \sum_g s_g P_g (\frac{cal}{rupee})_g \tag{3.11}$$

where P_f and X_f are the food price index and nominal food expenditures, g indexes the different food groups, P_g is the group price index, s_g is the group budget share, and $(\frac{cal}{rupee})_g$ are the calories per rupee of nominal expenditure on a group. Our three scenarios then correspond to changing s_g, $P_f/\{P_g\}$ and $(cal/rupee)_g$ holding the other variables constant.

Changing group shares turn out to be the most important factor. We use the nominal food expenditures and group calorie/rupee conversion factors at the rural 1983 average and vary only the group budget shares. Table 3.3 shows that the changes in group shares explain most of the differences in calorie per rupee of real expenditure over time but only part of the rural-urban difference. Households in the later periods spend a larger share of their food budget on food groups with lower calories per rupee.

We next vary relative group prices while holding shares and within-group composition (calories/rupee) constant at the rural 1983 level. This involves multiplying the rural 1983 nominal group expenditures by a conversion factor P_g/P_f constructed with base-weighted price indexes (normalized to one in base period). Table 3.3 shows that this exercise actually tends to increase calorie consumption because the prices of groups with lower calories/rupee (meat, vegetables, processed food and beverages) had more relative inflation than groups with higher calories/rupee (grains, oils, and

fruits).

Finally, we examine the effect of intra-group prices and composition by holding group shares and nominal expenditures fixed at rural 1983 levels and deflating the calories/rupee in each sector and period by a base-weighted group price index. Differences in the calorie/rupee conversion factors are then due to intra-group relative price and composition effects. Table 3.3 indicates that intra-group composition plays a large role in explaining the rural-urban gap but is less important over time.

The evidence on compositional shifts argues against an interpretation of declining calories due to a rising relative price for calories versus food quality. The base-weighted relative price of the cheapest staple group grains declined while higher quality groups became relatively expensive. The pattern of inflation and substitutions is more consistent with story in which changes in food demand towards higher quality (lower calorie/rupee) goods due to lower energy requirements or increased the relative prices of those goods. Though some of the compositional changes are caused by rising real food expenditure, the growth in real food expenditure is quite modest over the 1983-1993 period and most of the change in composition reflects a downward shift in the CF curve.

3.4 Caloric intake and requirements in the cross-section

3.4.1 Imputation of energy requirements

We use the India Time Use Survey (TUS) to imputed energy requirements. The NSS Organization interviewed 18,620 households in six Indian states (Haryana, Madhya Pradesh, Gujarat, Tamil Nadu, Orissa, and Meghalaya) between July 1998-June 1999. Every household member over age five was asked their time-use over the previous 24 hours as well as for abnormal and variant days (e.g. weekends, trips into town). Time-use is recorded in 20 minute increments and is classified into 154 different types of activities. A major advantage of this survey is that it records many household variables in the same format as the NSS consumption surveys, including monthly household expenditures, age, gender, education, and occupation.

To convert activities in the TUS into caloric requirements we use age/gender minimum caloric requirements (corresponding to the Basal Metabolic Rate or BMR) multiplied by one of four scaling factors based on our own classification of the intensity of the 154 different survey activities. The data appendix contains the details of our imputation procedure. Children under 6 do not have time-use recorded in our data so we use caloric requirements from the India Council for Medical Research (ICMR). For most of our comparisons we aggregate to the household level as we do not have individual caloric intake.

Other than the subjectivity involved in classifying the intensity of activities, there

are two major limitations of using time-use data to impute caloric requirements. First, we do not observe the heights and weights of individuals that affect baseline caloric requirements. Second, we are unable to capture differences in energy intensity *within* specific activities, particularly those that feature an important margin of substitution for animal and mechanical power. This issue arises primarily for transport but also for some agricultural activities and may bias up the energy requirements of richer households that use more capital than labor. Despite these limitations, we believe our measure of caloric requirements is superior that of the ICMR, which only feature three classes of energy intensity for adults - heavy, moderate, and sedentary. The ICMR also does not allow for different activity levels of children or allow age to affect the energy requirements of adults. Our measure allows significantly more variation along the extensive (length of the work day) and intensive (type of activity) margin, captures differences in home production, and allows us to link households to consumption data using variables with common definitions.

Table 3.4 presents summary statistics for the 1999-2000 NSS consumption data and the 1998-1999 Time Use data by sector. By restricting both samples to households without missing linking variables in the six common states, we are left with 29,415 NSS households and 18,571 TUS households. The first two rows show that average calorie intake (using our preferred group imputation) appears to above calorie requirements in urban areas but below in rural areas. The next two use alternate measures of intake ("all food imputation") and requirements based on ICMR data. The ICMR calorie requirement measure is higher than our due to greater calorie requirements for children aged 6-13 and for adults doing heavy activity. For example, the daily recommended intake of 3800 calories for adults doing heavy labor is much higher than our measures for rural (3275) and urban (2865) adult workers with similar occupations (see appendix for details).

The other variables in the two datasets are quite similar, though there is some discrepancy in household size and monthly per capita expenditure (MPCE), both a bit higher in the NSS. The difference in household size is driven by children but remains a bit of a mystery. The difference in expenditures may be partly due to inflation and real expenditure growth between 1998-1999 and 1999-2000, but another factor is recall bias. Measured expenditures are typically higher when households are asked for either (a)a specific list of items or (b)shorter recall periods, both features of the 55th NSS round but not the TUS. The bias caused by (b) may also affect imputation of calorie consumption (biasing it down relative to requirements recorded over a 24 hour period). These effects will probably bias the *level* of calories and expenditures across the two surveys making calculation of absolute calorie adequacy suspect. In most of the analysis that follows we focus on *percentage* differences in caloric intake and requirements across sectors and household characteristics so our results would be unaffected by scaling expenditures or energy requirements by a uniform factor. We revisit measurement of levels in section 3.6.

3.4.2 Household composition and life-cycle effects

Household composition affects consumption through multiple channels including total expenditures, economies of scale, and intra-household bargaining. Another channel is caloric requirements, which we investigate by regressing household caloric intake and requirements on the number of household members in different age/gender cells - seniors (aged 60 plus), adults (19-59), youth in the 16-18 and 13-15 range, children in the 10-12, 7-9,4-6, and 1-3 ranges and infants under 1 year of age. We use village/block dummies so we only use variation in household composition within areas to identify the effects. We also include a cubic in per capita expenditure to identify effects of household composition on calorie intake at constant total expenditure.

Table 3.5 present the results normalized by the coefficient on adult males. Calorie demand ratios by gender/age class are related to calorie requirement ratios with a correlation coefficient is 0.83. The third column presents average caloric requirements calculated directly for individuals, but the effects are similar to aggregating to the household level and using a linear regression. Columns four and five present results controlling for a cubic in monthly per capita expenditures (MPCE) and the patterns are similar. Seniors and children contribute less to caloric intake and requirements than adults. Compared to adult males, females and senior males contribute relatively more to intake than requirements. Female infants appear to contribute less to caloric intake than male infants, but gender discrimination against children not evident for other age groups. We do not observe individual caloric intake so any assessment of distributional issues is necessarily speculative.

The results in table 3.5 assume linear effects of household size for each gender/age class, but larger households may experience economies of scale. Holding expenditure per capita constant, larger households may have more leisure time if there are economies of scale in home production activities. For example, if two household members take turns shopping, cooking, cleaning, or taking care of children it may decrease their combined average energy requirements. This superior home production technology may also affect food expenditure patterns by leading household to purchase less prepared meals on the margin.

Table 3.6 confirms both of these hypotheses using OLS regressions of log household caloric intake or requirements on household composition variables in ratios (the ratios of each cell/class from 3.5), village/block dummies, a cubic in log MPCE and the log of household size. Column one shows that doubling household size holding composition and per capita expenditures constant lowers caloric intake by about 2.6%. This is directly related to the Barten paradox discussed by Deaton and Paxson [1998]. Deaton and Paxson [1998] note that according to Barten's model, if calories (or food expenditures generally) are an exludable, private good we would expect an elasticity greater than one with respect to household size holding per capita expenditure constant. Households should economize on certain shareable goods like housing and durables and spend a greater share on food (and potentially calories).

The second column suggests that economies of scale in caloric requirements may explain why the Barten model does not hold for food in many developing countries. The elasticity of household calorie requirements to household size is negative and virtually identical to the elasticity of calorie intake. Larger households appear to benefit from economies of scale in caloric requirements for market work (with more productive household members supplying more labor) and/or home production. Most of the effect is driven by food expenditures, with a small negative elasticity of calories conditional on food expenditure to household size doubles. This could occur if the reduction in energy requirements (predicting a fall in the CF) is accompanied by a superior home production technology for converting unprocessed food into tasty calories, with the two effects cancel out.[12]

We also examine life-cycle effects for adults. There is a growing literature on life-cycle consumption in developed (Aguair and Hurst [2005]) and developing (Hicks [2010]) countries that uses time-use data to study substitution of time for expenditures. In the United States and Mexico households appear to smooth calorie consumption over the life-cycle (in spite of the usual hump-shaped profile of food and total expenditures) through home production and shopping intensity.

We use the average age of household members over age 18 as our life-cycle measure but obtain similar results for households with one male and female adult. We control for household size and the ratio of men, women and children (in each gender/age cell) and regress log household calories on these controls and dummy variables for average adult age beginning with 23-27 and ending with 73-77. The omitted category is 19-22 year olds.[13]

Figure 3.8(a) plots the coefficients on the dummies with no expenditure controls. The solid (dotted) line represents percent deviation in caloric intake (requirements) relative to households with mean age 19-22. Calorie intake and requirements are highly correlated over the life-cycle. From peak to trough calorie consumption falls about 20% while requirements fall about 30%. As the decline in calorie requirements begins earlier and is steeper, older households are relatively better off in terms of net intake than younger and middle-aged households. Figure 3.8(b) presents the results when a cubic in log expenditure is included, netting out life-cycle variation in expenditures. This is especially important if expenditure-selective mortality changes the composition of households in the later years (e.g. the poorest households with low intake are negatively selected). The decline in caloric intake is smaller at older ages, implying that the main mechanism behind the decline is the fall in total expenditures of about 20% over these ages.

There is relatively little adjustment over the life-cycle in food expenditure per total

[12]We could interpret the latter effect in the model as higher p_c. We have not explored this finding in greater detail as it takes us far from the main subject of the paper, but we plan to explore it in future work.

[13]The TUS only includes a single cross-section so we do not control for cohort effects. Pooling our cross-sections and controlling for cohort effects yields similar life-cycle patterns for calorie intake.

expenditure (figure 3.9(a)) or in calories per food expenditure (figure 3.9(b)). The margin for substitution towards home production over the life-cycle thus appears to be small in India. This may be due to already high home production in poor counties like India - very little processed food or meals out are consumed, so there is little scope for older households with less market work to substitute time for expenditures on food and calories. The decline in caloric intake in late-life should not necessarily be interpreted as impoverishment or an inability to smooth consumption, as a significant portion coincides with a decline in caloric requirements. The causality is not obvious and there may be an old-age calorie-poverty trap, though Deaton and Subramanian [1996] argue that calorie traps are unlikely to exist given the price of cheap staples. Whether these patterns are common to the poorest countries or unique to India is an interesting question for further research.

3.4.3 Occupation and education

Household occupation and education affect household total expenditures but may also affect the caloric requirements and intake conditional on expenditure. When the income-generating process is more dependent on energy inputs, we expect higher caloric requirements and intake conditional on expenditures. Comparing households with different occupations in the same area/period, or similar occupations in different areas/periods, allows us to isolate the importance of these structural factors calorie intake and requirements. Given that there have been significant increases in education and shifts from more to less labor intensive tasks this would be an obvious candidate for explaining the missing calories.

The data provide two occupational classifications based on the economic activity that brings the most income or in kind value to the household. The first is household type, divided into five rural types - self-employed non-agriculture, agricultural laborer, other laborer, self-employed in agriculture, and other - and four urban types - self-employed, wage/salary worker, casual laborer, and other. This measure does not capture the length of the work day or the occupation of spouses and extended family, but provides a simple way to link the NSS and TUS data using a common classification and examine whether caloric intake and requirements are correlated across types.

The top panel of table 3.7 presents the results of regressions that control for household composition ratios, household size, a cubic in expenditure and village/block dummies. We report the coefficients on household type dummies that represent the percent difference in caloric intake or requirements from the omitted categories (other rural and other urban). Conditional on total expenditures the household types with the highest caloric requirements - self-employed agriculture, agricultural laborer, other rural labor and casual labor - also have the highest caloric intake. Wage/salary earner, self-employed in urban areas, and the omitted categories have both the lowest intake and requirements. The correlation of household type dummy coefficients across intake and requirements is 0.87. The patterns are even stronger when conditioning on food

expenditures instead of total expenditures (the CF curve). Consistent with our model the household types with higher caloric requirements tend to consume lower quality calories. Some of the adjustment also occurs on the FE margin, especially for self-employed in agriculture households.[14]

The middle panel of table 3.7 uses occupational classifications from the National Classification of Occupations (NCO). The classification ignores industries so there are service workers in agriculture and clerical workers in manufacturing. We use the broad classifications - professional, administrative, clerical, sales, service, primary, secondary and other. The correlation of caloric intake and requirements across occupations is 0.88, with primary, secondary, and service workers consuming and requiring more calories than sedentary professional, administrative and clerical workers.

Education variables provide an alternative to household types or occupations, as education is correlated with occupational classes and we expect more educated households to perform more sedentary tasks. While educated households have higher expenditures and calories in an absolute sense, conditional on expenditures we would expect them to require less calories. We first analyze dummies for the education level of the head of the household, divided into 7 discrete categories - illiterate (the omitted category), literate but not attending primary, some primary, primary completed, middle completed, secondary completed, and college completed.

The bottom panel of table 3.7 confirms our expectation that conditional on expenditures households with more educated heads have lower caloric requirements. They may work less hours for a given total expenditure due to higher hourly wages, or the caloric requirements per hour of work may be lower. Differences in caloric requirements are reflected in caloric intake almost one for one. Households with a college educated head consume 10% less calories and require 12% less calories than illiterate households, conditional on expenditure. Similar to household type and occupation, most of the adjustment in caloric intake occurs along the CF "quality" margin rather than the FE margin.

We can also generate a continuous variable by assigning years of schooling to each education class and taking an average over all adult household members. In addition to our usual controls we use dummies for each one year interval of average adult education. Figure 3.10(a) presents the results for total calories, which echo the findings in the bottom panel of table 3.7. Figure 3.10(b) confirms that similar patterns hold for the CF and FE curves.

3.4.4 Other factors

There are several other variables recorded in the NSS data that are likely to have an impact on energy requirements but cannot be linked to the Time Use Survey.

[14] One possibility is that they face a lower shadow-price for food since they can consume their own production without middle-man and retail markups.

Some NSS rounds ask questions about the main source of energy for home lighting and cooking, income from non-labor sources (pensions, remittances, rent and interest payments), and ownership of labor-saving durables like motorcycles/cars, washing machines, or fans/air conditioning[15]. Interpreting the effect of these variables on expenditure patterns is more difficult than with the exogenous household characteristics considered above, as service-flows from durables should arguably be included in household expenditure measures and some of these variables have a direct impact on expenditures (e.g. electricity for home appliances, the expense of wood versus other cooking/heating fuels, the cost of petrol for motorcycles and cars). However, if we are willing to assume that food is separable from other expenditures except through their effects on energy requirements, then we can interpret the effect of durable ownership on calories conditional on food expenditures as an effect operating through energy requirements.

The results in table 3.8 strongly suggest that variables we expect to reduce household energy requirements do lead households to substitute food quality for food calories. Caloric intake conditional on food expenditure is lower for households with non-wood fuel sources (6.1%), electricity (3%), washing machines (2.8%), motorized transport (2%), fan/air conditioning (1.1%), or a source of non-labor income (1.1%). Television ownership, which might be expected to affect household tastes and encourage consumption of more expensive calories, has little impact. The substitution of non-human energy sources for human sources, whether through fuel to operate motor vehicles, electricity for household appliances, or denser cooking fuels with a lower complementary human energy input, thus appears to have a large effect on caloric intake, an effect that may not be adequately captured by simply controlling for household expenditures. From the TUS we know that free collection of goods (including wood for cooking) and household maintenance take 53 and 459 minutes per day for the average household, so the potential household energy savings from replacing human power with non-human power sources is large.[16]

3.5 Quantitative assessment of energy requirements hypothesis

In this section we answer the central question of the paper - can the energy requirements hypothesis explain the 'missing" calories over time and between sectors? We use a calorie Engel curve regression and a dummy for the missing calories (in later periods or urban areas) and see how the missing calorie dummy changes when

[15]Energy requirements are higher at extreme temperatures

[16]According to the National Planning Commission more than 85 million households in India spend 30 billion hours a year gathering firewood. See http://www.thaindian.com/newsportal/uncategorized/86-percent-rural-indians-use-dung-cakes-firewood_10087695.html

we control for variables that are closely related to energy requirements. Reducing the missing calories requires variables that (1)have a large effect on caloric intake and requirements and (2)vary significantly across rural/urban areas or over time. While all of our controls satisfy the first criterion based on the evidence presented above, the demographic variables do not appear to vary much over time or across sectors (see table 3.9).

3.5.1 Rural-Urban Gap

We begin by analyzing the unexplained rural-urban calorie gap for the six states in the TUS. The first row of table 3.10 presents baseline estimates by regressing log household calorie intake on log household size, a cubic in log per capita real expenditure and an urban dummy that represents the "missing" urban calories (equivalent to the area between calorie Engel curves in figure 3.4). The gap is 17.6% without controlling for higher urban prices, but conditioining on log per capita real expenditure reduces the gap to 12.5%. When we repeat this regression using calorie requirements instead of consumption, we find a 9.8% gap.

We add the controls cumulatively. The second row of table 3.10 adds controls for demographic composition (ratios of male and female adults and several age/gender classes for children) and dummies for adult average age. The third row adds dummies for average adult years of schooling and head of household education class. The fourth row adds dummies for household type and NCO occupation.[17] The fifth row adds the household variables absent from the TUS like electricity, firewood for cooking, and ownership of labor-saving durables.

Controlling for demographics has little effect on the urban dummy for caloric intake or requirements, but adding the education variables decrease the gap for intake (requirements) by 20%(25%) compared to baseline and adding occupation reduces the gap by 56%(90%). Controlling for our other variables only reduces the intake gap a bit more (61%). The last two columns decompose the calorie gap into calories conditional on food expenditure (CF) and food expenditure conditional on total expenditure (FE). As we add controls the reduction in 'missing' urban calories occurs mainly on the CF margin, with the urban coefficient falling by over 100% and becoming slightly positive. The coefficient on the unexplained gap in the FE curve falls by only 42%. This is consistent with our model and earlier findings, as a large number of factors may influence the division of expenditures between food and non-food but with separability only energy requirements or within-food relative prices affect the CF curve.

Table 3.11 uses the time-use data to shed more light on the source of rural/urban energy requirement differences. Rural households spend over 200 minutes less on

[17]The omitted categories imply that the rural-urban dummy is comparing self-employed or wage earning professionals.

leisure activities and 200 minutes more on market activities per day. Broken down by gender, we see that urban and rural males spend similar time on market, non-market, and leisure activities but the composition of market activities is different. Urban males work mostly in the tertiary/service sector, while rural males work mostly in the calorie-intensive primary sector. The story is different for females, as rural females spend an extra 2 hours per day on market activities (mostly primary activities and free collection) of which about 90 minutes comes out of leisure and 30 minutes out of non-market/home production. The extra leisure time is mostly spent on socializing and watching television. While males and females in rural areas have higher energy requirements, for males this is due to a labor intensive margin and for females this is due to a labor extensive margin. These differences are consistent with table 3.10 as controlling for occupation and factors that affect free collection and home production reduces most of the gap in caloric requirements and intake.

3.5.2 Over time gap

Table 3.12 presents the missing calories over time. Each column presents the coefficient on the dummy for a different year/sector relative to the rural sector in 1983. We include dummies for all sector/years but only report them for 1983, 1993-1994 and 2004-2005. Each row cumulatively adds another set of variables - these are identical to the ones used in the rural-urban comparison, except that we are forced to exclude some variables that are not present in all survey rounds.[18] Unlike the rural-urban comparison we use data from the 17 biggest states in India and Delhi. The baseline estimates are the similar and use log household size and a cubic in log real expenditure.[19]

Row one of the top panel shows that the unexplained decline for rural areas between 1983 and 2005 was about 18% and the decline for urban areas was about 14%. When we add in all of our controls these fall to 14.5% and 12.2%, implying that our variables can only explain 20% and 15% of the decline in calories over time. The rural-urban fall by well over 50% for all years. The biggest effects come from education, occupation and other variables with little role for demographics.

The middle panel shows that our controls account for only 7.6% (urban) and 16% (rural) of the gap in FE curves over time. There appear to be other forces depressing food expenditures, especially in the 1993-2005 period, some of which were discussed earlier. The bottom panel shows that our variables do a much better job of capturing changes in the CF "quality" curve - 42% in rural areas and 60% in urban areas. This

[18]The 1983 NSS does not contain casual labor or wage/salary as household types. It also does not contain ownership data for durables, though we are able to construct a motor vehicle dummy based on petrol expenditure. Non-labor income is also not available for all survey rounds and are excluded.

[19]The base year/period is the rural sector of 1983 and we use a Tornqvist index though the results are largely unaffected by using different base years or indexes.

suggests that the energy requirements hypothesis still has a significant bite for changes over time, albeit not nearly enough to explain the dramatic fall in food budget shares conditional on total expenditures.

We consider three other specifications. First, we drop the 38th round allowing us to use a larger set of controls. When we do this, we find that the share of the missing calories that we can explain rises to 23% (41%) in rural areas and 35% (100%) in urban areas conditional on total (food) expenditure. Second, we estimate our equations separately by sector, allowing the control variables to have different effects for rural and urban households. This has little effect on the results. Third, we estimate the decline in calories separately by different household types and professions, conditioning on all of our other controls. The decline is 4-7% greater for primary sector households relative to administrative, sales, clerical, service or secondary sector workers. We see a similar pattern for self-employed agricultural households relative to self-employed non-agricultural households and other rural labor households. This suggests to us that significant changes in energy requirements *within* agricultural occupations are important to changes over time.[20]

3.6 Measuring poverty and welfare with caloric requirements

Although our primary purpose has been to explain demand patterns we briefly consider the implications of our analysis for poverty and welfare. We begin in table 3.13 by presenting poverty headcounts and rural/urban poverty line ratios for the five major Indian states in the 1999-2000 NSS and 1998-1999 TUS. According to official measures, poverty is highest in Orissa and Madhya Pradesh, is quite comparable between rural and urban sectors of each state, and the rural poverty lines are far below the urban ones for all states except Haryana. The second column presents modified estimates from Deaton and Dreze [2002] that use the NSS unit values to construct price indexes. Rural poverty remains similar but urban poverty falls dramatically as urban poverty lines are revised down. The third column presents a 'nutritional adequacy' poverty measure computed by Karan and Mahal [2005], who take daily recommended intakes of various nutrients from the ICMR and use linear programming methods to determine the least-cost bundle of food that meets requirements (subject to some palatability constraints).[21] The Karan and Mahal [2005] headcounts are higher in rural areas but lower in urban areas. The fourth column presents the share of undernourished children below age 5 in each state from the 2007 National Family

[20] As the NSS data do not allow us to compare the length of work days, the composition of crops or the use of inputs into agricultural production we are unable to explore this possibility further but this would be a promising area for further research.

[21] In their calculations all rural adults are manual workers and all urban adults are non-manual workers.

and Health Survey, which is significantly higher than official headcounts[22].

The official poverty lines were calculated in 1973 using the cost of consuming 2400 calories in rural areas and 2100 in urban areas in basic grains. The poverty lines were then updated over time based on broad-inflation measures. Defining poverty as calorie intake below these cutoffs leads to much higher head counts in all states and sectors except urban Orissa. We can improve on the 2400/2100 cutoffs by using calorie-cutoffs that vary by state/sector/occupation. Using state/sector averages of our imputed calorie requirements, we tend to find higher poverty in urban areas but similar poverty in rural areas. There is some variation across states, with a 6 point rise in poverty in urban Gujarat and an 8 point decline in rural Haryana. Using a 20% below requirement cutoff brings headcounts closer to official estimates, but we still find significantly greater poverty in most state/sectors with the exceptions of Orissa and urban Madhya Pradesh. While the poverty rates in the first three columns all point to Orissa as the state with greatest poverty, child undernourishment in Orissa is similar to Haryana and below Gujarat. Thus while our measure of poverty (like any calorie-based measure) may fail to capture balanced diets and non-food consumption, it is more highly correlated with some measures of extreme deprivation like child wasting.

We can also examine the impact of occupation-specific caloric requirements on poverty rates. Table 3.14 presents results for the state of Madhya Pradesh, comparing sedentary occupations (professional, administrative, clerical, and other occupations) to primary, secondary, and service/sales. We then compute poverty headcounts under (a)average sectoral energy requirements and (b)average occupation-specific requirements. With common energy requirements primary and secondary occupations have higher rates of poverty due to the distribution of per capita calories (and expenditures). When we use occupation specific energy requirements poverty rates for the primary sector rise and secondary and sedentary occupations fall. Accounting for different energy-intensities across occupations changes the distribution of poverty across occupations significantly when food and calories are used to measure poverty.

The model from section 3.2 provides a simple Engel equivalence scale for measuring welfare gains from lower caloric requirements. Holding relative prices and tastes constant utility is proportional to $Y[S_{nf}]$, with S_{nf} depending only on expenditures and caloric requirements. We regress the log of non-food budget share on log per capita expenditures (deflated by the appropriate price deflator), household size and dummies for different periods, sectors, and occupations. The coefficients on the dummies represent expenditure equivalent welfare gains under the assumption that relative prices and tastes are constant. Table 3.15 presents the results. While there appears to be a 5.6% (expedenditure-equivalent) welfare gain for urban areas, we cannot interpret this effect as a welfare gain because relative prices, variety and

[22]Undernourishment is defined as weight-for-age two standard deviations below the WHO reference growth charts (see Menon et al. [2009])

tastes may also differ. Because the coefficient on urban areas falls to about 0.5% when we add energy requirement controls, if these are orthogonal to the unobserved price, taste, and variety differences the implied welfare gain of lower calorie requirements is 5.1% for urban areas.

Over the 1983-2005 period there is a large percentage rise in non-food budget share at any level of expenditures, but our energy requirement proxies only have limited success explaining the increase so the welfare gain we attribute to energy requirements is only 3% in rural areas and 2.3% in urban areas. These welfare gains are of course tenative, depending on both the assumed functional form for demand and the exogeneity of our proxies for energy requirements. Perhaps more reliably we can calculate welfare effects across occupations, as prices, tastes and variety are similar across occupations within an area. Relative to sedentary professions we find that primary, secondary and service sector workers are significantly worse off, by 3-6% of total expenditure equivalent.

3.7 Conclusion

We find that energy requirements explain most of the missing calories for urban versus rural areas but are insufficient to explain the missing calories over time. This is largely because our energy requirement proxies are better at explaining gaps in quality (calories per food expenditure) than food expenditure per total expenditure, and this margin explains most of the urban-rural gap but a much smaller share of the gap over time. Other factors like relative prices, variety and quality that favor non-food and rising returns to education are likely important for explaining the downward drift in food expenditure Engel curves. We cannot definitively rule out a greater role for energy requirements as we have ignored the role of health improvements and our measure of energy requirements is limited and does not allow us to analyze changes in the energy inputs required *within* the 154 time-use categories.

Whether the pattern we observe in India - i.e. a downward shift in calorie Engel curves large enough to offset real expenditure growth - applies to other countries and time periods is an open question. Downward drift in food Engel curves is certainly widespread, but we do not know whether these (combined with changes in calories per food expenditure) are generally enough to outweigh income growth and generate a decline in calories per capita.

An extension of our approach could consider other nutrients. The link between household activity levels and food requirements is less clear cut outside of calories, but nutritional adequacy and hunger obviously encompass more than calories. A diet rich in cheap staples solves some health problems but causes others. The lowest expenditure per capita state in the TUS, Orissa, has relatively low calorie deficits by our measures (and relatively low child undernutrition), but Karan and Mahal [2005] find the highest rates of nutritional inadequacy.

Another direction we plan to explore is the relationship between household expenditures and caloric requirements through the household production function. This could add another dimension to general equilibrium models of structural change that use Stone-Geary preferences (e.g. Lagakos and Waugh [2009]). Our findings provide a reason why the degree of non-homotheticity caused by energy requirements may be a cause and consequence of economic growth.

Our potential resolution of the Barten puzzle through economies of scale in calorie requirements, this preliminary finding warrants further examination. More generally, the role of market/non-market substitution is critical for analyzing differences between life-cycle and business-cycle consumption in rich and poor countries. Our evidence for calories is quite different from what others have found for the United States and Mexico. Is there a general pattern and economic theory that would explain why poorer countries (and perhaps individuals) experience larger drops in calorie intake (and/or requirements) in old age and limited opportunities for substitution towards cheaper calories? A framework that allows market/home substitution to vary endogenously with household expenditures and/or economic growth and systematic cross-country evidence could be informative.

3.8 Tables and Figures

Table 3.1: Estimates of mean per capita calorie consumption in India

Authors	Sector	1983	1988	1994	2000	2005	1983-2005
Deaton and Drèze	Rural	2240	2233	2153	2148	2047	-193
(2009)	Urban	2070	2095	2073	2155	2021	-49
Chatterjee, Rae	Rural		2135	2100	2097		
and Ray (2007)	Urban		2073	2091	2169		
Kumar and Dey	Rural	2205			2332		
(2007)	Urban	1972			2440		
Meenakshi and	Mean	2219			2132		
Vishwanathan(2003)	Median	2076			2024		
NSS for NNMB states		2131	2139	2076	2020	1960	-171
(Deaton and Drèze)	Year	1975-79	1988-90	1996-97	2000-01	2004-05	
NNMB		2340	2283	2108	1954	1907	-405
				Our estimates			
Group imputation	Rural	2313	2285	2234		2140	-172
	Urban	2230	2234	2214		2136	-94
All food imp.	Rural	2320	2293	2244		2154	-166
	Urban	2178	2180	2192		2121	-58

Meenakshi and Vishwanathan (2003) report data by state for both sectors combined.
NNMB are the independent estimates from the National Nutritional
Monitoring Bureau reported
in Deaton and Dreze [2009], which cover a subset of states. Above are NSS
estimates from Deaton and Dreze [2009] for the same set of states in nearby years.

Table 3.2: Ratio of calories per rupee of expenditure, relative to rice and wheat

	1983		1993-94		2004-05	
	Rural	Urban	Rural	Urban	Rural	Urban
Grains	1.20	1.08	1.13	1.03	1.09	1.03
Pulses	0.62	0.58	0.41	0.41	0.35	0.40
Milk	0.47	0.37	0.33	0.28	0.28	0.27
Oil	0.48	0.53	0.45	0.49	0.44	0.55
Meat	0.16	0.13	0.12	0.11	0.11	0.11
Veg	0.29	0.24	0.23	0.20	0.19	0.19
Fruit	0.96	0.86	0.71	0.63	0.61	0.60
Sugar	0.50	0.52	0.35	0.41	0.32	0.39
Bev.	0.06	0.06	0.04	0.04	0.04	0.04
Proc.	0.25	0.23	0.19	0.19	0.19	0.19
All	0.82	0.61	0.64	0.51	0.56	0.49

Table 3.3: Ratios of mean real expenditures and calories per real expenditure relative to 1983 rural sector

	Sector	1983	1993-94	2004-05	%Δ 1983-1993	%Δ 1983-2005
Real food exp.	Rural	1.00	1.08	1.07	0.08	0.07
(fixed base)	Urban	1.31	1.40	1.42	0.07	0.09
Real food exp.	Rural	1.00	1.08	1.06	0.08	0.06
(Tornqvist)	Urban	1.34	1.41	1.44	0.05	0.07
Calories at	Rural	1.00	0.87	0.84	-0.13	-0.16
const. real exp.	Urban	0.73	0.68	0.66	-0.08	-0.11
Scenario 1	Rural	1.00	0.89	0.83	-0.11	-0.17
(Change s_g)	Urban	0.81	0.75	0.72	-0.08	-0.10
Scenario 2	Rural	1.00	1.07	1.09	0.07	0.09
(Change P_g,P)	Urban	0.99	1.05	1.06	0.06	0.07
Scenario 3	Rural	1.00	0.94	0.98	-0.06	-0.02
(Change $(cal/exp)_g$)	Urban	0.91	0.90	0.90	-0.01	-0.01

Table 3.4: Comparison of Time Use Survey and NSS

		TUS		NSS Consumption	
	Sector	Mean	Median	Mean	Median
		Requirements		Intake	
Per capita calories	Rural	2363	2323	2236	2088
	Urban	2091	2122	2327	2180
Per capita calories alt.	Rural	2491	2473	2232	2095
	Urban	2200	2274	2277	2163
MPCE	Rural	459	400	505	429
	Urban	804	694	947	734
Hhsize	Rural	4.07	4	4.82	5
	Urban	4.10	4	4.40	4
Age of head	Rural	43.20	40	44.40	42
	Urban	42.32	40	43.43	42
Male head	Rural	0.90		0.90	
	Urban	0.91		0.90	
Adult males	Rural	1.10		1.10	
	Urban	1.19		1.10	
Adult females	Rural	1.10		1.19	
	Urban	1.10		1.16	
Years schooling	Rural	3.58	3	3.20	2
	Urban	8.33	8.8	7.65	7.5

Table 3.5: Contributions to household caloric intake and requirements relative to male adult

				Controlling for HH MPCE	
	HH Intake	HH Req.	Ind. Req.	HH Intake	HH Req.
Male 60+	0.96	0.81	0.75	0.97	0.79
Male 19-59	1.00	1.00	1.00	1.00	1.00
Male 16-18	0.86	0.96	0.91	0.93	0.94
Male 13-15	0.80	0.82	0.76	0.89	0.80
Male 10-12	0.69	0.70	0.62	0.79	0.67
Male 7-9	0.68	0.55	0.50	0.77	0.52
Male 4-6	0.55	0.63	0.55	0.67	0.60
Male 1-3	0.38	0.53	0.41	0.53	0.49
Male < 1	0.38	0.52	0.41	0.52	0.49
Female 60+	0.87	0.59	0.60	0.94	0.57
Female 19-59	0.92	0.81	0.83	0.97	0.80
Female 16-18	0.80	0.84	0.80	0.90	0.82
Female 13-15	0.74	0.77	0.73	0.83	0.74
Female 10-12	0.76	0.68	0.64	0.85	0.66
Female 7-9	0.64	0.59	0.55	0.74	0.57
Female 4-6	0.56	0.62	0.56	0.68	0.59
Female 1-3	0.41	0.52	0.41	0.53	0.48
Female < 1	0.30	0.50	0.41	0.46	0.46

Table 3.6: OLS - effects of household size on caloric intake and requirements

Dep. var.	log cal intake	log cal req.	log food exp.	log cal intake
Exp. Var.	tot. exp.	tot. exp.	tot. exp.	food. exp.
log hhsize	0.974***	0.977***	0.960***	0.997
s.e.	(0.003)	(0.002)	(0.002)	(0.002)

*** denotes significantly below 1 at the 1% level.
All regressions include household composition ratios, cubics in log real expenditures, and village/block dummies.

Table 3.7: Coefficients on household type, occupation and education dummies

Dep. var	Cal. intake	Cal. req.	Food exp.	Cal. intake
Exp. control	Tot. exp.	Tot. exp.	Tot. exp.	Food exp.
		Types		
Self-emp. non-ag.	0.052 ***	0.055 ***	0.024 ***	0.036 ***
(Rural)	(0.007)	(0.005)	(0.006)	(0.006)
Ag. laborer	0.104 ***	0.160 ***	0.023 ***	0.091 ***
(Rural)	(0.006)	(0.003)	(0.005)	(0.005)
Other laborer	0.079 ***	0.092 ***	0.014 *	0.071 ***
(Rural)	(0.009)	(0.004)	(0.008)	(0.007)
Self-emp. ag.	0.094 ***	0.122 ***	0.050 ***	0.062 ***
(Rural)	(0.006)	(0.003)	(0.005)	(0.005)
Self-emp.	0.059 ***	0.031 ***	0.008	0.056 ***
(Urban)	(0.008)	(0.005)	(0.007)	(0.007)
Wage/salary	0.042 ***	-0.033 ***	-0.006	0.048 ***
(Urban)	(0.008)	(0.003)	(0.007)	(0.007)
Casual laborer	0.090 ***	0.050 ***	-0.005	0.098 ***
(Urban)	(0.009)	(0.004)	(0.008)	(0.008)
		National Classification of Occupations (NCO)		
Admin.	0.005	0.053 ***	0.007	0.001
	(0.008)	(0.007)	(0.007)	(0.007)
Clerical	0.007	0.001	0.002	0.004
	(0.008)	(0.005)	(0.007)	(0.006)
Sales	0.007	0.074 ***	0.010	-0.001
	(0.007)	(0.004)	(0.006)	(0.006)
Service	0.039 ***	0.041 ***	0.021 ***	0.027 ***
	(0.008)	(0.006)	(0.007)	(0.007)
Primary	0.075 ***	0.167 ***	0.029 ***	0.059 ***
	(0.006)	(0.003)	(0.006)	(0.005)
Secondary	0.042 ***	0.074 ***	0.008	0.038 ***
	(0.006)	(0.004)	(0.006)	(0.005)
Other occ.	-0.039 ***	-0.054 ***	0.001	-0.038 ***
	(0.008)	(0.008)	(0.007)	(0.007)
		Education levels		
Literate/no primary	-0.005	-0.012 **	0.000	-0.005
	(0.011)	(0.006)	(0.010)	(0.010)
Some primary	-0.015 ***	-0.024 ***	0.000	-0.015 ***
	(0.004)	(0.003)	(0.004)	(0.003)
Primary complete	-0.026 ***	-0.028 ***	-0.006	-0.025 ***
	(0.004)	(0.003)	(0.004)	(0.004)
Middle complete	-0.041 ***	-0.055 ***	-0.015 ***	-0.035 ***
	(0.004)	(0.003)	(0.004)	(0.004)
Secondary complete	-0.068 ***	-0.092 ***	-0.016 ***	-0.062 ***
	(0.004)	(0.003)	(0.004)	(0.004)
College complete	-0.099 ***	-0.116 ***	-0.026 ***	-0.084 ***
	(0.006)	(0.005)	(0.005)	(0.005)

OLS, s.e. in parentheses, ***/** denote sig. at 1%/5% levels.
Hh. comp., size, log exp. and village dummies included.
Ommitted categories: "other," "professional" and illiterate.

Table 3.8: Coefficients on other NSS household variables

Dep. var	Calorie intake	Calorie intake
Control for	Total exp.	Food exp.
Motorcycle/car	-0.051 ***	-0.020 ***
	(0.004)	(0.004)
AC/Fan	-0.008	-0.011***
	(0.004)	(0.003)
TV	-0.011 ***	-0.005
	(0.004)	(0.003)
Washing machine	-0.030 ***	-0.028 ***
	(0.007)	(0.006)
Electricity	-0.036 ***	-0.030 ***
	(0.004)	(0.004)
Wood for cooking	0.067 ***	0.061 ***
	(0.004)	(0.004)
Non-labor income	-0.013 ***	-0.011 ***
	(0.004)	(0.003)

OLS regression, standard errors in parentheses, ***, ** and * denote significance at 1%,5%, and 10% levels.
Household composition and size included with log expenditure controls and village/block dummies.

Table 3.9: Means of household characteristics by year and sector

Year	1983		1987-88		1993-94		1999-00		2004-05	
Sector	R	U	R	U	R	U	R	U	R	U
Adult age	38.8	36.1	38.4	36.2	38.7	36.2	38.8	37.1	39.1	37.6
Age	26.4	26.2	26.5	26.5	27.4	27.2	27.6	28.2	28.4	29.4
Size	5.18	4.81	5.08	4.70	4.90	4.46	5.03	4.52	4.89	4.37
Male adult	1.31	1.37	1.32	1.36	1.33	1.34	1.38	1.41	1.36	1.41
Male child	1.33	1.15	1.29	1.10	1.19	1.00	1.21	0.98	1.14	0.87
Female adult	1.34	1.24	1.33	1.25	1.33	1.23	1.37	1.29	1.38	1.32
Female child	1.21	1.05	1.14	1.00	1.05	0.89	1.07	0.85	1.01	0.76
Adult years sch.	1.97	5.79	2.23	6.19	2.77	6.95	3.28	7.74	3.98	8.39
Frac. Lit. adult	0.32	0.63	0.36	0.65	0.42	0.70	0.48	0.74	0.54	0.77
Self non-agg.	0.12	0.00	0.12	0.00	0.12	0.00	0.13	0.00	0.16	0.00
Agr. Labor	0.31	0.00	0.31	0.00	0.31	0.00	0.34	0.00	0.27	0.00
Oth. Labor	0.06	0.00	0.09	0.00	0.08	0.00	0.08	0.00	0.11	0.00
Self agg.	0.41	0.00	0.38	0.00	0.37	0.00	0.34	0.00	0.36	0.00
Self. Urban	0.00	0.33	0.00	0.34	0.00	0.33	0.00	0.34	0.00	0.38
Wage Urban	.	.	0.00	0.44	0.00	0.43	0.00	0.41	0.00	0.41
Labor Urban	.	.	0.00	0.13	0.00	0.13	0.00	0.14	0.00	0.12
Professional	0.02	0.08	0.02	0.08	0.06	0.17	0.02	0.09	0.03	0.08
Admin.	0.00	0.04	0.01	0.05	0.01	0.06	0.01	0.08	0.02	0.09
Clerical	0.02	0.11	0.02	0.11	0.02	0.11	0.02	0.10	0.02	0.09
Sales	0.04	0.14	0.04	0.15	0.05	0.15	0.04	0.15	0.05	0.16
Service	0.02	0.09	0.02	0.09	0.02	0.08	0.02	0.08	0.02	0.08
Primary	0.74	0.09	0.71	0.08	0.71	0.08	0.69	0.07	0.63	0.06
Secondary	0.11	0.37	0.13	0.36	0.13	0.35	0.13	0.34	0.17	0.35
Electric light	0.15	0.64	0.24	0.72	0.37	0.83	0.48	0.89	0.55	0.92
Cook with wood	0.77	0.46	0.79	0.37	0.78	0.30	0.76	0.22	0.75	0.22
Nonlabor inc.	.	.	0.11	0.15	0.12	0.19	0.14	0.21	.	.
Motor vehicle	.	.	0.02	0.09	0.03	0.14	0.05	0.21	0.09	0.28
AC or fan	.	.	0.01	0.05	0.13	0.42	0.26	0.69	0.38	0.83
Television	.	.	0.00	0.04	0.06	0.30	0.18	0.60	0.25	0.66
Washing machine	.	.	0.00	0.00	0.00	0.03	0.01	0.09	.	.

Table 3.10: Coefficients on urban dummy for caloric intake and requirements, 1998-1999 TUS and 1999-2000 NSS

Dep. var. (logs)	Cal. intake	Cal. req.	Food exp.	Cal. intake
Control for	Tot. exp.		Tot. exp.	Food exp.
Variables incl.				
Baseline	-0.125***	-0.098***	-0.077***	-0.093***
	(0.003)	(0.003)	(0.003)	(0.003)
Demographics	-0.123***	-0.105***	-0.076***	-0.093***
	(0.003)	(0.003)	(0.003)	(0.003)
Education	-0.100***	-0.074***	-0.068***	-0.066***
	(0.003)	(0.003)	(0.003)	(0.003)
Occupation	-0.054***	-0.010***	-0.054***	-0.025***
	(0.007)	(0.004)	(0.006)	(0.006)
Durables/other	-0.049***		-0.045***	0.013**
	(0.005)		(0.006)	(0.006)

OLS regressions, standard errors in parentheses.
***,**, and * denote significance at 1%,5% and 10% levels.
Baseline controls are cubic in log real expenditure and household size.
Controls are added cumulatively for each row.

Table 3.11: Minutes per day on various activities, by sector and gender

Activity	Household		Male adult		Female adult	
Sector	Rural	Urban	Rural	Urban	Rural	Urban
Primary	649.75	73.94	314.13	38.14	153.28	18.13
Free collection	73.33	17.13	12.22	2.39	38.10	10.25
Secondary	98.22	192.74	52.46	107.57	15.27	20.10
Tertiary	113.87	485.63	69.81	305.96	12.62	41.83
Total Market	935.16	769.44	448.62	454.06	219.27	90.32
Cook	229.04	233.30	5.40	5.99	161.51	171.63
Other hh maint.	230.15	241.47	23.06	19.56	137.30	157.02
Care for others	65.84	71.04	10.07	10.62	47.27	55.23
Total Nonmarket	525.02	545.81	38.54	36.18	346.09	383.88
Learning	248.41	317.16	7.83	18.46	2.31	12.12
Social	262.69	515.60	56.81	118.41	34.55	113.95
Sleep	1841.55	1817.84	528.54	503.76	515.28	511.11
Television	104.39	313.43	27.27	74.14	23.51	91.94
Other	1024.33	747.90	332.41	235.48	298.99	236.69
Total Leisure	3481.37	3711.93	952.86	950.24	874.64	965.80

Note: children under 6 are excluded from the household measure.

Table 3.12: Coefficients on sector/year dummies for caloric intake

Year	1983	1993-1994		2004-2005	
Sector	Urban	Rural	Urban	Rural	Urban
Total calories conditional on total real expenditure					
Baseline	-0.148	-0.078	-0.205	-0.183	-0.290
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Demographics	-0.146	-0.080	-0.205	-0.187	-0.295
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Education	-0.120	-0.075	-0.172	-0.175	-0.262
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Occupation	-0.087	-0.073	-0.140	-0.162	-0.228
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Durables/other	-0.063	-0.063	-0.100	-0.145	-0.185
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Total real food exp. conditional on total real expenditure					
Baseline	-0.035	-0.019	-0.048	-0.177	-0.239
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Demographics	-0.036	-0.019	-0.049	-0.177	-0.239
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Education	-0.033	-0.019	-0.044	-0.176	-0.235
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Occupation	-0.031	-0.018	-0.043	-0.173	-0.232
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Durables/other	-0.026	-0.013	-0.029	-0.163	-0.214
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Total calories conditional on total real food expenditure					
Baseline	-0.130	-0.068	-0.181	-0.092	-0.170
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Demographics	-0.128	-0.070	-0.182	-0.099	-0.179
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Education	-0.101	-0.064	-0.147	-0.079	-0.135
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Occupation	-0.070	-0.063	-0.115	-0.067	-0.102
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Durables/other	-0.049	-0.055	-0.082	-0.053	-0.065
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)

OLS regression, standard errors in parentheses.
All coefficients significant at 1% level.
All regressions include household size, cubic
in log real expenditure, and full set of dummies
for sector/period (only selected coefficients reported).
Coefficients are % differences relative
to Rural 1983 (ommitted dummy category).
Controls are added cumulatively for each row.

Table 3.13: Poverty headcounts

State	Sec.	Official	Deaton Drèze	Karan Mahal	% children under-nourish	2400/2100 below	Avg. cal req. below	20% below
Gujarat	Ru.	12.4	20	29.8	44.7	69.2	71.5	43.8
	Ur.	14.8	6.4	16.9		53.3	59.2	25.7
Haryana	Ru.	7.4	5.7	13.7	39.7	58.3	50.3	24.0
	Ur.	10	6.4	14.7		54.0	55.0	28.2
Madhya	Ru.	37.2	31.3	46.2	59.8	79.3	79.3	50.8
Pradesh	Ur.	38.5	13.9	25.6		60.8	64.3	32.5
Orissa	Ru.	47.8	43.5	54.4	40.9	70.0	69.9	33.2
	Ur.	43.5	15.6	30.3		34.6	40.7	12.4
Tamil	Ru.	20	24.3	45	30	73.2	71.8	44.5
Nadu	Ur.	22.5	11.3	24.5		47.3	56.6	27.1

Table 3.14: Calorie-based poverty headcounts (by occupation) for Madhya Pradesh

Occupation	Sector	State/sector average calorie requirements below	State/sector average calorie requirements 20% below	Occupation specific calorie requirements below	Occupation specific calorie requirements 20% below
		Poverty Headcounts			
Sedentary	Rural	68.9	24.2	58.9	17.6
	Urban	44.5	13.2	39.3	10.3
Primary	Rural	89.0	52.3	89.9	55.4
	Urban	82.0	49.1	81.0	49.2
Secondary	Rural	89.4	46.8	87.7	38.4
	Urban	81.5	36.6	79.0	34.8
Service/sales	Rural	80.6	32.7	87.7	29.5
	Urban	68.0	21.9	75.3	27.6

Table 3.15: Welfare gains/losses from caloric requirements

	$\%\Delta S_{nf}$	s.e.	Due to cal. req.
Urban vs. Rural 1999-2000	0.056	(0.004)	0.051
Rural 1983-2005	0.327	(0.002)	0.030
Urban 1983-2005	0.282	(0.002)	0.023
Occupations in 1999/2000			
Admin	-0.014	(0.010)	
Clerical	-0.012	(0.010)	
Sales	-0.018	(0.009)	
Service	-0.060	(0.011)	
Primary	-0.051	(0.008)	
Secondary	-0.028	(0.008)	
Other	-0.015	(0.010)	

Welfare gain is the effect of the dummy on log non-food share.
Controls for household size, composition, cubic real exp.
The last column measures the extent to which dummy is reduced by our variables that proxy for caloric requirements.
Omitted category for occupations is professional.
Occupation regressions include controls for urban/rural.

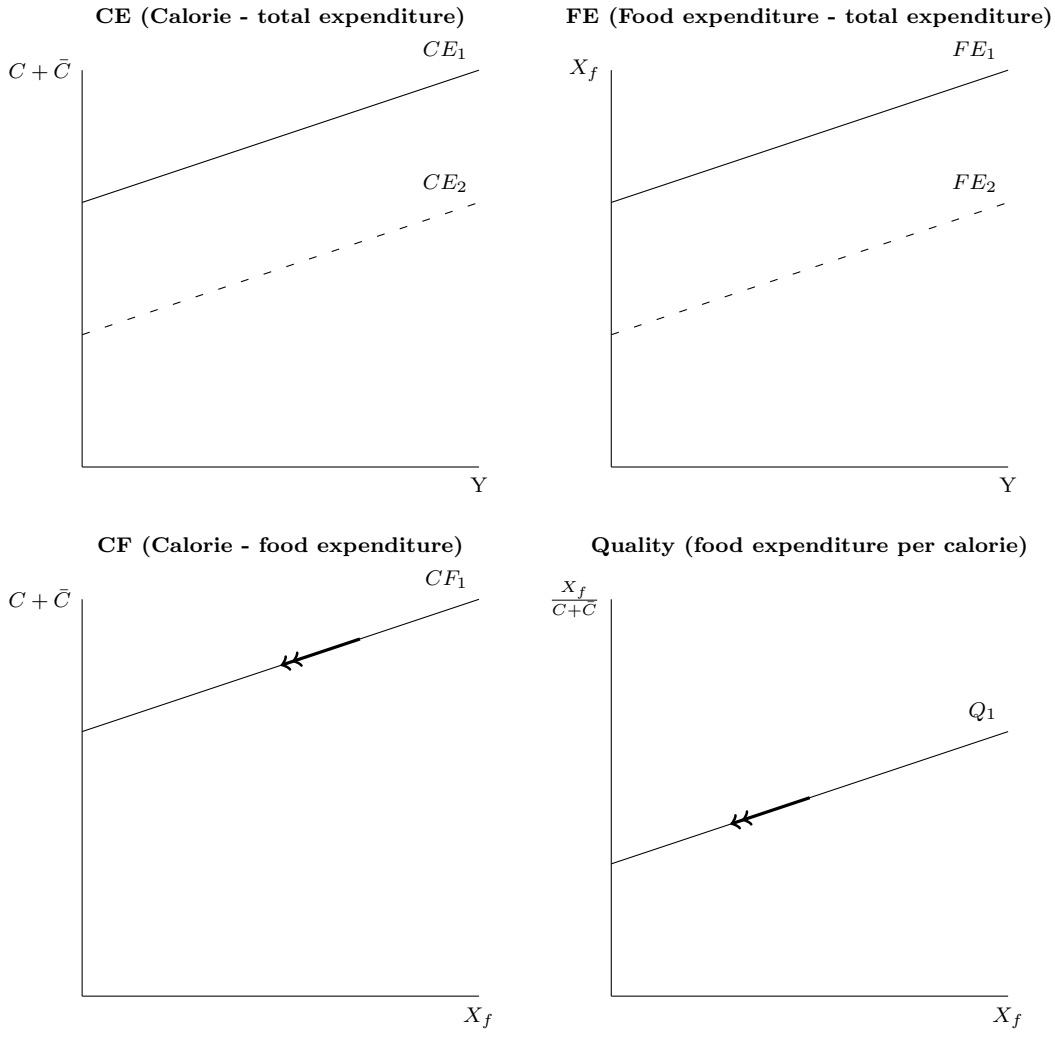


Figure 3.1: Fall in price of non-food p_{nf} (or rise in non-food taste/quality/variety p_f)

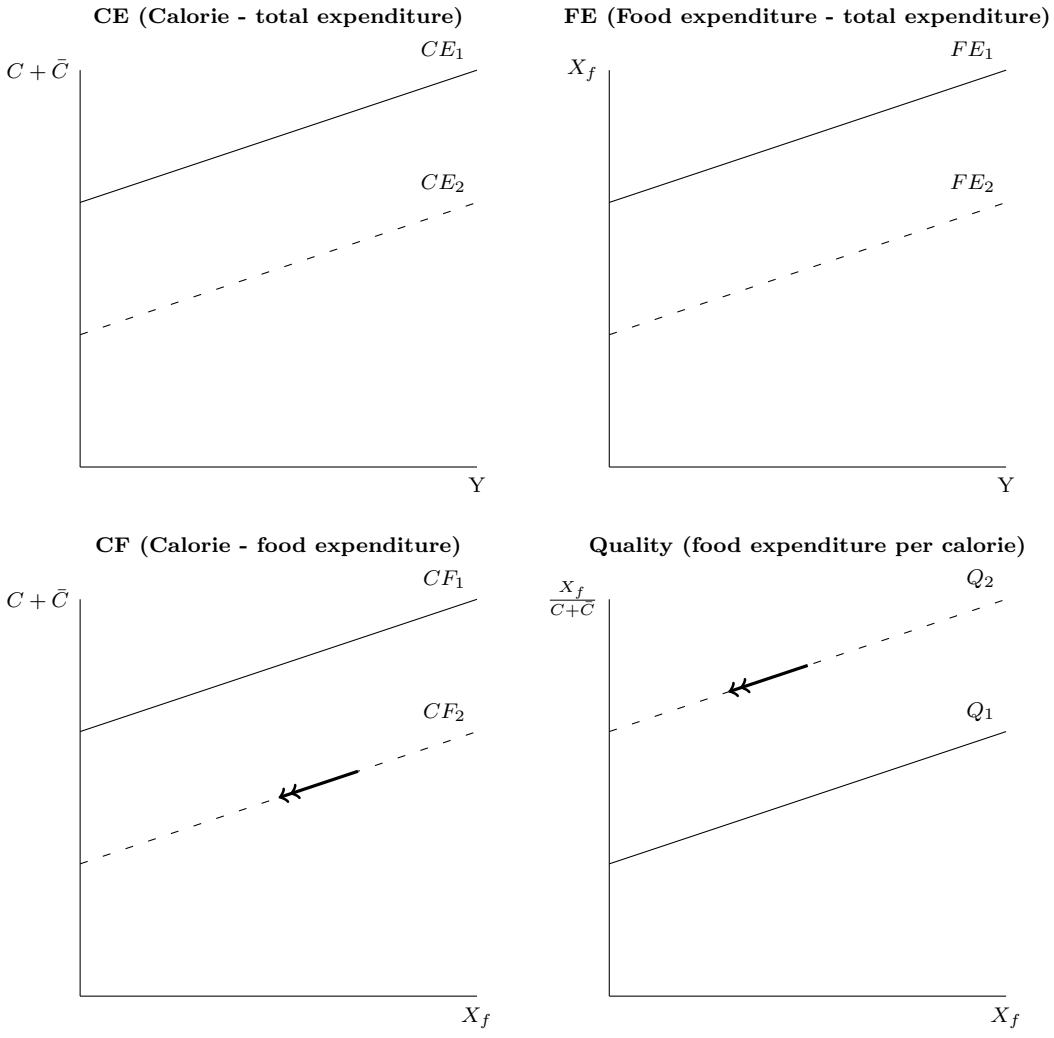


Figure 3.2: Fall in energy requirements $\bar{C}$

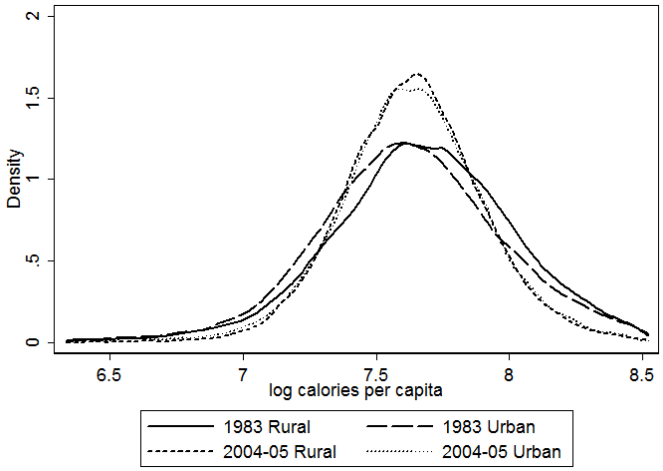


Figure 3.3: Kernel density estimation of log calories per capita

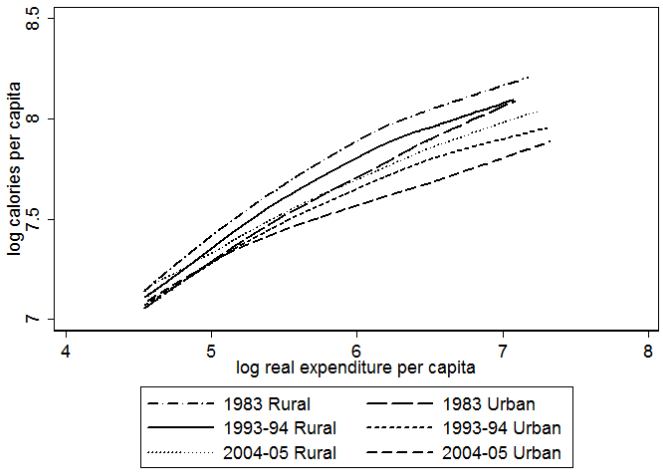
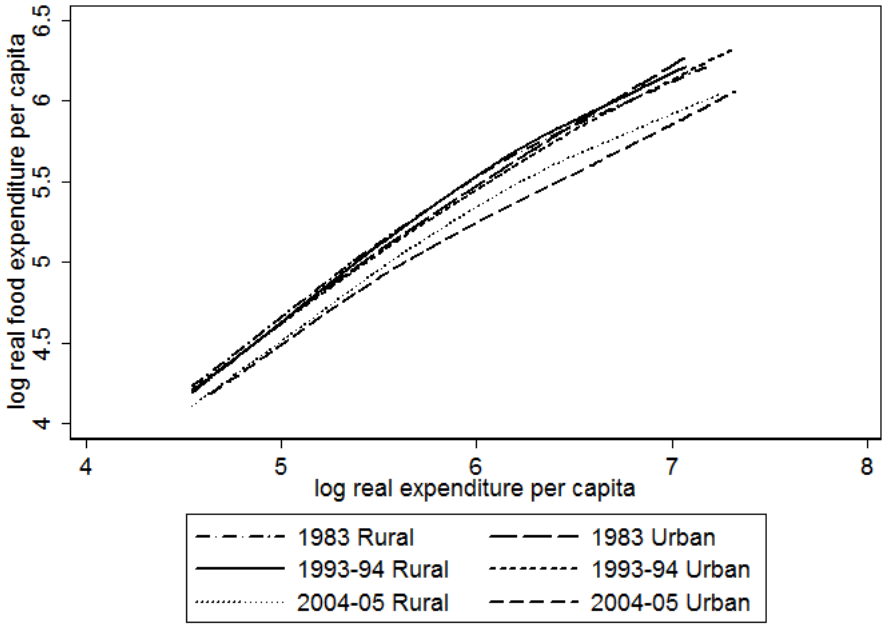
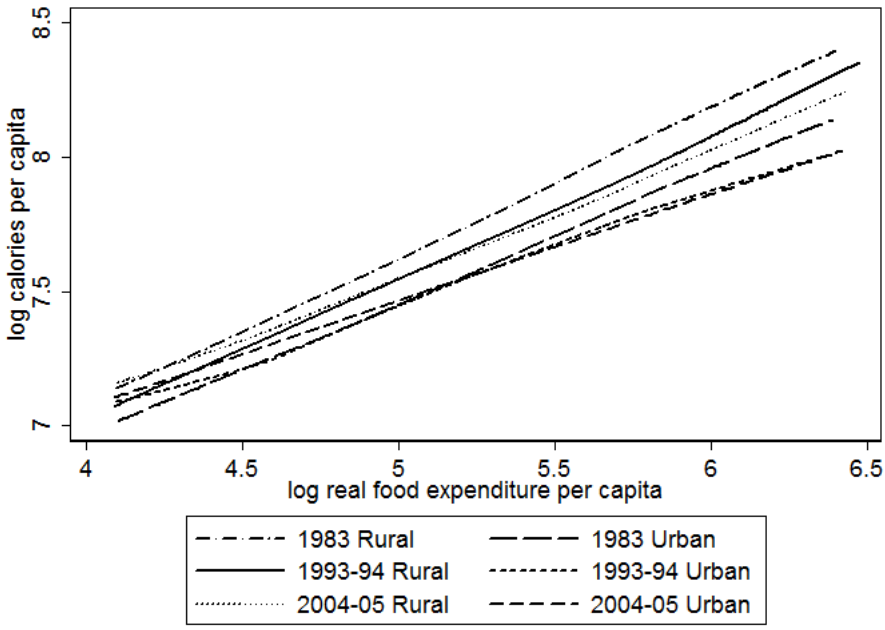


Figure 3.4: Locally weighted regression of log per capita calories on log per capita real expenditure (two adult, 3 child household)



(a) Food Expenditure - Total Expenditure



(b) Calories - Food expenditure

Figure 3.5: Locally weighted regressions for two adult, three child households (logs)

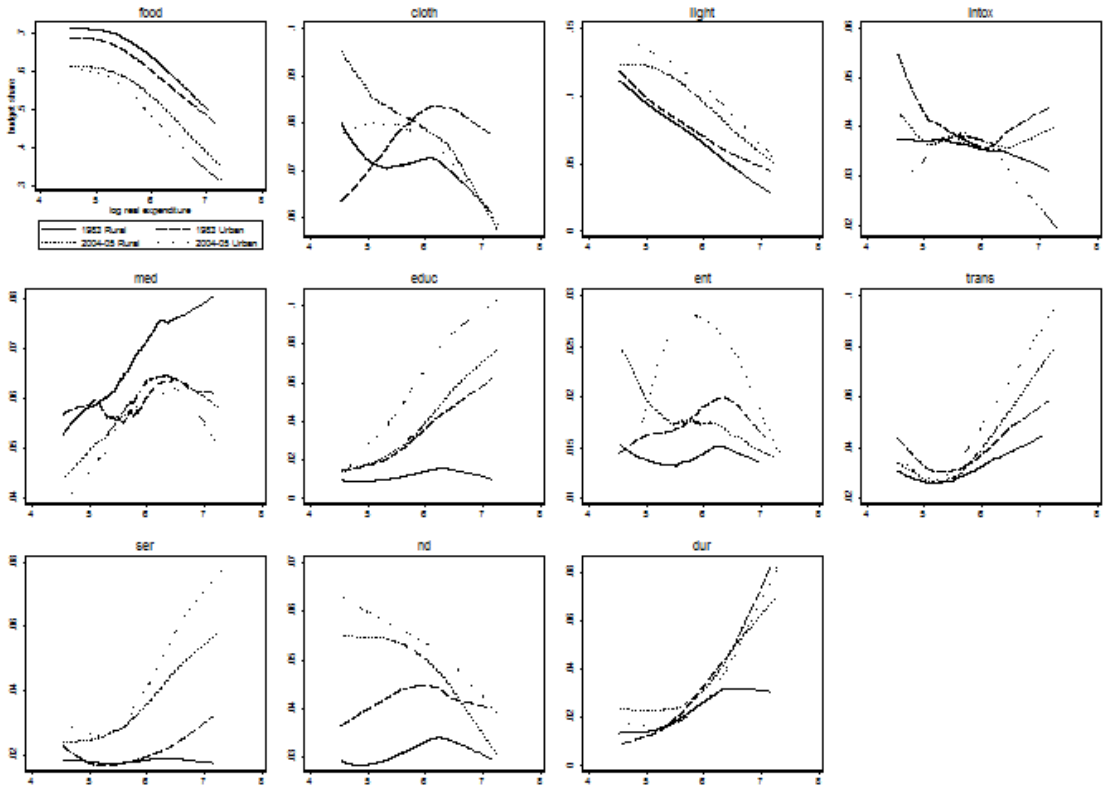
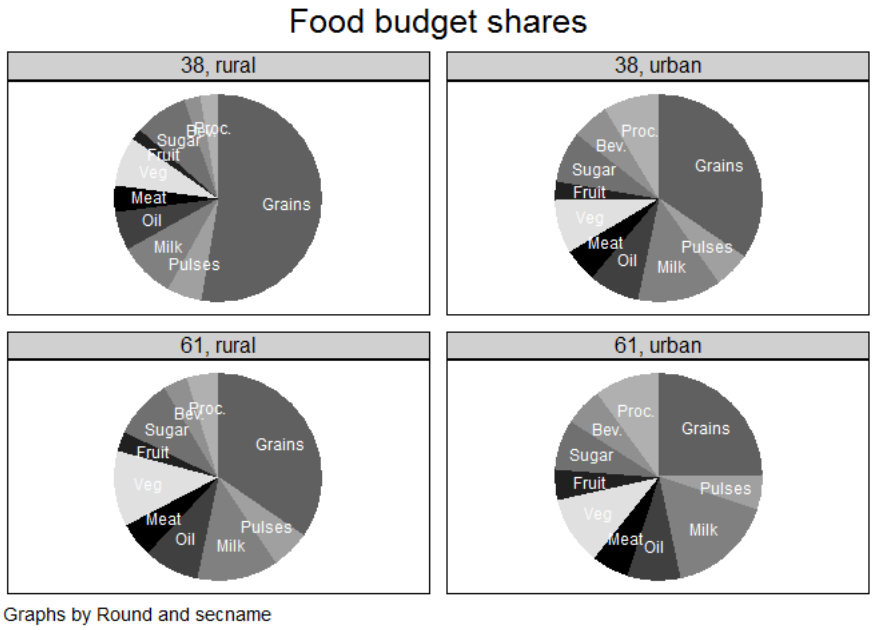
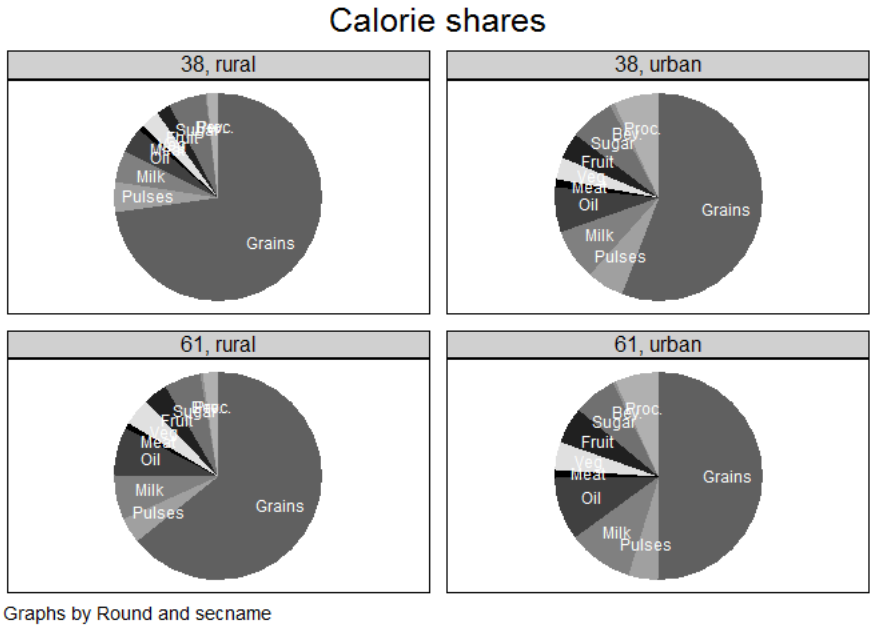


Figure 3.6: Locally weighted regression of budget share on log per capita real expenditure (two adult, 3 child household)



(a) Food expenditures

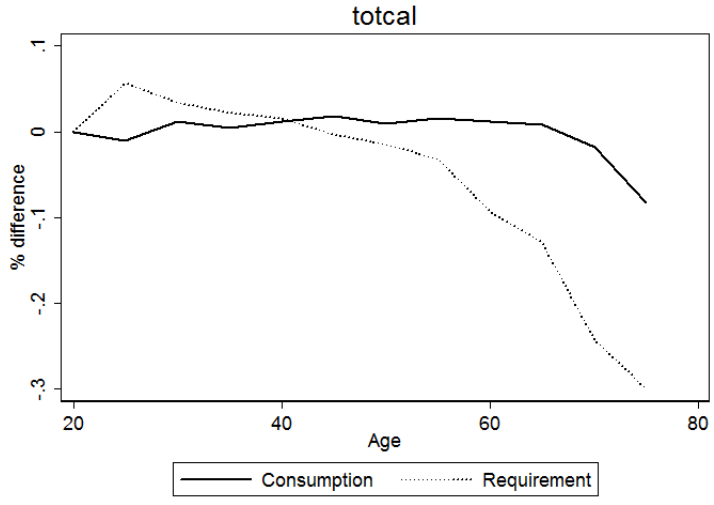


(b) Calories

Figure 3.7: Food group shares by year/sector

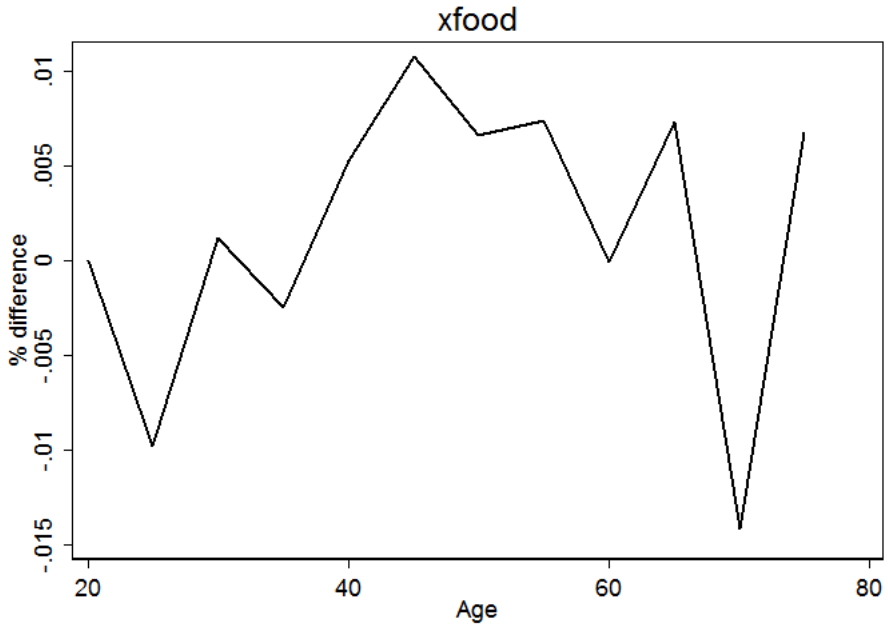


(a) Unconditional

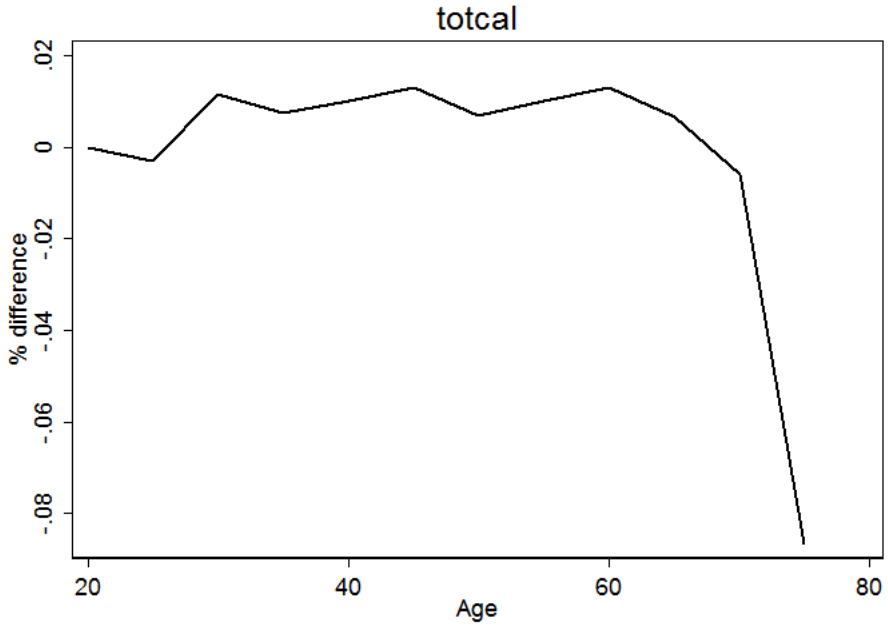


(b) Conditional on total expenditures

Figure 3.8: Life-cycle calorie consumption and requirements, relative to 19-22 year old adults

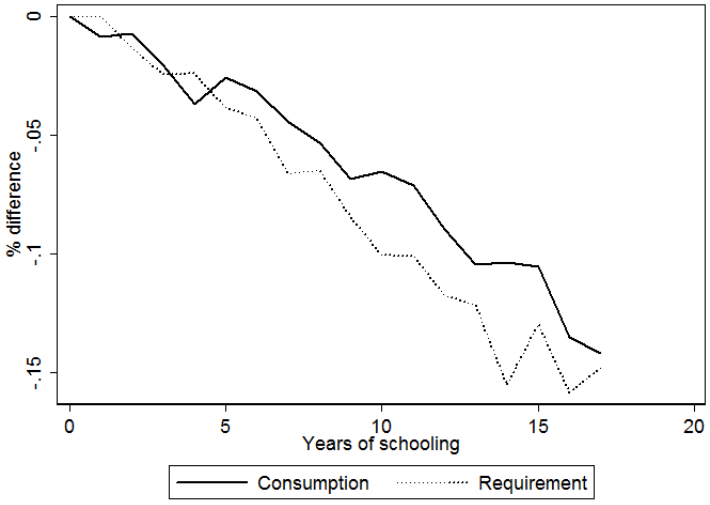


(a) Food exp. cond. on tot. exp.

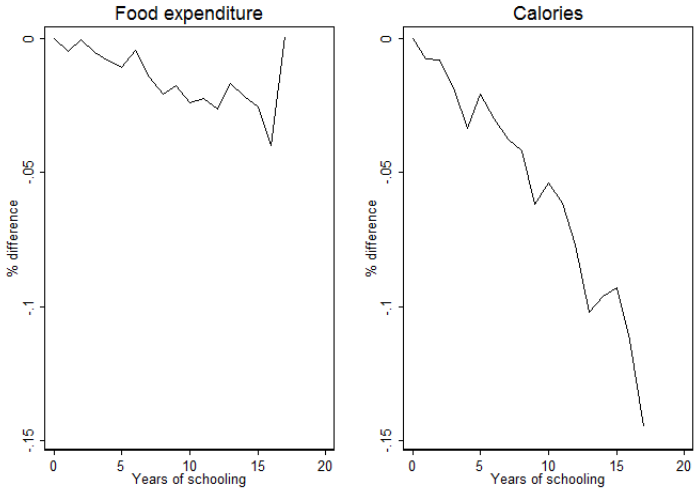


(b) Cal. cond. on food exp.

Figure 3.9: Life-cycle coefficients, relative to 19-22 year olds



(a) Cal. conditional on tot. exp.



(b) Cal. cond. on food exp. and food exp. cond. on tot. exp.

Figure 3.10: Effect of average adult years of schooling on calorie and food consumption/requirements

Chapter 4

Bibliography

Bibliography

Mark Aguair and Erik Hurst. Consumption vs. expenditure. *Journal of Political Economy*, 113(5):919–948, 2005.

Ingvild Almas. International income inequality: Measuring ppp bias by estimating engel curves for food. *Luxemburg Income Study Working Paper No. 473*, 2008.

Howarth E. Bouis. A food demand system based on demand for characteristics. *IFPRI Food Consumption and Nutrition Division Discussion Paper No. 7*, pages 1–111, 1995.

Anne Case. Does money protect health status? evidence from south african pensions. In David Wise, editor, *Perspectives on the Economics of Aging*. University of Chicago Press for NBER, Chicago, 2004.

Srikanta Chatterjee, Allan Rae, and Ranjan Ray. Food consumption and calorie intake in contemporary india. *Working Paper*, 2007.

Dora L Costa. Pensions and retirement: Evidence from union army veterans. *Quarterly Journal of Economics*, 110(2):297–319, 1995.

Dora L Costa. Displacing the family: Union army pensions and elderly living arrangements. *Journal of Political Economy*, 105(6):1269–1292, 1997.

Dora L. Costa. Estimating real income in the us from 1888 to 1994: Correcting cpi bias using engel curves. *Journal of Political Economy*, 109(6):1288–1310, 2001.

Dora L Costa. Pensions and retirement among black union army veterans. *Journal of Economic History*, 70(3):567–592, 2010.

Dora L Costa and Matthew E Kahn. *Heroes and Cowards: The Social Face of War*. Princeton University Press, Princeton, 2008.

Dora L. Costa, Lorens A. Helmchen, and Sven Wilson. Race, infection, and arteriosclerosis in the past. *The National Academy of Sciences of the USA*, 104(3): 1321913224, 2007.

David Cutler and Grant Miller. The role of public health improvements in health advances: The twentieth-century united states. *Demography*, 42(1):1–22, 2005.

David Cutler, Angus Deaton, and Adriana Lleras-Muney. The determinants of mortality. *Journal of Economic Perspectives*, 20(3):97–120, 2006.

Angus Deaton and Jean Dreze. Poverty and inequality in india: A re-examination. *Economic and Political Weekly*, September 7:3729–3748, 2002.

Angus Deaton and Jean Dreze. Food and nutrition in india: Facts and interpretations. *Economic and Political Weekly*, 44(7):42–65, 2009.

Angus Deaton and Valerie Kozel. Data and dogma: The great indian poverty debate. *World Bank Research Observer*, 2005.

Angus Deaton and Christina Paxson. Economies of scale, household size, and the demand for food. *Journal of Political Economy*, 106(5):897–930, 1998.

Angus Deaton and S Subramanian. The demand for food and calories. *Journal of Political Economy*, 104(1):133–162, 1996.

Michael Dubin. *United States Congressional Elections, 1788 - 1997*. McFarland & Company Inc., Jefferson, 1998.

Frederick Dyer. *A Compendium of the War Rebellion, Vol. 1*. Thomas Yoseloff, New York, 1959.

Shari Eli. Wealth is health: Pensions and disease onset in the gilded age. *Working Paper*, 2010.

Lia C Fernald, Xiaohui Hou, and Paul J Gertler. Oportunidades program participation and body mass index, blood pressure and self-reported health in mexican adults. *Preventing Chronic Disease*, 5(3), 2008.

Joseph P Ferrie and Werner Troesken. Water and chicago's mortality transition, 1850-1925. *Explorations in Economic History*, 45(1):1–16, 2005.

Robert Fogel. *The Escape from Hunger and Premature Death, 1700 - 2100*. Cambridge University Press, Cambridge, 2004.

William H Glasson. *Federal Military Pension in the United States*. Oxford University Press, New York, 1918.

C Gopalan, B V Rama Sastri, and S C Balasubramanian. *The Nutritive Value of Indian Foods*. Indian Council of Medical Research, Hyderabad: National Institute of Nutrition, 2004.

Michael R Haines. The urban mortality transition in the united states, 1800 - 1940. *NBER Working Paper Series on Historical Factors in Long Run Growth*, No. 134, 2001.

Bruce W. Hamilton. Using engel's law to estimate cpi bias. *American Economic Review*, 91(3):619–630, 2001.

Daniel L. Hicks. Consumption volatility, marketization, and expenditure in emerging market economies. *Working Paper*, 2010.

Frederick Hoffman. *Race Traits and Tendencies of the American Negro*. American Economic Association, New York, 1896.

Margaret Humphreys. *Intensely Human: The Health of the Black Soldier in the American Civil War*. Johns Hopkins University Press, Baltimore, 2008.

Guido Imbens and Jeffrey M Wooldridge. Control functions and related methods. *NBER Lecture Notes 6, Summer 2007*, 2007.

Seema Jayachandran, Adriana Lleras-Muney, and Kimberly V Smith. Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs. *American Economic Journal: Applied Economics*, 2(2):118–146, 2010.

Anup Karan and Ajay Mahal. Health, nutrition and poverty: Linking nutrition to consumer expenditures. In *Financing and Delivery of Health Care Services in India*. National Commission on Macroeconomics and Health, Ministry of Health and Family Welfare, Government of India, 2005.

Martin Kaufman. *American Medical Education: The Formative Years, 1765 - 1910*. Greenwood Press, Westport, 1976.

Praduman M Kumar and Madan M Dey. Long-term changes in indian food basket and nutrition. *Economic and Political Weekly*, September 1st:3567–3572, 2007.

David Lagakos and Michael Waugh. Specialization, economic development and aggregate productivity differences. *Working Paper*, 2009.

Arthur Lewbel and Krishna Pendakur. Equivalence scales. *New Palgrave Dictionary of Economics*, (2nd edition), 2007.

Nicholas Li. An engel curve for variety: Household food consumption and welfare in india. *Working Paper*, 2010.

Claudia Linares. The civil war pension law. *CPE Working Paper 2001-6*, 2001.

Trevon Logan. The transformation of hunger: The demand for calories past and present. *The Journal of Economic History*, 69(2):388–408, 2009.

Kenneth Martis. *The Historical Atlas of the United States Congressional Districts: 1789 - 1983*. The Free Press, New York, 1982.

Thomas McKeown. *The modern rise of population*. Edward Arnold, London, 1976.

JB McKinlay and SM McKinlay. The questionable contribution of medical measures to the decline of mortality in the u.s. in the 20th century. *The Milbank Memorial Fund Quarterly: Health and Society*, 55(3):405–428, 1977.

Purnima Menon, Anil Deolalikar, and Anjor Bhaskar. *India State Hunger Index: Comparisons of Hunger Across States*. IFPRI, 2009.

National Democratic Committee. *Republican Abuses at the Pension Bureau*. Brentanos, New York, 1888.

RO Polkinhorn. *The Campaign Book of the Democratic Party, 1886*. Office of the Librarian of Congress, Washington, D.C., 1886.

Samuel H Preston and Michael R Haines. *Fatal Years: Child Mortality in Late-nineteenth Century America*. Princeton University Press, Princeton, 1991.

Charles E Rosenberg. *The Care of Strangers: The Rise of America's Hospital System*. Basic Books, New York, 1987.

Donald R Shaffer. *After the Glory: The Struggles of Black Civil War Veterans*. University Press of Kansas, Kansas, 2004.

Theda Skocpol. *Protecting Soldiers and Mothers*. The Belknap Press of Harvard University Press, Cambridge, 1992.

Werner Troesken. *Water, Race, and Disease*. MIT Press, Cambridge, 2004.

U.S. Census Bureau. *Historical Statistics of the United States, Colonial Times to 1957*. U.S. Census Bureau, Washington, D.C., 1960.

U.S. Pension Bureau. *Instructions to Examining Surgeons for Pensions*. Department of the Interior, Washingotn, D.C., 1882.

Sven E Wilson. Institutional and discretional discrimination in public sector programs: The union army disability pension. *Working Paper*, 2007.

Appendix A

Appendices

A.1 Age-based Laws

Age	Income per month ($)
Over 62	12
Over 70	15
Over 75	20

	Monthly Pension Income ($) given Length of Service						
Age	90 days	6 months	1 year	1.5 years	2 years	2.5 years	3 years
Over 62	13	13.5	14	14.5	15	15.5	16
Over 66	15	15.5	16	16.5	17	18	19
Over 70	18	19	20	21.5	23	24	25
Over 75	21	22.5	24	27	30	30	30

	Monthly Pension Income ($) given Length of Service						
Age	90 days	6 months	1 year	1.5 years	2 years	2.5 years	3 years
Under 72	30	30	30	30	30	30	30
72 and over	30	32	35	38	40	40	40

A.2 Imputing caloric intake

A.2.1 Calorie Intake

As there is some disagreement about the direction and magnitude of the trend in calorie consumption and as quantitative evaluation is important to us we delve deeper into the construction of calorie intake measures. We cannot address the issue of systematic under-reporting or over-reporting using the NSS data alone, and the 30 day recall period and reliance on a single informant may bias measured food consumption in several ways. Beyond measurement error in the data itself, there are also several important assumptions and imputations that affect the calorie estimates. These can be broadly divided into 3 categories - (1)food items with no quantity data or imprecise quantity units (even though caloric conversion factors may be accurate), (2)composite food items with unknown calorie conversions (even though the quantity measures may be precise), and (3)meals received and given by the household that are not accounted for in total calories or household size (and hence bias estimates of calories per capita). Several items - most notably processed foods, beverages, and cooked meals - suffer from both the first and second problems, and there are some items with inconsistent measurement of quantity and different units across the five survey years we examine. The third problem takes two forms - meals received for free by household members (which are not recorded in the household consumption data but are sometimes recorded on the household roster) and meals given by the household to non-members. A fourth but less important issue is treatment of alcoholic beverages, which are typically not factored into food expenditures or calorie consumption but are potentially an important source of both for some households.

A.2.2 Data issues

To get a sense of the magnitude of these issues, table A.1 reports some summary statistics for consumption of the different sets of "problem" goods.[1] The first row reports the share of food expenditures on goods with no quantity data, which has been increasing over time and is higher in urban areas. Many of these goods fall into the

[1]Unless otherwise noted, all summary statistics reported are weighted using the multiplier factors provided by the surveys. We use the combined central and state samples and use data from the 17 biggest states, urban Delhi, and Meghalaya.

processed food and "other" categories. The second row reports the share of expenditures on composite commodities - defined as those commodities with "other" in the description (with three exceptions - "palak/other leafy vegetables" and "other edible oils" are excluded, as their caloric content is likely to be very similar to other products in that category, and cereal substitutes are included since they include varied goods like tapioca, jackfruit suits, and sago). This narrow definition of composite commodities excludes some processed foods that could be considered composite commodities, like biscuits or salted refreshments but includes categories like "other vegetables" and "other animals" and "other dairy products" that contain quantity information. The expenditure share of the composite commodity categories has risen over time and is higher in urban areas. The third row reports spending on all items in the processed foods and beverages categories, which contain several notable composite items, items lacking quantity data, and uncertain caloric conversions - the food expenditure share of this category is much larger for urban households and it has increased by about 3 percentage points for rural and urban households over the sample period, almost doubling for rural households. The fourth row presents expenditures on cooked meals, a subset of the expenditures on processed foods, which is higher in urban areas but has actually decreased over time. Cooked meals include both restaurant meals and transfers in kind from employers so this decline need not imply a decline in restaurant meals - it could also imply increased formalization of employee-employer relations and a shift in wage versus in-kind payment. Note that the expenditure share on cooked meals remains very low compared to what is observed in wealthy countries and middle-income developing countries (CITE). The fifth row shows expenditures on alcohol as a share of food expenditures, and while there has been a 25% increase the level remains low but slightly higher in rural areas.

The sixth and seventh rows of table A.1 show the share of expenditures that can be directly converted to calories using either a conservative or a liberal imputation criteria. The conservative criteria only converts calories directly for goods that both include (a)quantity units in weights or volumes (as opposed to units or missing quantities, as is the case for most beverages, processed food, cooked meals and some fruits and other goods) and (b)obvious calorie conversions (which rules out most composite commodities even if they are measured in KG). The liberal criteria attempts to convert virtually all goods directly and only excludes goods with no quantity measures. Goods with discrete units are converted to masses, and the published caloric conversion tables (from Gopalan et al. [2004] or Karan and Mahal [2005]) are supplemented with data from the IndiaMD website and other sources. The conservative criteria only covers 80% of food expenditures in urban areas and about 90% in rural areas, and the share covered declined by 2-4% over the sample period. The liberal criteria covers over 95% and 90% of urban and rural food expenditures respectively, with a 1.4%-2.3% decrease in expenditure share. There is thus an intrinsic trade-off between measurement error induced by attempting to broaden the coverage for direct calorie conversion and the error induced by imputing the caloric content of the unconverted

part of food expenditures.

We next turn to measurement of unrecorded meals to the household and meals provided to others. The expenditure data records all expenditures by the household on food and this includes food that is given to guests, as part of ceremonies, or to employees - provided they do not live with the household and therefore do not qualify as household members. An accurate measure of per capita calorie consumption by the household requires a downward adjustment to calorie consumption due to these meals to others. Conversely, each household receives free meals as guests of other households, through school or other public programs, or from employers. The NSS instructions require that these free meals not be recorded under household consumption (with their value imputed at market rates), unless their is some payment. Thus subsidized meal purchases would be recorded but free meals from school or employers would not. There is some ambiguity as meals from employers would constitute transfers in kind and should technically be recorded in the consumption data but due to uncertain valuation this is often not the case. Since some meals are received from institutional employers or schools it is not necessary that these free meals given to others and those received balance out on average.

Table A.2 provides summary statistics on the share of households giving or receiving free meals, the mean number of meals given and received in the last month, and the median number of meals given or received conditional on giving or receiving meals. There is a clear pattern with rural households providing more meals to others than urban households and a reverse pattern for free meals received (until the last survey round). The pattern over time is less clear and a bit inconsistent, with some implausibly large jumps. As expected on average meals given exceed meals received, since all of the meals given would typically be recorded for both the giving and receiving household, while meals given by non-household employers, schools, government programs would not be recorded. While the distribution of meals given and received is quite skewed - with a few households hosting large ceremonies and a few households heavily dependent on free food received - the average effect is not quite large and is unlikely to significantly bias estimates of calorie consumption per capita. Table A.2 also includes the quantity of purchased cooked meals consumed, with the main lesson being that cooked meals are much more important to urban than rural households and their consumption has declined, particularly in urban areas. Thus the decline in expenditure share from table A.1 is not simply due to the availability of cheaper cooked meals.

A.2.3 Calorie estimates

In light of these issues we construct several different measures of calorie consumption using different imputation schemes, which helps to clarify which basic facts are quite robust and which depend on contestable assumptions. Table A.3 presents calories per capita per day using several different imputation schemes. There are three

steps to the imputation procedure. We begin with either the conservative or liberal direct conversion of calories. For goods that normally have quantities reported but are sometimes missing quantities we use the median unit value (expenditure/quantity) to impute quantity, and we also censor quantities so that no household purchases a good for a unit value more than 20 times more or less than median unit value. These two steps ensure that the calorie measurements for categories with relative few quantity observations - especially processed foods - are not biased by the presence of outliers. Next we impute the non-converted part of food expenditures using either (a)calorie/rupee for directly converted goods by household, (b)the average calorie/rupee for directly converted goods across all households, or (c)the group average calorie/rupee averaged across all households. Imputation (a) allows the calorie per rupee of expenditure to vary across households, with richer households typically having lower calories per rupee of directly converted expenditure and hence less imputed calories per rupee of non-converted expenditure. Imputations (b) and (c) remove this idiosyncrasy by averaging across all households, by sector and survey round to control for differences in prices. Measure (c) allows differences in average calorie/rupee conversion rates across different food groups, which is important given the large range in calories/rupee documented later. When performing this imputation we can also consider an adjustment factor - for example, to take account of the fact that most of the unmeasured calories come from goods with generally high cost per calorie (e.g. processed foods, beverages, other meats, ice cream) we might apply a factor of 0.5 to the calories/rupee measure.[2] Finally, having imputed the calories of the missing food, we also need to consider outliers in the data, so we calculate both the uncensored mean, the median, or the trimmed mean which drops households in the top and bottom 1% of food expenditures and direct calories imputed.

The first row of table A.3 presents the uncensored mean calories per capita per day using the liberal direct conversion and imputing the rest of the calories by multiplying the rest of expenditures by half of the calorie per rupee of expenditures directly converted for each household. This captures the fact that most of the imputed calories come from foods with a generally higher cost per calorie than the average directly converted basket, and allows the cost per calorie to rise with household budgets. The next five row each change one parameter at a time. The second row does uses the conservative direct conversion, meaning that a greater share of expenditures are imputed. The third row uses a one to one adjustment factor instead of a a one half factor, thereby assuming that the non-converted foods have a similar price per calorie as the directly converted expenditure. The fourth row imputes the non-converted expenditure using the sectoral annual average rather than the household-specific calorie/rupee factor. The fifth and sixth rows report the median and censored mean, which trims the 1% tails of the food expenditure and converted calorie distributions.

[2]Deaton and Dreze [2009] do this explicitly for cooked meals, implying that a cooked meal is equivalent to the aggregate food consumption basket with a markup of 100%.

The seventh row of table A.3 imputes the unmeasured calories using group-specific conversion factors equal to the average calorie per rupee for each group, averaged across all households. Direct imputation is done using the liberal conversion criteria (which ensures that there are at least 4 goods in each group with direct calorie conversion). Since imputation is now done by each group there is less concern about imputing the low cost per calorie of grains or pulses to goods like 'cooked meals,' 'other processed food' and 'other beverages' so we do not multiply by one half. For comparison the eighth row assumes that the imputed goods have a calorie/rupee rate half as high as the rest of the goods in the group - this might be more reasonable for some categories, such as ice-cream (which could have twice the cost per calorie as milk), other fruit (given that coconut, singara, and dried fruits and nuts are directly converted and have high calories per rupee), and cooked meals (compared to pickles, sauces, jam/jelly, and cakes). The ninth row presents the group results of row seven but trimming the 1% tails of expenditure and calories.

Altogether, the estimates presented in table A.3 strongly suggest that there has been a large decline in calories per capita for rural households and that rural households in 1983 consumed significantly more calories than urban households on average. However, there is some uncertainty about whether urban calories per capita have risen or declined and whether calories per capita in urban areas exceed those in rural areas in 2004-05. These results are sensitive to the imputation method. Using medians we sometimes find a modest increase in calories per capita in urban areas, though the range in table A.3 is quite small at -74 to 18. Using group-specific, average or higher calorie/rupee adjustment factors also tends to shift the rural-urban gap in 2004-05 in favor of urban households.

The bottom two rows of table A.3 present our two preferred specifications, corresponding to row (6) and row (9), but adding in calories from alcohol and the effect of a 'household adjustment factor.' This factor accounts for free meals and meals to others by assuming that they have the same calories per capita of other meals consumed by the household. The precise formula used is

$$hh.\ adj.\ factor = \frac{pay\ meals\ at\ home\ + pay\ meals\ outside\ + free\ meals}{pay\ meals\ at\ home\ + pay\ meals\ outside\ + meals\ to\ others} \tag{A.1}$$

Note that the 55th survey round (1999-2000) did not record meals to others so it is excluded from this calculation, even though one can include a positive inflation factor accounting for free meals consumed. Comparing rows (10) and (11) to (6) and (9) we see that these last two adjustments have a minimal effect. The adjustments tend to increase calories per capita in urban areas but by a greater amount in the early period. In rural areas the pattern is reversed, with a slightly negative adjustment in the early period and positive in the later period. The net effect is thus to decrease the fall in calories in rural areas and increase in the fall in calories in urban areas, and a modest reduction in the rural-urban gap. The magnitude of the effect overall

is at most 20%. Throughout the rest of the paper we use the estimate of row 10 as our baseline measure of caloric intake and check it against the other measures, noting the differences only if they are economically significant.

A final issue that we cannot address with our data is that the nutritional content of particular foods may vary over time and space. Many foods lose some of their nutritional content with transportation over longer distances and storage, the composition of the 'other' goods may vary systematically over different areas and periods, and the caloric content of processed foods may also vary. To the extent that transportation lowers caloric content for goods that we measure this would tend to decrease urban relative to rural calories and might also lower caloric intake further over time. For goods with unknown caloric content, our imputation procedure may capture some of these effects, as areas and periods with higher calories per rupee for directly converted goods might also have higher calories per rupee for imputed goods, but we cannot be certain.

A.3 Imputing caloric requirements

To record time-use information, the surveyors attempted to interview each member of the household over age 5 about their time-use over the preceding 24 hour period. Busy, reluctant or incapable members had their time-use recalled by another household member. Time-use was captured for up to three separate types of days - normal, abnormal, and variant - to capture variations in the weekly schedule including market days, weekend activities, etc. The measure we use is based on a weighted average of these three days based on how many days the household reported of each type in the preceding week. The interview team included both a male and female interviewer as the goal of the survey was to measure and validate the contribution of women to economic life in India.

The survey also records a number of other variables that are recorded in the same format as in the NSS consumption surveys - monthly expenditures, land ownership, religion, and scheduled caste/tribe at the household level and age, gender, education, and occupation for each household member. Unfortunately, the Time Use Survey was not carried out simultaneously with the NSS consumption survey, which means that comparable consumption data is only available for the July 1997-June 1998 period or the July 1999-June 2000 period. The closest geographical match is at the district level as individual villages and cities are not recorded or geocoded.

To go from time-use to caloric requirements, we begin by classifying the 154 different types of activities into four different levels energy requirement. This requires numerous judgment calls. The intensity level of each activity is relative to a complete state of rest (activity level 0), and we get the intensity factors from the calories per hour website at http://www.caloriesperhour.com/. The intensity ratios (relative to complete rest) are 1.7 for level 1 (playing cards), 3 for level 2 (cook-

ing/housework/walking 2mph), and 5 for level 3 (chopping wood/push-mowing). To give some concrete examples from our data, activity level 3 includes ploughing, preparing land, cleaning of land, wood cutting, chopping and stocking of firewood, and building and construction of dwellings. Activity level 2 includes cooking, sweeping, and assembling machines, equipment and other products. Activity level 1 corresponds to sedentary labor such as service in government, professional work, reading, and watching tv. Activity level 0 corresponds to sleeping or 'doing nothing, rest and relaxation.' Our classification of activities into different levels of intensity are available upon request. Assuming that households sleep 8 hours a day, spend 8 hours awake at intensity level 1 and then another 8 hours working at intensity levels 1/2/3 for heavy, moderate, or sedentary market work a 26 year old man weighing 70 kilograms would require 3952/2928/2272 calories. This lines up roughly with the ICMR recommendations of 3800/2875/2424 calories.

We take as the baseline caloric requirements those corresponding to a 70 KG 26 year old man. We then convert this energy requirement by a multiplicative factor corresponding to the relative Basal Metabolic Rate (BMR) for a person of a given age and gender. The BMR captures the energy consumed by the body at a complete state of rest for a given age and gender, and it multiplicatively scales the energy requirement of different activities that consume more energy than resting. BMR rises and falls in age, starting out higher for women but peaking earlier. Our baseline female is 62 KG. For children under age 6 we use the daily energy requirements from the India Council for Medical Research (ICMR). For infants aged 0-6 months and 7-12 months, for which the ICMR gives energy requirements by weight, we use energy requirements for 1-3 year olds. These should be a reasonable approximation of calorie requirements based on an average child growth chart plus an extra energy requirement for lactating mothers. The NSS data do not report pregnancy status so we are likely to underestimate the calorie requirements for pregnant women by about 300 calories per day according to the ICMR.

The ICMR provides daily energy requirements for adult men and women as well children of different ages, but adult caloric requirements are only divided into three activity cells - heavy, moderate, sedentary. They also do not take account of activity levels by children, an important omission given that they have separate age/gender cells for boys/girls aged 13-15 and 16-18, age ranges where child labor inside and outside the household is likely to be quite important in some areas. The ICMR theoretically provides us with an alternative set of energy requirements for analysis but we prefer our measure for several reasons - it allows us to account for household age and life-cycle effects for adults and labor by children and adolescents, we can match energy requirements to a variety of household characteristics rather than industry or occupation (which would be the only way of imputing household calorie requirements in the NSS using the ICMR recommendations), and we have a much more fine-grained measure of energy requirements that has both an extensive margin - number of hours working on different activities - and an intensive margin - requirements for activities

of different intensity.

The most important limitation of the TUS data is that we do not have a measure of the intensity of individual activities. While many agricultural tasks are likely to be highly labor intensive some may have assistance from mechanical and animal energy sources. This issue also occurs for all transportation related activities - since the TUS does not record mode of transport, we assume an activity level of 2 which would tend to overstate energy requirements for motorized vehicular transport but understate energy requirements for walking and cycling. Other limitations include the lack of data on height and weight for individuals or systematic biases in activity recall.

Table A.1: Problem foods for calorie imputation

		1983	1987-88	1993-94	1999-00	2004-05
Share of food expenditure by problem category						
		38	43	50	55	61
No quantity	Rural	0.011	0.016	0.021	0.022	0.036
	Urban	0.028	0.035	0.042	0.041	0.057
"Other"	Rural	0.028	0.028	0.028	0.028	0.035
	Urban	0.025	0.026	0.031	0.033	0.045
Proc. food and bev.	Rural	0.052	0.064	0.066	0.073	0.085
	Urban	0.144	0.150	0.157	0.146	0.159
Cooked meals	Rural	0.016	0.019	0.012	0.014	0.013
	Urban	0.059	0.060	0.056	0.046	0.046
Alcohol	Rural	0.012	0.013	0.013	0.014	0.015
	Urban	0.009	0.011	0.011	0.012	0.013
Share of food exp. with calorie conversions						
Conservative	Rural	0.906	0.891	0.889	0.880	0.865
	Urban	0.808	0.797	0.788	0.799	0.787
Liberal	Rural	0.972	0.965	0.967	0.963	0.949
	Urban	0.909	0.904	0.900	0.912	0.895

Table A.2: Cooked meals, meals to other households and free meals (per 30 days)

		1983	1987-88	1993-94	1999-00	2004-05
		Cooked meals				
Mean number	Rural	1.985	2.512	1.561	1.606	1.537
	Urban	6.574	6.557	5.577	4.580	4.468
Share consuming	Rural	0.074	0.092	0.063	0.049	0.058
	Urban	0.154	0.171	0.144	0.122	0.127
Cond. Median	Rural	12	12	12	16	12
	Urban	30	28	30	27	20
		Meals to guests, employees, ceremonies				
Mean number	Rural	14.650	10.311	10.429	.	7.862
	Urban	10.483	12.178	6.208	.	6.039
Share consuming	Rural	0.407	0.377	0.141	.	0.447
	Urban	0.354	0.350	0.104	.	0.382
Cond. Median	Rural	10	10	12		8
	Urban	10	10	12		7
		Free meals				
Mean number	Rural	7.717	6.572	6.005	6.220	11.473
	Urban	8.152	6.993	7.040	6.342	7.836
Share consuming	Rural	0.261	0.228	0.193	0.179	0.329
	Urban	0.235	0.218	0.196	0.177	0.233
Cond. Median	Rural	14	12	16	20	24
	Urban	18	16	20	20	22

Table A.3: Daily calories per person: different imputations

Direct conv.	Cal./rupee + adj. fact.	Stat.	Sect.	1983	1987-88	1993-94	1999-00	2004-05	Δ 1983 to 2005
Lib.	Ind x0.5	Mean	Rural	2350	2302	2226	2217	2121	-229
			Urban	2156	2165	2128	2201	2085	-70
Cons.	Ind x0.5	Mean	Rural	2305	2295	2213	2201	2105	-200
			Urban	2124	2150	2107	2170	2057	-67
Lib.	Ind x1	Mean	Rural	2377	2337	2261	2255	2171	-206
			Urban	2227	2254	2223	2287	2189	-39
Lib.	Avg. x0.5	Mean	Rural	2358	2312	2233	2223	2130	-229
			Urban	2214	2235	2189	2254	2159	-55
Lib.	Ind x0.5	Median	Rural	2158	2150	2107	2099	2027	-131
			Urban	2007	2046	2045	2114	2025	18
Lib.	Ind x0.5	Mean	Rural	2328	2297	2229	2217	2124	-205
		1% trim	Urban	2141	2154	2147	2203	2092	-49
Lib.	Gr.avg. x1	Mean	Rural	2341	2293	2215	2209	2104	-237
			Urban	2159	2208	2141	2243	2095	-64
Lib.	Gr. avg. x0.5	Mean	Rural	2327	2274	2200	2188	2081	-246
			Urban	2106	2135	2078	2169	2032	-74
Lib.	Gr. Avg. x1	Mean	Rural	2321	2289	2219	2212	2110	-211
		1% trim	Urban	2168	2202	2169	2250	2106	-62
Including calories from alcohol and hh. adj. factor									
Lib.	Ind x0.5	Mean	Rural	2320	2293	2244		2154	-166
		1% trim	Urban	2178	2180	2192		2121	-58
Lib.	Gr.avg. x1	Mean	Rural	2313	2285	2234		2140	-172
		1% trim	Urban	2230	2234	2214		2136	-94