Essays in the Economics of Aging

Karolos Panagiotis Arapakis

Thesis submitted for a Ph.D. in Economics.

Department of Economics

University College London

June 2022

Declaration

I, Karolos Panagiotis Arapakis, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Karolos Arapakis, June 2022

Statement of Conjoint Work

Two out of the three primary chapters that form this thesis involve conjoint work, as specified below.

Chapter II "On the Distribution and Dynamics of Medical Expenditure among the Elderly" is conjoint work with Eric French, John Bailey Jones and Jeremy McCauley. Overall, my contribution amounts to two thirds of the total.

Chapter III "Dementia and Disadvantage in the United States and England" is conjoint work with Eric Brunner, Eric French, and Jeremy McCauley. Overall, my contribution amounts to two thirds of the total. A version of this essay has been published as:

> Arapakis, K., E. Brunner, E. French, and J. McCauley (2021). "Dementia and disadvantage in the USA and England: population-based comparative study". *BMJ Open*, 11(10).

Acknowledgements

Throughout my PhD I have been helped, encouraged and supported by multiple individuals, and I apologize for not being able to name and thank each individually.

First and foremost, I would like to thank my principal supervisor Eric French, who has provided me with invaluable guidance and insights, given me many opportunities, and has taught me a lot about research. I would also like to thank my secondary supervisor, Richard Blundell, for his guidance and advice.

I have been fortunate to interact with many during my PhD, each of whom have taught me something valuable and are owed my thanks. To name a few: John Bailey Jones, Jeremy McCauley, Eric Brunner, Peter Spittal, Mara Solomon and Charisios Grivas.

The Economics Department at UCL has been a great environment to learn and conduct research. I have greatly benefited from the comments and questions of many faculty members, as well as fellow students.

Funding Acknowledgements

I gratefully acknowledge the funding that made my PhD possible. I was funded by the Alexander S. Onassis Foundation and ESRC through a fee waiver (+3 funding) and the grant Inequality and the insurance value of transfers across the life cycle (ES/P001831/1).

Abstract

This thesis is made up of three main essays that aim to develop a deeper understanding of issues involving the public insurance programs for the elderly, and the risks they insure against.

In the first essay (Chapter 2), using data from the Health and Retirement Study linked to administrative Medicare and Medicaid records, along with the Medical Expenditure Panel Survey, we estimate the stochastic process for total and out-of-pocket medical spending. By focussing on dynamics, we consider not only the risk of catastrophic expenses in a single year, but also the risk of moderate but persistent expenses that accumulate into a catastrophic lifetime cost. We also assess the reduction in out-of-pocket medical spending provided by public insurance schemes such as Medicare or Medicaid. We find that although Medicare and Medicaid pay the majority of medical expenses, households at age 65 will on average incur $59,000 in out-of-pocket costs with 10 percent of households incurring more than $121,000 in out-of-pocket expenses over their remaining lives.

In the second essay (Chapter 3), we compare dementia prevalence and how it varies by socioeconomic status (SES) in the United States and England. We compare between country differences in age-gender standardized dementia prevalence, across the SES gradient. Dementia prevalence was estimated in each country using an algorithm based on an identical battery of demographic, cognitive, and functional measures. Dementia prevalence is higher among the disadvantaged in both countries, with the United States being more unequal according to four measures of SES. Once past health factors and education were controlled for, most of the within country inequalities disappeared; however, the cross-country difference in prevalence for those in the lowest income decile remained disproportionately high. This provides evidence that disadvantage in the United States is a disproportionately high risk factor for dementia.

In the final essay (Chapter 4), we assess the optimal structure the U.S. Social Security system, taking into account the current system's unfunded liabilities, transition dynamics and political feasibility constraints. We base the assessment on an estimated overlapping generations general equilibrium model that features both aggregate and idiosyncratic uncertainty. The quantitative analysis establishes that although transition costs greatly restrict the U.S. government's ability to move away from the current Social Security system, ignoring the political feasibility constraints allows the government to increase welfare by transitioning to a more progressive and less costly to operate system. However, taking into account the political feasibility constraints overturns this result, as no reform is simultaneously welfare increasing and politically feasible.

Impact Statement

This thesis has direct policy relevance for agencies providing public insurance in both the US and UK. Each chapter presents evidence that can inform the wider research community and help shape future policy.

Chapter 2 has policy relevance for healthcare provision in the US. We show that although Medicare and Medicaid pay the majority of medical expenses, households at age 65 will on average incur $59,000 in out-of-pocket costs with 10 percent of households incurring more than $121,000 in out-of-pocket expenses over their remaining lives. Therefore, out-of-pocked medical spending can accumulate to a catastrophic lifetime cost, despite Medicaid and Medicare's nearly universal coverage.

Chapter 3 has direct policy implications for healthcare provisions in the US and England that aim to insure against dementia. We show that dementia prevalence is higher among the disadvantaged in both countries, with the United States being more unequal according to four measures of socio-economic status. We also show that although controlling for past health factors and education eliminates most of the within country inequalities, the cross-country difference in prevalence for those in lowest income decile remains disproportionately high. The implications of our results are that interventions designed to prevent dementia should be targeted towards the most disadvantaged. This is especially true in the US.

Chapter 4 has direct policy implications for the design and reform of the US Social Security system. The chapter presents a tractable model that captures the value of insurance, along with the distortions imposed by the existence and reform of the US Social Security system. Results suggest that although transition costs greatly restrict the U.S. government's ability to move away from the current Social Security system, ignoring the political feasibility constraints allows the government to increase welfare by transitioning to a more progressive and less costly to operate system. However, political feasibility constraints are far more restrictive than transition costs. Our results highlight the competing interests of different generations, that manifest even for minor Social Security reforms. While future generations value these reforms, those alive at the time of the reform suffer the costs. Hence, policy makers should take these competing interests into account.

I have disseminated this research to academic audiences in the UK and abroad. I will further disseminate this research through scholarly publications.

Contents

1 Introduction **12**

Bibliography 13

2 On the Distribution and Dynamics of Medical Expenditure among the Elderly **15**

2.1 Introduction 15

2.2 Previous Literature 16

2.3 Data and Descriptive Statistics 17

2.3.1 The HRS 18

2.3.2 Administrative Medicare and Medicaid Records 19

2.3.3 Imputations Using MEPS Data 19

2.3.4 Our Medical Spending Measure 20

2.4 The Model 21

2.4.1 Total Medical Spending 22

2.4.2 Health and Mortality 26

2.5 Results 26

2.5.1 Health and Longevity 26

2.5.2 The Cross Sectional Distribution of Medical Spending and Parameter Estimates 27

2.5.3 The Distribution of Lifetime Medical Spending 28

2.5.4 Out-of-Pocket Medical Spending 30

2.6 Discussion and Conclusions 31

2.A Appendix . . . 32
2.A.1 Our Medicare and Medicaid Data . . . 32
2.A.2 Sample Selection . . . 33
2.A.3 Imputing Missing Medical Expenditures . . . 33
2.A.4 Validating the Administrative Medical Spending Data . . . 35
2.A.5 Plotting the Moments of Medical Spending . . . 37
2.A.6 Estimation methodology . . . 40
2.A.7 Demographic Transition Probabilities in the HRS . . . 41
Bibliography . . . 43

3 Dementia and Disadvantage in the United States and England 46

3.1 Introduction . . . 46
3.2 Strengths and limitations of this study . . . 47
3.3 Methods . . . 47
3.3.1 Description of Surveys . . . 47
3.3.2 Cohort Description . . . 48
3.3.3 Patient and Public Involvement . . . 48
3.3.4 Dementia Case Definition . . . 48
3.3.5 Measures of Socioeconomic Status . . . 49
3.4 Statistical Analysis . . . 49
3.5 Results . . . 50
3.6 Discussion . . . 53
3.6.1 Main findings . . . 53
3.6.2 Comparison with previous studies and how findings are an advance on current literature . . . 54
3.6.3 Implications (wider interpretation) . . . 55
3.6.4 Strengths and Weaknesses of Analysis . . . 55

3.7 Conclusion . . . 56

3.A Appendix . . . 57

3.A.1 HRS Data . . . 58

3.A.2 ELSA Data . . . 58

3.A.3 Harmonised Variables HRS and ELSA . . . 58

3.A.4 Hurd et al. Dementia Prediction Algorithm . . . 59

Bibliography . . . 71

4 Transition Costs and the Optimal U.S. Social Security System 74

4.1 Introduction . . . 74

4.2 The Model . . . 77

4.2.1 Demographics . . . 78

4.2.2 Households . . . 78

4.2.3 Firms . . . 81

4.2.4 The Government . . . 81

4.2.5 Equilibrium Definition . . . 82

4.3 Estimation . . . 82

4.3.1 First Step . . . 83

4.3.2 Second Step . . . 84

4.4 Data . . . 85

4.5 Calibration and Estimation Results . . . 85

4.5.1 Calibration . . . 85

4.5.2 First Step Estimates . . . 86

4.5.3 Second Step Estimates . . . 87

4.6 The Quantitative analysis . . . 91

4.6.1 The Government's Problem . . . 91

4.6.2 The transition optimum . . . 92
4.6.3 Political Support . . . 95
4.6.4 Welfare increasing and politically feasible reforms . . . 96
4.7 Sensitivity Analysis . . . 97
4.8 Conclusions . . . 98
4.9 Appendix A - Model details . . . 100
4.9.1 Model Timing . . . 100
4.9.2 Private Pension Payments . . . 100
4.9.3 Social Security Payments . . . 101
4.10 Appendix B - Computational details . . . 101
4.10.1 Household problem . . . 101
4.10.2 Algorithm accuracy checks . . . 102
4.11 Appendix C - Estimation details . . . 104
4.11.1 First step . . . 104
4.11.2 Second step . . . 105
Bibliography . . . 107

List of Figures

2.1 Mean and 90^{th} Percentile of Household Medical Spending by Age 21

2.2 Moments of the Medical Spending Distribution 24

2.3 Persistence of medical spending 28

2.4 Mean and 90^{th} Percentile of Annual and Remaining Lifetime Medical Spending. 29

2.5 Mean and 90^{th} Percentile of Annual and Remaining Lifetime Out-of-Pocket Medical Spending. . . . 30

3.1 SES Gradient of Dementia, USA versus England 53

3.2 SES Gradient of Dementia, US vs England, Controlling for Past Health 67

4.1 Wage life-cycle profile 87

4.2 Data and model simulated life-cycle profiles 88

4.3 The Social Security benefit functions 92

4.4 The transition welfare function 93

4.5 Life-cycle profiles for the status quo and the transition optimum 95

4.6 Household voting 96

4.7 Welfare increasing and politically feasible reforms 97

List of Tables

2.1 Medical Spending by Total Expenditures and Payor 20

2.2 Life expectancy in years, conditional on reaching age 65 27

2.3 Sample Selection 33

2.4 Individual Medical Spending Percentiles: HRS versus MCBS 36

2.5 Income, Out-of-pocket Spending, and Medicaid Recipiency Rates: HRS versus MCBS 37

3.1 Prevalence of dementia, USA versus England, 2017 52

3.2 Difference in prevalence of dementia, USA vs England, 2017 54

3.3 Studies on Dementia Prevalence by Socioeconomic Status (SES) 57

3.4 HRS and ELSA Response Rates 60

3.5 Summary Statistics, Raw Sample, England vs United States 61

3.6 Summary Statistics, Age-Gender Standardized, England vs United States 62

3.7 Summary Statistics for Predictor Variables, England vs United States 63

3.8 Indexes of Inequality 64

3.9 Difference in Prevalence of Dementia with Controls, United States vs England, 2017 65

3.10 Difference in Prevalence of Dementia with Controls, by Income Decile 66

3.11 Prevalence of Dementia, United States vs. England, Full-Sample 68

3.12 Difference in Prevalence of Dementia with Controls, Full Sample 69

3.13 Prevalence of Dementia in United States, Whites versus Non-whites 70

4.1 Ffirst step GMM estimates . . . 86

4.2 Second Step MSM estimates . . . 87

4.3 The model's parameters . . . 90

4.4 Difference in aggregate statistics . . . 94

4.5 CEV decomposition . . . 94

4.6 The table shows three statistics on the household forecast errors. . . . 102

Chapter 1

Introduction

In this thesis, I use applied micro and macroeconomic techniques to address a range of questions pertaining to public insurance programs and the risks that they insure against, drawing evidence from the US and UK. There are three primary chapters. The chapters address topics that should be of interest to policy makers. Each chapter can be read as a free-standing essay, however there are important connections between each and all highlight the risks that the elderly face.

The first two chapters focus on modelling and analysing the medical spending and dementia risks that the elderly face. In Chapter 2 (co-authored with Eric French and John Bailey Jones and Jeremy McCauley) we estimate the stochastic process for total medical spending. We also assess the reduction in out-of-pocket medical spending risk provided by public insurance schemes such as Medicare or Medicaid. In Chapter 3 (based on co-authored work with Eric Brunner, Eric French and Jeremy McCauley) we we compare dementia prevalence and how it varies by socio-economic status (SES) across the United States and England. We also compare between country differences in age-gender standardized dementia prevalence, across the SES gradient.

In Chapter 2 our results suggest that lifetime medical spending is high and uncertain. The spending's level and the dispersion diminish slowly with age. Most of these expenditures will be covered by Medicare, Medicaid, or other private and public insurers. Our data suggest that out-of-pocket spending rises more or less linearly with total spending, with households covering about 22% of their total expenditure. We find that at age 65, households will on average incur over $59,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of $121,000. A household alive at age 90 will on average spend more than $21,000 out-of-pocket before they die, highlighting the slow decline of expected lifetime costs as households age.

In Chapter 3 we find that Dementia prevalence is lower in England at 9.7% than the US at 11.2%, a difference of 1.4 percentage points that is highly statistically significant. For all four SES measures, income, education, wealth, and non-housing wealth, there is a clear gradient in dementia prevalence, with the most disadvantaged groups in both England and the US having higher dementia prevalence. Also, we find that the poorest individuals in the US face a disproportionately high burden. We find that the high burden can be partly but not fully explained by past health factors. Specifically, once past health factors and education were controlled for, althought most of the within country inequalities disappeared, the cross-country difference in prevalence for those in lowest income decile remained disproportionately high.

In the final chapter of the thesis, Chapter 4, we assess the US government's ability to reform the current Social Security system. We base the assessment on an estimated overlapping generations general equilibrium model that features both aggregate and idiosyncratic uncertainty[1]. The model is estimated using the PSID data, and it includes an infinitely lived government, a Pay-as-You-Go Social Security system, a private pension fund, households of multiple generations and competitive firms. This setup allows us to study both the short term and the longer term implications of the reform.

The findings suggest that the U.S. government's ability to move away from the current system is greatly restricted by transition costs. However, the U.S. government can implement a new Social Security system that is more progressive and less costly to operate. The new system imposes fewer distortions in the model economy and increases welfare on both the transition path and the longer-term. Specifically, a simultaneous reduction in the average level of benefits combined with a reform of the benefit function can deliver welfare gains, that provide an increase of 0.7% in terms of consumption equivalent variation. The life-cycle profiles of the current and new systems are similar.

Then we consider household votes at the start of the transition and find that reforms that reduce the level of benefits do not gather a household majority. Given the model's results, we create the sets of welfare increasing and politically feasible Social Security reforms. We find that the intersection between the two sets is equal to the empty set. This implies that transitioning to a new Social Security system is no longer possible, as there is no reform that is simultaneously welfare increasing and politically feasible.

Bibliography

Conesa, J. C. and D. Krueger (1999). Social security reform with heterogeneous agents. *Review of Economic Dynamics 2*(4), 757–795.

Harenberg, D. and A. Ludwig (2019). Idiosyncratic risk, aggregate risk, and the welfare effects of social security. *International Economic Review 60*(2), 661–692.

Huggett, M. and J. Parra (2010). How well does the u.s. social insurance system provide social insurance? *Journal of Political Economy 118*(1), 76–112.

Jones, J. and Y. Li (2020, 07). Social security reform with heterogeneous mortality. *Federal Reserve Bank of Richmond Working Papers 20*, 1–65.

[1]There is a large literature on Social Security reforms. See for example Nishiyama and Smetters (2007); Huggett and Parra (2010); Kitao (2014); Jones and Li (2020). We selected to model both risks as Conesa and Krueger (1999) emphasize the importance of heterogeneity for household voting, and Harenberg and Ludwig (2019) the importance modelling both risks for welfare analysis.

Kitao, S. (2014). Sustainable social security: Four options. *Review of Economic Dynamics 17*(4), 756–779.

Nishiyama, S. and K. Smetters (2007). Does social security privatization produce efficiency gains? *The Quarterly Journal of Economics 122*(4), 1677–1719.

Chapter 2

On the Distribution and Dynamics of Medical Expenditure among the Elderly

2.1 Introduction

Despite nearly universal enrollment in Medicare, most elderly Americans still face the risk of catastrophic medical expenses. This is because Medicare does not pay for long hospital and nursing home stays and requires co-payments for many other treatments. Medicaid fills many of these gaps, but only for households that pass a means test. Medical spending is thus a major financial concern among elderly households. In a recent survey, more affluent individuals were worried about rising health care costs than about any other financial issue (Merrill Lynch Wealth Management 2012).

In this paper we document patterns of medical spending among older households, distinguishing between spending covered by public insurance programs such as Medicare or Medicaid and the out-of-pocket expenses borne by the households themselves. Even though numerous papers have estimated the medical spending risks that older Americans face in any given year, very few studies have estimated the distribution of cumulative lifetime spending. These lifetime totals, however, are critical when assessing the income and savings adequacy of older households. Households care not only about the risk of catastrophic expenses in a single year, but also about the risk of moderate but persistent expenses that accumulate into catastrophic lifetime costs. We use new data and methods to improve the measurement and assessment of this risk. In particular, we make three contributions.

First, to the best of our knowledge, this is the first paper to estimate the dynamic process for total spending by all payors among the population aged 65+. Our main dataset is the Health and Retirement Study (HRS) which has high quality information on out-of-pocket medical spending, linked to Medicare and Medicaid administrative data. To this we add data from the Medical Expenditure Panel Survey (MEPS), which allow us to impute private insurance and other payments not measured directly in the HRS. This yields an estimate of medical spending for

HRS households that accounts for all payors, giving us the first long panel measures of comprehensive medical spending for the age 65+ population.

Second, we estimate the stochastic process for medical spending with a specification far more flexible than those used in previous studies. Specifically, we use the framework developed by Arellano et al. (2017), which allows for non-linear persistence and non-normal shocks. Using this specification allows us to more accurately predict the distribution of medical spending, both annual and cumulative, after age 65. This allows us to better understand the risks facing elderly households over the entirety of their remaining life.

Third, we model the share of medical spending paid for by Medicare and Medicaid, and calculate the extent to which Medicare and Medicaid reduce lifetime medical spending. Using detailed data and an advanced methodology allow us to better understand who benefits from Medicaid and Medicare.

In our framework medical spending depends on household structure and health, among other factors. We estimate dynamic models of these variables. Simulating our estimated models allows us to construct household histories and compute the distribution of total lifetime medical spending. Thus we can calculate the share of people who face catastrophic medical spending over the course of their lives.

We find that lifetime medical spending during retirement is high and uncertain. Over their remaining lives, households at age 65 will incur, on average, $272,000 in total medical spending, of which $59,000 will be paid out of pocket. At the top tail, 10% of households will incur more than $563,000 in total medical spending of which $121,000 will be paid out-of-pocket. The level and the dispersion of remaining lifetime spending diminishes only slowly with age. For example, a household alive at age 90 will on average spend more than $99,000 in total and $21,000 out-of-pocket before they die. The reason for this is that as households age, surviving individuals on average have fewer remaining years of life, but are also more likely to live to extremely old age when medical spending is very high. Although initial health, and initial marital status have large effects on this spending, much of the dispersion in lifetime spending is due to events realized in later years. We find that Medicare and Medicaid cover a large amount of lifetime medical spending, substantially lowering the risk of catastrophic medical bills. However, on average, households still pay for 22% of all medical spending out of pocket.

The rest of the paper is organized as follows. Section 2.2 contains a literature review. In section 2.3 we discuss some key features of the data sets that we use in our analysis, the HRS and MEPS, and describe how we construct our measure of medical spending. In section 2.4 we introduce our model and describe our simulation methodology. We discuss our results in section 2.5 and conclude in section 2.6.

2.2 Previous Literature

The life-cycle implications of health and medical spending, especially at older ages, are an area of considerable interest. Several papers (De Nardi et al. 2010, Kopecky and Koreshkova 2014, Ameriks et al. 2020, De Nardi et al. 2021) show that health care costs that rise with age and income can explain much of the U.S. elderly's saving behavior. Other work suggests that medical spending risk is important in explaining cross-country differences in the consumption (Banks et al. 2019) and savings decisions (Nakajima and Telyukova 2018) of elderly households. Still other studies have considered the role of medical expenses in: bankruptcy (Livshits et al. 2007); the adequacy of savings at retirement (Scholz et al. 2006, Skinner 2007); and annuitization (Lockwood 2012, Pashchenko 2013, Reichling and Smetters 2015).

Relative to the work exploring the consequences of medical spending risk, the number of papers attempting to measure this risk is limited. Most studies in this literature analyze out-of-pocket spending (e.g., Feenberg and Skinner 1994, French and Jones 2004, Fahle et al. 2016, Hurd et al. 2017); the studies analyzing total spending in the U.S. use data from individual insurance plans (Alemayehu and Warner 2004, Hirth et al. 2016) or short panels of two years or less (Pashchenko and Porapakkarm 2016).

Several studies have found that out-of-pocket medical spending shocks are reasonably well-described as the sum of a persistent AR(1) process and a white noise shock (Feenberg and Skinner 1994, French and Jones 2004, Jones et al. 2018, De Nardi et al. 2021). French and Jones (2004) also find that the innovations to this process can be modelled with a normal distribution that has been adjusted to capture the risk of catastrophic health care costs. However, they also note that their procedure must be applied with care. They point out, for example, that estimates are sensitive to bottom coding decisions, in part because of left skewness in the data that the normal distribution does not capture. De Nardi et al. (2010, 2016) extend French and Jones's (2004) spending model to account for additional covariates, and embed the medical spending model in a model of lifecycle savings decisions, but consider only singles and do not control for end-of-life events.[1] Jones et al. (2018) and De Nardi et al. (2021) extend the model further to account for the risk of end-of-life spending and the joint spending of couples. Our paper differs from these studies in two ways. First, it analyzes spending by all payors, as opposed to just out-of-pocket spending. This allows us to assess the extent to which Medicare and Medicaid reduce the medical spending risk facing older households. Second, we replace the common specification, which assumes a Gaussian AR(1)-plus-white noise, with a more advanced semi-parametric framework developed by Arellano et al. (2017). This framework allows for non-Gaussian shocks and persistence that varies with age and the value of the shock.[2]

2.3 Data and Descriptive Statistics

We use HRS data matched with administrative Medicare and Medicaid records. We use MEPS data to impute payments for payors missing in the HRS. The result is a version of HRS medical spending data that is representative of all payors. We use data from 1999-2012, when we also have the Medicare and Medicaid records.

[1] See French et al. (2006) and Poterba et al. (2017) on the importance of these events.

[2] Yet another model of out-of-pocket spending appears in Webb and Zhivan (2010). They estimate a rich model of stochastic morbidity and mortality with multiple health indicators and assume that medical expenditures are a function of these health conditions, along with a collection of socioeconomic indicators. In their framework, all of the variation in medical spending is due to variation in these controls; there are no residual shocks.

2.3.1 The HRS

The HRS is a representative biennial survey of the population ages 51 and older, and their spouses. To focus on the Medicare population, we focus on those ages 65 and older, using data starting in 1999. Although drawn from the non-institutionalized population when first interviewed, these individuals are tracked and reinterviewed as they enter nursing homes and other institutions. Consistency with our demographic model also leads us to drop a small number of households who, for example, are "partnered" or whose partner reports conflicting marital status. Furthermore, we drop households who do not consent to provide their Medicare and Medicaid information or there appear to be problems with the matching. This leaves us with 4,391 households comprising 39,002 household year observations. Appendix 2.A.2 documents our sample selection criteria.

The HRS conducts interviews every other year. Households are followed until both members die; attrition for other reasons is low. When the respondent for a household dies, in the next wave an "exit" interview with a knowledgeable party – usually another family member – is conducted. This allows the HRS to collect data on end-of-life medical conditions and spending (including burial costs). Fahle et al. (2016), compare the medical spending data from the "core" and exit interviews, show that out-of-pocket spending rises significantly in the last year of life.

The HRS has a variety of health indicators. We assign individuals to the nursing home state if they were in a nursing home at least 120 days since the last interview or if they spent at least 60 days in a nursing home before the next scheduled interview and died before that scheduled interview. We assign the remaining individuals a health status of "good" if their self-reported health is excellent, very good or good and a health status of "bad" if their self-reported health is fair or poor.

The HRS collects data on all out-of-pocket medical expenses, including private insurance premia and nursing home care. The HRS medical spending measure is backward-looking: medical spending at any wave is measured as total out-of-pocket spending over the preceding two years. French et al. (2017) compare out-of-pocket medical spending data from the HRS, MEPS, and the Medicare Current Beneficiary Survey (MCBS). They find that the HRS matches up well with the MEPS for items that MEPS covers, but that the HRS is more comprehensive than the MEPS in terms of the items covered.

To control for socioeconomic status, we follow De Nardi et al. (2021) and construct a measure of lifetime earnings or "permanent income" (PI). We first find each household's "non-asset" income, a pension measure that includes Social Security benefits, defined benefit pension benefits, veterans benefits and annuities. Because there is a roughly monotonic relationship between lifetime earnings and these pension variables, post-retirement non-asset income is a good measure of lifetime permanent income. Since these income sources tend to change only in response to inflation and death of a family member, we assume this income is a deterministic function of a household specific effect (which is a measure of PI) and household composition. We then use fixed effects regression to convert non-asset income, which depends on age and household composition as well as lifetime earnings, to a scalar measure comparable across all households. In particular, we assume that the log of non-asset income for household i at age t follows

$$\ln y_{it} = \alpha_i + \kappa(t, f_{it}) + \omega_{it}, \tag{2.1}$$

where: α_i is a household-specific effect; $\kappa(t, f_{it})$ is a flexible function of age and family structure f_{it} (i.e., couple, single man, or single woman); and ω_{it} represents measurement error. The percentile ranks of the estimated fixed effects, $\widehat{\alpha}_i$, form our measure of permanent income, $\widehat{I}_i$.

2.3.2 Administrative Medicare and Medicaid Records

The Centers for Medicare and Medicaid Services (CMS) have confidential administrative spending records for Medicare (Parts A, B and D) and Medicaid, that we link to the survey responses of consenting HRS respondents. We have Medicare data for each year between 1991 and 2016. These records include reimbursement amounts for inpatient, skilled nursing facility, home health, and hospice claims made under Medicare Part A, as well as outpatient, carrier (non-institutional medical care providers such as individual or group practitioners, non-hospital labs, and ambulances), and durable-medical-equipment payments made under Medicare Part B. To this total we add drug-related spending made under Medicare Part D, which began in 2006.

As with the Medicare records, we link restricted Medicaid data for those who give permission, allowing us to measure Medicaid spending for each year between 1999 and 2012. The Medicaid files contain information on enrollment, service use, and spending. Appendix 2.A.1 describes the Medicare and Medicaid data in more detail. Linking these data to the HRS results in a broad set of spending measures for the years 1999-2012. These will be the years used in our main analysis.

2.3.3 Imputations Using MEPS Data

While the HRS contains accurate measures of out-of-pocket medical spending and can be linked to Medicare and Medicaid records, it does not contain information on the payments made by private and smaller public insurers (such as the Veterans Administration or state and local health departments). To circumvent this issue, we use data from the 1996-2017 waves of the Medical Expenditure Panel Survey (MEPS).

The MEPS is a nationally representative survey of non-institutionalized households. MEPS respondents are interviewed up to 5 times over a 2 year period, forming short panels. We aggregate the data to an annual level. MEPS respondents are asked about their (and their spouse's) health status, health insurance, and the health care expenditures paid out-of-pocket, by Medicaid, by Medicare, private insurance and by other sources. The survey responses are matched to medical spending information provided by health care providers. Although the MEPS does not capture certain types of medical expenditures, such as nursing home expenditures, it captures sources of medical spending extremely well.[3]

[3]Pashchenko and Porapakkarm (2016) compare MEPS data to the aggregate statistics and show that MEPS captures types of

Table 2.1: Medical Spending by Total Expenditures and Payor

Total Spending Percentiles	Total Spending		Percent paid by			
	Average Exp.	Pct. of Total	Out-of-Pocket	Medicare	Medicaid	Other
All	24,000	100.0	21.6	64.0	10.7	3.7
95-100%	138,700	28.9	24.7	61.0	11.2	3.1
90-95%	74,200	15.5	17.9	63.9	16.2	2.0
70-90%	39,200	32.7	16.8	68.7	11.6	2.9
50-70%	16,000	13.4	23.3	64.7	6.0	6.1
0-50%	4,600	9.5	32.2	56.3	4.0	7.6

Notes: Total spending is the sum of annual household Medicare, Medicaid, out-of-pocket and other spending, age 65+. Other includes private insurance and other government payors. Expenditures are expressed in 2014 dollars.

To impute medical spending not captured in the HRS, we proceed in two steps. First, we use the MEPS data to regress these payments on a set of observable variables found in both data sets. Variables include household income, a fourth order age polynomial, labor force participation status, education, marital status, doctor and hospital visits, race indicators, health measures, out-of-pocket spending and interactions. This regression has an R^2 statistic of 0.13 for private insurance payments and 0.02 for other payors. Second, we impute these expenses in the HRS data using a conditional mean-matching procedure, a procedure very similar to hot-decking. Applying the MEPS regression coefficients to the HRS data yields predicted values for each HRS household, to which we add residuals drawn from MEPS households with similar levels of predicted spending. We describe our approach in more detail in Appendix 2.A.3.

2.3.4 Our Medical Spending Measure

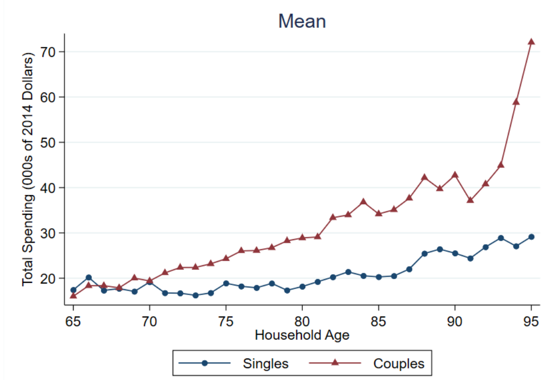
Table 2.1 summarizes the medical spending data contained in our HRS sample. The first two columns of the table show total spending by all payors for older households, sorted by total spending percentile, while the remaining 4 columns decompose this total by payor. Table 2.1 shows that Medicare and especially Medicaid benefits are concentrated among the top half of the total spending distribution, helping to ensure against catastrophic expenses. Another way to see this outcome is to note that at the bottom half of the spending distribution, about 32% of medical expenses are paid out of pocket, while at the top, out-of-pocket expenditures comprise between about 17% and 25% of the total.

In Appendix 2.A.4 we compare our medical spending measures against the MCBS. French et al. (2017) finds that the data on out-of-pocket and insurance premia in the two surveys match up well. We extend their exercise by comparing Medicare and Medicaid payments in the HRS restricted data to Medicare and Medicaid payments in the MCBS.

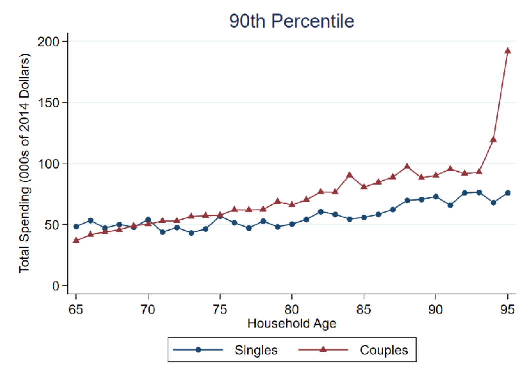
Figure 2.1 shows how the mean and 90^{th} percentile of both total and out-of-pocket medical spending change with

spending very well.

the age of the household head. Medical spending rises rapidly with age for both singles and couples. As noted in Table 2.1, medical spending is very concentrated. Thus, it should come as no surprise that the 90^{th} percentile is significantly higher than the mean. Spending for couples is higher than for singles, but in most cases is much less than double the spending of singles. It bears noting that as households age, they tend to transition from couples to singles; at older ages, the number of couples is small, leading to erratic profiles.

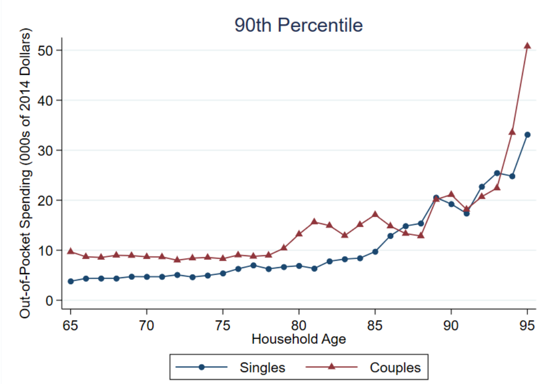


(a) Total Spending: Means



(b) Total Spending: 90^{th} Percentile

(c) Out-of-Pocket: Means



(d) Out-of-Pocket: 90^{th} Percentile

Figure 2.1: Mean and 90^{th} Percentile of Household Medical Spending by Age

2.4 The Model

We estimate the distribution of lifetime medical spending in five steps. In the first step we estimate the log of total medical spending as a function of age, health, family structure and PI, using Ordinary Least Squares (OLS). In the second step, we estimate the stochastic process for the unexplained component of medical spending – the residuals from the first step regression – using the methodology developed by Arellano et al. (2017). In the third step, we estimate the mapping from total medical spending to out-of-pocket medical spending. In the fourth step we estimate a Markov Chain model of health and mortality. In the final step, we use the estimated models to simulate health, mortality, and lifetime medical spending.

2.4.1 Total Medical Spending

Underlying framework

Let $M_{i,t}$ denote total medical spending for household i at time t ,and let $m_{i,t}$ denote its logarithm net of the observed variables contained in the vector $X_{i,t}$:

$$\ln M_{i,t} = X_{i,t}\gamma + m_{i,t}. \tag{2.2}$$

We assume that $m_{i,t}$ can be expressed as the sum of the persistent first-order Markov component $\eta_{i,t}$ and the transitory component $\varepsilon_{i,t}$:

$$m_{i,t} = \eta_{i,t} + \varepsilon_{i,t}, \quad \forall i \in \{1, ..., N\}, \forall t \in \{1, ..., T\}. \tag{2.3}$$

We assume the transitory component is i.i.d., but require only that it be zero-mean and satisfy the regularity conditions set forth in Arellano et al. (2017). In practice the vector $X_{i,t}$ includes an age polynomial, health indicators, household structure and death-year indicators, PI percentile, and interactions among the aforementioned variables.

A common specification

Many previous studies, following Feenberg and Skinner (1994) and French and Jones (2004), have used the following specification for the shocks:

$$\eta_{i,t} = \phi\eta_{i,t-1} + \zeta_{i,t}, \tag{2.4}$$
$$\eta_{i,init} \overset{iid}{\sim} N\Big(0, \sigma^2_{\eta_{init}}\Big),\ \zeta_{i,t} \overset{iid}{\sim} N\Big(0, \sigma^2_{\zeta}\Big),\ \varepsilon_{i,t} \overset{iid}{\sim} N(0, \sigma^2_{\varepsilon}). \tag{2.5}$$

Thus, the persistent component $\eta_{i,t}$ is an AR(1) with the innovation $\zeta_{i,t}$ independent of $\eta_{i,t-1}$, whereas the transitory component is homoskedastic white noise. We will refer to this specification as the "standard model".

Equations (2.4) - (2.5) impose three types of restrictions.

1. Linearity of the process for the persistent component. In this context linearity means that the conditional expectation of $\eta_{i,t-1}$ is linear in its lagged value $\eta_{i,t-1}$. Moreover, it is straightforward to show that the expectation of $m_{i,t}$ given $m_{i,t-1}$ is linear, $m_{i,t} = \delta m_{i,t-1} + e_{i,t}$, with $E(e_{i,t} \,|\, m_{i,t-1}) = 0$.

2. Stationarity of the distributions of $\zeta_{i,t}$, $\varepsilon_{i,t}$ and $e_{i,t}$. The distribution of $\zeta_{i,t}$, $\varepsilon_{i,t}$, is unconditionally stationary, and the distribution of $e_{i,t}$ is stationary conditional on t.

3. Normality, not only of the shocks $\zeta_{i,t}$ and $\varepsilon_{i,t}$, but also of the forecast error $e_{i,t}$.

See Appendix 2.A.5 for details.

We assess the implications of the standard model in Figure 2.2, following the graphical analysis of earnings found in De Nardi et al. (2020). Figure 2.2 shows moments of the medical spending residuals conditional on their lagged percentile ranks. To construct the residual $m_{i,t}$, we first regress the log of medical spending on age, household composition, and health variables as in equation (2.2). We next regress each household's time-t residual on its lagged value, estimating $m_{i,t} = \delta m_{i,t-1} + e_{i,t}$ and calculating the "second-order" residual $e_{i,t}$.

Figure 2.2 plots the mean, standard deviation, skewness and kurtosis of $e_{i,t}$ as a function of the percentile rank of $m_{i,t-1}$. The top left panel plots means. This graph shows that at the 95th percentile of lagged medical spending, $m_{i,t-1}$, the current value of $e_{i,t}$ is on average -0.1. Put differently, if last period's medical spending is high, this period's value is also high, but it is 10% lower than what would be implied by a linear regression on lagged medical spending. The data thus reject the first restriction, linearity, of the standard model.

The top right panel of Figure 2.2 plots the standard deviation of medical spending. This is U-shaped, declining until the 40th percentile of last period's spending and increasing after the 60th percentile. In other words, dispersion in medical spending is highest for those with low or high medical spending. The data thus reject the prediction that the standard deviation will be constant in $m_{i,t-1}$.

The bottom left panel shows skewness. Normality implies that skewness is zero across the distribution, but in the data skewness is positive for low values of spending and is below -0.5 for high values. This means for people with high medical spending last period, this period's medical spending tends to fall a lot when it falls. The bottom right panel shows kurtosis: normality would imply that this is three, but in the data it is often higher, implying a distribution with tails fatter than those found in the normal distribution. In short, Figure 2.2 shows that data rejects all three predictions of the standard model.

The Arellano et al. (2017) estimator

Our finding that the data are strongly at odds with the standard model leads us to seek a more general framework that relaxes the previous three restrictions yet fits within our two-component structure. Thus we use the quantile-based panel data model proposed by Arellano et al. (2017) and extended by Arellano et al. (2021).

To apply Arellano et al.'s (2017) framework, rewrite the conditional distribution for the persistent component $\eta_{i,t}$ as:

$$\eta_{i,t} = Q_\eta(\nu_{i,t} \,|\, \eta_{i,t-1}, a_{i,t}), \quad \nu_{i,t} \overset{iid}{\sim} U[0,1], \tag{2.6}$$

where $Q_\eta(\nu \,|\, \eta_{i,t-1}, a_{i,t})$ denotes the νth quantile of $\eta_{i,t}$ conditional on its lagged value and age ($a_{i,t}$). The quantile function Q_η maps $\eta_{i,t}$'s conditional rank, $\nu_{i,t}$, into a value of $\eta_{i,t}$ itself. To fix ideas, if we draw $\nu_{i,t} = 0.1$, the realized value of $\eta_{i,t}$ will equal the 10th percentile of the conditional distribution of $\eta_{i,t}$ at age $a_{i,t}$, given $\eta_{i,t-1}$. As a rank, $\nu_{i,t}$ is distributed uniformly over the $[0, 1]$ interval.

In the standard model, it follows from equation (2.4) that the quantile function takes the linearly separable form $\eta_{i,t} = Q_\eta(\nu_{i,t} \,|\, \eta_{i,t-1}, a_{i,t}) = \phi\eta_{i,t-1} + \sigma_\zeta\Phi^{-1}(\nu_{i,t})$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative

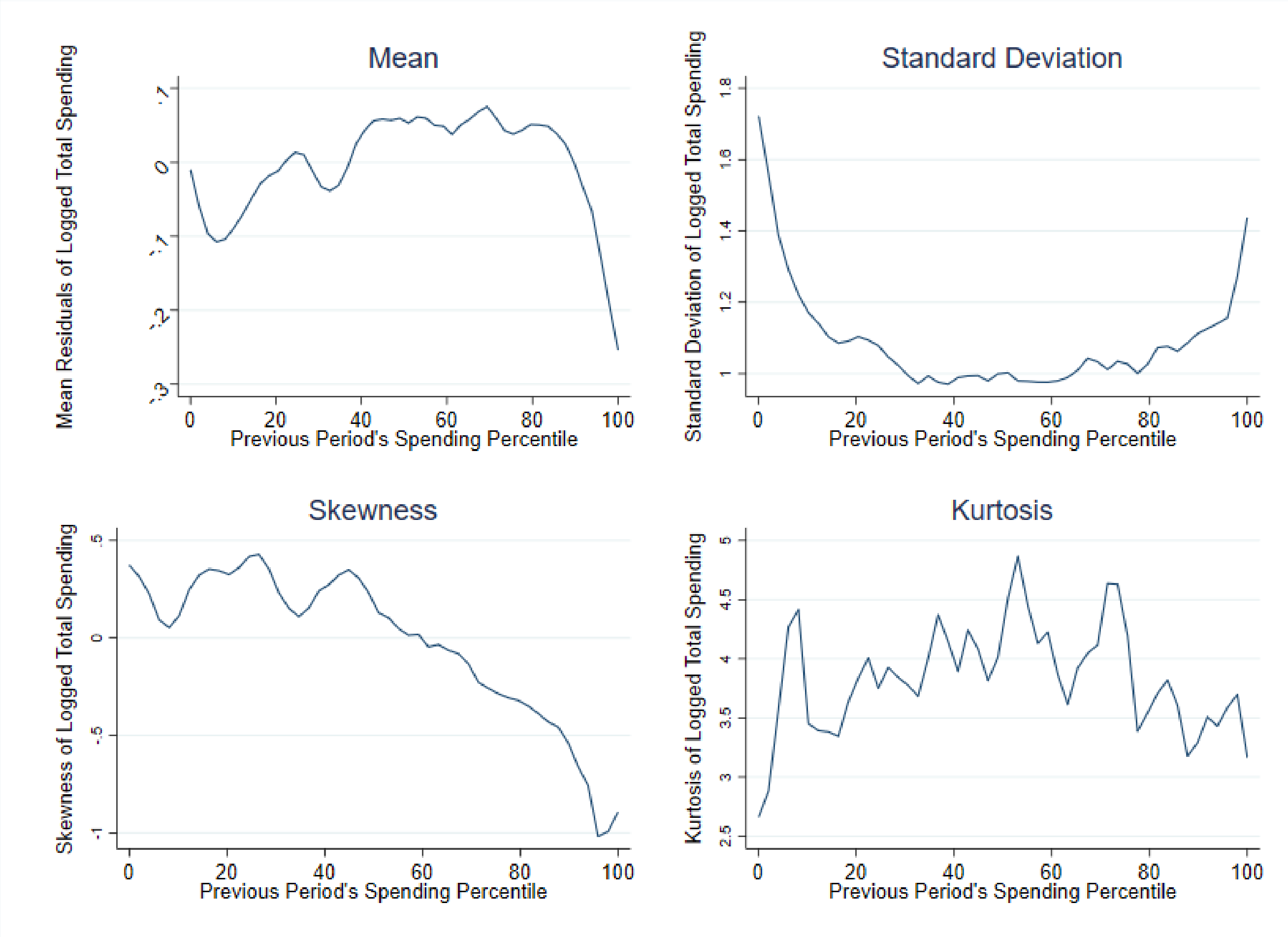


Figure 2.2: Moments of the Medical Spending Distribution

distribution function and σ_ζ is the standard deviation of $\zeta_{i,t}$. (Conversely, with $\nu_{i,t} \sim U[0,1]$, we have $\sigma_\zeta \Phi^{-1}(\nu_{i,t}) \sim N(0, \sigma_\zeta^2)$). Age-independence, normality, and linearity can thus be expressed as restrictions on the quantile function in equation (2.6).

In its most unrestricted form, this specification allows for a great degree of flexibility. One way to see this is to construct the persistence measure

$$\phi_\tau(\eta_{i,t-1}, a_{i,t}) = \frac{\partial Q_\eta(\tau \,|\, \eta_{i,t-1}, a_{i,t})}{\partial \eta_{i,t-1}}, \tag{2.7}$$

with τ denoting the conditional rank of interest. $\phi_\tau(\eta_{i,t-1}, a_{i,t})$ measures the effect of $\eta_{i,t-1}$ on the τth conditional quantile of $\eta_{i,t}$. Persistence can vary by rank (τ), age ($a_{i,t}$) and prior realization ($\eta_{i,t-1}$). In contrast, in the standard model persistence always equals the constant ϕ.

In estimation we parametrically approximate the conditional quantile function by low-order Hermite polynomials. Let $h_k^\eta(\cdot)$ denote the kth Hermite polynomial used in the approximation of $\eta_{i,t}$, with $\{h_k^\eta(\cdot)\}_{k=0}^{K_\eta}$ forming the polynomial basis for the approximation. $Q_\eta(\tau \,|\, \eta_{i,t-1}, a_{i,t})$ is thus a linear combination of the K_η Hermite polynomials, with the coefficients on the polynomials, $\left\{\beta_k^{\eta}(\tau)\right\}_{k=0}^{K_\eta}$ themselves functions of the quantile rank τ. We thus have

$$Q_\eta(\tau \mid \eta_{i,t-1}, a_{i,t}) = \sum_{k=0}^{K_\eta} \beta_k^{\eta}(\tau) \cdot h_k^{\eta}(\eta_{i,t-1}, a_{i,t}), \quad \tau \in (0,1]). \tag{2.8}$$

The distributions of the initial shock $\eta_{i,1}$ and the transitory shocks $\{\varepsilon_{i,t}\}_t$ are handled in ways analogous to how we handle the persistent component:

$$Q_1(\tau \mid a_{i,1}) = \sum_{k=0}^{K_1} \beta_k^{1}(\tau) \cdot h_k^{1}(a_{i,1}), \quad \tau \in (0,1), \tag{2.9}$$

$$Q_\varepsilon(\tau \mid , a_{i,t}) = \sum_{k=0}^{K_\varepsilon} \beta_k^{\varepsilon}(\tau) \cdot h_k^{\varepsilon}(a_{i,t}), \quad \tau \in (0,1). \tag{2.10}$$

For these distributions we do not condition on $\eta_{i,t-1}$ but only only on age.

Each of the coefficient functions $\left(\{\beta_k^{\eta}(\tau)\}_{k=0}^{K_\eta}, \{\beta_k^{1}(\tau)\}_{k=0}^{K_1}, \{\beta_k^{\varepsilon}(\tau)\}_{k=0}^{K_\varepsilon}\right)$ in equations (2.8)-(2.10) is modelled with a set of polynomial splines defined over the intervals $\left\{[\tau_{\ell-1}, \tau_\ell]\right\}_{\ell=1}^{L}$, along with two low-dimensional tail functions defined over $(0, \tau_1]$ and $[\tau_L, 1)$. It is the parameters for these weighting functions that we must estimate.

As both the persistent and transitory shocks are unobserved, we cannot estimate the parameters of the weighting functions directly using quantile regressions. Furthermore, our data are unbalanced because sample members die. We therefore follow the extension of the *E-M* algorithm described in and applied by Arellano et al. (2021).

- In the *E-step* we find the posterior distribution of the unobserved persistent shocks $(\{\eta_{i,t}\}_t)$ implied by the data and the current parameterization of the model. In particular, we use the coefficients of the Hermite polynomials $\left(\{\beta_k^{\eta}(\tau)\}_{k=0}^{K_\eta}, \{\beta_k^{1}(\tau)\}_{k=0}^{K_1}, \{\beta_k^{\varepsilon}(\tau)\}_{k=0}^{K_\varepsilon}\right)$, which fully determine the distributions of the shocks, and a Monte Carlo method to simulate draws from the distributions of the initial shock $\eta_{i,1}$ and the subsequent shocks $\{\eta_{i,t}\}_t$.[4] This part of the procedure is a special case of the Sequential Monte Carlo methods described in greater detail in Creal (2012).
- In the *M-step* we use quantile regressions to update the coefficient functions for the Hermite polynomials, using the distribution of $\{\eta_{i,t}\}_t$ found in the *E-step*. Once the coefficients have been updated, we return to the *E-step* and simulate new draws.

We iterate between the E and M steps until the parameters converge. See Appendix 2.A.6 for a more detailed description of the methodology.

[4] This approach takes advantage of the Markovian structure of the model and has been shown to perform well in low-dimensional models.

2.4.2 Health and Mortality

Let $hs_{i,g,t}$ denote the health of member $g \in \{h, w\}$ in household i at age t. Health has four mutually exclusive possible values: dead; in a nursing home; in bad health; or in good health. We assume that the transition probabilities for an individual's health depend on his or her current health, age, permanent income I, and gender g.[5] It follows that the elements of the health transition matrix are given by

$$\pi_{q,r}(t, I_i, g) = \Pr\left(hs_{i,g,t+1} = r \,\middle|\, hs_{i,g,t} = q; t, I_i, g\right), \tag{2.11}$$

with the transitions covering a one-year interval. Although the HRS interviews every other year, we adopt the approach in De Nardi et al. (2016), who fit annual models of health to the HRS data for singles. We extend their approach to account for the dynamics of two-person households. We estimate health/mortality transition probabilities by fitting the transitions observed in the HRS to a multinomial logit model.[6] See Appendix 2.A.7 for further details.

2.5 Results

2.5.1 Health and Longevity

Table 2.2 shows the life expectancies implied by our demographic model for those still alive at age 65. The first panel of the table shows the life expectancies for singles under different configurations of gender, PI percentile, and age-65 health. The healthy live longer than the sick, the rich (higher PI) live longer than the poor, and women live longer than men. For example, a single man at the 10th PI percentile in a nursing home expects to live only 4.14 more years, while a single woman at the 90th percentile in good health expects to live 22.5 more years. The second panel of the table shows life expectancies for married *households*, that is, the average length of time that at least one member of the household is still alive or, equivalently, the life expectancies for the oldest survivors. While wives generally outlive husbands, a non-trivial fraction of the oldest survivors are men, and the life expectancy for a married household is roughly three years longer than that of a married woman with same initial health and PI quantile.

The results shown in Table 2.2 are disaggregated by PI and initial health. When we average over all these factors, we find that a man alive at age 65 will on average live an additional 17.50 years, while a woman alive at age 65 will on average live an additional 20.96 years.[7]

[5] We do not allow health transitions to depend on medical spending. The empirical evidence on whether medical spending improves health, especially at older ages, is surprisingly mixed (De Nardi et al. 2016). Likely culprits include reverse causality – sick people have higher expenditures – and a lack of insurance variation – almost every retiree gets baseline insurance through Medicare. We also do not allow health transitions to depend on marital status. De Nardi et al. (2021) find that after controlling for income and past health, marital status has little added predictive power.

[6] We do not control for cohort effects. Instead, our estimates are a combination of period (cross-sectional) and cohort probabilities. This may lead us to underestimate the lifespans expected by younger cohorts as they age. Nevertheless, lifespans have increased only modestly over the sample period. Accounting for cohort effects would have at most a modest effect on our estimates.

[7] We construct these distributions with bootstrap draws of people aged 63-67 in HRS.

Table 2.2: Life expectancy in years, conditional on reaching age 65

	Men			Women			
Permanent Income Percentile	Nursing Home	Bad Health	Good Health	Nursing Home	Bad Health	Good Health	All*
Individuals							
10	4.14	15.25	17.49	7.32	18.70	20.22	18.06
50	4.89	16.99	19.37	8.71	20.52	22.04	20.68
90	4.91	17.35	19.86	8.83	20.97	22.52	20.96
Couples (oldest survivors)†							
10	9.35	21.95	23.15				22.50
50	11.27	24.04	25.17				24.98
90	11.46	24.56	25.69				25.52
All Men							17.50
All Women							20.96
All Couples (oldest survivor)							24.83

* Averages taken over initial health found in the data. Results indexed by PI percentile are taken over the associated PI quintile.

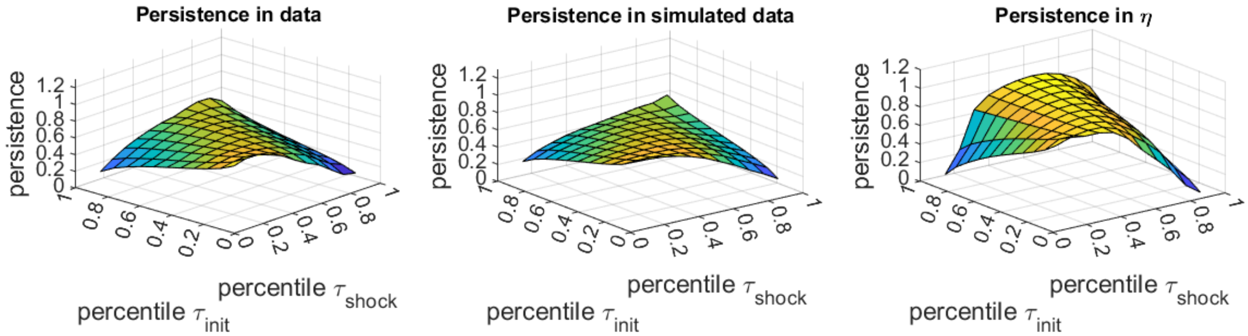
† Health-specific results for couples based on the assumption that the spouses have the same health at age 65.

Another key statistic for our analysis is the probability that a 65 year old will spend significant time (a stay of more than 120 days) in a nursing home before he or she dies. We estimate this probability to be 24.4% for men and 36.9% for women. Nursing home incidence differs relatively modestly across the PI distribution. Although high-income people are less likely to be in a nursing home at any given age, they live longer, and older individuals are much more likely to be in a nursing home. The nursing home risk also varies relatively little with initial health status (good or bad), for similar reasons.

Although all households in the HRS are initially non-institutionalized, the HRS does a good job of tracking individuals as they enter in nursing homes. French and Jones (2004) show that the HRS sample matches very well the aggregate statistics on the share of the elderly population in a nursing home by 1999 when the sample begins. In 1999 roughly 2.8% of men and 6.1% of women in the data had entered nursing homes. We also understate the number of nursing home visits because we exclude short-term visits: as Friedberg et al. (2014) and Hurd et al. (2017) document, many nursing home stays last only a few weeks and are associated with lower expenses. We focus only on the longer and more expensive stays faced by households.

2.5.2 The Cross Sectional Distribution of Medical Spending and Parameter Estimates

The Persistence of Medical Spending: Panel (a) of Figure 2.3 shows the estimated persistence of medical spending in the data. Formally, it is the derivative of the conditional quantile function, defined in equation (2.7), averaged across age ($a_{i,t}$). It shows how persistent next period's medical spending is as a function of next period's shock (indexed by its percentile rank τ_{shock}) and the percentile rank of today's medical spending (τ_{init}).



(a) HRS data (b) Model simulation (c) Persistent component

Figure 2.3: Persistence of medical spending

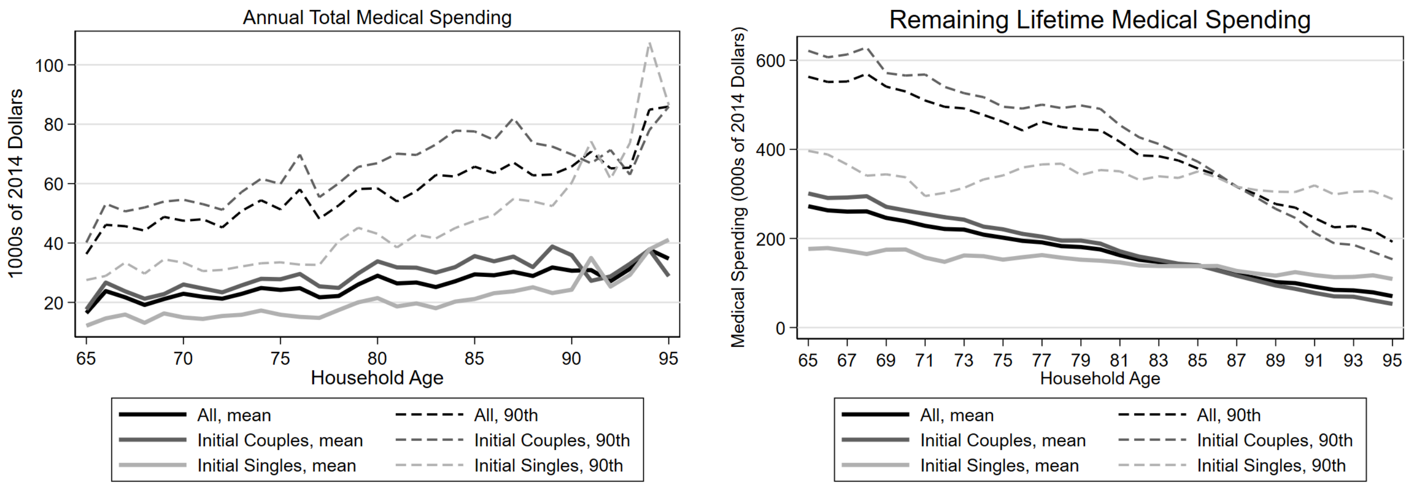
Panel (a) shows that persistence is higher near the middle of the distribution. This means that those in the middle may face moderate but persistent expenses that can accumulate into a catastrophic lifetime cost. In contrast, persistence is quite low (less than 0.1) for large medical shocks among those with low initial medical spending. It is much higher, almost 0.6, for those with high initial spending and large shocks (the top point on the graph). What this means is that catastrophic medical spending is particularly persistent. Persistence is also higher for those with low initial spending and small shocks (the bottom point on the graph). In short, when medical spending is low, it is likely to stay low, and when medical spending is high, it is likely to stay high. However, it follows from Figures 2.2 and 2.3 that when medical spending is high, it can fall a lot (as the conditional distribution is left-skewed when current spending is high) and when medical spending is low, it can rise a lot (as the conditional distribution is right-skewed when current spending is low). Recall that the AR(1)-plus-white noise specification often assumed in the literature implies that persistence would be constant. Panel (b) shows persistence in the simulated data. The similarity of Panels (a) and (b) shows that the model matches the persistence of medical spending observed in the data.

Panel (c) shows persistence in the permanent component of medical spending η. The persistence of η is much higher than for medical spending overall, as the latter includes the transitory shock ε. Over much of its distribution, the persistence of η is close to 1, indicating that for severe health conditions, such as dementia, medical spending is very persistent. It is these shocks, which cause high spending for many periods, that can drain a family's resources.

Budget sets: Table 2.1 shows that the mapping from total expenditures to the out-of-pocket expenditures paid by households is approximately linear. We therefore assume that everyone pays 21.6% of total costs out of pocket.

2.5.3 The Distribution of Lifetime Medical Spending

Panel (a) of Figure 2.4 shows the model-predicted mean and 90th percentile of annual health care spending for surviving households (including those who died during the year). Medical spending rises rapidly with age. For example, mean spending rises from \$22,900 per year at age 70 to \$34,700 at age 95. The upper tail rises even more rapidly, with the 90th percentile increasing from \$47,200 to \$85,200. These patterns are broadly similar to the raw



(a) (b)

Figure 2.4: Mean and 90^{th} Percentile of Annual and Remaining Lifetime Medical Spending.

Notes: Mean and 90^{th} Percentile of Annual (Panel a) and Remaining Lifetime (Panel b) Total Medical Spending for Surviving Households, Initial Singles and Initial Couples.

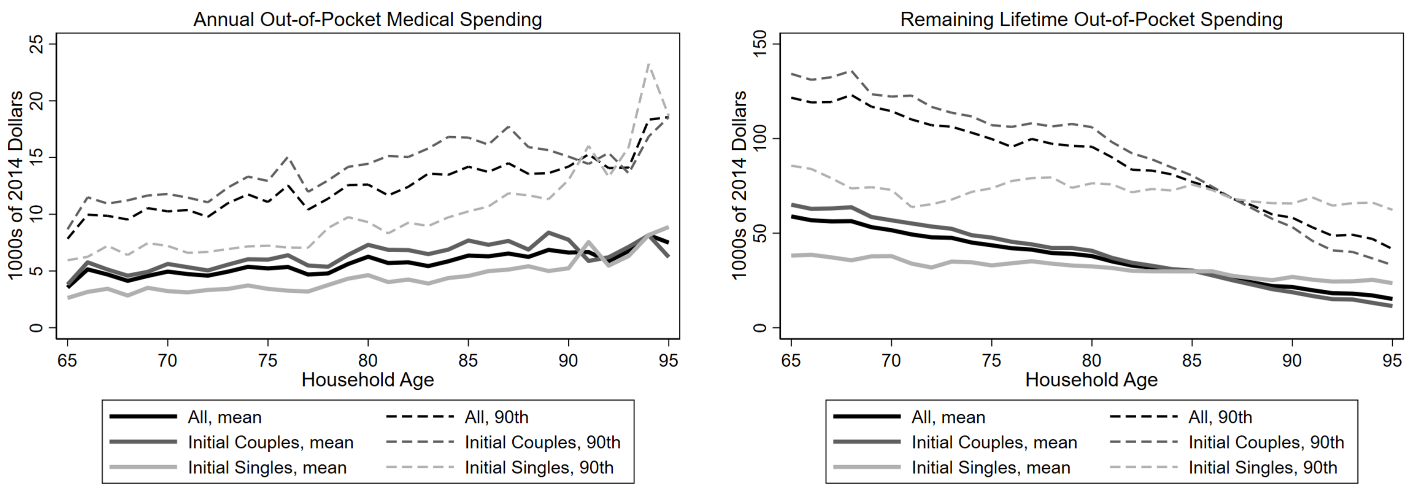
data in Figure 2.1. One difference is that the medical spending gap between singles and couples is smaller at older ages than in Figure 2.1. This is because here we are conditioning on initial household structure whereas Figure 2.1 conditions on current household structure. Virtually all initial couples have lost a spouse and are single if either household member is still alive at age 90. Thus it is unsurprising that medical spending of initial singles and initial couples is similar at this age. Here we focus on initial household structure so we can calculate the present value of lifetime spending.

Panel (b) of Figure 2.4 plots our main variable of interest, lifetime total spending. At each age, we calculate the present discounted value of remaining medical spending from that age forward, using an annual real discount rate of 3 percent. These lifetime totals are considerable. At age 65, households will incur, on average, over $272,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of $563,000. One might expect the lifetime totals to fall rapidly as households age and near the ends of their lives. This is not the case. A household alive at age 90 will on average spend more than $99,000 before they die. The slow decline of lifetime costs is due mostly to the tendency of medical costs to rise with age. Households that live to older ages have shorter remaining lives but higher annual spending rates.

A number of papers have considered whether most of the rise in medical spending with age is due rising mortality and thus end-of-life expenses: see the discussion in De Nardi et al. (2016). In our spending model we allow for for the jump in medical spending prior to death as well as age related medical spending.

2.5.4 Out-of-Pocket Medical Spending

Our baseline measure is total medical spending, which is the sum of payments made out-of-pocket, by Medicare, by Medicaid, and by other private and public payors. Figure 2.5 shows the distributions of out-of-pocket spending predicted by the model. To construct these values, we take the total medical spending amounts, then calculate the out-of-pocket amount using the budget sets discussed in Section 2.5.2. Parallel to Figure 2.4, Panel (a) of Figure 2.5 shows the distribution of annual out-of-pocket spending and Panel (b) shows remaining lifetime out-of-pocket medical spending.



(a) (b)

Figure 2.5: Mean and 90^{th} Percentile of Annual and Remaining Lifetime Out-of-Pocket Medical Spending.
Notes: Mean and 90^{th} Percentile of Annual (Panel a) and Remaining Lifetime Out-of-Pocket Medical Spending for Surviving Households (Panel b), Initial Singles and Initial Couples.

Predicted out-of-pocket spending in Figure 5 mirrors the patterns seen in Figure 4, although the levels are lower. At age 65, households will on average incur over $59,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of $121,000. A household alive at age 90 will on average spend more than $21,000 out-of-pocket before they die, again highlighting the slow decline of expected lifetime costs as households age.

A number of recent papers have argued that Medicaid significantly reduces the out-of-pocket spending risk faced by older households. Brown and Finkelstein (2008) conclude that Medicaid crowds out private long-term care insurance for about two-thirds of the wealth distribution. De Nardi et al. (2016) find that most single retirees, including those at the top of the income distribution, value Medicaid at more than its actuarial cost. While both of these papers model Medicaid formally, they lack data to estimate the underlying distributions and thus rely heavily on functional form assumptions. Furthermore, previous papers have tended to abstract away from Medicare payments, in large part because Medicare spending data have not been available. We find that despite these payors, medical spending risk is high in old age.

2.6 Discussion and Conclusions

In this paper we use health and spending models to simulate the distribution of lifetime medical expenditures as of age 65, adding to the handful of studies on this topic. We also assess the importance of Medicare and Medicaid in reducing lifetime medical spending risk. The simulations show that lifetime medical spending is high and uncertain, and that the level and the dispersion of this spending diminish only slowly with age. Most of these expenditures will be covered by Medicare, Medicaid, or other private and public insurers. Our data suggest that out-of-pocket spending rises more or less linearly with total spending, with households covering about 22% of their total expenditure.

We find that at age 65, households will on average incur over $59,000 of medical spending over the remainder of their lives. Those at the 90th percentile of the spending distribution will incur spending of $121,000. A household alive at age 90 will on average spend more than $21,000 out-of-pocket before they die, highlighting the slow decline of expected lifetime costs as households age.

We conclude by pointing out some caveats to our analysis. We assume, as do many other empirical papers, that medical spending is exogenous, while in reality it is a choice variable. Although the demand for some medical goods and services is extremely inelastic, the demand for others might be elastic. Nursing home care, for example, is a bundle of medical and non-medical commodities, and the latter can vary greatly in quality, with the choice between a single and a shared room being just one example. Thus households can reduce their medical spending risk by purchasing fewer medical goods. While our assumption of exogenous spending arguably leads us to overstate out-of-pocket spending risk, our assumption that the effective co-pay rate is always 22% arguably leads us to understate the risk. For example, most households lack private nursing home insurance, Medicare coverage is extremely limited, and Medicaid coverage is subject to means-testing. Thus many households (such as those in nursing homes whose wealth is too high to qualify for Medicaid) pay more than 22% and others (who are in hospital for modest periods of time) will spend less than 22%. This leaves many households potentially exposed to significant long-term care expenditure risk. Our approach can be extended to consider these types of risks.

2.A Appendix

2.A.1 Our Medicare and Medicaid Data

Medicare

We link restricted Medicare fee-for-service (Parts A and B), and Part D data for the years 1999-2012 (2006 was the first year of Medicare Part D and thus our Part D data begins then) to our HRS survey data for respondents who consent to allow their Medicare data to be linked to their survey responses (approximately 64.7% percent of persons in our study population). These records have enrollment information and data on reimbursement amounts for inpatient, skilled nursing facility, home health, and hospice claims (Medicare Part A), as well as outpatient, carrier (non-institutional medical care providers such as individual or group practitioners, non-hospital labs, and ambulances), and durable medical equipment claims for Medicare Part B.

We use the Beneficiary Annual Summary File (BASF), which summarizes information from the micro-level claims records. The BASF contains annual information for each individual on the number of months of enrollment in Medicare Part A, Part B, and non-fee-for-service plans. The BASF has information on Medicare fee-for-service (FFS) claims. Almost all claims for services used by non-FFS Medicare patients are not observed in these data, so all analyses exclude an individual in a given year if they were enrolled in a non-FFS Medicare plan for more than half the year.

Medicare Part D is the prescription drug benefit. We calculate the Medicare Part D payment using the Part D event files. For the Part D Medicare contribution we subtract from the gross drug cost the payments paid by the beneficiary, family, or friends.

Medicaid

As with the Medicare data, we are able to link restricted Medicaid data (CMS Medicaid Analytic eXtract, or "MAX" files) for those in the HRS who gave permission, allowing us to measure Medicaid expenditures for the Medicaid beneficiaries in our data set for the years 1999-2012. The MAX files contain personal summaries (which contain eligibility, enrollment, and demographic information) and claims data across four service categories (inpatient, long-term care, prescription drugs, and other services). Other services include a variety of services (e.g., physician services and lab work) that do not fit under the other three service categories. The inpatient, long-term care, prescription drugs, and other services files contain the primary variable of interest, "Medicaid Payment Amount," which is the total amount of money paid by Medicaid for a particular service. We sum over all the claims for all the different service categories for a particular individual in each year.

2.A.2 Sample Selection

We drop households for various reasons. These include disagreement between household members, problematic mortality transitions, households with multiple members of the same sex, and refusal to provide Medicare and Medicaid spending records. Table 3 below denotes the starting public HRS sample size and the sample size after every drop.

Table 2.3: Sample Selection

Drop reason	Post-drop sample size
	26,598
One spouse dead and the other claims never married or divorced	26,596
One spouse claims married and the other something else	25,343
Two or more members of the same sex	24,897
One spouse claims never married an the other is not missing	24,818
One spouse claims divorced and the other widowed	24,545
People who are partnered	23,466
People who "died", then came back to life	23,161
Transitions with widowed or both widowed	23,055
Transitions to got married or divorced	22,719
Household split etc. based on sub-household identifier	22,695
Households that appear once and have missing marital status	22,683
No domestic partner in sample but married in some wave	21,903
No living member the year household joined HRS	21,713
Households dead or under 65 in the estimation period (1999-2012)	9,969
2+ years post death Medicare and Medicaid spending	6,783
Incomplete Medicare and Medicaid records	4,391

Notes: Sample size refers to the number of unique households.

2.A.3 Imputing Missing Medical Expenditures

Our goal is to measure all medical spending: the variable M_{it} in equation (2.2) of the main text is defined to include out-of-pocket spending, Medicare and Medicaid payments, and private and other public (such as Veterans Administration benefits, and care provided by local and state health departments) insurance payments. While the HRS includes information on out-of-pocket spending and can be linked to Medicare and Medicaid payments, it does not include Medicare Part C, private, or other public insurance payments. In this appendix we describe how we use data from the Medical Expenditure Panel Survey (MEPS) to impute these payments in the HRS. Although the MEPS has extremely high quality information on all payors for all household members, it lacks the long panel dimension of the HRS. Our imputation procedures allow us to exploit the best of both data sets.

Our imputation procedure has two steps. First, we use the MEPS to infer private and other public insurance payments, conditional on variables observed in both datasets. Second, we impute private and other public insurance payments in the HRS data using a conditional mean matching procedure (which is a procedure very similar to hot-decking).

First Step of Imputation Procedure

We use the MEPS to infer payments of other payors, conditional on the observable variables that exist in both the MEPS and the HRS datasets.

Let i index individuals in the HRS and j index individuals in the MEPS. Define M_{it}^{obs} as out of pocket, Medicaid, and Medicare (Part A, B, and D, but not Part C) payments observed in both the HRS and MEPS data sets, M_{it}^{miss} as the components of medical spending missing in the HRS but observed in the MEPS, and $M_{it} = M_{it}^{miss} + M_{it}^{obs}$ as total medical spending. To impute M_{it}^{miss}, which is missing in the HRS, we follow David et al. (1986), French and Jones (2011), and De Nardi et al. (2020) and use a predictive mean-matching regression approach. There are two steps to our procedure. First, we use the MEPS data to regress M_{it}^{miss} on observable variables that exist in both data sets. Second, we impute M_{it}^{miss} in the HRS data using a conditional mean-matching procedure, a procedure very similar to hot-decking.

First, for every member of the the MEPS sample, we regress the variable of interest M^{miss} on the vector of observable variables z_{jt}, yielding $M_{jt}^{miss} = z_{jt}\beta + \varepsilon_{jt}$. Second, for each individual j in the MEPS we calculate the predicted value $\widehat{M^{miss}}_{jt} = z_{jt}\hat{\beta}$, and for each member of the sample we calculate the residual $\widehat{\varepsilon}_{jt} = M_{jt}^{miss} - \widehat{M^{miss}}_{jt}$. Third, we sort the predicted value $\widehat{M^{miss}}_{jt}$ into deciles and keep track of all values of $\hat{\varepsilon}_{jt}$ within each decile. We use this procedure separately to impute Medicare Part C benefits, private payments, and other payments.

In practice we include in z_{jt} a fourth-order age polynomial, marital status, gender, self-reported health (=1 if self reported health is good, very good, or excellent), race, visiting a medical practitioner (doctor, hospital or dentist), out-of-pocket medical spending, education of head (high school, some college, college), death of an individual, and total household income. We estimate this regression two times: once for the privately insured, and once for other payors.

Because the measure of medical spending in the HRS is medical spending over two years, we divide HRS out-of-pocket medical spending by 2 and assume that medical spending is equal across the two years.

Second Step of Imputation Procedure

For every observation in the HRS sample with a positive Medicaid indicator, we impute $\widehat{Med}_{it} = z_{it}\widehat{\beta}$, using the values of $\widehat{\beta}$ estimated from the MEPS. Then we impute ε_{it} for each observation of this subsample by finding a random observation in the MEPS with a value of $\widehat{Med}_{jt}$ in the same decile as $\widehat{Med}_{it}$, and setting $\hat{\varepsilon}_{it} = \hat{\varepsilon}_{jt}$. The imputed value of Med_{it} is $\widehat{Med}_{it} + \hat{\varepsilon}_{it}$.

As David et al. (1986) point out, our imputation approach is equivalent to hot-decking when the "z" variables are discretized and include a full set of interactions. The advantages of our approach over hot-decking are two-fold. First, many of the "z" variables are continuous. Second, to improve goodness of fit we use a large number of "z" variables. We find that adding extra variables are very important for improving goodness of fit when imputing payments. Because hot-decking uses a full set of interactions, this would result in a large number of hot-decking cells relative to our sample size. Thus, in this context, hot decking is too data intensive.

2.A.4 Validating the Administrative Medical Spending Data

Here, we examine in greater detail the accuracy of the administrative medical spending data, as well as the out-of-pocket spending found in the Assets and Health Dynamics of the Oldest Old (AHEAD) cohort of the HRS, comparing them to data from the MCBS. See De Nardi, French, and Jones (2016) and De Nardi et al. (2016) for more details of the MCBS data and (for example) Nicholas et al. (2011) for details of the HRS linked data.

The MCBS is a nationally representative survey of Medicare beneficiaries, consisting of Disability Insurance recipients and Medicare recipients aged 65 and older. The survey contains an over-sample of beneficiaries older than 80 and disabled individuals younger than 65. Respondents are asked about health status, health insurance, and health care spending (from all sources). The MCBS data are matched to Medicare records, and medical spending data are created through a reconciliation process that combines information from survey respondents with Medicare administrative files. As a result, the survey is thought to give extremely accurate data on Medicare payments and fairly accurate data on out-of-pocket and Medicaid payments. As in the HRS survey, the MCBS survey includes information on those who enter a nursing home or die. Respondents are interviewed up to 12 times over a 4 year period. We aggregate the data to an annual level. In both samples, we applied only modest sample selection restrictions. The key sample selection issue shown in Table 2.3 is that in the HRS we drop households with missing or erroneous Medicare or Medicaid records.

Here we compare distributions of total, out-of-pocket, Medicare, and Medicaid payments between the MCBS are the HRS data. Medical spending the HRS is measured at an individual level (rather than household) to be comparable with the MCBS. The comparison can be seen in Table 2.4. Medical spending is higher in our HRS sample than in the MCBS sample. Furthermore, this higher level of spending is driven by higher out-of-pocket spending, Medicare, and Medicaid spending. These differences potentially are an advantage of the HRS data since, as noted in De Nardi et al. (2016), the MCBS clearly understates aggregate Medicare and especially Medicaid spending, potentially due to the issue that the MCBS does not have administrative data on Medicaid spending, and thus relies heavily on imputation.

The next set of benchmarking exercises that we perform is for out-of-pocket medical spending, Medicaid recipiency and income between the AHEAD cohort of the HRS and MCBS. For both the HRS and MCBS, we restrict the sample to singles (over the sample period) who meet the HRS/AHEAD age criteria (at least 70 in 1994, 72 in 1996, ...) and who are not working over the sample period. Because the MCBS sample lacks spousal information, for this analysis we focus only on singles. We use De Nardi et al.'s (2016) measure of permanent income and construct a measure of permanent income, which is the percentile rank of total income over the period we observe these individuals (the MCBS asks only about total income). The first four columns of Table 2.5 show sample statistics from the full HRS/AHEAD sample while the final three columns of the table show sample statistics from the MCBS sample. The first statistics we compare are income. Total income in the HRS/AHEAD data (including asset and other non-annuitized income) lines up well with total income in the MCBS data, although income in the top quintile of the MCBS is higher than in the HRS/AHEAD. Next, we compare out-of-pocket medical spending in the MCBS and HRS/AHEAD. Out-of-pocket medical spending (including insurance payments) averages $2,360 in the bottom

Table 2.4: Individual Medical Spending Percentiles: HRS versus MCBS

	Total Spending				OOP			
Total Spending	HRS		MCBS		HRS		MCBS	
Percentiles	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total
All	17,091	100	14,120	100	3,825	100	2,740	100
95-100%	114,238	33.4	97,880	34.6	45,643	59.6	26,930	49.1
90-95%	59,000	17.3	48,890	17.3	8,619	11.3	6,700	12.2
70-90%	26,870	31.4	20,540	29.1	3,480	18.2	2,920	21.3
50-70%	9,025	10.6	7,750	11	1,394	7.3	1,360	9.9
0-50%	2,502	7.3	2,250	8	178	3.6	420	7.6

	Medicare				Medicaid			
Total Spending	HRS		MCBS		HRS		MCBS	
Percentiles	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total	Average Exp.	Pct. Total
All	11,343	100	7,720	100	1,896	100	1,320	100
95-100%	85,268	37.6	67,560	43.7	33,773	89.1	24,980	94.7
90-95%	41,731	18.4	28,370	18.4	4,092	10.8	1,360	5.2
70-90%	17,251	30.4	10,280	26.6	230	0.2	10	0.1
50-70%	5,031	8.9	2,980	7.7	0	0	0	0
0-50%	1,076	4.7	550	3.5	0	0	0	0

Table 2.5: Income, Out-of-pocket Spending, and Medicaid Recipiency Rates: HRS versus MCBS

	HRS/AHEAD Data				MCBS Data		
Income Quintile	Total Income	Annuity Income	Out-of-pocket Expenses	Medicaid Recipiency	Total Income	Out-of-pocket Expenses	Medicaid Recipiency
1	7,740	4,820	2,550	60.9	6,750	4,050	69.9
2	10,290	8,270	4,270	28.1	10,020	5,340	41.8
3	15,500	10,900	5,050	11.0	13,740	6,470	15.5
4	19,290	14,390	6,360	5.6	19,710	7,300	8.0
5	33,580	26,300	7,000	3.0	44,150	8,020	5.4

Notes:1996-2010, for those age 72 and older in 1996.

PI quintile and \$6,340 in the top quintile in the HRS/AHEAD. In comparison, the same numbers in the MCBS data are \$3,540 and \$7,020. Overall, out-of-pocket medical spending in the MCBS and HRS/AHEAD are similar, which may be surprising given that the two surveys each have their own advantages in terms of survey methodology.[8] The share of the population receiving Medicaid transfers is also very similar in the HRS/AHEAD and MCBS. 61% and 70% of those in the bottom PI quintile are on Medicaid in the the HRS/AHEAD and MCBS, respectively. In the top quintile, 3% of people are on Medicaid in the HRS/AHEAD whereas 5% are in the MCBS.

2.A.5 Plotting the Moments of Medical Spending

This appendix shows how we plot the moments of medical spending in Figure 2.2. We follow closely De Nardi et al.'s (2020) approach.

Assume that the log of medical spending for household (or person) i at period t is given by equation (2.2), which we replicate here:

$$\ln M_{i,t} = X_{i,t}\gamma + m_{i,t}, \tag{2.12}$$

where $X_{i,t}$ is a vector of conditioning variables such as age, household composition, and health. As long as we impose no restrictions on $m_{i,t}$, equation (2.12) is an identity that is true by construction. It is often commonly assumed that $m_{i,t}$ is independent of $X_{i,t}$, but this is not necessary for our purposes. Likewise, we can construct the autoregression

$$m_{i,t} = \delta m_{i,t-1} + e_{i,t}. \tag{2.13}$$

Our procedure is as follows:

1. Using OLS, estimate the coefficient γ in equation (2.12) and construct $\{m_{i,t}\}_{i,t}$.
2. Find $r_{i,t}$, the percentile rank of $m_{i,t}$, using a single ranking across all periods.

[8] There are more detailed questions underlying the out-of-pocket medical expense questions in the HRS, including the use of "unfolding brackets". Respondents can give ranges for medical expense amounts, instead of a point estimate or "don't know" as in the MCBS. The MCBS has the advantage that forgotten medical out-of-pocket medical expenses will be imputed if Medicare had to pay a share of the health event.

3. Using OLS, estimate the coefficient δ in equation (2.13) and construct $\{e_{i,t}\}_{i,t}$.
4. Construct the variable $y_{i,t} = f(e_{i,t})$, where $f(e)$ is a function that can be used to construct summary statistics.
   (a) For means we use $f(e) = e$.
   (b) For variances we use $f(e) = e^2$. (To get standard deviations, we need to find the square root of the predicted variances.) Let $\widehat{\sigma}^2$ denote the estimated variance, the average value of e^2.
   (c) For skewness we use $f(e) = e^3/\widehat{\sigma}^3$, where $\widehat{\sigma}^3$ utilizes the standard deviation from step (b).
   (d) For kurtosis we use $f(e) = e^4/\widehat{\sigma}^4$.
5. For each summary statistic, regress the appropriate $y_{i,t}$ against $r_{i,t-1}$, and plot the predicted values. In practice, we use a kernel-weighted local polynomial regression.

If the medical spending residuals follow an AR(1), then equation (2.13) is equivalent to equation (2.4) and the residuals $\{\hat{e}_{i,t}\}_{i,t}$ from equation (2.13) will deliver a consistent estimates of the residuals $\{\zeta_{i,t}\}_{i,t}$ from equation (2.4). Given that $\zeta_{i,t}$ is assumed to be iid and normally distributed (see equation (2.5)), it will at any value of $m_{i,t-1}$ have a mean of zero, constant variance, no skewness, and kurtosis of three.

In the more complex case where medical spending residuals follow an AR(1) plus white noise, the probability limit of $\hat{\delta}$ from equation (2.13) will be:

$$\begin{aligned}
\text{plim } \hat{\delta} &= \frac{Cov(m_{i,t}, m_{i,t-1})}{Var(m_{i,t-1})} \\
&= \frac{Cov(\eta_{i,t}, \eta_{i,t-1})}{Var(\eta_{i,t-1}) + Var(\varepsilon_{i,t-1})} \\
&= \phi \frac{Var(\eta_{it-1})}{Var(\eta_{it-1}) + Var(\varepsilon_{i,t-1})},
\end{aligned} \tag{2.14}$$

so that

$$\begin{aligned}
e_{i,t} &= m_{i,t} - \delta m_{i,t-1} \\
&= \eta_{i,t} + \varepsilon_{i,t} - \delta(\eta_{i,t-1} + \varepsilon_{i,t-1}) \\
&= \phi\eta_{i,t-1} + \zeta_{i,t} + \varepsilon_{i,t} - \delta(\eta_{i,t-1} + \varepsilon_{i,t-1}) \\
&= (\phi - \delta)\eta_{i,t-1} + \zeta_{i,t} + \varepsilon_{i,t} - \delta\varepsilon_{i,t-1}.
\end{aligned}$$

All the objects above are normally-distributed and zero-mean, and thus unconditionally e_{it} is also normally-distributed and zero-mean. This is also true conditional on $m_{i,t-1}$, as

$$
\begin{aligned}
\mathbb{E}[e_{i,t}\,|\,m_{i,t-1}] &= \mathbb{E}[(\phi-\delta)\eta_{i,t-1}+\zeta_{i,t}+\varepsilon_{i,t}-\delta\varepsilon_{i,t-1}\,|\,m_{i,t-1}] \\
&= (\phi-\delta)\,\mathbb{E}[\eta_{i,t-1}\,|\,m_{i,t-1}]-\delta\,\mathbb{E}[\varepsilon_{i,t-1}\,|\,m_{i,t-1}] \\
&= (\phi-\delta)\Big(\frac{Cov(\eta_{i,t-1},m_{i,t-1})}{Var(m_{i,t-1})}m_{i,t-1}\Big)-\delta\Big(\frac{Cov(\varepsilon_{i,t-1},m_{i,t-1})}{Var(m_{i,t-1})}m_{i,t-1}\Big) \\
&= (\phi-\delta)\frac{Var(\eta_{i,t-1})}{Var(m_{i,t-1})}m_{i,t-1}-\delta\frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}m_{i,t-1} \\
&= \phi\frac{Var(\eta_{i,t-1})Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})^2}m_{i,t-1}-\phi\frac{Var(\eta_{i,t-1})Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})^2}m_{i,t-1} \\
&= 0,
\end{aligned}
$$

with the fifth line following from equation (2.14), which implies that

$$
\phi-\delta \;=\; \phi\Bigg(1-\frac{Var(\eta_{i,t-1})}{Var(\eta_{i,t-1})+Var(\varepsilon_{i,t-1})}\Bigg) \;=\; \phi\frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}.
$$

The conditional variance of $e_{i,t}$ is given by:

$$
\begin{aligned}
Var(e_{i,t}\,|\,m_{i,t-1}) &= \mathbb{E}\left[e_{i,t}^2\,|\,m_{i,t-1}\right]-\mathbb{E}[e_{i,t}\,|\,m_{i,t-1}]^2 \\
&= \mathbb{E}\left[\Big((\phi-\delta)\eta_{i,t-1}+\zeta_{i,t}+\varepsilon_{i,t}-\delta\varepsilon_{i,t-1}\Big)^2\,|\,m_{i,t-1}\right] \\
&= (\phi-\delta)^2\,\mathbb{E}\left[\eta_{i,t-1}^2\,|\,m_{i,t-1}\right]+\mathbb{E}\left[\zeta_{i,t}^2\,|\,m_{i,t-1}\right]+\mathbb{E}\left[\varepsilon_{i,t}^2\,|\,m_{i,t-1}\right] \\
&\quad +\delta^2\,\mathbb{E}\left[\varepsilon_{i,t-1}^2\,|\,m_{i,t-1}\right]-2\delta(\phi-\delta)\,\mathbb{E}\left[\eta_{i,t-1}\varepsilon_{i,t-1}\,|\,m_{i,t-1}\right] \\
&= (\phi-\delta)^2\Big(Var(\eta_{i,t-1}\,|\,m_{i,t-1})+\mathbb{E}[\eta_{i,t-1}\,|\,m_{i,t-1}]^2\Big)+\sigma_\zeta^2+\sigma_\varepsilon^2 \\
&\quad +\delta^2\Big(Var(\varepsilon_{i,t-1}\,|\,m_{i,t-1})+\mathbb{E}[\varepsilon_{i,t-1}\,|\,m_{i,t-1}]^2\Big)-2\delta(\phi-\delta)\,\mathbb{E}\Big[\eta_{i,t-1}\varepsilon_{i,t-1}\,|\,m_{i,t-1}\Big] \\
&= (\phi-\delta)^2Var(\eta_{i,t-1})\Bigg(1-\frac{Var(\eta_{i,t-1})}{Var(m_{i,t-1})}\Bigg)+\delta^2Var(\varepsilon_{i,t-1})\Bigg(1-\frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}\Bigg) \\
&\quad +(\phi-\delta)^2\frac{Var(\eta_{i,t-1})^2}{Var(m_{i,t-1})^2}m_{i,t-1}^2+\delta^2\frac{Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^2}m_{i,t-1}^2 \\
&\quad -2\delta(\phi-\delta)\frac{Var(\eta_{i,t-1})Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})^2}m_{i,t-1}^2+\sigma_\zeta^2+\sigma_\varepsilon^2 \\
&= (\phi-\delta)^2Var(\eta_{i,t-1})\frac{Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}+\delta^2Var(\varepsilon_{i,t-1})\frac{Var(\eta_{i,t-1})}{Var(m_{i,t-1})} \\
&\quad +\phi^2\frac{Var(\eta_{i,t-1})^2Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^4}m_{i,t-1}^2+\phi^2\frac{Var(\eta_{i,t-1})^2Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^4}m_{i,t-1}^2 \\
&\quad -2\phi^2\frac{Var(\eta_{i,t-1})^2Var(\varepsilon_{i,t-1})^2}{Var(m_{i,t-1})^4}m_{i,t-1}^2+\sigma_\zeta^2+\sigma_\varepsilon^2 \\
&= ((\phi-\delta)^2+\delta^2)\frac{Var(\eta_{i,t-1})Var(\varepsilon_{i,t-1})}{Var(m_{i,t-1})}+\sigma_\zeta^2+\sigma_\varepsilon^2.
\end{aligned}
$$

Hence, the mean of $e_{i,t}$ conditional on $m_{i,t-1}$ is zero and the variance of $e_{i,t}$ conditional on $m_{i,t-1}$ is constant. For the conditional skewness and kurtosis, note that the distribution of $e_{i,t}$ conditional on $m_{i,t-1}$ is normal, as it is a linear combination of variables that themselves follow a multivariate normal distribution, conditional on $m_{i,t-1}$ which is normal. It follows that the skewness and kurtosis are constant.

2.A.6 Estimation methodology

We estimate the model using the extension of the *E-M* algorithm employed by Arellano et al. (2017). Recall from equations (2.8)-(2.10) that the functions $Q_\eta(\tau \,|\, \eta_{i,t-1}, a_{i,t})$, $Q_1(\tau \,|\, a_{i,1})$ and $Q_\varepsilon(\tau \,|\, , a_{i,t})$ are constructed from Hermite polynomials $\left(\{h_k^\eta(\cdot)\}_{k=0}^{K_\eta}, \{h_k^1(\cdot)\}_{k=0}^{K_1}, \{h_k^\varepsilon(\tau)\}_{k=0}^{K_\varepsilon}\right)$, using the coefficient functions $\{\beta_k^\eta(\tau)\}_{k=0}^{K_\eta}$, $\{\beta_k^1(\tau)\}_{k=0}^{K_1}$, and $\{\beta_k^\varepsilon(\tau)\}_{k=0}^{K_\varepsilon}$. The coefficient functions are in turn modeled with a set of polynomial splines defined over the intervals $\left\{[\tau_{\ell-1}, \tau_\ell]\right\}_{\ell=1}^{L}$, along with two tail functions for $(0, \tau_1]$ and $[\tau_L, 1)$. It is the parameters for these weighting functions that we must estimate.

Define θ as the vector of all parameters (the β parameters) in equations (2.8)-(2.10). The procedure to estimate θ is as follows. Starting with the vector $\hat{\theta}^{(0)}$ we iterate between the following two steps until $\hat{\theta}^{(j)}$ converges:

1. *Stochastic E-Step*: For each observation i, draw S values of $\eta_i^{(s)} = (\eta_{i1}^{(s)}, ..., \eta_{iT}^{(s)})$ from $f_i(\,.\,; \hat{\theta}^{(j)})$ (derived from $Q_1^{(j)}(\cdot)$, $Q_\eta^{(j)}(\cdot)$ and $Q_\varepsilon^{(j)}(\cdot)$).

2. *M-step*: Find

$$\underset{\beta_{\ell 0}^\eta, \ldots, \beta_{\ell K_\eta}^\eta}{\operatorname{argmin}} \quad \sum_{i=1}^N \sum_{s=1}^S \sum_{t=2}^T \rho_{\tau_\ell}\left(\eta_{it}^{(s)} - \sum_{k=1}^{K_\eta} \beta_{\ell k}^\eta \, h_k^\eta(\eta_{i,t-1}^{(s)}, a_{it})\right), \quad \ell = 1, ..., L.$$

   We use $\rho_\tau(\cdot)$ to denote Koenker and Bassett Jr's (1978) quantile "check" function. To identify the full set of splines, this function is minimized at each point ℓ on the grid over τ. The coefficients for $\varepsilon_{i,t}$ and $\eta_{i,1}$ likewise solve

$$\underset{\beta_{\ell 0}^\varepsilon, \ldots, \beta_{\ell K_\varepsilon}^\varepsilon}{\operatorname{argmin}} \quad \sum_{i=1}^N \sum_{s=1}^S \sum_{t=1}^T \rho_{\tau_\ell}\left(m_{i,t} - \eta_{it}^{(s)} - \sum_{k=1}^{K_\varepsilon} \beta_{\ell k}^\varepsilon \, h_k^\varepsilon(a_{it})\right), \quad \ell = 1, ..., L,$$

$$\underset{\beta_{\ell 0}^1, \ldots, \beta_{\ell K_1}^1}{\operatorname{argmin}} \quad \sum_{i=1}^N \sum_{s=1}^S \rho_{\tau_\ell}\left(\eta_{i1}^{(s)} - \sum_{k=1}^{K_1} \beta_{\ell k}^1 \, h_k^1(a_{i1})\right), \quad \ell = 1, ..., L.$$

   There are also moment conditions related to the tails of the distribution: See Arellano et al. (2017). These estimates give us $\hat{\theta}^{(j+1)}$.

For longer panels, settings with unbalanced data, or when estimating more complicated models the *E-step* can perform poorly when using standard samplers (e.g., Metropolis-Hastings). We therefore employ the sequential Monte-Carlo (SMC) approach implemented by Arellano et al. (2021). Comprehensive surveys of these methods can be found in Doucet et al. (2009) and Creal (2012).

We will use the Gaussian analogues to equations (2.8), (2.9) and (2.10) as importance distributions.

Step 1: SMC Stochastic E-Step to sample from $f(\eta_{i,1}, ..., \eta_{i,T} | Y_i^T, a_i^T)$. For $i = 1, ..., N$:

At $t = 1$:

1. Sample S particles $\eta_1^{(s)} \sim g(\eta_1 | y_1)$, where $g(\cdot)$ is the closed form posterior from the Gaussian model.
2. Compute the weights $w_1(\eta_1^{(s)})$ and apply a self-normalization to obtain $W_1^{(s)} \propto w_1(\eta_1^{(s)})$.
3. If $Var(W^{(s)})$ exceeds some threshold, re-sample $\{W_1^{(s)}, \eta_1^{(s)}\}$ to obtain S equally weighted particles.

At $t > 1$:

1. Sample S particles $\eta_t^{(s)} \sim g(\eta_t | \eta_{t-1}, y_t)$, where $g(\cdot)$ is the closed-form posterior from the Gaussian model.
2. Compute the weights $w(\eta_{1:t}^{(s)})$ and apply a self-normalization to obtain $W_t^{(s)} \propto w_t(\eta_{1:t}^{(s)})$.
3. If $Var(W^{(s)})$ exceeds some threshold, re-sample $\{W_t^{(s)}, \eta_t^{(s)}\}$ to obtain S equally weighted particles.
4. If $t = T$, sample P particles to be used in the *M-Step*. (We set $P = 1$).

Step 2: M-Step

1. Update quantile regressions for equations (2.8), (2.9) and (2.10) .
2. Update Laplace parameters for the tail functions.
3. Update parameters for Gaussian proposal distributions.

2.A.7 Demographic Transition Probabilities in the HRS

Let $hs_{i,j,t} \in \{0, 1, 2, 3\}$ denote death ($hs_{i,j,t} = 0$) and the 3 mutually exclusive health states of the living (nursing home = 1, bad = 2, good = 3, respectively) of household member j, household i, time t. Let $x_{i,j,t}$ be a vector that includes a constant, age, permanent income, gender, and powers and interactions of these variables, and indicators for previous health and previous health interacted with age. Our goal is to construct the likelihood function for the transition probabilities.

Using a multivariate logit specification, we have, for $q \in \{1, 2, 3\}$, $r \in \{0, 1, 2, 3\}$, we rewrite equation (2.11) as

$$\begin{aligned} \pi_{q,r,t} &= \Pr(hs_{i,g,t+1} = r | \, hs_{i,g,t} = q; x_{i,g,t}) \\ &= \gamma_{qr} \Big/ \sum_{s \in \{0,1,2,3\}} \gamma_{qs}, \\ \gamma_{qs} &\equiv 1, \quad s = 0 \\ \gamma_{qs} &= \exp\left(x_{i,g,t}\beta_s\right), \quad s \in \{1, 2, 3\}, \end{aligned}$$

where $\{\beta_s\}_{s=1}^3$ are coefficient vectors for each future state s and $x_{i,g,t}$ is the explanatory variable vector which depends on the current state q.

The formulae above give 1-period-ahead transition probabilities, whereas what we observe in the HRS data set are 2-period ahead probabilities, $\Pr(hs_{i,g,t+2} = r | \, hs_{i,g,t} = q; x_{i,g,t})$. The two sets of probabilities are linked, however, by

$$
\begin{aligned}
\Pr(\, hs_{i,g,t+2} = r |\, hs_{i,g,t} = q; x_{i,g,t}) &= \sum_s \Pr(\, hs_{i,g,t+2} = r |\, hs_{i,g,t+1} = s; x_{i,g,t}) \Pr(\, hs_{i,g,t+1} = s |\, hs_{i,g,t} = q; x_{i,g,t}) \\
&= \sum_s \pi_{sr,t+1} \pi_{qs,t},
\end{aligned}
$$

imposing $\pi_{00,t+1} = 1$. This allows us to estimate $\{\beta_k\}$ directly from the data using maximum likelihood.

Bibliography

Alemayehu, B. and K. E. Warner (2004). The lifetime distribution of health care costs. *Health Services Research 39*(3), 627–642.

Ameriks, J., J. S. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy 128*(6), 2375–2451.

Arellano, M., R. Blundell, and S. Bonhomme (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica 85*(3), 693–734.

Arellano, M., R. Blundell, S. Bonhomme, and J. Light (2021). Heterogeneity of consumption responses to income shocks in the presence of nonlinear persistence.

Banks, J., R. Blundell, P. Levell, and J. P. Smith (2019). Life-cycle consumption patterns at older ages in the united states and the united kingdom: Can medical expenditures explain the difference? *American Economic Journal: Economic Policy 11*(3), 27–54.

Brown, J. and A. Finkelstein (2008). The interaction of public and private insurance: Medicaid and the long term care insurance market. *American Economic Review 98*(5), 837–880.

Creal, D. (2012). A survey of sequential monte carlo methods for economics and finance. *Econometric reviews 31*(3), 245–296.

De Nardi, M., G. Fella, and G. Paz-Pardo (2020). Nonlinear household earnings dynamics, self-insurance, and welfare. *Journal of the European Economic Association 18*(2), 890–926.

De Nardi, M., E. French, and J. B. Jones (2010). Why do the elderly save? the role of medical expenses. *Journal of Political Economy 118*(1), 39–75.

De Nardi, M., E. French, and J. B. Jones (2016). *American Economic Review 106*(11), 3480–3520.

De Nardi, M., E. French, J. B. Jones, and J. McCauley (2016). Medical spending on the US elderly. *Fiscal Studies 37*(3-4), 717–747.

De Nardi, M., E. French, J. B. Jones, and R. McGee (2021). Why do couples and singles save during retirement? Technical report, National Bureau of Economic Research.

Doucet, A., A. M. Johansen, et al. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering 12*(656-704), 3.

Fahle, S., K. McGarry, and J. Skinner (2016). Out-of-pocket medical expenditures in the United States: Evidence from the Health and Retirement Study. *Fiscal Studies 37*(3-4), 785–819.

Feenberg, D. and J. Skinner (1994). The risk and duration of catastrophic health care expenditures. *The Review of Economics and Statistics 76*(4), 633–647.

French, E., M. De Nardi, J. B. Jones, O. Baker, and P. Doctor (2006). Right before the end: Asset decumulation at the end of life. *Economic Perspectives 30*(3), 2–13.

French, E. and J. B. Jones (2004). On the distribution and dynamics of health care costs. *Journal of Applied Econometrics 19*(6), 705–721.

French, E., J. B. Jones, and J. McCauley (2017). The accuracy of economic measurement in the Health and Retirement Study. *Forum for Health Economics and Policy 20*(2).

Friedberg, L., W. Hou, W. Sun, A. Webb, and Z. Li (2014, November). New evidence on the risk of requiring long-term care. Center for Retirement Research at Boston College Working Paper 2014-12.

Hirth, R. A., S. Calonico, T. B. Gibson, H. Levy, J. Smith, and A. Das (2016). Long-term health spending persistence among the privately insured in the US. *Fiscal Studies 37*(3-4), 749–783.

Hurd, M. D., P.-C. Michaud, and S. Rohwedder (2017). Distribution of lifetime nursing home use and of out-of-pocket spending. *Proceedings of the National Academy of Sciences 114*(37), 277–297.

Jones, J. B., M. De Nardi, E. French, R. McGee, and J. Kirschner (2018, May). The lifetime medical spending of retirees. Working Paper 24599, National Bureau of Economic Research.

Koenker, R. and G. Bassett Jr (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

Kopecky, K. and T. Koreshkova (2014, July). The impact of medical and nursing home expenses on savings and inequality. *American Economic Journal: Macroeconomics 6*(3), 29–72.

Livshits, I., J. MacGee, and M. Tertilt (2007). Consumer bankruptcy: A fresh start. *American Economic Review 97*(1), 402–418.

Lockwood, L. M. (2012). Bequest motives and the annuity puzzle. *Review of economic dynamics 15*(2), 226–243.

Merrill Lynch Wealth Management (2012, February). Merrill Lynch Affluent Insights Survey fact sheet.

Nakajima, M. and I. A. Telyukova (2018). Medical expenses and saving in retirement: The case of U.S. and Sweden. Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute.

Nicholas, L. H., K. M. Langa, T. J. Iwashyna, and D. R. Weir (2011). Regional variation in the association between advance directives and end-of-life medicare expenditures. *Jama 306*(13), 1447–1453.

Pashchenko, S. (2013). Accounting for non-annuitization. *Journal of Public Economics 98*, 53–67.

Pashchenko, S. and P. Porapakkarm (2016). Medical spending in the US: facts from the Medical Expenditure Panel Survey data set. *Fiscal Studies 37*(3-4), 689–716.

Poterba, J. M., S. F. Venti, and D. A. Wise (2017). The asset cost of poor health. *Journal of the Economics of Ageing 9*, 172–184.

Reichling, F. and K. Smetters (2015). Optimal annuitization with stochastic mortality probabilities. *American Economic Review 105*(11), 3273–3320.

Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006). Are americans saving optimally for retirement? *Journal of Political Economy 114*, 607–643.

Skinner, J. (2007). Are you sure you're saving enough for retirement? *Journal of Economic Perspectives 21*(3), 59–80.

Webb, A. and N. A. Zhivan (2010, February). How much is enough?: The distribution of lifetime health care costs. Center for Retirement Research at Boston College Working Paper 2010-1.

Chapter 3

Dementia and Disadvantage in the United States and England

3.1 Introduction

Dementia, a severe and irreversible decline in memory and other cognitive functions, is a major and increasing global health challenge. It is the fifth leading cause of death globally and is one of the most common comorbidities for Covid-19 morbidity.(Nichols et al., 2019; Kuo et al., 2020) It results in large social and economic costs.(Hurd et al., 2013; Hudomiet et al., 2019) Americans are more likely to be in poor health than their English counterparts in multiple dimensions, including heart disease and diabetes.(Banks et al., 2006) These differences are large along all points of the socioeconomic status (SES) gradient, although the gradient is generally steeper in the US. While the SES gradient for many diseases has been well established,(Marmot, 2006; Brunner, 2016; Marmot, 2020) only a few studies have focused specifically on dementia.(Rusmaully et al., 2017; George et al., 2020) The available evidence is summarized in Table 3.3 in the online appendix. The evidence of the SES gradient for dementia is also less clear, as in England while a strong association has been established between wealth and dementia incidence, the same was not observed for education level.(Cadar et al., 2018) The Global Burden of Disease Study (GBD) reported that in 2017 among those aged over 70, the US had a lower overall prevalence of dementia at 7.89% compared with the UK at 8.91%.(of Disease, 2020) However, the GBD has identified substantial heterogeneity in case-ascertainment methods throughout the dementia literature, resulting in location-specific inconsistencies and potentially biased cross-country comparisons. This has led to calls for analyses with more consistent and comparable measures of dementia to inform policy makers, researchers, and clinicians about global differences in dementia.(Launer, 2018) In this study, we compared dementia prevalence in England and the US among non-Hispanic whites aged 70+, and how it varied across the SES gradient of each country. Location-specific inconsistencies caused by differences in diagnostic practices were not an issue in our study because we used an identical case definition for dementia, and the surveys in the analysis shared the same design and sampling techniques. More specifically, we used two large surveys, the US Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA), that contain a battery of the same demographic, cognitive, and functional measures, and we applied the same prediction algorithm in both countries to detect undiagnosed as well as diagnosed cases. We compared dementia prevalence within and across England and the US using important indicators of SES, specifically: income; education; wealth; and non-housing wealth.

3.2 Strengths and limitations of this study

- This is the first study to compare dementia prevalence across countries using the same survey methodology and the exact same measure of dementia. The surveys have similar sample selection and questionnaire design. We standardise our estimates by age and gender to the English population aged over 70 in 2016. Any differences in overall prevalence across the two countries should represent true differences.
- We measure the SES gradient of dementia across four different measures of SES: income; education; wealth; and non-housing wealth.
- Dementia disproportionately affects the most disadvantaged in in both countries, although the gradient is steeper in the US according to all four measures of SES.
- We do not ascertain dementia directly, but predict cases using a common battery of measures in ELSA and HRS. One of the SES measures, education, is also used as a predictor of dementia.

3.3 Methods

3.3.1 Description of Surveys

Data were extracted from the 2016 and earlier waves of the HRS and ELSA, which are nationally representative biennial surveys of the US and English populations, respectively.(Sonnega et al., 2014; Steptoe et al., 2012) Both the HRS and ELSA follow respondents longitudinally until death, with new cohorts entering to maintain population representativeness as the study sample gets older. The design of ELSA was based on the HRS, making the two surveys analogous, with both collecting data on health, ability, demographics, employment, and wealth. In addition to measuring health conditions and difficulties respondents have with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs), sample members also have their cognitive function assessed. A range of tests adapted from the Telephone Interview for Cognitive Status (TICS) have been carried out in HRS since 1996 and ELSA since 2014. If a sample member was unable to respond in person, a proxy respondent was asked questions about the respondent's change in memory. Both surveys have a high response rate, which is displayed in the online appendix Table 3.4. We describe these surveys in more detail in the online appendix.

3.3.2 Cohort Description

Our samples are restricted to non-Hispanic whites over the age of 70 years old that live in the community or in nursing homes in 2016. This provides a study sample of 5,330 participants in the HRS and 3,147 participants in ELSA. We restrict our sample to non-Hispanic whites to ensure estimates are comparable across countries. Summary statistics of both the raw and selected samples are displayed in online appendix Tables 3.5 and 3.6, and specifications that include ethnic minorities are also displayed.

3.3.3 Patient and Public Involvement

No study participants were involved in setting the research question or outcome measures, nor were they involved in any other area of the design, implementation, and analysis of the study. There are no direct plans to disseminate the results of the research to study participants.

3.3.4 Dementia Case Definition

The HRS included a detailed clinical substudy (ADAMS: Aging, Demographics and Memory Study) of 856 sample members aged 70+ who completed an in-depth in-home assessment of cognitive status conducted by experienced teams at the Duke University Dementia Epidemiology Research Center who diagnosed each participant as normal, cognitively impaired but not demented (CIND), or demented.(Hurd et al., 2013) Data from ADAMS is regarded as the gold-standard dementia diagnoses against which to train algorithms to predict dementia.(Gianattasio et al., 2018) Hurd et al. estimated separate ordered probit models in the ADAMS subsample for self- and proxy-respondents to generate a predictive algorithm for cognitive status, based on the ADAMS diagnoses, for the whole HRS sample. The algorithm uses a range of variables including demographic information, Activities of Daily Living (ADLs), Instrumental Activities of Daily Living (IADLs), TICS questionnaire, as well as the change in these variables across waves.(Hurd et al., 2013) Proxy respondents had a separate predictive algorithm as they were asked a different set of questions from self-respondents, which included the short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE). The use of a proxy to assess cognitive decline and dementia in elderly people is a recognised accepted standard method for identifying severe cognitive impairment and has been validated many times.(Jorm et al., 2000) Importantly, the same set of questions used in the Hurd et al. algorithm is asked of self- and proxy-respondents in both HRS and ELSA. Summary statistics for a variety of predictors are displayed in online appendix Table 3.7.

We applied Hurd et al.'s predictive algorithm to estimate the probably of dementia for those in the HRS sample in 2016 and extended the prediction to the ELSA sample. The algorithm predicts the probability of dementia in the following year; therefore, we predicted dementia prevalence in 2017. Hurd et al's predictive algorithm has been shown to have an accuracy (percentage correctly classified as demented or nondemented) of 94%, sensitivity of 65%, and a specificity of 98% in the estimation sample.(Gianattasio et al., 2018) An in-depth discussion of the predictive algorithm procedure can be found in the online appendix

Non-response for people unable or unwilling to participate in the survey is important when attempting to estimate dementia prevalence across the population. While attrition exists in both surveys, it is unlikely to significantly affect our estimates since among older ELSA and HRS respondents there is no statistically significant correlation between attrition and prior health or the SES indicators of education, income and wealth.(Banks et al., 2010)

3.3.5 Measures of Socioeconomic Status

We considered four measures of SES: income; education; wealth; and non-housing wealth. Income is measured as current household income from all sources. For education we used total years of schooling. Wealth is measured as the sum of all household reported savings, stocks, bonds, business wealth, other assets, and the value of housing assets (e.g., properties) after financial debt and mortgage debt has been subtracted. Non-housing wealth is the same measure as wealth but excludes housing assets and mortgage debt, and therefore measure wealth that can be more easily converted to cash. Wealth and non-housing wealth are both measured from 4 years prior to minimize reverse causality, as medical expenses associated with dementia are high and may run down wealth.(Hudomiet et al., 2019) For each measure, we created a SES gradient by ranking individuals based on that measure. For income, wealth, and non-housing wealth, we assigned everyone to a decile in their respective country. For education, we ranked individuals according to their number of years of schooling.

3.4 Statistical Analysis

We created a pooled dataset of the two surveys. In our statistical analysis we used HRS and ELSA sampling weights to adjust for nonresponse and for the sampling design of the surveys. To make both within country estimates along the SES gradient and cross-country estimates directly comparable, estimates were agegender standardized to the English population aged over 70 in 2016 using direct standardization, categorising the population into ten groups: five age bands (70-74; 75-79; 80-84; 85-89; and 90+) by gender. We estimated the prevalence of dementia in each country, their difference, and compared the prevalence along the four SES gradients. For each estimate presented we computed the corresponding 95% confidence interval (CI), and for any differences we computed the corresponding p-values. For each SES factor, as well as estimating the age-gender standardized prevalence along the gradient, we calculated the Relative Index of Inequality (RII) and Slope Index of Inequality (SII) using generalized linear models (log binomial regression) with logarithmic and identity link functions, respectively. The RII can be interpreted as the relative likelihood of dementia prevalence of those in the lowest SES group compared to those in the highest, and the SII can be interpreted as the absolute effect on dementia probability of moving from the lowest SES group to the highest.(Wagstaff et al., 1991) To assess whether any observed differences could be explained by disparities in past health risk factors across countries we conditioned on a variety of risk factors and assessed how our estimates changed. Where possible, when conditioning on these factors we used past health instead of current health to address the problem of reverse causality: i.e., the problem that dementia may cause health problems such as low weight. Statistical analyses were performed using STATA software.

3.5 Results

Table 3.1 shows the age-gender standardized prevalence of dementia for the aged over 70 white nonHispanic population in both England and the US. Dementia prevalence is lower in England at 9.7% (95% CI, 8.9% to 10.6%) than the US at 11.2% (95% CI, 10.6% to 11.8), a difference of 1.4 percentage points (pp) that is highly statistically significant (P = 0.0055). Table 3.1 also shows dementia prevalence for different SES groups, in terms of income, education, wealth, and non-housing wealth. Regardless of the measure of the SES, there is a clear gradient in dementia prevalence, with the most disadvantaged groups in both England and the US having higher dementia prevalence. The gradient is steeper in the US and is driven by significantly higher dementia prevalence for those at the very bottom of the distribution. In the US, those in the lowest income decile have a dementia prevalence of 18.7% (95% CI, 16.6% to 20.8%), which is considerably higher than in England, with a prevalence among those in the lowest decile of 11.4% (95% CI, 8.9% to 13.9%). The difference is highly statistically significant ($P < 0.0001$). For income deciles above the lowest, the difference across the two countries is much smaller and not statistically significant. This same general pattern is evident across the other measures of SES that we consider, although when using wealth, the difference between those in the bottom decile is not statistically significant. Figure 3.1 presents the same dementia prevalence information shown in Table 3.1, but in graphical format. It also reports the Slope Index of Inequality (SII) for the four different measures of the SES for both countries. In both the US and England, dementia is more prevalent among the more disadvantaged. The gradient tends to be steeper in the US, corresponding to a larger (in absolute value) SII in the US for each SES measure. For income, the SII is -0.062 (95% CI, -0.097 to -0.028) and -0.085 (95% CI, -0.114 to -0.057) for England and the US, respectively. The SIIs are not statistically different. If the lowest income decile is excluded, the SII for England becomes slightly steeper (-0.067 (95% CI, -0.107 to -0.027)) whereas the SII for the US becomes less steep than England (-0.060 (95% CI, -0.093 to -0.027)). Next, we attempted to understand the potential drivers of these gradients and the differences in the gradients across countries. We extended the analysis to account for cardiometabolic diseases (i.e. diabetes, heart disease and stroke), and behaviours (i.e. smoking and body mass index) as dementia risk factors.(Nichols et al., 2019; Yaffe et al., 2013) Previous research showed these factors to be more prevalent in the US than England, especially among the most disadvantaged.(Banks et al., 2006) Table 3.2 displays the percentage point difference in dementia prevalence after we controlled for various measures of past health and behaviours. The results are split into three panels: the whole sample; the whole sample excluding those in the lowest income decile; and the lowest income decile. As was shown in Table 3.1, the difference in the prevalence of dementia between England and the US was 1.43 pp (P = 0.0055). Table 3.2 shows this difference declined to 0.894 pp (P = 0.11) when we excluded the lowest income decile. Controlling for past health and behaviours modestly reduced this cross-country difference further: the difference declined by a maximum of 19%. For the lowest income decile, controlling for past health and behaviour reduced the cross-country difference of 7.27 pp ($P < 0.0001$) by a more substantial 33%.

In the online appendix Tables 3.8-3.10 we investigated whether past health and behaviour explained the SES gradient within each country. We found that in England and the US these factors accounted for most of the SES gradient, as shown in online appendix Figure 3.3. However, in the US, prevalence in the lowest income decile remained disproportionately high.

Education has also been shown to be a risk factor for dementia. Table 3.1 shows that in both the US and England, the less educated have higher dementia prevalence. Controlling for education increased the estimated difference across countries, from 1.43 pp to 2.82 pp, as can be seen in the online appendix Table 3.9. Education cannot explain these differences since the English have lower educational attainment. In our main analysis we exclude ethnic minorities, who have higher dementia prevalence and comprise a larger share of the US than the English population. Including minorities increased the estimated difference across countries, from 1.43 pp to 2.42 pp, as can be seen in the online appendix Tables 3.2 and 3.11. This is largely caused by the high prevalence of dementia among minorities in the US as displayed in online appendix Table 3.12.

Table 3.1: Prevalence of dementia, USA versus England, 2017

	England Age-gender standardised prevalence (95%CI)	USA Age-gender standardised prevalence (95%CI)	Difference	P value
All	0.097 (0.089 to 0.106)	0.112 (0.106 to 0.118)	0.014	0.0055
Household income decile				
1 (lowest)	0.114 (0.089 to 0.139)	0.187 (0.166 to 0.208)	0.073	<0.0001
2	0.113 (0.090 to 0.136)	0.141 (0.119 to 0.163)	0.028	0.090
3	0.124 (0.097 to 0.151)	0.111 (0.095 to 0.127)	-0.013	0.42
4	0.099 (0.071 to 0.126)	0.118 (0.099 to 0.137)	0.019	0.26
5	0.094 (0.072 to 0.116)	0.086 (0.069 to 0.102)	-0.008	0.56
6	0.098 (0.070 to 0.127)	0.108 (0.088 to 0.128)	0.010	0.59
7	0.068 (0.042 to 0.093)	0.100 (0.078 to 0.122)	0.032	0.060
8	0.083 (0.053 to 0.114)	0.082 (0.066 to 0.097)	-0.002	0.92
9	0.082 (0.041 to 0.122)	0.093 (0.071 to 0.116)	0.012	0.62
10 (highest)	0.059 (0.035 to 0.083)	0.077 (0.052 to 0.103)	0.019	0.30
Years of schooling				
9 or fewer	0.128 (0.101 to 0.154)	0.190 (0.162 to 0.218)	0.062	0.0015
10	0.095 (0.074 to 0.116)	0.137 (0.109 to 0.165)	0.042	0.018
11	0.096 (0.077 to 0.115)	0.109 (0.080 to 0.139)	0.013	0.471
12	0.071 (0.042 to 0.100)	0.124 (0.114 to 0.133)	0.053	0.0006
13	0.061 (0.038 to 0.083)	0.116 (0.090 to 0.141)	0.055	0.0013
14 or more	0.056 (0.039 to 0.073)	0.085 (0.076 to 0.093)	0.029	0.0031
Household wealth decile				
1 (lowest)	0.165 (0.132 to 0.198)	0.187 (0.162 to 0.211)	0.022	0.31
2	0.117 (0.092 to 0.143)	0.149 (0.129 to 0.169)	0.031	0.061
3	0.100 (0.073 to 0.127)	0.107 (0.091 to 0.122)	0.006	0.68
4	0.093 (0.071 to 0.115)	0.115 (0.098 to 0.132)	0.022	0.12
5	0.110 (0.085 to 0.134)	0.091 (0.077 to 0.106)	-0.018	0.21
6	0.080 (0.057 to 0.103)	0.089 (0.074 to 0.103)	0.008	0.55
7	0.070 (0.046 to 0.094)	0.103 (0.084 to 0.123)	0.034	0.034
8	0.092 (0.066 to 0.118)	0.089 (0.072 to 0.106)	-0.003	0.85
9	0.082 (0.048 to 0.116)	0.103 (0.084 to 0.123)	0.021	0.28
10 (highest)	0.060 (0.038 to 0.081)	0.067 (0.051 to 0.084)	0.008	0.58
Household non-housing wealth decile				
1 (lowest)	0.136 (0.103 to 0.168)	0.201 (0.176 to 0.226)	0.065	0.0019
2	0.109 (0.084 to 0.134)	0.137 (0.119 to 0.156)	0.029	0.074
3	0.123 (0.095 to 0.151)	0.131 (0.111 to 0.150)	0.008	0.66
4	0.096 (0.073 to 0.118)	0.101 (0.085 to 0.116)	0.005	0.71
5	0.108 (0.082 to 0.134)	0.107 (0.091 to 0.123)	-0.001	0.95
6	0.093 (0.067 to 0.119)	0.086 (0.070 to 0.101)	-0.007	0.64
7	0.099 (0.072 to 0.125)	0.086 (0.070 to 0.102)	-0.013	0.42
8	0.078 (0.051 to 0.105)	0.090 (0.073 to 0.106)	0.011	0.49
9	0.058 (0.036 to 0.081)	0.092 (0.074 to 0.111)	0.034	0.024
10 (highest)	0.063 (0.043 to 0.084)	0.079 (0.062 to 0.095)	0.015	0.26

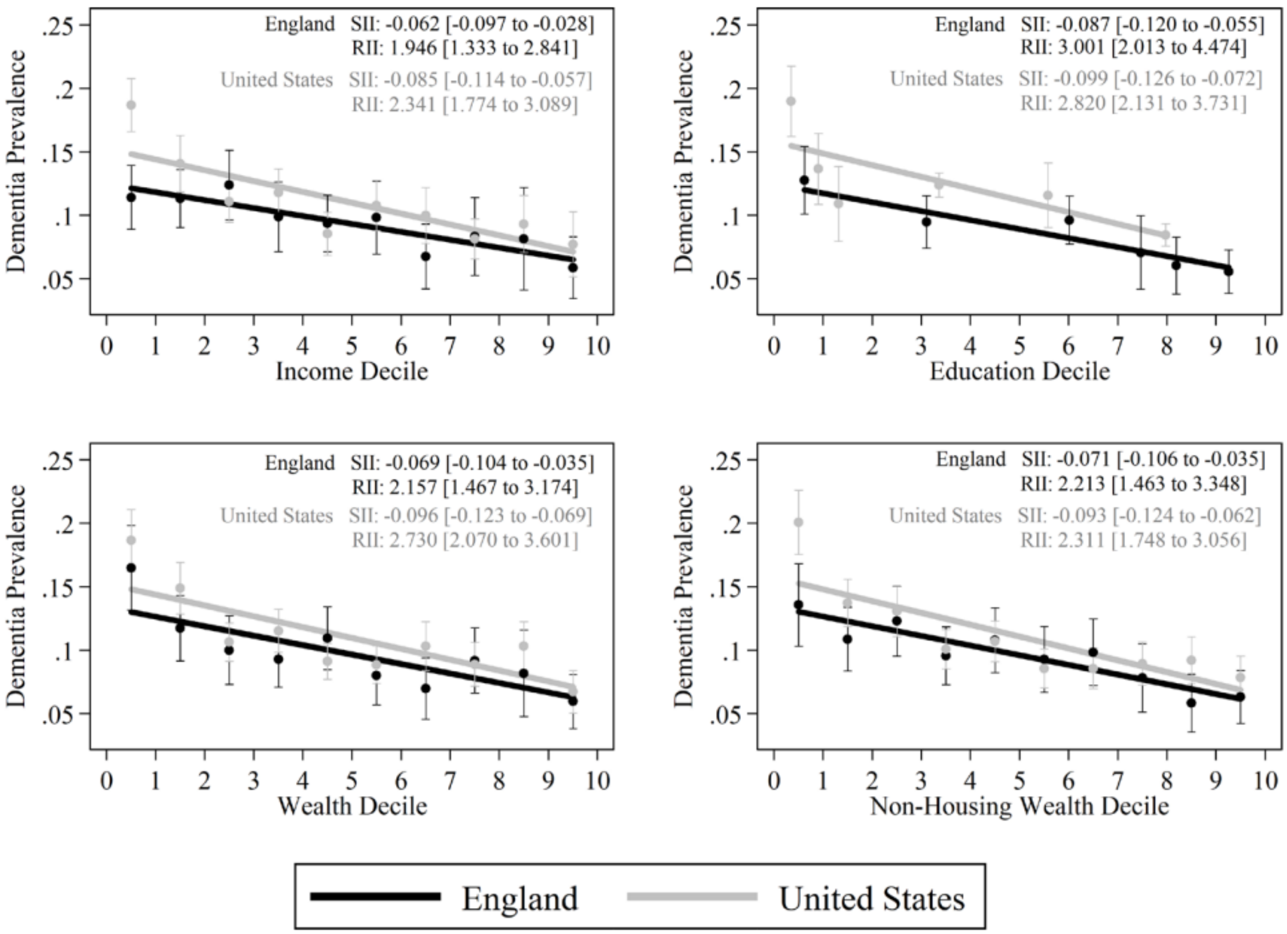


Figure 3.1: SES Gradient of Dementia, USA versus England

SES Gradient of Dementia, USA versus England, 2017, according to four measures of SES. Absolute and relative inequality shown with 95% CI. The points in this figure represent the mean age-gender standardised dementia prevalence for each country by SES, along with 95% CI for these predictions. The solid lines represent the fitted Slope Index of Inequality (SII: absolute inequality) for each country. The values of the SII and the corresponding Relative Index of Inequality (RII: relative inequality) are listed in the top right of each figure, with 95% CI in brackets. For education, individuals are ranked based on their years of schooling within each country, and as the USA has higher educational attainment, individuals with 14+ years of schooling are the 80th percentile of the US education distribution, but at the 90th percentile of the English educational distribution. SES, socioeconomic status.

3.6 Discussion

3.6.1 Main findings

Using nationally representative samples of older individuals from England and the US, and applying the same algorithmic procedure to predict dementia in both samples, we showed that in both the US and England dementia is more prevalent among the disadvantaged, and the SES gradient of dementia is steeper in the US. The steeper gradient in the US is largely driven by those in the lowest decile. In both countries, most of the SES gradient disappeared when we controlled for past health related factors, although prevalence for those in lowest income decile in the US remained disproportionately high. If the lowest income decile is excluded from our sample the difference in dementia prevalence across the countries is statistically insignificant, and the remaining SES gradient of dementia is remarkably similar across countries. While poorer individuals face a higher burden of dementia in both England and the US, the extremely poor in the US face a disproportionately high burden of dementia. Controlling for past health-related factors can explain some, but not all, of the cross-country difference. It can explain up to around one third of the difference for those in the lowest income decile. While past health factors such as adiposity and smoking are correlated with dementia, those in the lowest income decile in the US do not smoke more or have higher BMIs than their English counterparts. Therefore, this cannot explain their disproportionately higher prevalence of dementia. Education also cannot explain the difference, as the US population is more educated at every income decile, and in fact the educational difference masks some of the underlying dementia risk difference between countries. Adding minorities increased estimated dementia prevalence, especially in the US, because dementia prevalence is higher among minorities, who comprise a higher share of the US than the English population. This fits with prior research which showed dementia prevalence is higher for non-whites and Hispanics.(Yaffe et al., 2013; Vega et al., 2017) We did not observe higher dementia prevalence among Hispanics and non-whites in the US for those in relatively high socioeconomic groups.

Table 3.2: Difference in prevalence of dementia, USA vs England, 2017

Whole sample					
Percentage point difference	1.43	1.34	1.15	1.21	1.18
p value	0.0055	0.0091	0.025	0.020	0.034
% Difference from baseline	-	-6%	-19%	-16%	-17%
Excluding lowest income decile					
Percentage point difference	0.89	0.88	0.74	0.75	0.81
p value	0.11	0.11	0.18	0.18	0.18
% difference from baseline	-	-2%	-17%	-16%	-10%
Lowest income decile					
Percentage point difference	7.27	6.14	5.45	5.93	4.85
p value	<0.0001	0.0003	0.002	0.0011	0.012
% difference from baseline	-	-15%	-25%	-18%	-33%
Control for					
Past cardiometabolic diseases		✓	✓	✓	✓
Past psychiatric conditions			✓	✓	✓
Ever smoked				✓	✓
Past BMI					✓

3.6.2 Comparison with previous studies and how findings are an advance on current literature

Previous studies have shown cross-country variation in dementia prevalence. However, substantial heterogeneity in case-ascertainment methods across countries and studies makes interpreting any observed differences difficult. We believe this is the first study to compare dementia prevalence in England and the US using the exact same measure of dementia, thus overcoming previous difficulties in making comparisons across the two countries due differences in diagnostic practices and case definitions. We also compared the SES gradient of dementia in both countries. While some studies have shown in both England and the US those with lower education and less wealth have been found to have higher rates of dementia,(Hudomiet et al., 2019; Rusmaully et al., 2017; Cadar et al., 2018; Langa et al., 2017; Rocca et al., 2011) there are no systematic comparative studies. We compared prevalence along the SES gradient using almost identical measures of income, wealth, and education. Further, we standardized the cross-country comparison for age and gender, using the English over 70 population as the standard population. We found that dementia prevalence is higher and more concentrated among the poorest in the US than England. Detailed disaggregation according to SES measures shows the true extent of the excess burden of dementia in the very poorest group in the US.

We showed that risk factors for dementia such as cardiometabolic diseases, psychiatric conditions, high BMI, smoking have similar affects across countries. Accounting for these risk factors removes most of the SES gradient for both countries, but disproportionately high prevalence remains for the most disadvantaged in the US.

3.6.3 Implications (wider interpretation)

Much research has shown that low income Americans are more likely to be in poor health and die younger than their high-income counterparts.(Chetty et al., 2016) We show that these health differences also extend to dementia prevalence.

While risk factors contribute to higher prevalence among those who are more disadvantaged, those in the US appear to have an undue burden that is caused by risk factors for which we cannot account. One possible explanation is differential access to healthcare. The NHS provides broadly equitable care according to education in the older population after accounting for health status.(Stoye et al., 2020) In the US, the poor often go uninsured, and although virtually every American aged 65 or older is eligible for Medicare, around 20% of Medicare beneficiaries healthcare must be financed out of pocket.(De Nardi et al., 2016) The extent to which health care provision below and above aged 65 may account for the relative excess dementia burden in the US is unclear.

The implications of our results are that interventions designed to attempt to prevent dementia should be targeted towards the most disadvantaged. This is especially true in the US. As yet, we are unable to advocate specific measures as we do not yet understand the specific nature of disadvantage in respect to dementia risk

3.6.4 Strengths and Weaknesses of Analysis

This study has a strong design. Results are directly comparable across England and the US. The same predictive algorithm was applied to both countries, addressing the problem of heterogeneity in case ascertainment which has affected the literature.(Nichols et al., 2019; Launer, 2018) Further, because ELSA and HRS share sample selection and questionnaire design, any differences in overall prevalence and SES gradients in prevalence across the two countries should represent true differences. In contrast to the Global Burden Disease study, we find higher dementia prevalence in the US. Furthermore, we measured the SES gradient of dementia across four different measures of SES, with consistent results. Our work also highlights the usefulness of the standardized measure of dementia to allow for meaningful comparisons across countries.

This study had three limitations. First, we do not ascertain dementia directly, but predict cases using a common battery of measures in ELSA and HRS. Importantly, Hurd's prediction algorithm has high accuracy and although our case definition lacks a clinical point of reference in England, it is based on a detailed clinical substudy in the US.(Gianattasio et al., 2018) Further, cross-cultural subjectivities in reporting of impairment severity are likely to be similar in the US and England (see online appendix Table 3.7). It would be of great value for future work to use the Harmonized Cognitive Assessment Protocol (HCAP) data to provide a standard clinical point of reference to validate and verify the cross-country dementia prevalence estimates.(Langa et al., 2020) Second, education is one of the factors in the predictive algorithm for dementia and also one of our measures of SES. The dementia algorithm takes account of the well-documented correlation between education level and cognitive function in adult life. Nevertheless, we found substantial absolute and relative inequalities in dementia prevalence according to education level in the US and UK. Education may be a successful approach for reducing dementia risk. (Nguyen et al., 2016) Third, while we show risk factors explain a large proportion of the differences in dementia between the England and the US - although cannot account for the difference in the lowest income decile - there are likely other unmeasured confounding factors that impact dementia prevalence which we do not observe.

3.7 Conclusion

Given the large social and economic costs of dementia, there is great value in understanding the scope and burden of dementia in the population along the SES gradient. This study indicates that more disadvantaged individuals face a higher burden of dementia and that the poorest individuals in the US face a disproportionately high burden. The high burden faced by these individuals can be partly but not fully explained by past health factors. We lacked data on other important possible contributing factors such as habitual drug use. Further research is needed to fully understand this issue using data from multiple sources.

3.A Appendix

Table 3.3: Studies on Dementia Prevalence by Socioeconomic Status (SES)

Authors	Country	Socioeconomic Indicator	Years	Findings
Langa et al. (2017)	United States	Education	2000 and 2012	Link between education and dementia risk
Hudomiet et al, (2019)	United States	Education and Social Security Benefits	1998 to 2014	Link between education and dementia risk
Rusmaully et al, (2017)	United Kingdom	Education, height, occupational position	Various years between 1985 and 2015	High cognitive reserve associated with lower risk for dementia
Cadar et al, (2018)	United Kingdom	Education, wealth and the index of multiple deprivation	2002 to 2015	Lower wealth in but not education associated with increased risk for dementia
Rocca et al,(2011)	United States	Education, net worth	1993 and 2002	Higher education, higher net worth protected against cognitive impairment.
Basu (2013)	United States	Education	2000 - 2002	Education has causal effect on dementia risk
Nguyen et al, (2016)	United States	Education	1998 to 2010	Education protective against dementia risk
Crimmins et al, (2018)	United States	Education	2000 and 2010	More education linked to lower dementia prevalence
Garcia et al, (2018)[9]	United States	Race, ethnicity, nativity, and education	2012	Education reduces the odds for CIND
Weden et al, (2018)	United States	Race, ethnicity, total number of children, marital status, highest educational attainment, and net total assets in 2000	2000 and 2010	Strong protective role of educational attainment and persisting rural disadvantages for dementia

3.A.1 HRS Data

The Health and Retirement Study (HRS) is a nationally representative, biennial longitudinal survey of adults in the United States.(Cadar et al., 2018) It started in 1992 and since it collects a wide range of questions on income, wealth, employment, health, cognition, and demographics. It utilises a steady-state sampling design, with a new cohort aged 51-56 entering every 6 years. In total, 43,478 individuals have been interviewed to date. We used data from the 2016 and earlier waves to predict dementia and focused on the over the age of 70 years old that live in the community or in nursing homes in 2016. This left a sample of 7,165 individuals. We restricted our attention to non-Hispanic whites for comparability with ELSA. These restrictions generated a main study sample of 5,330 participants with 4,932 being self-respondents, and 398 proxy interviews.

3.A.2 ELSA Data

The English Longitudinal Study of Ageing (ELSA) is a biennial longitudinal survey of adults in England, developed as a companion study to the HRS.(of Disease, 2020) ELSA was also designed to be nationally representative of the non-institutionalised population. Respondents remain the study if they become institutionalised. While it is has been shown to be representative of the English population in terms of sociodemographic characteristics, the proportion of non-white people in the survey is very small.[12] We used data from the 2016 and earlier waves and focused on the aged over 70 years old that live in the community or in nursing homes in 2016. This left a sample of 3,224 individuals. We restricted our attention to non-Hispanic whites for comparability. These restrictions generated a study sample of 3,147 participants with 3,007 being selfrespondents, and 140 proxy interviews.

3.A.3 Harmonised Variables HRS and ELSA

We use the same set of cognitive and demographic variables available in both the HRS and ELSA.

For cognitive and demographic measures, we used the same questions as Hurd et al.(Launer, 2018) Specifically, we used variables on demographics, difficulties respondents have with Activities of Daily Living (ADLs), Instrumental Activities of Daily Living (IADLs), and a range of cognitive function tests adapted from the Telephone Interview for Cognitive Status (TICS) and their change across waves. For Proxy respondents we also used the shortened 16 question form of IQCODE.

We used four variables to measure socioeconomic status (SES). These are household income, education, total household wealth, and total household non-housing wealth. For the HRS we used variables from HRS RAND dataset. To establish each respondent's SES in terms of income, wealth, and non-housing wealth, we assigned everyone to a decile rank for each variable in their respective country. For education, we ranked respondents according to their number of years of schooling. We used respondents' reported wealth and non-housing wealth from 4 years prior. If a respondent had a missing value of income, wealth, or nonhousing wealth we used the most recent non-missing observation from previous waves. All amounts of income, wealth and non-housing wealth were deflated to 2016 GBPs, and USD amounts were converted to GBPs using 1 USD = 0.75 GBP.

Finally, in our analysis, we investigated how variables that have been associated with dementia affect our results. These include Body Mass Index (BMI), past smoking behaviour, past stroke. Respondents are asked whether they have any health conditions, of which we use whether they had any cardiometabolic diseases (diabetes, heart disease, and stroke) and/or psychiatric conditions 4 years prior. BMI is collected by a nurse for every ELSA respondent when they first enter the survey, which for everyone over 70 in the 2016 survey means their BMI was collected over 10 years prior. BMI is recorded in each wave of HRS based on self-reported height and weight. To make the BMI measures in HRS comparable with ELSA, we used respondents first ever recorded BMI in HRS, which for our sample was at least 10 years prior.

3.A.4 Hurd et al. Dementia Prediction Algorithm

To estimate the probability of dementia for each individual present in our pooled data we follow the method of Hurd et al. that is based on the HRS supplement Aging, Demographics and Memory Study (ADAMS).(Launer, 2018) It is a representative subsample of HRS members aged over 70 that received a detailed in-home assessment of their cognitive status by experienced teams at the Duke University Dementia Epidemiology Research Center.(Sonnega et al., 2014) Consensus conferences were used to establish a final diagnosis of dementia or a cognitive impairment with no dementia (CIND) for each participant. Hurd et al. applied separate ordered probit models to self- and proxy-respondents to generate a predictive algorithm, based on the ADAMS diagnoses, for the whole HRS sample, using a range of variables including demographic information, ADLs, IADLs, the TICS questionnaire, IQCODE, lagged variables and differences in variables between waves. The algorithm predicts the probability of dementia in the following year.

Most of the coefficients used in the predictive algorithm can be found in the online appendix of the Hurd et al. paper. To account for the small number of missing variables in HRS and ELSA we either used coefficients provided by the authors or estimated them by a conditional minimum distance estimator using the publicly available HRS data and the predicted values of the researcher contribution Dementia Predicted Probability Files (DPPF), that follows the method of Hurd et al. To verify our predictions in HRS, we compared our predictions with the DPPF data available on the HRS website for the year 2008 and found that we accurately matched predicted dementia.

Hurd et al's predictive algorithm has been shown to have high specificity and accuracy, and generally performs well in terms of sensitivity compared to other predictive algorithms for dementia.(Steptoe et al., 2012)

Table 3.4: HRS and ELSA Response Rates

	Year	2010	2012	2014	2016
HRS					
	Wave	10	11	12	13
	Response rate (%)	81.0	89.1	87.1	84.3
ELSA					
	Wave	5	6	7	8
	Fieldwork response rate (%)	79.1	80.2	80.1	82.4

Notes: The response rates are calculated differently in HRS and ELSA. For the HRS, the response rate includes all individuals who were determined to be eligible for HRS who completed a baseline interview.(Gianattasio et al., 2018) For ELSA, the fieldwork response rate is the proportion of eligible survey units who participate in the research study, where 'eligible' means not having been found to be ineligible through death or moving out of Great Britain.(Jorm et al., 2000)

Table 3.5: Summary Statistics, Raw Sample, England vs United States

	Raw Full Sample		With Population Weights	
	Non-Standardized Mean		Non-Standardized Mean	
	England	USA	England	USA
Total Sample Size	3,147	5,330	3,147	5,330
Proxy Respondents (%)	0.044	0.075	0.050	0.072
Age	78.8	80.5	79.5	79.7
Female	0.551	0.590	0.553	0.561
Married	0.580	0.517	0.584	0.536
Current Income and Wealth (£)				
Income	21,070	28,800	20,215	30,240
Wealth	280,000	210,000	270,000	230,000
Non-Housing Wealth	40,000	88,275	36,000	96,150
Wealth 4 Years Prior (£)				
Wealth	270,000	220,000	250,000	230,000
Non-Housing Wealth	42,300	96,800	35,727	100,000
Education				
Less than High School	0.362	0.174	0.424	0.161
High-school or Some College	0.473	0.573	0.444	0.560
College	0.165	0.253	0.132	0.279
Current Health				
ADLs (Out of 6)	0.510	0.583	0.561	0.548
IADLs (Out of 5)	0.348	0.513	0.411	0.482
Arthritis	0.569	0.728	0.578	0.715
Cancer	0.190	0.251	0.189	0.244
Lung disease	0.117	0.135	0.118	0.127
Diabetes	0.162	0.246	0.166	0.240
Heart Disease	0.390	0.383	0.396	0.365
High Blood Pressure	0.576	0.684	0.583	0.671
Psychiatric Condition	0.134	0.171	0.132	0.175
Stroke	0.092	0.103	0.096	0.095
Health 4 Years Prior				
ADLs	0.346	0.274	0.381	0.267
IADLs	0.159	0.228	0.185	0.219
Diabetes	0.131	0.215	0.135	0.212
Heart Disease	0.309	0.316	0.312	0.303
High Blood Pressure	0.513	0.653	0.519	0.643
Psychiatric Condition	0.120	0.157	0.118	0.162
Stroke	0.063	0.074	0.065	0.067
Other Health Factors				
Ever Smoked	0.660	0.553	0.674	0.561
BMI: <20 (Underweight)	0.016	0.038	0.015	0.036
BMI: 20-24.9 (Normal Weight)	0.266	0.352	0.262	0.343
BMI: 25-29.9 (Overweight)	0.461	0.405	0.461	0.410
BMI: 30+ (Obese)	0.256	0.205	0.262	0.211

Table 3.6: Summary Statistics, Age-Gender Standardized, England vs United States

	Full Sample		Lowest Income Decile	
	Age-Gender Standardized Mean		Age-Gender Standardized Mean	
	England	USA	England	USA
Age	79.5	79.3	79.7	79.7
Female	0.553	0.553	0.553	0.553
Married	0.584	0.547	0.148	0.170
Current Income and Wealth (£)				
Income	20,547	31,927	8,247	8,645
Wealth	271,511	236,431	120,261	46,762
Non-Housing Wealth	38,715	100,996	7,778	5,509
Wealth 4 Years Prior (£)				
Wealth	257,459	242,655	111,830	55,953
Non-Housing Wealth	40,604	110,530	6,476	13,724
Education				
Less than High School	0.424	0.160	0.619	0.323
High-school or Some College	0.444	0.559	0.318	0.556
College	0.132	0.281	0.063	0.122
Current Health				
ADLs (Out of 6)	0.561	0.523	0.559	0.827
IADLs (Out of 5)	0.411	0.457	0.421	0.679
Arthritis	0.578	0.715	0.617	0.750
Cancer	0.189	0.244	0.157	0.196
Lung disease	0.118	0.127	0.137	0.217
Diabetes	0.166	0.242	0.157	0.283
Heart Disease	0.396	0.364	0.402	0.427
High Blood Pressure	0.583	0.671	0.619	0.703
Psychiatric Condition	0.132	0.174	0.158	0.262
Stroke	0.096	0.094	0.111	0.135
Health 4 Years Prior				
ADLs	0.381	0.255	0.424	0.546
IADLs	0.185	0.205	0.218	0.393
Diabetes	0.135	0.214	0.131	0.254
Heart Disease	0.312	0.302	0.345	0.352
High Blood Pressure	0.519	0.640	0.559	0.662
Psychiatric Condition	0.118	0.162	0.142	0.271
Stroke	0.065	0.066	0.064	0.099
Other Health Factors				
Ever Smoked	0.674	0.565	0.741	0.594
BMI: <20 (Underweight)	0.015	0.036	0.025	0.032
BMI: 20-24.9 (Normal Weight)	0.261	0.342	0.361	0.364
BMI: 25-29.9 (Overweight)	0.461	0.410	0.397	0.339
BMI: 30+ (Obese)	0.262	0.212	0.216	0.265

Table 3.7: Summary Statistics for Predictor Variables, England vs United States

	Full Sample Age-Gender- Standardized Mean		Lowest Income Decile Age-Gender- Standardized Mean	
	England	USA	England	USA
Cognitive Scores				
Dates (Out of 4)	3.663	3.387	3.573	3.198
Backward counting 20	0.912	0.875	0.935	0.830
Serial 7 (Out of 5)	3.787	3.364	3.667	2.580
Scissor	0.978	0.920	0.993	0.912
Cactus	0.915	0.891	0.913	0.842
PM/Vice-President	0.846	0.900	0.828	0.854
Immediate recall (Out of 10)	5.318	4.631	5.010	3.972
Delayed recall (Out of 10)	3.797	3.684	3.185	2.894
Limitations				
ADLs (Out of 6)	0.561	0.523	0.559	0.827
IADLs (Out of 5)	0.411	0.457	0.421	0.679
Proxy Respondent	0.050	0.068	0.038	0.075

Notes: Sample includes non-Hispanic white population aged 70+ only. The sample size is 3,147 participants in England and 5,330 participants in the United States. The Cognitive Scores are from a range of tests adapted from Telephone Interview for Cognitive Status (TICS). Only non-proxy respondents answered the TICS questions. ADL and IADLs are Activities of Daily Living and Instrumental Activities of Daily Living, respectively.

Table 3.8: Indexes of Inequality

	Slope Index of Inequality (SII)			Relative Index of Inequality (RII)		
	England	USA	Difference	England	USA	Difference
Income	-0.062	-0.085	-0.023	1.946	2.341	1.203
	[-0.097 to -0.028]	[-0.114 to -0.057]	(0.31)	[1.333 to 2.841]	[1.774 to 3.089]	(0.44)
Education	-0.087	-0.099	-0.011	3.001	2.820	0.940
	[-0.120 to -0.055]	[-0.126 to -0.072]	(0.56)	[2.013 to 4.474]	[2.131 to 3.731]	(0.80)
Wealth	-0.069	-0.096	-0.026	2.157	2.730	1.265
	[-0.104 to -0.035]	[-0.123 to -0.069]	(0.24)	[1.467 to 3.174]	[2.070 to 3.601]	(0.33)
NH Wealth	-0.071	-0.093	-0.022	2.213	2.311	1.044
	[-0.106 to -0.035]	[-0.124 to -0.062]	(0.36)	[1.463 to 3.348]	[1.748 to 3.056]	(0.87)
Excluding Most Disadvantaged Group						
Income	-0.067	-0.060	0.007	2.104	1.833	0.871
	[-0.107 to -0.027]	[-0.093 to -0.027]	(0.79)	[1.343 to 3.296]	[1.313 to 2.560]	(0.63)
Education	-0.070	-0.062	0.008	2.509	1.932	0.770
	[-0.107 to -0.032]	[-0.094 to -0.030]	(0.76)	[1.548 to 4.065]	[1.378 to 2.708]	(0.39)
Wealth	-0.065	-0.059	0.005	2.092	1.887	0.902
	[-0.104 to -0.025]	[-0.091 to -0.028]	(0.84)	[1.319 to 3.317]	[1.349 to 2.639]	(0.72)
NH Wealth	-0.064	-0.075	-0.012	2.077	1.981	0.954
	[-0.107 to -0.020]	[-0.110 to -0.040]	(0.69)	[1.217 to 3.544]	[1.433 to 2.740]	(0.88)
Full Sample with Controls						
Income	-0.040	-0.051	-0.011	1.499	1.698	1.133
	[-0.075 to -0.004]	[-0.079 to -0.023]	(0.63)	[1.028 to 2.187]	[1.275 to 2.263]	(0.61)
Education	-0.0336	-0.056	-0.022	1.308	1.742	1.332
	[-0.0754 to 0.008]	[-0.086 to -0.026]	(0.40)	[0.899 to 1.903]	[1.306 to 2.324]	(0.24)
Wealth	-0.060	-0.045	0.015	1.777	1.483	0.834
	[-0.097 to -0.023]	[-0.074 to -0.015]	(0.53)	[1.224 to 2.582]	[1.131 to 1.945]	(0.44)
NH Wealth	-0.050	-0.051	-0.002	1.562	1.614	1.033
	[-0.087 to -0.012]	[-0.080 to -0.023]	(0.94)	[1.078 to 2.262]	[1.231 to 2.116]	(0.89)

Notes: This table shows the Slope Index of Inequality (SII) and Relative Index of Inequality (RII), calculated using generalized linear models (log binomial regression) with identity and logarithmic link functions, respectively. The SIIs are displayed graphically in Figure 1. The estimates of RII are risk ratios. P-values are displayed in parentheses under differences. The panel ”Excluding Most Disadvantaged Group” excludes individuals in the bottom income decile. The panel ”Full Sample with Controls” controls for whether an individual had any cardiometabolic diseases (diabetes, heart disease, and stroke) and/or psychiatric conditions 4 years prior, whether an individual has ever smoked, and Body Mass Index (BMI) from at least 10 years prior.

Table 3.9: Difference in Prevalence of Dementia with Controls, United States vs England, 2017

Whole Sample:					
England Prevalence	0.097	0.088	0.086	0.089	0.092
	(0.089 to 0.106)	(0.078 to 0.098)	(0.073 to 0.099)	(0.072 to 0.106)	(0.072 to 0.111)
United States	0.112	0.100	0.098	0.101	0.120
	(0.106 to 0.118)	(0.092 to 0.108)	0.087 to 0.109)	(0.087 to 0.114)	(0.104 to 0.135)
Percentage Point Difference	1.43	1.15	1.21	1.18	2.82
p-value	0.0055	0.025	0.020	0.034	<0.0001
Excluding Lowest Income Decile:					
England Prevalence	0.094	0.086	0.086	0.090	0.094
	(0.085 to 0.103)	(0.076 to 0.096)	(0.073 to 0.1)	(0.073 to 0.108)	(0.074 to 0.114)
United States	0.103	0.093	0.094	0.098	0.115
	(0.097 to 0.109)	(0.085 to 0.102)	(0.083 to 0.105)	(0.084 to 0.113)	(0.099 to 0.131)
Percentage Point Difference	0.89	0.74	0.75	0.81	2.09
p-value	0.11	0.18	0.18	0.18	0.0021
Lowest Income Decile:					
England Prevalence	0.114	0.107	0.088	0.082	0.095
	(0.089 to 0.139)	(0.074 to 0.14)	(0.036 to 0.14)	(0.021 to 0.144)	(0.025 to 0.164)
United States	0.187	0.162	0.148	0.131	0.163
	(0.166 to 0.208)	(0.129 to 0.195)	(0.103 to 0.192)	(0.078 to 0.184)	(0.103 to 0.223)
Percentage Point Difference	7.27	5.45	5.93	4.85	6.85
p-value	<0.0001	0.0020	0.0011	0.012	0.0015
Control for:					
Past Health Conditions		✓	✓	✓	✓
Ever Smoked			✓	✓	✓
Past BMI				✓	✓
Education					✓

Notes: Sample includes non-Hispanic white population aged 70+ only. The sample size is 3,147 participants in England and 5,330 participants in the United States. All estimates are age-gender standardized to the overall 2016 aged 70+ white population in England. The difference is calculated as the prevalence in the US minus prevalence in England. Differences are displayed as percentage points. 'Past Health Conditions' controls for whether an individual had any cardiometabolic diseases (diabetes, heart disease, and stroke) and/or psychiatric conditions 4 years prior. 'Ever Smoked' controls for whether an individual has ever smoked. BMI stands for Body Mass Index. 'Past BMI' includes dummy variables to control for whether an individual is classed as underweight, normal weight, overweight or obese. BMI measurements are based on when an individual first entered the survey, which is at least 10 years prior. We assume the effect of each control is constant across countries and age/gender groups. The reference groups are: No past health conditions; never smoked; normal weight; and high-school education.

Table 3.10: Difference in Prevalence of Dementia with Controls, by Income Decile

	England			United States		
Whole Sample:						
Prevalence	0.097	0.092	0.089	0.112	0.105	0.101
CI	(0.089-0.106)	(0.082-0.102)	(0.072-0.106)	(0.106-0.118)	(0.098-0.113)	(0.087-0.114)
Income decile:						
1 Prevalence	0.114	0.112	0.082	0.187	0.173	0.131
	(0.089-0.139)	(0.079-0.144)	(0.021-0.144)	(0.166-0.208)	(0.142-0.205)	(0.078-0.184)
2 Prevalence	0.113	0.119	0.118	0.141	0.145	0.148
	(0.09-0.136)	(0.088-0.149)	(0.05-0.187)	(0.119-0.163)	(0.113-0.176)	(0.086-0.21)
3 Prevalence	0.124	0.107	0.130	0.111	0.089	0.093
	(0.097-0.151)	(0.074-0.139)	(0.073-0.187)	(0.095-0.127)	(0.067-0.11)	(0.048-0.137)
4 Prevalence	0.099	0.095	0.094	0.118	0.114	0.111
	(0.071-0.126)	(0.065-0.126)	(0.041-0.148)	(0.099-0.137)	(0.088-0.14)	(0.062-0.16)
5 Prevalence	0.094	0.087	0.061	0.086	0.084	0.063
	(0.072-0.116)	(0.062-0.112)	(0.024-0.098)	(0.069-0.102)	(0.063-0.104)	(0.035-0.09)
6 Prevalence	0.098	0.097	0.116	0.108	0.107	0.100
	(0.07-0.127)	(0.064-0.131)	(0.056-0.177)	(0.088-0.128)	(0.081-0.134)	(0.053-0.148)
7 Prevalence	0.068	0.066	0.073	0.100	0.101	0.104
	(0.042-0.093)	(0.038-0.094)	(0.021-0.124)	(0.078-0.122)	(0.073-0.129)	(0.054-0.154)
8 Prevalence	0.083	0.084	0.097	0.082	0.078	0.085
	(0.053-0.114)	(0.051-0.117)	(0.054-0.14)	(0.066-0.097)	(0.058-0.098)	(0.054-0.117)
9 Prevalence	0.082	0.076	0.100	0.093	0.085	0.107
	(0.041-0.122)	(0.034-0.118)	(0.048-0.151)	(0.071-0.116)	(0.06-0.111)	(0.07-0.145)
10 Prevalence	0.059	0.058	0.051	0.077	0.077	0.068
	(0.034-0.083)	(0.029-0.087)	(0.015-0.087)	(0.052-0.103)	(0.051-0.103)	(0.036-0.1)
Control for:						
Past Health Conditions		✓	✓		✓	✓
Ever Smoked			✓			✓
Past BMI			✓			✓

Notes: Sample includes non-Hispanic white population aged 70+ only. The sample size is 3,147 participants in England and 5,330 participants in the United States. All estimates are age-gender standardized to the overall 2016 aged 70+ white population in England. 'Past Health Conditions' controls for whether an individual had any cardiometabolic diseases (diabetes, heart disease, and stroke) and/or psychiatric conditions 4 years prior. 'Ever Smoked' controls for whether an individual has ever smoked. BMI stands for Body Mass Index. 'Past BMI' includes dummy variables to control for whether an individual is classed as underweight, normal weight, overweight or obese. BMI measurements are based on when an individual first entered the survey, which is at least 10 years prior. We assume the effect of each control is constant across countries and age/gender groups. The reference groups are: No past health conditions; never smoked; and normal weight.

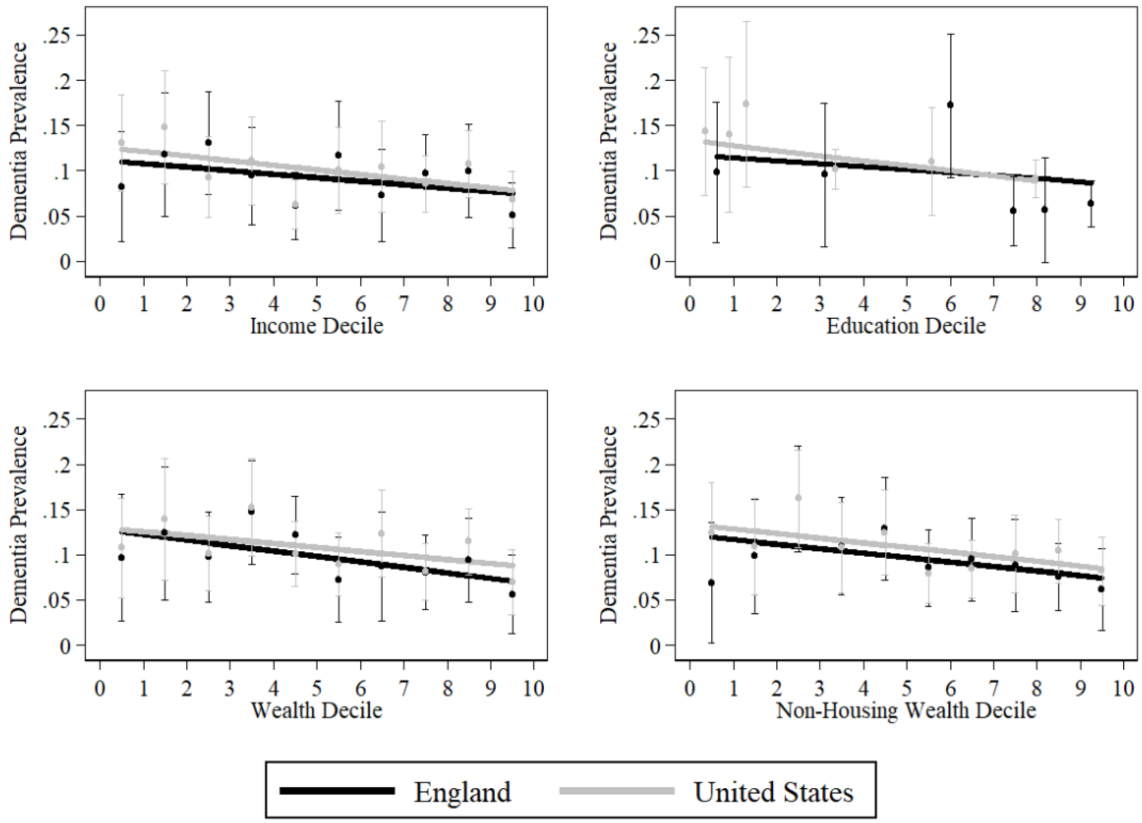


Figure 3.2: SES Gradient of Dementia, US vs England, Controlling for Past Health

Notes: The points in this figure represent the mean age-gender standardized dementia prevalence for each country by socioeconomic status (SES) after controlling for past health and behaviour, along with 95% confidence interval for these predictions. The solid lines represent the fitted Socioeconomic Index of Inequality (SII) for each country.

Table 3.11: Prevalence of Dementia, United States vs. England, Full-Sample

	England Age-gender standardised prevalence (95%CI)	USA Age-gender standardised prevalence (95%CI)	Difference	P value
All	0.099 [0.091,0.108]	0.124 [0.118,0.129]	0.024	<0.0001
Household income decile				
1 (Lowest)	0.116 [0.091,0.140]	0.211 [0.192,0.230]	0.095	<0.0001
2	0.126 [0.101,0.152]	0.161 [0.144,0.178]	0.035	0.025
3	0.124 [0.098,0.150]	0.141 [0.120,0.162]	0.017	0.32
4	0.100 [0.072,0.127]	0.143 [0.124,0.162]	0.044	0.0094
5	0.101 [0.078,0.124]	0.097 [0.082,0.111]	-0.004	0.76
6	0.102 [0.072,0.131]	0.104 [0.086,0.121]	0.002	0.91
7	0.061 [0.043,0.079]	0.108 [0.088,0.127]	0.047	0.0006
8	0.088 [0.057,0.118]	0.085 [0.069,0.102]	-0.002	0.91
9	0.081 [0.041,0.122]	0.089 [0.070,0.108]	0.008	0.74
10 (Highest)	0.057 [0.034,0.080]	0.076 [0.055,0.098]	0.020	0.22
Years of schooling				
9 or less	0.128 [0.103,0.153]	0.212 [0.195,0.230]	0.084	<0.0001
10	0.099 [0.078,0.119]	0.143 [0.118,0.167]	0.044	0.0072
11	0.097 [0.079,0.116]	0.119 [0.093,0.144]	0.021	0.19
12	0.069 [0.041,0.097]	0.134 [0.124,0.143]	0.064	<0.0001
13	0.063 [0.040,0.085]	0.118 [0.095,0.141]	0.055	0.00071
14 or more	0.055 [0.039,0.071]	0.086 [0.078,0.094]	0.031	0.00071
Household wealth decile				
1 (Lowest)	0.172 [0.136,0.207]	0.219 [0.198,0.241]	0.047	0.024
2	0.118 [0.093,0.144]	0.169 [0.151,0.188]	0.051	0.0014
3	0.105 [0.078,0.132]	0.138 [0.120,0.155]	0.032	0.049
4	0.090 [0.068,0.111]	0.120 [0.105,0.135]	0.030	0.023
5	0.118 [0.093,0.143]	0.116 [0.100,0.132]	-0.002	0.89
6	0.077 [0.056,0.099]	0.097 [0.084,0.110]	0.020	0.13
7	0.073 [0.047,0.098]	0.092 [0.076,0.107]	0.019	0.21
8	0.093 [0.067,0.118]	0.103 [0.086,0.120]	0.010	0.52
9	0.080 [0.047,0.113]	0.101 [0.084,0.118]	0.021	0.27
10 (Highest)	0.060 [0.040,0.081]	0.074 [0.058,0.089]	0.013	0.30
Household non-housing wealth decile				
1 (Lowest)	0.147 [0.113,0.180]	0.215 [0.195,0.235]	0.068	0.00055
2	0.114 [0.089,0.140]	0.184 [0.165,0.203]	0.070	<0.0001
3	0.121 [0.095,0.148]	0.148 [0.129,0.168]	0.027	0.11
4	0.100 [0.076,0.125]	0.129 [0.112,0.147]	0.029	0.054
5	0.110 [0.084,0.135]	0.104 [0.088,0.119]	-0.006	0.70
6	0.093 [0.067,0.118]	0.099 [0.085,0.113]	0.006	0.66
7	0.098 [0.072,0.124]	0.085 [0.071,0.098]	-0.013	0.37
8	0.079 [0.052,0.106]	0.090 [0.075,0.105]	0.011	0.48
9	0.057 [0.035,0.078]	0.094 [0.077,0.110]	0.037	0.0076
10 (Highest)	0.065 [0.045,0.085]	0.083 [0.067,0.099]	0.018	0.15

Table 3.12: Difference in Prevalence of Dementia with Controls, Full Sample

Whole Sample:					
England Prevalence	0.099	0.088	0.088	0.084	0.085
	[0.091 to 0.108]	[0.079 to 0.098]	[0.076 to 0.1]	[0.068 to 0.1]	[0.068 to 0.103]
United States	0.124	0.107	0.107	0.105	0.123
	[0.118 to 0.129]	[0.1 to 0.114]	[0.097 to 0.117]	[0.093 to 0.118]	[0.109 to 0.137]
Percentage Point Difference	2.44	1.87	1.90	2.14	3.80
p-value	<0.0001	<0.0001	0.00015	0.00011	<0.0001
Excluding Lowest Income Decile:					
England Prevalence	0.096	0.086	0.086	0.087	0.090
	[0.087 to 0.105]	[0.076 to 0.096]	[0.073 to 0.099]	[0.071 to 0.103]	[0.072 to 0.108]
United States	0.113	0.098	0.099	0.100	0.117
	[0.107 to 0.118]	[0.091 to 0.106]	[0.089 to 0.108]	[0.086 to 0.113]	[0.103 to 0.132]
Percentage Point Difference	1.64	1.25	1.27	1.24	2.70
p-value	0.0024	0.020	0.019	0.030	<0.0001
Lowest Income Decile:					
England Prevalence	0.116	0.107	0.106	0.078	0.099
	[0.091 to 0.140]	[0.077 to 0.138]	[0.062 to 0.15]	[0.026 to 0.13]	[0.036 to 0.162]
United States	0.211	0.187	0.186	0.161	0.195
	[0.192 to 0.230]	[0.16 to 0.214]	[0.149 to 0.223]	[0.117 to 0.206]	[0.138 to 0.252]
Percentage Point Difference	9.54	7.93	7.96	8.37	9.60
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Control for:					
Past Health Conditions		✓	✓	✓	✓
Ever Smoked			✓	✓	✓
Past BMI				✓	✓
Education					✓

Notes: Full sample, population aged 70+, including minorities. The sample sizes are 3,224 participants in England, and 7,165 participants in the United States. All estimates are age-gender standardized to the overall 2016 aged 70+ white population in England. The difference is calculated as the prevalence in the US minus prevalence in England. Differences are displayed as percentage points.'Past Health Conditions' controls for whether an individual had any cardiometabolic diseases (diabetes, heart disease, and stroke) and/or psychiatric conditions 4 years prior. 'Ever Smoked' controls for whether an individual has ever smoked. BMI stands for Body Mass Index. 'Past BMI' includes dummy variables to control for whether an individual is classed as underweight, normal weight, overweight or obese. BMI measurements are based on when an individual first entered the survey, which is at least 10 years prior. We assume the effect of each control is constant across countries and age/gender groups. The reference groups are: No past health conditions; never smoked; normal weight; and high-school education.

Table 3.13: Prevalence of Dementia in United States, Whites versus Non-whites

	England Age-gender standardised prevalence (95%CI)	USA Age-gender standardised prevalence (95%CI)	Difference	P value
All	0.182 [0.169 to 0.195]	0.112 [0.106 to 0.118]	-0.070	<0.0001
Household income decile				
1 (Lowest)	0.224 [0.199 to 0.249]	0.204 [0.178 to 0.231]	-0.020	0.28
2	0.21 [0.180 to 0.243]	0.143 [0.123 to 0.163]	-0.068	0.00034
3	0.22 [0.158 to 0.282]	0.121 [0.102 to 0.140]	-0.099	0.0023
4	0.202 [0.165 to 0.240]	0.133 [0.111 to 0.155]	-0.069	0.0018
5	0.116 [0.076 to 0.156]	0.094 [0.079 to 0.109]	-0.022	0.31
6	0.164 [0.097 to 0.230]	0.098 [0.081 to 0.116]	-0.065	0.065
7	0.129 [0.080 to 0.178]	0.107 [0.086 to 0.128]	-0.022	0.42
8	0.10 [0.047 to 0.166]	0.084 [0.067 to 0.101]	-0.023	0.47
9	0.072 [0.026 to 0.118]	0.090 [0.069 to 0.110]	0.017	0.50
10 (Highest)	0.084 [0.038 to 0.129]	0.077 [0.055 to 0.099]	-0.006	0.81
Years of schooling				
9 or less	0.235 [0.212 to 0.258]	0.190 [0.162 to 0.218]	-0.045	0.015
10	0.172 [0.127 to 0.218]	0.137 [0.109 to 0.165]	-0.035	0.19
11	0.144 [0.109 to 0.179]	0.109 [0.080 to 0.139]	-0.035	0.14
12	0.198 [0.170 to 0.227]	0.124 [0.114 to 0.133]	-0.074	<0.0001
13	0.159 [0.101 to 0.218]	0.116 [0.090 to 0.141]	-0.043	0.18
14 or more	0.099 [0.079 to 0.119]	0.085 [0.076 to 0.093]	-0.015	0.20
Household wealth decile				
1 (Lowest)	0.228 [0.201 to 0.255]	0.208 [0.176 to 0.240]	-0.020	0.36
2	0.208 [0.179 to 0.237]	0.154 [0.131 to 0.177]	-0.054	0.0043
3	0.184 [0.142 to 0.227]	0.120 [0.102 to 0.138]	-0.064	0.0063
4	0.142 [0.111 to 0.174]	0.115 [0.098 to 0.132]	-0.027	0.13
5	0.159 [0.117 to 0.202]	0.108 [0.092 to 0.124]	-0.051	0.027
6	0.120 [0.084 to 0.155]	0.093 [0.080 to 0.107]	-0.026	0.18
7	0.156 [0.086 to 0.226]	0.088 [0.072 to 0.103]	-0.068	0.063
8	0.197 [0.134 to 0.259]	0.099 [0.081 to 0.116]	-0.098	0.0031
9	0.122 [0.065 to 0.178]	0.099 [0.082 to 0.117]	-0.022	0.46
10 (Highest)	0.045 [0.018 to 0.073]	0.075 [0.059 to 0.090]	0.029	0.071
Household non-housing wealth decile				
1 (Lowest)	0.231 [0.207 to 0.256]	0.197 [0.167 to 0.227]	-0.034	0.085
2	0.204 [0.177 to 0.231]	0.177 [0.151 to 0.203]	-0.027	0.16
3	0.187 [0.142 to 0.232]	0.135 [0.114 to 0.155]	-0.052	0.036
4	0.158 [0.128 to 0.188]	0.123 [0.104 to 0.143]	-0.034	0.059
5	0.106 [0.069 to 0.143]	0.102 [0.086 to 0.118]	-0.004	0.85
6	0.099 [0.063 to 0.134]	0.098 [0.084 to 0.113]	0.000	0.99
7	0.124 [0.066 to 0.182]	0.083 [0.069 to 0.097]	-0.040	0.18
8	0.140 [0.069 to 0.210]	0.088 [0.073 to 0.103]	-0.051	0.16
9	0.127 [0.054 to 0.201]	0.092 [0.075 to 0.108]	-0.035	0.36
10 (Highest)	0.069 [0.050 to 0.088]	0.084 [0.068 to 0.101]	0.016	0.23

Bibliography

Banks, J., M. Marmot, Z. Oldfield, and J. P. Smith (2006, 05). Disease and Disadvantage in the United States and in England. *JAMA 295*(17), 2037–2045.

Banks, J., A. Muriel, and J. Smith (2010, 01). Attrition and health in ageing studies: Evidence from elsa and hrs. *Institute for the Study of Labor (IZA), IZA Discussion Papers 2*.

Brunner, E. (2016, 05). Social factors and cardiovascular morbidity. *Neuroscience & Biobehavioral Reviews 74*.

Cadar, D., C. Lassale, H. Davies, D. J. Llewellyn, G. D. Batty, and A. Steptoe (2018, 07). Individual and Area-Based Socioeconomic Factors Associated With Dementia Incidence in England: Evidence From a 12-Year Follow-up in the English Longitudinal Study of Ageing. *JAMA Psychiatry 75*(7), 723–732.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016, 04). The association between income and life expectancy in the united states, 2001-2014. *JAMA 315*.

De Nardi, M., E. French, J. Jones, and J. McCauley (2016, 09). Medical spending of the us elderly. *Fiscal Studies 37*, 717–747.

George, K. M., P. L. Lutsey, A. Kucharska-Newton, P. Palta, G. Heiss, T. Osypuk, and A. R. Folsom (2020, 05). Life-Course Individual and Neighborhood Socioeconomic Status and Risk of Dementia in the Atherosclerosis Risk in Communities Neurocognitive Study. *American Journal of Epidemiology 189*(10), 1134–1142.

Gianattasio, K., Q. Wu, M. Glymour, and M. Power (2018, 11). Comparison of methods for algorithmic classification of dementia status in the health and retirement study. *Epidemiology 30*, 1.

Hudomiet, P., M. D. Hurd, and S. Rohwedder (2019). The relationship between lifetime out-of-pocket medical expenditures, dementia, and socioeconomic status in the U.S. *The Journal of the Economics of Ageing 14*(C).

Hurd, M. D., P. Martorell, A. Delavande, K. J. Mullen, and K. M. Langa (2013). Monetary costs of dementia in the united states. *New England Journal of Medicine 368*(14), 1326–1334. PMID: 23550670.

Jorm, A., H. Christensen, A. Korten, P. Jacomb, and S. Henderson (2000, 07). Informant ratings of cognitive decline in old age: Validation against change on cognitive tests over 7 to 8 years. *Psychological medicine 30*, 981–5.

Kuo, C.-L., L. C. Pilling, J. L. Atkins, J. A. H. Masoli, J. Delgado, G. A. Kuchel, and D. Melzer (2020, 07). ApoE e4e4 Genotype and Mortality With COVID-19 in UK Biobank. *The Journals of Gerontology: Series A 75*(9), 1801–1803.

Langa, K. M., E. B. Larson, E. M. Crimmins, J. D. Faul, D. A. Levine, M. U. Kabeto, and D. R. Weir (2017, 01). A Comparison of the Prevalence of Dementia in the United States in 2000 and 2012. *JAMA Internal Medicine 177*(1), 51–58.

Langa, K. M., L. H. Ryan, R. J. McCammon, R. N. Jones, J. J. Manly, D. A. Levine, A. Sonnega, M. Farron, and D. R. Weir (2020, 2019). The health and retirement study harmonized cognitive assessment protocol project: Study design and methods. *Neuroepidemiology*.

Launer, L. (2018, 11). Statistics on the burden of dementia: need for stronger data. *The Lancet Neurology 18*.

Marmot, M. (2020). Health equity in england: the marmot review 10 years on. *BMJ 368*.

Marmot, M. G. (2006, 03). Status SyndromeA Challenge to Medicine. *JAMA 295*(11), 1304–1307.

Nguyen, T. T., E. J. T. Tchetgen, I. Kawachi, S. E. Gilman, S. Walter, S. Y. Liu, J. J. Manly, and M. M. Glymour (2016, 2016 Jan). Instrumental variable approaches to identifying the causal effect of educational attainment on dementia risk. *Ann Epidemiol 26*, 71–6.e1–3. Times Cited: 0 0.

Nichols, E., C. Szoeke, S. Vollset, N. Abbasi, F. Abd-Allah, J. Abdela, M. Aichour, R. Akinyemi, F. Alahdab, S. Asgedom, A. Awasthi, S. Barker-Collo, B. Baune, Y. Béjot, A. Belachew, D. Bennett, B. Biadgo, A. Bijani, M. Bin Sayeed, C. Brayne, D. Carpenter, F. Carvalho, F. Catalá-López, E. Cerin, J. Choi, A. Dang, M. Degefa, S. Djalalinia, M. Dubey, E. Duken, D. Edvardsson, M. Endres, S. Eskandarieh, A. Faro, F. Farzadfar, S. Fereshtehnejad, E. Fernandes, I. Filip, F. Fischer, A. Gebre, D. Geremew, M. Ghasemi-Kasman, E. Gnedovskaya, R. Gupta, V. Hachinski, T. Hagos, S. Hamidi, G. Hankey, J. Haro, S. Hay, S. Irvani, R. Jha, J. Jonas, R. Kalani, A. Karch, A. Kasaeian, Y. Khader, I. Khalil, E. Khan, T. Khanna, T. Khoja, J. Khubchandani, A. Kisa, K. Kissimova-Skarbek, M. Kivimäki, A. Koyanagi, K. Krohn, G. Logroscino, S. Lorkowski, M. Majdan, R. Malekzadeh, W. März, J. Massano, G. Mengistu, A. Meretoja, M. Mohammadi, M. Mohammadi-Khanaposhtani, A. Mokdad, S. Mondello, G. Moradi, G. Nagel, M. Naghavi, G. Naik, L. Nguyen, T. Nguyen, Y. Nirayo, M. Nixon, R. Ofori-Asenso, F. Ogbo, A. Olagunju, M. Owolabi, S. Panda-Jonas, V. Passos, D. Pereira, G. Pinilla-Monsalve, M. Piradov, C. Pond, H. Poustchi, M. Qorbani, A. Radfar, R. Reiner, S. Robinson, G. Roshandel, A. Rostami, T. Russ, P. Sachdev, H. Safari, S. Safiri, R. Sahathevan, Y. Salimi, M. Satpathy, M. Sawhney, M. Saylan, S. Sepanlou, A. Shafieesabet, M. Shaikh, M. Sahraian, M. Shigematsu, R. Shiri, I. Shiue, J. Silva, M. Smith, S. Sobhani, D. Stein, R. Tabarés-Seisdedos, M. Tovani-Palone, B. Tran, T. Tran, A. Tsegay, I. Ullah, N. Venketasubramanian, V. Vlassov, Y. Wang, J. Weiss, R. Westerman, T. Wijeratne, G. Wyper, Y. Yano, E. Yimer, N. Yonemoto, M. Yousefifard, Z. Zaidi, Z. Zare, T. Vos, V. Feigin, and C. Murray (2019, January). Global, regional, and national burden of alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurology 18*(1), 88–106.

of Disease, G. B. (2020, 07). Global Burden of Disease. GBD Compare. IHME Viz Hub. GBD Comp. Viz Hub. https://vizhub.healthdata.org/gbd-compare/.

Rocca, W. A., R. C. Petersen, D. S. Knopman, L. E. Hebert, D. A. Evans, K. S. Hall, S. Gao, F. W. Unverzagt, K. M. Langa, E. B. Larson, and L. R. White (2011). Trends in the incidence and prevalence of alzheimer's disease, dementia, and cognitive impairment in the united states. *Alzheimer's & Dementia 7*(1), 80–93. Prevalence and Trends of Alzheimer's Disease and Other Age-Related Cognitive Impairment in the United States.

Rusmaully, J., A. Dugravot, J.-P. Moatti, M. G. Marmot, A. Elbaz, M. Kivimaki, S. Sabia, and A. Singh-Manoux (2017, 06). Contribution of cognitive performance and cognitive decline to associations between socioeconomic factors and dementia: A cohort study. *PLOS Medicine 14*(6), 1–17.

Sonnega, A., J. D. Faul, M. B. Ofstedal, K. M. Langa, J. W. Phillips, and D. R. Weir (2014, 03). Cohort Profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology 43*(2), 576–585.

Steptoe, A., E. Breeze, J. Banks, and J. Nazroo (2012, 11). Cohort Profile: The English Longitudinal Study of Ageing. *International Journal of Epidemiology 42*(6), 1640–1648.

Stoye, G., B. Zaranko, M. Shipley, M. McKee, and E. J. Brunner (2020). Educational inequalities in hospital use among older adults in england, 2004-2015. *The Milbank Quarterly 98*(4), 1134–1170.

Vega, I., L. Cabrera, C. Wygant, D. Velez Ortiz, and S. Counts (2017, 05). Alzheimer's disease in the latino community: Intersection of genetics and social determinants of health. *Journal of Alzheimer's disease : JAD 58*.

Wagstaff, A., P. Paci, and E. van Doorslaer (1991). On the measurement of inequalities in health. *Social Science & Medicine 33*(5), 545–557.

Yaffe, K., C. Falvey, T. B. Harris, A. Newman, S. Satterfield, A. Koster, H. Ayonayon, and E. Simonsick (2013). Effect of socioeconomic disparities on incidence of dementia among biracial older adults: prospective study. *BMJ 347*.

Chapter 4

Transition Costs and the Optimal U.S. Social Security System

4.1 Introduction

The role of the U.S. Social Security system is to provide insurance by pooling risk between and within generations. However, this potentially valuable risk sharing comes at the cost of economic distortions. The balance between insurance and distortions by means of a partial privatization is a topic of active debate in the literature[1]. On the one hand, a partially privatized (mixed) system will increase output, savings, long term wages and ameliorate labour market distortions. On the other, a partial privatization will entail transition costs that will be paid by generations alive at or after the reform and households will be exposed to more risk.

Previous research suggests that the long-term optimal U.S. Social Security system is smaller than the current system (Feldstein, 1985; Imrohoroglu et al., 1995; Krueger and Kubler, 2006; Hong and Rios-Rull, 2007; Bagchi, 2015)[2]. However, transitioning from the current unfunded system to the long-term optimum entails transition costs that are large enough to potentially make the reform welfare decreasing (Nishiyama and Smetters, 2007; Olovsson, 2010). These costs arise due to the repayment of the unfunded liability and the post-reform capital accumulation that will reduce consumption and increase hours of work over the transition. These costs also impose political constraints on the government, as the majority of households alive at the time of the reform may find themselves worse off (Conesa and Krueger, 1999). Hence, what is the Social Security system that the U.S. should transition to, in light of these restrictions, remains an open question.

The contribution of this article is to provide a candidate answer to this question using a model that captures

[1] See Feldstein and Liebman (2002), Shiller (1999), Krueger and Kubler (2006)

[2] Harenberg and Ludwig (2019) find a larger system than the rest of the literature. Results on the progressiveness of the optimal system are not as decisive. Although Huggett and Parra (2010); Golosov et al. (2013); Jones and Li (2020) argue that the optimum it is more progressive than the current system, Nishiyama and Smetters (2008); Ndiaye (2018) argue that is less.

transition costs and household voting. Using the current U.S. Social Security system as the starting point, we consider the case where the government announces an unanticipated once-and-for-all reform and enforces a new optimal two pillar mixed system. While the first pillar is a based on an unfunded Pay-as-You-Go system, the second is based on an defined contribution investment based program[3]. As a bi-product of this assessment, we establish the differences in terms of welfare, life-cycle decision profiles and aggregate statistics between the optimal and the current Social Security systems.

To quantitatively characterise the reform, we introduce a non-stationary, estimated overlapping generations general equilibrium model, where the government, a Pay-as-You-Go Social Security system, a private pension fund, households of multiple generations and competitive firms coexist and interact. This setup allows us to study both the short term and the longer term implications of the reform. The model builds on a large structural analysis literature of retirement behaviour and insurance[4]. Households make endogenous consumption, savings (both liquid and illiquid defined contribution pension savings), and labour supply decisions, in the presence of both aggregate and idiosyncratic uncertainty. As a result, while each generation is subject to a different sequence of aggregate shocks, within each generation households are ex-post heterogeneous.

The model captures the value of insurance, along with the distortions imposed by the existence and reform of a realistic Pay-as-You-Go Social Security system. On the insurance side, the existence of a Pay-as-You-Go Social Security system provides partial insurance against both aggregate and idiosyncratic risk. Aggregate risk is shared intergenerationally by giving retired households a claim on the labour income of working households[5] and idiosyncratic risk is shared intragenerationally by redistributing resources from those with high lifetime income to low lifetime income, and from those that have shorter lifespans to those with longer. Also, both types of risk interact in important ways, as the welfare benefit from insurance against the combined aggregate and idiosyncratic risk is greater than the insurance benefit against the sum of the risks in isolation (See Harenberg and Ludwig 2019).

In terms of distortions, the existence of a Pay-as-You-Go Social Security system can negatively affect efficiency and welfare in the model economy through four channels. First, Social Security crowds out household savings, reducing the aggregate capital stock and thus output and wages in the long run. Second, beyond the lower capital stock, the crowding out of savings also reduces bequest transfers from parents to their dependents[6]. The model includes a non-stationary, heterogeneous bequest transfer mechanism that captures these transfers. Third, the internal rate of return provided by Pay-as-You-Go Social Security system is on average lower than the market returns in the model economy. Due to the redistributiveness of Social Security, the internal rate of return is lower for those with above average incomes, causing heterogeneous implicit taxes that distort labour supply[7,8]. Finally, the transition

[3]See Feldstein and Liebman (2002) for a discussion on the advantages of mixed systems.

[4]Some papers include Conesa and Krueger (1999); French (2005); Nishiyama and Smetters (2007); Krueger and Kubler (2006); Nishiyama and Smetters (2007); Hong and Rios-Rull (2007); Olovsson (2010); Kitao (2014); Harenberg and Ludwig (2019).

[5]See Shiller (1999) for a discussion on how the imperfect correlation between wages and returns affects intergenerational risk sharing.

[6]Caliendo et al. (2014) show that under certain conditions, the decline in bequest transfers can offset the benefits of risk sharing, even when factor prices are exogenous. Hong and Rios-Rull (2007) also assess the importance of bequests.

[7]The implicit tax arises as for every dollar contributed, benefits increase by less than a dollar in present value. The exact benefit increase per dollar contributed depends on demographic factors, the benefit function specification and average earnings. See French (2005) for details on the computation of average earnings and Goda et al. (2011) for a discussion of the implicit taxes induced by the current U.S. system. On the aggregate level these implicit taxes go to the repayment of the interest on the unfunded pension liability. See Shiller (1999); Nishiyama and Smetters (2007).

[8]See Imrohoroglu and Kitao (2009); French and Jones (2012); Bagchi (2015); Blundell et al. (2016) for more on labour supply and

from the current Pay-as-You-Go, to a new smaller system, entails the repayment of the implicit debt inherited by previous generations that received more than they contributed, and the unfavourable transition dynamics due to the post-reform capital accumulation. The model economy features this implicit debt, that needs to be repaid by households during the transition.

Findings suggest transition costs greatly restrict the U.S. government's ability to move away from the status quo. However, the U.S. government can implement a new Social Security system that is more progressive and less costly to operate. The new system imposes fewer distortions in the model economy and increases welfare on both the transition path and the longer-term. Specifically, a simultaneous reduction in the average level of benefits combined with a reform of the benefit function can deliver welfare gains, that provide an increase of 0.7% in terms of consumption equivalent variation. Considering voting at the start of the transition, we find that most reforms that reduce the level of benefits do not gather a household majority. More progressive systems do gather a majority, but households have to be compensated by an increase in the level of average benefits. Combining both transition costs and household voting, we surprisingly find that transitioning to a new Social Security system is no longer possible, as no reform is simultaneously welfare increasing and politically feasible.

We perform sensitivity analysis and find that the results are robust with respect to changes in the model's environment and the household preferences. We find that alternative population growth rates ("2050 scenario"), the inclusion of technological progress and higher Frisch elasticities have a small impact on the results. In all cases the politically feasibility constraint binds before any move to a welfare increasing Social Security system is permitted.

The first strand of literature this article relates to, conducts analysis of old age Social Security systems in non-stationary environments. This literature focuses on household populations that are not in a deterministic steady state. This can be caused by a deterministic transition between two steady states Auerbach et al. (1987); Huang et al. (1997); Conesa and Krueger (1999); Nishiyama and Smetters (2007); Kitao (2014); Nishiyama (2015); Peterman and Sommer (2019a,b), or the presence of aggregate uncertainty in the model economy Krueger and Kubler (2006); Olovsson (2010); Harenberg and Ludwig (2019). While the deterministic transition models abstract from aggregate shocks and hence cannot consider how aggregate risk is spread intergenerationally, models with aggregate uncertainty are not estimated and abstract from modelling the institutional features of Social Security and the endogeneity of labour supply.

This article contributes in this literature in two ways. First, we estimate a retirement model in a non-stationary environment, that generates household lifecycle decision profiles and aggregate equilibrium statistics similar to the U.S. data. Second, this article combines aggregate shocks with endogenous labour supply and a realistic Pay-as-You-Go Social Security system. This allows the model to balance the value of insurance with the distortions created by the existence and reform of an unfunded Social Security system.

The second strand of literature this article relates to, concerns the optimal design of unfunded Social Security

retirement.

systems[9,10]. In partial equilibrium environments, Huggett and Parra (2010); Golosov et al. (2013) and Ndiaye (2018) compute the optimal benefit function, while O'Dea (2018) evaluates the optimal means-tested old age income floor and argues that an increase would be welfare enhancing. In general equilibrium, Imrohoroglu et al. (1995) evaluate the optimal replacement ratio, Olovsson (2010) evaluates unfunded Social Security systems that can share aggregate risk efficiently, Moser and Olea de Souza e Silva (2019) addresses the optimal design of pensions when agents are presently biased, Hosseini and Shourideh (2019) assess Pareto optimal policy reforms, Jones and Li (2020) consider multiple parametric reforms, one of which is the optimal shape of the Social Security benefit function and Ludwig and Grevenbrock (2021) consider the joint optimal design of Social Security and unemployment insurance.

This article adds to the literature in three ways. First, we conduct optimal design analysis in an environment that features both idiosyncratic and aggregate uncertainty. Previous research restricted its attention to only one type of risk sharing (either aggregate or idiosyncratic). In this paper, the model's unfunded Social Security system is selected to share risk optimally, both intra and inter-generationally. Second, we consider optimal design, allowing the government to impose a mixed system (illiquid savings accounts combined with an unfunded Social Security system) and third, we assess questions that concern the optimal design of Social Security in light of the transition path and the unfunded liability of the current U.S. Social Security system. Thus, this paper takes into account the transition costs each reform entails, in a heterogeneous agent model.

The third and final strand of literature this article relates to, assesses the implications of household voting in Social Security reforms. Conesa and Krueger (1999) consider three alternative stylized reforms and conclude that political support is decreasing in household heterogeneity and that reforms that do not entail smooth transitions (i.e. a sequence of Social Security systems) gather a larger percent of votes. This article features household voting in an environment with realistic household heterogeneity due to the interaction of both aggregate and idiosyncratic uncertainty. In addition, instead of comparing a small number of stylized reforms, we evaluate household votes for the entire space of potential parametric reforms.

The structure of the article is the following. Section 4.2 presents the model and the equilibrium definition. Section 4.3 presents the estimation process. Section 4.4 describes the data. Section 4.5 shows the calibration and estimation results. Section 4.6 presents the quantitative analysis. Section 4.7 presents the sensitivity analysis and Section 4.8 concludes.

4.2 The Model

The model features a rich environment where ex-post heterogeneous households of multiple generations, a Defined Contribution private pension fund, a Pay-as-You-Go Social Security system, a government, and competitive firms

[9]This strand is part of the more general and broad literature that studies the welfare and economic implications of pension systems. Hong and Rios-Rull (2007); Krueger and Kubler (2006); Harenberg and Ludwig (2019) consider unfunded systems. Nishiyama (2011); Imrohoroglu et al. (1998); Kitao (2010) consider funded tax-deferred accounts.

[10]Some papers use parsimonious models to derive optimal Social Security systems. See Feldstein (1985); Shiller (1999); Ball and Mankiw (2007).

coexist and interact. Time is discrete, denoted by $t = 0, ..., \infty$ and one model period corresponds to one calendar year. The model features both aggregate and idiosyncratic risk. The former is over the exogenous aggregate shock that hits the model economy in the beginning of every period, and the latter is over mortality, bequests and idiosyncratic productivity. By assumption, the markets for aggregate and idiosyncratic risk are closed. This permits the model's Social Security system to potentially increase welfare by providing partial insurance against both types of risks.

4.2.1 Demographics

Each household represents a household head that enters the economy at age 21 and can live up to age J. At every age, $j \in \{21, ..., J\}$, the conditional probability of surviving to age $j+1$ is denoted by s_{j+1} and $s_{j'|j} = \frac{1}{s_j} \prod_{\tau=j}^{j'} s_\tau$ denotes the conditional probability of an age j household surviving to age j'. By assumption, $s_{J+1} = 0$. Every period, a continuum (new generation) of households enters the model economy. We normalize the initial population to unity and assume that the size of each successive generation grows at a rate of λ_ν.

4.2.2 Households

Choices & Preferences In every period t, living households that have age j make a consumption $(c_{t,j})$ / savings $(a_{t+1,j+1})$ decision, a labour supply $(h_{t,j})$ decision and eligible households choose their private pension contributions $(b_{t,j})$, through salary sacrifice. Households care about consumption, leisure and knowledge that they will leave behind assets upon their death. They value consumption and leisure by a within period utility function $u(c_{t,j}, h_{t,j})$ and value bequests by a bequest function $u^{Beq}(a_{t+1,j+1})$. By assumption preferences take the following intertemporally separable form

$$u(c_{t,j},\, h_{t,j}) + \mathbb{E}_t \left\{ \sum_{\tau=1}^{J+1\text{-}j} \beta^{\tau-1} s_{j+\tau\text{-}1|j} \left[\beta s_{j+\tau} u(c_{t+\tau,j+\tau},\, h_{t+\tau,j+\tau}) + (1\text{-}s_{j+\tau}) u^{Beq}(a_{t+\tau,j+\tau}) \right] \middle| X_t \right\} \tag{4.1}$$

where β is the subjective discount factor and the expectation $\mathbb{E}_t$ is taken with respect to the cumulative distribution of future random events, conditional on all available information (X_t). The role of a household is to maximize equation (1) by choosing the sequence $\{c_{t+\tau,j+\tau}, h_{t+\tau,j+\tau}, b_{t+\tau,j+\tau}\}_{\tau=0}^{J+1\text{-}j}$, subject to the constraints described below.

Preference specification We parametrise the within period utility function as an additive CRRA over consumption and leisure.

$$\begin{aligned} u(c_{t,j}, h_{t,j}, j) &= \frac{c_{t,j}^{1-\sigma}}{1-\sigma} + \chi \frac{\left(1 - h_{t,j} - \theta_j^h \mathbb{I}[h_{t,j} > 0]\right)^{1-\gamma}}{1-\gamma} \\ u^{Beq}(a_{t+1,j+1}) &= \theta_B a_{t+1,j+1} \end{aligned} \tag{4.2}$$

where $\mathbb{I}[h_{t,j} > 0]$ is an indicator function assuming the value one if the agent is employed, θ_j^h is an age-dependent fixed cost of work, σ is the coefficient of relative risk aversion, γ is the utility curvature with respect to leisure and χ is the weight of leisure relative to consumption. We define the bequest function as a linear function similar to Kitao (2014) and denote its coefficient by θ_B.

Earnings Dynamics Households can participate in the labour market and be compensated per hour of work. Their wage $w^*_{t,j}$ is decomposed as the product of an endogenously determined aggregate component w_t (specified in Section 4.2.3) and an idiosyncratic component $\tilde{w}_{t,j}$. Household pre-tax labour earnings are defined as

$$y^l_{t,j} = w^*_{t,j} h_{t,j} = (w_t \tilde{w}_{t,j}) h_{t,j} \tag{4.3}$$

where $h_{t,j}$ denotes the household's annual hours of work. The idiosyncratic component is obtained by $log(\tilde{w}_{t,j}) = \dot{w}(h_{t,j}, j) + \eta_{t,j}$ where $\eta_{t,j}$ is idiosyncratic unobservable productivity and $\dot{w}(h_{t,j}, j)$ is a deterministic function that takes into account the age and annual hours of work of the household. Idiosyncratic productivity is evolving stochastically, following a first-order autoregressive process

$$\eta_{t,j} = \rho\, \eta_{t\text{-}1,j\text{-}1} + u_{t,j}, \quad u_{t,j} \sim \mathcal{N}(0, \sigma^2_u) \tag{4.4}$$

where ρ is the correlation coefficient and $u_{t,j}$ is a normally distributed innovation. See Section 4.3.1 and Appendix C for details on estimation and Appendix B for a technical note on the structural implementation.

Budget Constraint The budget constraint of the household is described by the law of motion for non-labour cash on hand, the asset evolution equation and the liquidity constraint.

$$\begin{aligned} m_{t+1,j+1} &= T^k_{t+1}(a_{t+1,j+1},\, beq_{t+1,j+1},\, r_{t+1},\, \mathfrak{T}_{t+1}) + ppb_{t+1,j+1} + ssb_{t+1,j+1} \\ a_{t+1,j+1} &= m_{t,j} + T^l_t(y^l_{t,j},\, b_{t,j},\, \mathfrak{T}_t) - c_{t,j} \\ a_{t+1,j+1} &\geq 0 \end{aligned} \tag{4.5}$$

where $1+r_{t+1}$ is next period's returns, $ppb_{t+1,j+1}$ is pension benefits, $ssb_{t+1,j+1}$ is Social Security benefits, $beq_{t+1,j+1}$ is bequest transfers, $\mathfrak{T}_t$ is the government's policy schedule defined in Section 4.2.4. T^l_t and T^k_{t+1} are functions that map to after tax labour and capital income. Under this specification, households can accumulate savings in the form of liquid assets to self-insure, but are not allowed to borrow.

Private Pension Savings In addition to liquid assets, households can self-insure by contributing to an illiquid defined-contribution private pension savings account. Contributions are paid through a salary sacrifice scheme and accumulate free of tax over the working life of the household according to

$$\begin{aligned} n_{t+1,j+1} &= (1 + r_t) n_t + y^l_{t,j} b_{t,j} - ppb_{t,j} \\ n_{t+1,j+1} &\geq 0 \end{aligned} \tag{4.6}$$

where $b_{t,j}$ is the contribution rate. Once households reach age J_{pp} they start receiving benefits $ppb_{t,j}$ until either death or their pension wealth is depleted ($n_{t,j} = 0$). See appendix A for the computation of pension benefits.

Social Security Benefits In the model, every household is entitled to Social Security benefits $ssb_{t,j}$ once it reaches the age of J_{ss}. Benefits are calculated using the household's age and its Average Indexed Monthly Earnings (AIME)[11], where the latter is denoted by $ss_{t,j}$ and evolves according to

$$ss_{t+1,j+1} = \begin{cases} ss_{t,j} + min\{\frac{y^l_{t,j}(1-b_{t,j})}{35}, \frac{y^s}{35}\} & \text{if } j \leq J_w \\ ss_{t,j} + max\{0, \frac{y^l_{t,j}(1-b_{t,j})-ss_{t,j}}{35}\} & \text{otherwise} \end{cases} \tag{4.7}$$

where y^s corresponds to the Social Security contributions cap and J_w denotes the household's age 35 years after entering the model economy. The benefit function that maps $ss_{t,j}$ to $ssb_{t,j}$ is specified by the government's policy schedule $\mathfrak{T}_t$. In the benchmark model specification, the benefit function mimics the actual U.S. formula. Note that unlike private pensions, Social Security benefit payments will continue until the household's death. Hence, proving partial-insurance against longevity.

Bequest Transfers The model includes non-stationary, heterogeneous bequest transfers, that link households of different generations. In every period t, a percent of the households die with potentially positive assets. These bequests are allocated to incoming age 21 households that enter the economy in period $t + 1$. The number of recipients and the size of bequests for each diseased household are determined by its savings decision and the model's population dynamics[12].

State variables To make optimal choices, households take into account their current characteristics, the economy's aggregate state and the structure of uncertainty. Each household is fully described by its non-labour cash on hand (m), private pension savings (n), idiosyncratic productivity (η), Average Indexed Monthly Earnings (ss), the age of the household (j), and the aggregate state of the economy. We denote the vector of household characteristics by $\omega = (m, n, \eta, ss, j)$ and its support by $\Omega = M \times N \times H \times SS \times \{1, ..., J\}$. The economy's aggregate state in period t is described by the aggregate shock (z_t), the period specific measure of households (μ_t) over their characteristics vector, and the policy schedule of the government ($\mathfrak{T}_t$) that fully describes the government's tax and benefit system from period t onwards. Hence, the state space vector of the households is defined as $X_t = (m, n, \eta, ss, j; \mu_t, z_t, \mathfrak{T}_t) = (\omega; \mu_t, z_t, \mathfrak{T}_t)$.

Recursive Formulation The household's problem is solved recursively. In period t, the policy functions of a

[11]In reality, the Social Security system uses the beneficiaries 35 highest earning years to calculate benefits. In the model after the first 35 years AIME is updated upwards only if annual earnings exceed the current AIME value.

[12]An assumption needs to be imposed, as including the state space of parents in the state space of their dependents renders the problem computationally infeasible. Following Hong and Rios-Rull (2007), we impose that bequests are received in the first period of life. Under this assumption, the model generates substantial heterogeneity in bequest transfers and the need for bequest forecasts in an environment with aggregate uncertainty is circumvented.

household with characteristics $\omega \in \Omega$ are obtained as the solution of

$$V(\omega;\mu_t,z_t,\mathfrak{T}_t) = \max_{\substack{c,a' \\ h,b}} \left\{ u(c,1\text{-}h,j) + \mathbb{E}\left[s_{j+1}\beta V(\omega';\mu_{t+1},z_{t+1},\mathfrak{T}_{t+1}) + (1\text{-}s_{j+1})u^{Beq}(a') \middle| \omega;\mu_t,z_t,\mathfrak{T}_t \right] \right\}$$

$$\begin{aligned}
& st. \\
& Equations\ (2)\text{-}(7) \\
& z_{t+1} = \Pi_z(z_t) \\
& \mu_{t+1} = \Gamma_\mu(\mu_t,z_t) \\
& b = 0, \quad \text{if } j \geq J_{pp}
\end{aligned}$$

where the prime denotes next period's variables. Since there is no closed form solution to the problem, it is solved numerically[13]. The solution algorithm builds on Krusell and Smith (1998), Carroll (2006), Young (2010), Kopecky and Suen (2010), Druedahl and Jørgensen (2017) and Clausen and Strub (2020). The details of the solution algorithm are presented in Appendix B.

4.2.3 Firms

The firms are competitive and produce under a constant returns to scale production function. Output is determined by

$$Y_t = F(K_t, L_t; z_t) = z_t K_t^\alpha (A\, L_t)^{1-\alpha}$$

where z_t is the aggregate shock, A is labour augmenting technological progress, K_t and L_t are the aggregate capital and labour inputs and α is the capital's share of output. We define the aggregate capital input as the sum of beginning of period liquid[14] and illiquid assets and the aggregate labour input as the sum of efficient labour units ($\eta h_{t,j}$). The profit maximizing conditions[15] are $w_t = F_2(K_t, L_t; z_t)$ and $r_t = F_1(K_t, L_t; z_t) - \delta$, where $\delta \in (0,1)$ is the capital's depreciation rate.

4.2.4 The Government

The government runs the Pay-as-You-Go Social Security system, makes public purchases of goods $\mathcal{C}_t$ and collects taxes. Taxes are determined by the policy schedule $\mathfrak{T}_t = \{\tau^l_{t+\tau}, \tau^s_{t+\tau}, \tau^k_{t+\tau}, ssb_{t+\tau}\}_{\tau=0}^{\infty}$, where $\tau^k_{t+\tau}$ is the capital gains tax rate, $\tau^s_{t+\tau}$ the Social Security tax rate (collected up to a cap of y^s) and $\tau^l_{t+\tau}$ is the linear labour income

[13] To access the accuracy of the numerical solution we follow Judd (1992), Santos (2000) and Den Haan (2010). These checks are required to assess whether the errors involved are sufficiently small, so that they have no effect on the inference of the theoretical model's behaviour (Santos, 1999). Also, multiple checks are carried to ensure that the support of the continuous state variables is sufficiently large. See Appendix B for technical details.

[14] Note that this is different than m. Specifically, beginning of period liquid assets are equal to the previous period's savings decision of the household ($j > 21$) or bequest transfers ($j = 21$) and are contained in non-labour cash on hand.

[15] The conditions imply firms make capital and labour choices after the state of the economy shock z_t is realised. See Appendix A for more details on the model's timing.

tax[16]. The policy schedule is known at the beginning of period t. The budget constraint of the government is defined as

$$\int_\Omega \tau_t^k r_t(a(X_t) + beq(X_t))\, d\mu_t(\omega) \;+\; \int_\Omega \tau_t^l y^l(X_t)(1-b(X_t))\, d\mu_t(\omega) \;+\; \int_\Omega \tau_t^s min\{y^l(X_t)(1-b(X_t)), y^s\}\, d\mu_t(\omega) = \mathcal{C}_t + \Delta SS_t \;\; \forall t$$

where a is previous period's savings and beq is bequest transfers, ΔSS_t is the Social Security surplus or deficit and $\mu_t(\omega)$ is the measure of households in state $\omega \in \Omega$. The constraint has to be satisfied every period. Thus, the government adjusts its labour income tax rate to balance it[17]. See Section 4.3.1 for the exact specification of the tax rates and benefits. The variable ΔSS_t corresponds to the Security deficit or surplus, defined as

$$\Delta SS_t = \int_\Omega ssb_t(X_t)\, d\mu_t(\omega) - \int_\Omega \tau_t^s \; min\Big\{y^l(X_t)(1-b(X_t)),\, y^{ss}\Big\}\, d\mu_t(\omega)$$

where $ssb(X_t)$ is the benefit function.

4.2.5 Equilibrium Definition

Let the tuple $(\mu_t, z_t, \mathfrak{T}_t)$ denote the aggregate state of the economy, where μ_t is a measure defined over the household characteristics, z_t is the period t aggregate shock and $\mathfrak{T}_t$ the policy schedule. A series of measures μ_t, bequest measures $\{beq_\tau\}_{\tau=t}^\infty$, factor prices $\{r(\mu_\tau, z_\tau, \mathfrak{T}_\tau), w(\mu_\tau, z_\tau, \mathfrak{T}_\tau)\}_{\tau=t}^\infty$, the policy schedule $\mathfrak{T}_t$, a law of motion $\Gamma_\mu(\mu_t, z_t)$, the household value functions $\{v(\omega;\mu_\tau, z_\tau, \mathfrak{T}_\tau)\}_{\tau=t}^\infty$ and the policy functions of the households

$$\Big\{c(\omega;\mu_\tau, z_\tau, \mathfrak{T}_\tau), h(\omega;\mu_\tau, z_\tau, \mathfrak{T}_\tau), a'(\omega;\mu_\tau, z_\tau, \mathfrak{T}_\tau), b(\omega;\mu_\tau, z_\tau, \mathfrak{T}_\tau)\Big\}_{\tau=t}^\infty$$

are in a competitive equilibrium for every period $\tau = t, ..., \infty$, if each household solves its utility maximization problem, the measure Γ_μ is generated by the policy functions, the firms solve their profit maximization problem, all the markets clear, bequests are consistent with the policy functions and model's population dynamics, and the government balances its budget.

4.3 Estimation

This section describes estimation process. The aim is to estimate the model's parameters that generate household lifecycle decision profiles and aggregate equilibrium statistics that mimic the U.S. data. Similar to Gourinchas and Parker (2002), French (2005) and Blundell et al. (2016) we follow a two step procedure. In the first step, we either calibrate following the literature or estimate outside of the model a subset of the model's parameters. In the second,

[16] Note that the tax rates contained in $\mathfrak{T}_t$ are different objects than the functions T_t^k and T_t^l, that map the capital and labour income of the households to their post tax and contribution dollar amount.

[17] Note that we have suppressed the dependence of the labour tax rate on the state of the economy.

conditional on the first step estimates and calibrations, we estimate the remaining parameters via the Method of Simulated Moments (MSM). A rigorous presentation of the model's estimation is located in Appendix C.

4.3.1 First Step

Idiosyncratic earnings process This subsection describes the estimation of the lifecycle profile for the idiosyncratic (and deterministic) component of wages and the parameters that govern the unobservable productivity process. The goal is to obtain estimates that take into account selection (in the form of a labour market participation decision), the presence of unobserved individual heterogeneity and the wage penalty of working part-time. Empirically, we parametrise the deterministic component of idiosyncratic wages as $\dot{w}(h_{t,j}, j) = \phi(j) + \alpha^h ln(h_{t,j})$, where $\phi(j)$ is an age polynomial and a^h is the part-time wage penalty. We calibrate the coefficient a^h (see the next subsection for the calibration), define the wage net of part-time work penalty, $z_{i,t,j} = log(\tilde{w}_{i,t,j}) - \alpha^h log(h_{i,t,j})$ and estimate the fixed-effect (FE) regression

$$z_{i,t,j} = \phi_0 + \phi_1 j + \phi_2 j^2 + \phi_3 j^3 + f_i + \varepsilon_{i,t,j}$$

where f_i is an unobservable individual specific fixed-effect, $\{\phi_s\}_{s=0}^3$ are the coefficients of the cubic age polynomial and $\varepsilon_{i,t,j}$ is the regression residual. We interpret the sum $\eta_{i,t,j} = f_i + \varepsilon_{i,t,j}$ as the idiosyncratic unobservable productivity of the household and estimate it by standard GMM methods[18].

Sample selection bias An issue with the FE regression presented above is that it does not account for non-random selection in the labour market. To address the sample selection bias in a model consistent way, we extend[19] the iterative method introduced by French (2005) to correct the effect that non-random sample selection has on the life-cycle profile of wages and the unobservable idiosyncratic productivity. At the end of this iterative process, the life-cycle profile of wages is equal between the FE estimates and the model simulated data (conditionally on being employed in the model world) and the parameters that govern idiosyncratic productivity are estimated using selection-corrected regression residuals.

The procedure proceeds as follows. i) The life-cycle model is estimated using the biased FE life-cycle wage profile and idiosyncratic productivity process. ii) Using the model simulated data where both workers and non-workers are observed, the wage profiles are iteratively adjusted[20] until the simulated wage profile (conditional on being employed in the model world) is equal to the estimated FE profile. iii) Using the updated wage profiles, the residuals are adjusted for every individual in the PSID and the idiosyncratic productivity process is estimated using the adjusted

[18] We specify the theoretical model as $\eta_{i,t,j} = AR(1) + MA(1)$, where we interpret the $AR(1)$ part as the "true" process and the $MA(1)$ part as measurement error. The GMM estimation is performed by minimizing the (weighted) distance between the theoretical and empirical variance-covariance matrices. We find that the $AR(1)$ estimates are very robust with respect to alternative measurement error specifications and weighting matrix choices. Appendix C describes the GMM estimation procedure in detail.

[19] Rigorously, although the object of interest is $\mathbb{E}[log(\tilde{w}_{i,t,j}) - \alpha^h log(h_{i,t,j})|j]$, the FE regression (that uses observations only for those employed) identifies $\mathbb{E}[log(\tilde{w}_{i,t,j}) - \alpha^h log(h_{i,t,j})|j, h_{i,t,j} > 0]$. As a consequence, the life-cycle profiles, the FE residuals and thus the coefficients that govern idiosyncratic productivity will be affected by the sample selection bias. See Appendix C for an extensive discussion on selection correction.

[20] The iterative procedure of French (2005) has a unique fixed-point in convex models and as French (2005) and O'Dea (2018) discuss, simulations suggest that there is a unique fixed-point in non-convex models as well.

residuals. iv) The model is estimated using the updated wage profile and idiosyncratic productivity process. Steps ii) to iv) are repeated until convergence.

4.3.2 Second Step

Method of simulated moments In the second step, conditional on the first step's estimates and calibrations, we estimate the coefficients $(\beta, \gamma, \chi, \theta_B)$ that enter the household's utility function via the MSM (McFadden, 1989; Pakes and Pollard, 1989). We match the life-cycle profiles of non-pension wealth, pension wealth, hours of work conditionally on being employed and participation rates. Before describing the estimation process, it is worth noting the identification strategy.

Identification The level of liquid and illiquid pension wealth at every age assist in the the identification of the subjective discount factor (β), as more patient households will tend to hold greater wealth at any age. The decline of wealth post-retirement assists in the identification of the bequest motive (θ_B), as households with a higher bequest motive (conditional on risk aversion) will hold more assets to bequeath to their descendants. The participation rate profiles assist in the identification of leisure's weight relative to consumption (χ), as households that have a lower preference for leisure relative to consumption will remain at work for more years. The annual hours of work life-cycle profile and its slope near retirement assists in the identification of the utility function's curvature with respect to leisure (γ), as households with higher values of leisure curvature will respond more to the work disincentives induced by Social Security.

Estimation process To obtain the simulated moments, we solve the model and simulate the lifecycle decision profiles for $T = 10,000$ overlapping generations. The initial conditions of each incoming generation are determined by the bequeathed savings of households that died in the previous period, the model's population dynamics and the idiosyncratic productivity process. We stack the simulated moments in a vector $M^{sim}(\theta)$, the equivalent data moments in a vector M^{data} and define the estimated[21] coefficient vector $\hat{\theta}$ by

$$\hat{\theta} = \underset{\theta \in \Theta}{argmin} \left\{ \sum_{k=1}^{K} (M_k^{sim}(\theta) - M_k^{data})^2 / Var(M_k^{data}) \right\}$$

where M_k corresponds to the kth element of vector M. In terms of the weighting matrix, although the optimal weighting matrix is asymptotically efficient, Altonji and Segal (1996) argue that it may be severely biased in small samples, a problem that tends to exacerbate as the tails of the empirical distribution get fatter. For this reason we weight each squared conditional moment with the inverse of its empirical variance. To correct for simulation noise when computing the asymptotic standard errors we follow Gourieroux et al. (1993).

[21] As the model includes discrete choices and fixed-cost of work, the expected value function features potentially a large number of downward kinks and is not globally concave (See Clausen and Strub (2020) and Fella (2014) for discussions on downward kinks). As a consequence the GMM objective function might i) not be smooth and ii) have many local minima. To address both issues, we commence the iterative derivative-free Bound Optimization BY Quadratic Approximation (BOBYQUA) algorithm of Powell (2009) from multiple starting values of $\theta \in \Theta$ and select the estimate $\hat{\theta}$ that generates the smallest GMM criterion (as the global minimum).

4.4 Data

We use the Panel Study of Income Dynamics (PSID), a biennial longitudinal survey that contains a representative sample of the U.S. household population and focus on the years 1999-2019. We deflate prices to 2014 dollars using the Consumer Price Index (cpi) and for each household we keep track of the household head. The initial sample consists of 17,489 unique households and we impose the following 5 restrictions.

We keep the Survey Research Center (SRC) sample (7,165 households dropped), keep only singles (8,028 households dropped), drop individuals that were not the household's head for one or more periods (751 households dropped), drop individuals that were not 18 at the first wave of the sample (163 households dropped), and drop individuals that answer the survey for one wave only and then drop out (170 households dropped). The final sample consists of 1,212 unique households and 13,332 person-year observations.

The hourly wage is computed as the annual labour earnings divided by total hours of work. We set the wage to missing if it is less than 9 dollars or higher than 500 dollars. When a household head has less than 300 total hours of work in a given year, he is considered as a non-participant in the labour market. We set hours of work to missing if a household reports more than 5500 hours (that correspond to more than twice the full time employment hours) or less than 300 hours. The PSID asset measure includes real estate, the value of a farm or business, vehicles, stocks, mutual funds, IRAs, Keoghs, liquid assets, bonds, other assets, and investment trusts less mortgages and other debts. It does not include pension or Social Security wealth. Pension wealth is defined as the sum of the Defined Contribution and Defined Benefit pension wealth. As we do not want results to be affected by wealthy individuals, we set the wealth variables to missing for households with assets over \$2,000,000 and pension wealth over \$500,000.

4.5 Calibration and Estimation Results

This Section presents the parameter calibration, the estimation results and the model's fit of the data. At the end of this Section, Table 4.3 is showing all the model parameters,n description and values.

4.5.1 Calibration

Demographics The model's population growth rate is set equal to 1.1%, a value close to the U.S. long-term average and assess the effect of a lower growth rates in Section 4.7. Households enter the economy at age 21 and we set the maximum household age J to 101, after which households die with probability one. We set the Social Security and private pension eligibility ages to 66 and 62 years old respectively. As the PSID has poor data on mortality, instead of directly estimating the conditional survival probabilities, we use the estimates of Bell and Miller (2005).

Preferences The relative risk aversion parameter for the baseline model is set to $\sigma = 2$. This value generates an inter-temporal elasticity of substitution (IES) of $1/\sigma = 0.5$, a value in line with the literature. The model features a fixed cost of work that is increasing with age. This allows the model to generate Frisch elasticities that are higher for households closer to retirement, conditional on annual hours of work. This cost is denoted[22] by θ_j^h and conditionally on age, it is determined by three parameters $\{k_i\}_{i=1}^3$ that are calibrated such that the fixed cost of work in terms of weekly hours is 6 at age 40 and 9 at age 66.

Earnings Following Aaronson and French (2004) we set the part-time wage penalty equal to 0.415, this implies that part-time workers earn approximately 25% less than the full time employed. The parameters that determine the idiosyncratic productivity process are estimated in the first step and corrected for non-random sample selection in the second.

The Government Social Security tax τ^s is set to 10.6%, with a cap of $105,000$\$. Labour taxes τ_t^l, are linear and adjust endogenously every period t to balance the government's budget constraint. The capital tax is equal to 18%. We calibrate government consumption and set it equal to 23% of the unconditional deterministic steady state GDP.

Firms The capital share of output is set equal to 32% and the capital depreciation rate is set equal to 8.1%, in line with the literature. The parameter A is a normalization. The transition matrix for TFP is described by the parameters $\{\pi_i^z\}_{i=1}^2$ that are set equal to $\{0.941, 0.059\}$ following Harenberg and Ludwig (2019), that calibrate the autocorrelation of TFP using the national income and products accounts (NIPA) data. Their target is an autocorrelation of 0.88 on an annual basis. The Total factor productivity (TFP) shock is set to 2.7%. This value generates fluctuations in the real side of the model economy that mimic the the U.S. aggregate data.

4.5.2 First Step Estimates

The first step estimates the life-cycle profile for idiosyncratic wages and the unobservable idiosyncratic productivity process. Table 4.1 shows the estimates for the coefficients that govern the idiosyncratic productivity process and their standard errors.

GMM Estimates		
coefficient	value	s.e.
ρ	0.981	(0.0057)
σ_η^2	0.014	(0.0044)

Table 4.1: Ffirst step GMM estimates
Notes: The table presents the first step GMM estimates for the unobservable idiosyncratic productivity process and their asymptotic standard errors (in parenthesis), after non-random selection in the labour market has been accounted for.

[22]The age-dependent fixed cost of work is modelled as $\theta^h(\{k_i\}_{i=1}^3, j) = \theta_j^h = k_1 + k_2(j/J_w)^{k_3}$. We set this specification following Kitao (2014).

In Appendix B we discuss the creation of idiosyncratic productivity grids and the Markov transition matrices using these estimates. Figure 4.1 displays the raw life-cycle profile for idiosyncratic wages minus the part-time wage penalty of the PSID data and the smooth FE life-cycle profile.

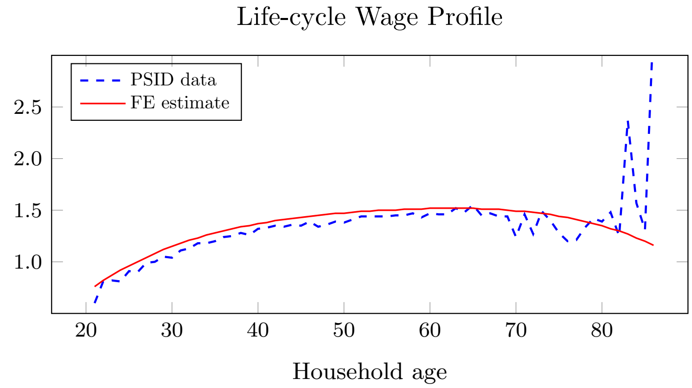


Figure 4.1: Wage life-cycle profile
Notes: Life-cycle profile for the logarithm of idiosyncratic wages minus the part-time wage penalty ($\mathbb{E}[log(\tilde{w}_{i,t,j}) - \alpha^h log(h_{i,t,j})|j]$), raw PSID data and the smoothed fixed effects estimate.

4.5.3 Second Step Estimates

The estimated coefficients and their standard errors are presented in Table 4.2. The estimates imply that the Frisch labour supply elasticity[23] for a 40 year old household is 0.38, which is significantly less than one as suggested by MaCurdy (1981); Altonji (1986); Blundell and Macurdy (1999). As the Frisch elasticity in this model depends on age, through the age dependent fixed-cost of work and annual work hours, older households have a higher elasticities.

coefficient	value	s.e.
$\hat{\chi}$	0.54	(—)
$\hat{\gamma}$	2.75	(—)
$\hat{\beta}$	0.97	(—)
$\hat{\theta}_B$	0.45	(—)

Table 4.2: Second Step MSM estimates
Notes: The table presents the second step MSM estimated preference coefficients along with their asymptotic standard errors (in parenthesis).

The simulated life-cycle profiles for liquid wealth, pension wealth, hours of work and the participation rates, along with the PSID counterparts are displayed in Figure 4.2. The top left right panel of Figure 4.2 shows the data and model simulated liquid wealth profiles. In the data assets gradually increase until the early 70s where they reach a peak at around \$250,000 and then they start to decline. In the model households save to self-insure against both aggregate and idiosyncratic risk and to leave bequests. The model is able to replicate this general pattern, slightly undershooting in the 40s and overemphasising the de-cumulation in the late 70s.

[23] The Frisch elasticity in this model is obtained by $\varepsilon_f = (1 - h - \theta^h_j \mathbb{I}[h > 0])/(h\gamma)$.

The top right right panel of Figure 4.2 shows the data and model simulated private pension wealth profiles. In the data private pension wealth increases gradually until the mid 70s where it reaches a peak at around $150,000 and then it starts to decline. Households save for the same reason as liquid wealth, however they cannot use their private pension wealth to smooth consumption before their mid 60s. The model is able to closely replicate this pattern slightly undershooting in the 30s and overshooting in the 60s.

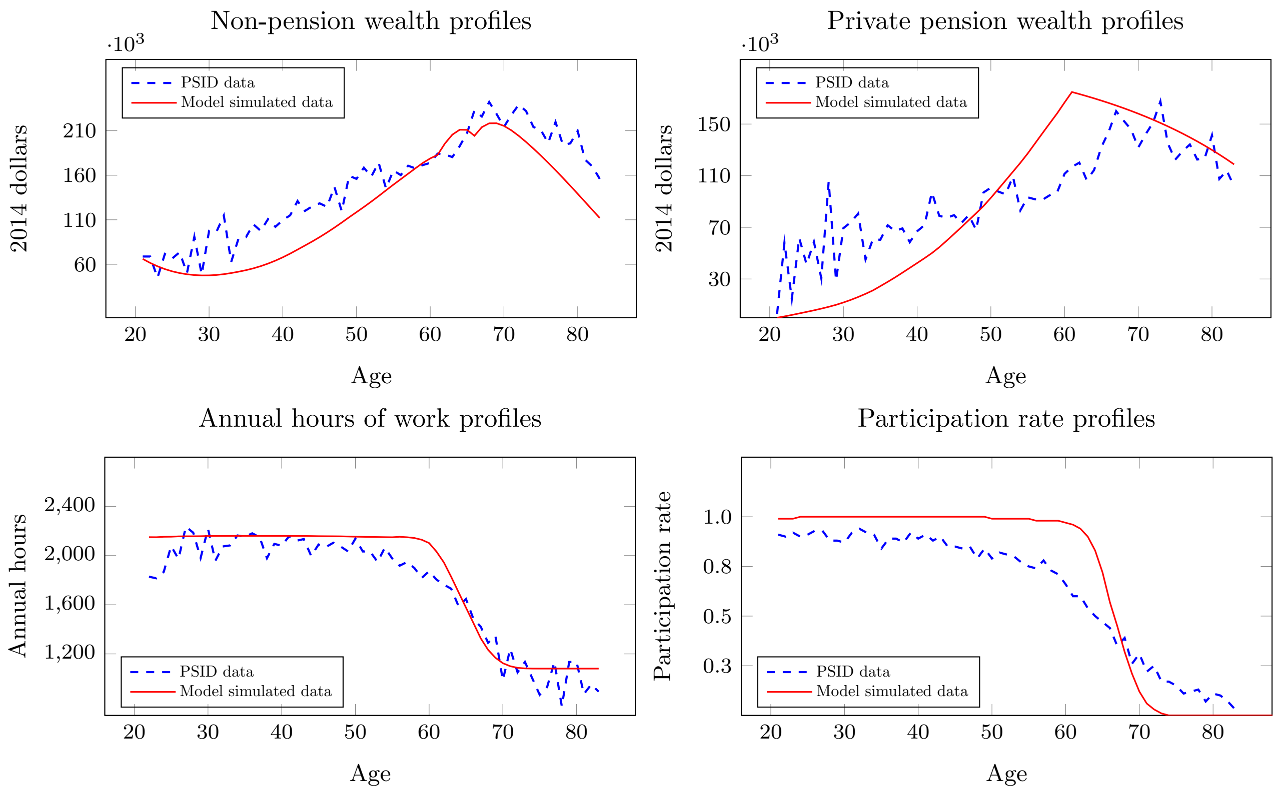


Figure 4.2: Data and model simulated life-cycle profiles
Notes: Data and model simulated life-cycle profiles for liquid wealth, pension wealth, annual hours and participation rates. The blue dashed curves correspond to the conditional moments obtained using household head data from the PSID and the red curve to the conditional moments generated by the simulated model.

The bottom left right panel of Figure 4.2 shows the data and model simulated annual hours of work profiles, conditionally on being employed. In the data hours of work are lower than full-time employment in the early 20s and then they become fairly flat at around the full-time employment level until the mid-50s. Then they gradually decline, until the age of 70 where they become fairly flat again. Although the model does not capture the lower hours of work in the early 20s, it does a good job after the mid 20s. It captures the flat profile up to the 50s, the gradual decline up to the 70s and the flat level in the post-retirement hours of work.

The bottom right panel of Figure 4.2 shows the data and model simulated participation rate profiles. In the data rates are fairly flat and close to 90% until the mid-40s. The rates start to decline gradually up to the early 60s and then the drop gets steeper until almost everyone is retired by the age of 80. The model's overall pattern is consistent with the data, although it overshoots for the pre-retirement years and it undershoots in the post-retirement years. In both the data and the model households spend on average approximately 43 years in the labour market.

In addition to the household life-cycle profiles, the model makes predictions about multiple untargeted aggregate statistics. The model endogenously generates a correlation between returns of wages of 0.43, that is close to the annual correlation of the NIPA data at 0.5 (Harenberg and Ludwig, 2019). The Social Security system is running a surplus of 0.41% of GDP, a value which close to the actual Social Security system (Kitao, 2014). The Social Security benefits amount to 5.4% of GDP. The labour taxes are adjusted every period to bind the government's budget constraint, on average they equal to 23.6%, excluding social security contributions.

The capital to output ratio (K/Y) is on average 2.71, which implies average returns that are equal to 5.1%. The model does a good job at replicating business cycle statistics observed in the U.S. data. The standard deviation of output is 3.7% on an annual basis. Consumption is pro-cyclical and less volatile than output with approximately 0.55 times its standard deviation. Aggregate capital is pro-cyclical and more volatile than output, its standard deviation relative to output is roughly 3.9 times larger. The model generated standard deviation for annual aggregate hours of work is roughly 0.45 times that of output. The standard deviation is lower than the output's, as the estimated Frisch labour supply elasticity is lower than what is used in the business cycle literature. In Section 4.7, we assess the impact that higher Frisch elasticities have on results.

Model Coefficients		
Parameter	Description	Value (S.E.)
Demographics		
λ_ν	Population growth	1.1%
J	Maximum age	101
J_w	Age after 35 years in the economy	56
J_{ss}	Social Security eligibility	66
J_{pp}	Private pension eligibility	62
$\{s_j\}_{j=21}^{J}$	Conditional survival probabilities	Bell and Miller (2005)
Preferences		
σ	Relative risk aversion	2.0
β	Subjective discount factor	0.98 (—)
γ	Utility curvature of leisure	2.60 (—)
χ	Weight of leisure relative to consumption	0.48 (—)
θ_B	Bequest motive parameter	0.45 (—)
$\{k_i\}_{i=1}^{3}$	Fixed cost of work	$\{0.053, 0.04, 2.98\}$
Earnings		
α^h	Part-time wage penalty	0.25
ρ	Productivity persistence	0.981 (0.0057)
σ_η^2	Productivity variance	0.014 (0.0044)
Government		
τ^k	Capital tax	18%
τ^s	Social Security tax	10.5%
τ^l	Labour tax	endogenous
y^s	Social Security cap	$105,000\$$
$\mathcal{C}$	Government consumption	23% of GDP
Production		
α	Capital share of output	32%
δ	Capital depreciation rate	8.1%
A	Normalization	0.98
z_t	TFP shock	$+/-2.7\%$
$\{\pi_i^z\}_{i=1}^{2}$	TFP shock transition coefficients	$\{0.941, 0.059\}$

Table 4.3: The model's parameters

Notes: The table presents the model's parameters, their description, values and standard errors if they are estimated. The first column shows the parameter notation, the second its description and the third its value.

4.6 The Quantitative analysis

In this Section we describe the government's problem. Then, we evaluate the optimal Social Security system, compare it with the status quo and discuss the differences in terms of aggregate statistics, life-cycle profiles and welfare.

4.6.1 The Government's Problem

The aim of the government is to improve the welfare of the households in the model economy, by altering aspects of its policy schedule $\mathfrak{T}_{t_0}$ and mandating a new partially privatized Social Security system $\mathfrak{T}'_{t_0}$.

Government's policy To ensure the computational feasibility of the problem, the government's degrees of freedom are restricted to i) the choice of a flexible two parameter Social Security benefit function, ii) the minimum mandatory private pension contribution level, b_{min} and iii) the level of the linear labour taxes, that adjust every period to bind the government's budget constraint. If ss denotes the household's average income monthly earnings, the flexible two parameter family is of the form

$$ssb(j, ss) = \mathbb{I}[j \geq J_{ss}] \cdot \phi_0 (ss)^{\phi 1}$$

where ϕ_0 controls the level and ϕ_1 controls how progressive (or regressive) the old age Social Security benefits are with respect to the AIME[24]. Lower values for ϕ_0 correspond to lower average benefits, and lower values for ϕ_1 correspond to more progressive systems (where lower AIME have higher replacement rates relative to high AIME households). This flexible specification approximates well the current U.S. Social Security system and allows the government to consider a wide range of alternative progressive, regressive and flat benefit functions. Figure 4.3 shows the U.S. Social Security benefit function and the best fix approximation of the flexible two parameter function.

Social welfare function In addition to the benefit function, one more object that needs to be specified is the social welfare function[25]. We follow Fehr and Kindermann (2015) and use a subcase of their general social welfare function that applies a utilitarian perspective to living households and an ex-ante perspective to future households. The social welfare function for each potential tuple (ϕ_0, ϕ_1) is defined as

$$\mathfrak{W} = \overbrace{\mathbb{E}\left[V(\omega; \mu_{t_0}, z_{t_0}, \mathfrak{T}'_{t_0}) \,\middle|\, \omega; \mu_{t_0}, z_{t_0}, \mathfrak{T}'_{t_0}\right]}^{\text{generations alive at } t_0} + \sum_{\tau=t_0+1}^{\infty} \psi(\tau)\, \overbrace{\mathbb{E}\left[V(\omega; \mu_\tau, z_\tau, \mathfrak{T}'_\tau) \,\middle|\, j = 21; \mu_{t_0}, z_{t_0}, \mathfrak{T}'_{t_0}\right]}^{\text{future generations}}$$

[24]Note that under this specification labour taxes adjust every period to bind the government's constraint and the benefit function remains constant over time. Krueger and Kubler (2006) discuss how adjusting the benefits vs taxes affects their results and conclude that it is of secondary importance.

[25]Note that the reform is implemented in period t_0 where the state of the economy is $(\mu_{t_0}, z_{t_0}, \mathfrak{T}_{t_0})$ and hence transition results will be anchored by the initial state. This issue is discussed in Fehr and Kindermann (2015). We commence the transition from a large number of different periods, all at least 10 years apart and then evaluate the average welfare. This strategy is also employed by Olovsson (2010).

where $\psi(\tau)$ is the discount factor for the welfare of future generations, $V(\omega;\mu_{t_0},z_{t_0},\mathfrak{T}'_{t_0})$ is the remaining lifetime utility of a household with age j in period t_0 and $V(\omega;\mu_\tau,z_\tau,\mathfrak{T}'_\tau)$ is the ex-ante lifetime utility of a household that will enter the model economy in some period $\tau \geq t_0+1$. Under this assumption, the government takes into consideration the remaining lifetime utility of households alive during the reform and the ex-ante lifetime utility of future households.

We assume that the discount factor takes the form $\psi(\tau)=\beta^{\tau-t_0}$. This specification has two advantages. First, it lies between the two extremes of either considering only living households or considering the long-term stochastic steady state. Second, it does not depend on market outcomes or a specific reform, as it only depends on an estimated preference parameter.

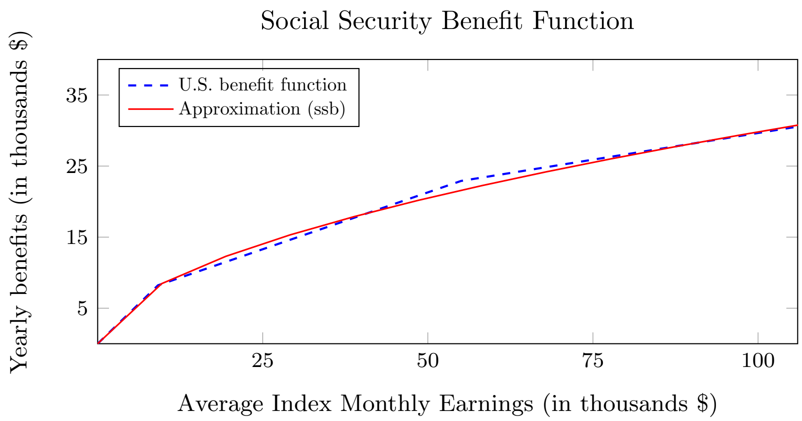


Figure 4.3: The Social Security benefit functions
Notes: This figure presents the current piecewise linear U.S old age Social Security benefit function that features three bend-points in blue and the flexible two parameter "best fit" benefit function, generated using the values $\phi_0 = 0.41$ and $\phi_1 = 0.54$.

Household voting Living households vote in the beginning of period t_0, after every shock apart from mortality has realized. We measure the percent of positive votes in favour of each potential reform by

$$\text{Positive Votes} = \int_\Omega \mathbb{I}\Big[V(\omega;\mu_{t_0},z_{t_0},\mathfrak{T}'_{t_0}) \geq V(\omega;\mu_{t_0},z_{t_0},\mathfrak{T}_{t_0})\Big]\, d\mu_{t_0}(\omega)$$

where the indicator function assumes the value of one when the condition is satisfied. This voting procedure implies that households compare their expected remaining lifetime utility in the status quo with the counterfactual scenario where the reform takes place.

4.6.2 The transition optimum

We consider once-and-for-all unanticipated Social Security reforms for every tuple (ϕ_0,ϕ_1) and pension contribution floor. The Social Security benefits and pension contribution floor are adjusted according to the reform and labour taxes adjust every period to bind the government's constraint.

Figure 4.4 shows how welfare changes as the U.S. moves away from the status quo. It is evident that the U.S. government's ability to move away from the current system is restricted by transition costs[26]. However, a substantial move is permitted and the new Social Security system increases welfare in both the transition and longer-term.

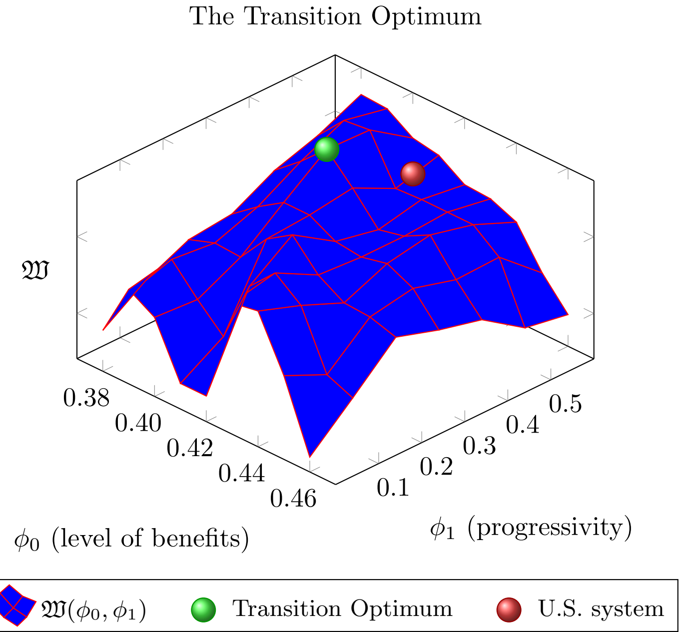


Figure 4.4: The transition welfare function
This figure presents the transition welfare function conditional on a zero private pension contribution floor. The green sphere is the transition optimum ($\phi_0 = 0.40, \phi_1 = 0.40$) and the red sphere is the current U.S. system ($\phi_0 = 0.41, \phi_1 = 0.54$). Adjustments to the private pension contribution floor have a negligible impact to the results.

The new system imposes fewer distortions in the model economy and is less costly to operate. The private pension contribution floor has a negligible impact on the results, with welfare gradually decreasing as the contribution floor rises. Table 4.4 shows the difference in terms of aggregate statistics between the current U.S. system and the new system. Output increases by 1.4%, labour supply by 1.1% and wages by 1%. Labour taxes are reduced by 3.6%. In terms of Consumption Equivalent Variation (CEV)[27], there is an increase of +0.7% between the two systems.

[26] In the longer-term, using an ex-ante perspective that is consistent with our social welfare function, we find that welfare is monotonically decreasing with the level and insensitive with the progressiveness of benefits.

[27] To obtain the CEV, we simulated the model for 30,000 periods. This statistic requires a very large simulation horizon to converge.

Aggregate differences	
Statistic	% Difference
Change in Output	+1.4%
Change in Capital	+2.2%
Change in Consumption	+1.2%
Change in Labour supply	+1.1%
Change in Wages	+1.0%
Change in Labour Taxes	−3.6%
Change in Bequests	+2.3%
Change in CEV	+0.7%

Table 4.4: Difference in aggregate statistics
Notes: The table presents the difference in aggregate statistics between the new and the current U.S. Social Security systems.

To understand the importance of labor supply responses, we follow Conesa et al. (2009) and decompose CEV to the component coming from changes in consumption and a component coming from changes in hours of work and bequests[28]. Households work more hours and receive higher bequests, with the former affecting welfare negatively and the latter positively. Also, labour responses are large enough to fully erase the gain stemming from the higher bequest transfers and reduce the component coming from consumption by roughly a third.

CEV Decomposition	
Statistic	% Difference
Change in CEV	+0.7%
Change in CEV_C	+1.1%
Change in $CEV_{L,B}$	−0.4%

Table 4.5: CEV decomposition
Notes: The table presents the decomposition of CEV in the components coming from the change in consumption and the change in leisure/bequest.

[28]As equation (1) shows, households draw utility from consumption, labour supply and bequests. Denoting by (c^0, h^0, beq^0) their benchmark allocation and by (c^1, h^1, beq^1) their allocation post reform, we obtain CEV_C and $CEV_{L,B}$ as the solutions of

$$\begin{aligned} \mathfrak{V}(c^1, h^0, beq^0) &= \mathfrak{V}(c^0(1 + CEV_C), h^0, beq^0) \\ \mathfrak{V}(c^1, h^1, beq^1) &= \mathfrak{V}(c^1(1 + CEV_{L,B}), h^0, beq^0) \end{aligned} \tag{4.8}$$

where $\mathfrak{V}$ is the social welfare function. It can be shown that the sum of the two components add to the total, $CEV = CEV_C + CEV_{L,B}$. In practise we obtain the values for CEV_C and $CEV_{L,B}$ numerically.

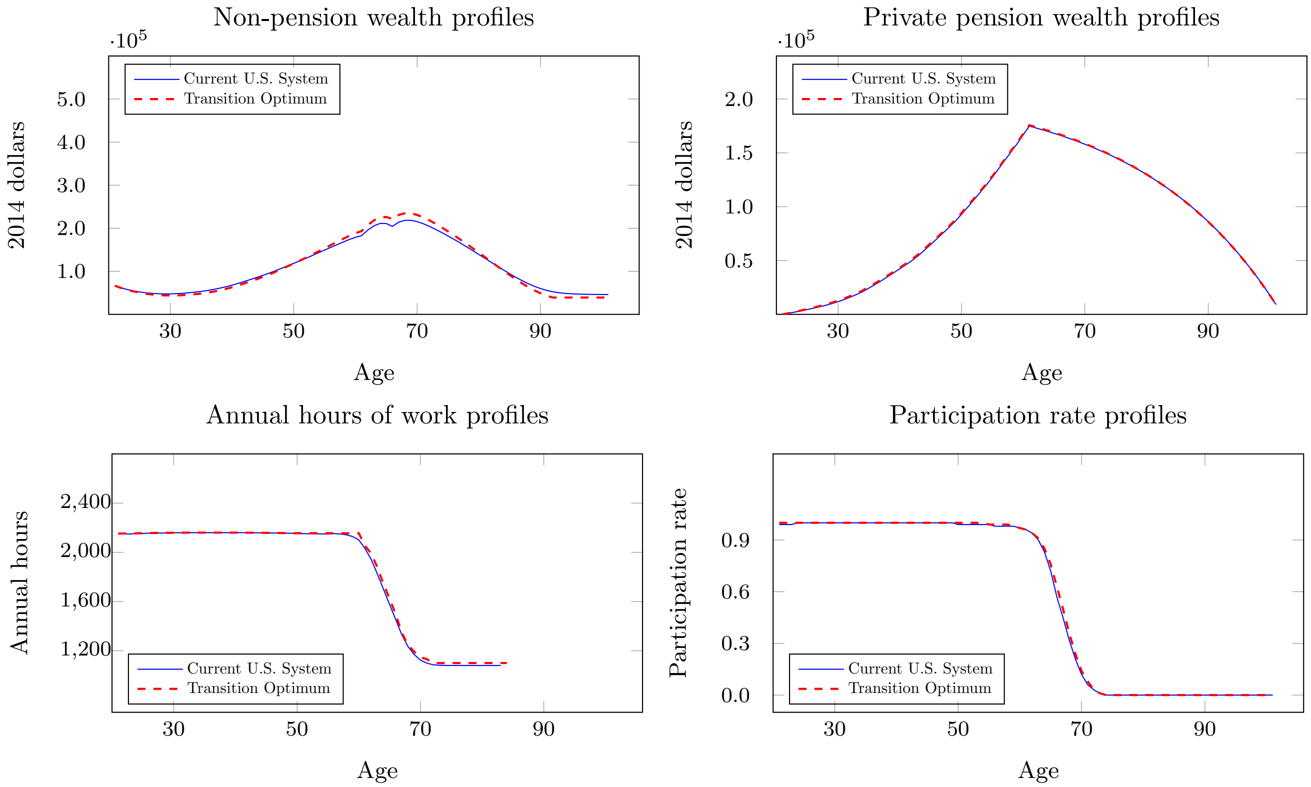


Figure 4.5: Life-cycle profiles for the status quo and the transition optimum
Notes: Comparison of the life-cycle profiles for the current and transition optimal Social Security systems. The top left sub-figure is showing non-pension wealth, the top right sub-figure pension wealth, the bottom left sub-figure annual hours of work and the bottom right sub-figure participation rates.

Figure 4.5 shows how the life-cycle profiles of the new system compare to the current U.S. system. Although the profiles for hours of work, participation rates and private pension wealth remain largely unchanged, households alter their savings decisions, by holding more capital close to their retirement and less once they are old. Note that Figure 4.5 shows the longer-term profiles, after the economy converges to the new stochastic steady state. For the transitional generations labor supply profiles differ significantly, with households working more hours and staying in the labor market longer to accumulate savings.

4.6.3 Political Support

From the previous analysis of the social welfare function, it is unclear for which reforms the government can gather political support. In this Section we consider the votes of households alive in period t_0 and evaluate the percent of positive votes for each potential tuple (ϕ_0, ϕ_1), and pension contribution floor. Figure 4.6 shows the percent of households that vote in favour of each reform. As emphasized by Conesa and Krueger (1999), there is a very strong status quo bias in favour of the current U.S. system.

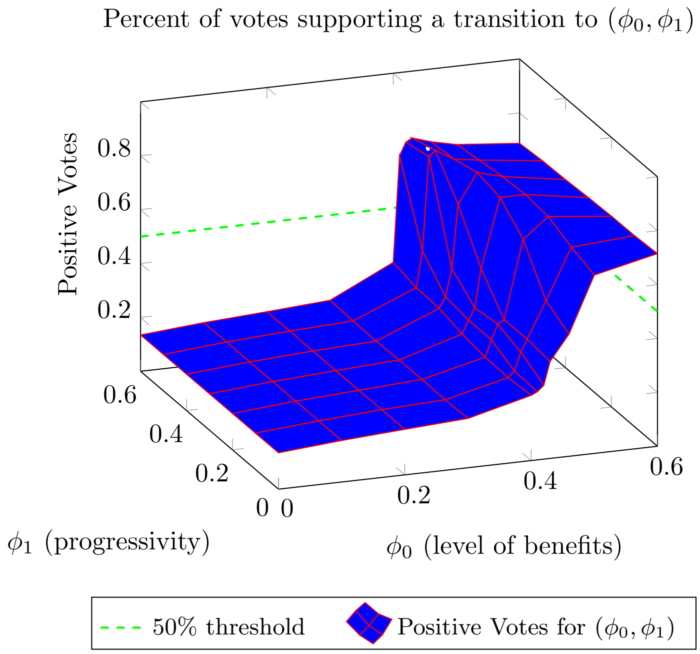


Figure 4.6: Household voting
Notes: This figure shows the percent of households voting in favour of a transition to a system characterised by the tuple (ϕ_0, ϕ_1). In this figure, we omit the transition to the status quo tuple $(\phi_0 = 0.41, \phi_1 = 0.54)$.

While households are unwilling to vote for reforms that reduce the level of benefits in the model economy, they are willing to vote for reforms that increase it. This is caused by i) the capital stock being partially consumed over the transition, ii) households enjoying higher Social Security benefits, but not living long enough to suffer the full consequences of the crowding out of savings and higher labour market distortions. Also, households are unwilling to vote for reforms that increase the Social Security benefits progressiveness, unless they are compensated by an increase in the level. However the drop in votes is not as steep as in the level decrease, as only households with high average index monthly earnings are affected by these reforms.

Note that every tuple receives a positive amount of votes. This highlights the competing interests of households with different ages. However, as the economy takes over 100 periods to converge to the new stochastic steady state, even young households close to the age of 40 are willing to vote against reforms that reduce the level of benefits.

4.6.4 Welfare increasing and politically feasible reforms

While political support take into account only households alive in period t_0, the social welfare function takes into account households alive in period t_0 and future generations. Hence, it is of interest to find the optimal Social Security system subject to both transition costs and politically feasibility constraints. Figure 4.7 shows which reforms are welfare increasing and which are politically feasible.

The blue shaded region shows reforms that the majority votes in favour of and the red shaded region shows reforms that are welfare increasing. The red shaded region with the black line pattern shows reforms that are welfare increasing in both the transition and the longer-term

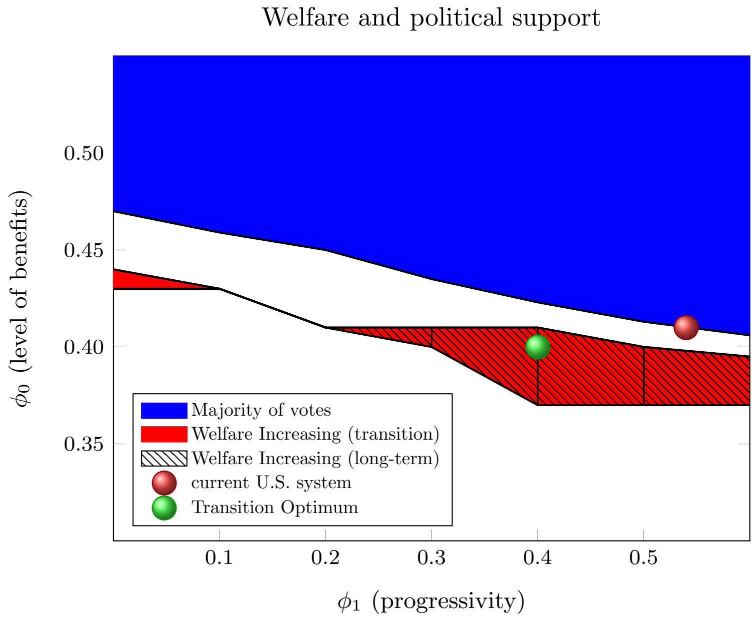


Figure 4.7: Welfare increasing and politically feasible reforms
The figure shows which reforms are welfare increasing and for which the government can gather political support. The blue shaded region denotes reforms that the majority of households vote in favour of. The red shaded region denotes reforms that increase welfare taking the transition into account. The black line pattern denotes reforms that increase welfare on both the transition and the longer-term.

Figure 4.7 shows that the intersection of the politically feasible and welfare increasing reform sets is equal to the empty set. As discussed in the previous subsections, although the U.S. can reduce the size of the Social Security system and increase welfare, the political feasibility constraint binds before such a move is possible. This results is mostly driven by large amount of risk in the model economy (Conesa and Krueger, 1999). Households are exposed to more aggregate and idiosyncratic risk once a reform takes place, but they don't live long enough to see the benefits.

4.7 Sensitivity Analysis

In this Section we present robustness checks to see if the model's results will alter with changes in the preference specification, technological growth rate and population growth rate.

Including technological progress In the benchmark specification, we use preferences that are inconsistent with a balanced growth path. To assess the impact of labour augmenting technological growth, we use the specification defined by

$$\mathbb{E}_t\left\{\sum_{\tau=0}^{J\text{-}j}\beta^{\tau}s_{j+\tau|j}\left[\frac{c_{t+\tau,j+\tau}^{1-\sigma}}{1-\sigma}+\xi_{t+\tau}^{1-\sigma}\frac{\chi}{1-\gamma}l_{t+\tau,j+\tau}^{1-\gamma}\right]\middle|X_t\right\} \tag{4.9}$$

where β is the subjective discount factor, σ is the coefficient of relative risk aversion, γ is the utility curvature with respect to leisure, χ is the weight of leisure relative to consumption and the expectation $\mathbb{E}_t$ is taken with respect to the cumulative distribution of future random events, conditional on all available information (X_t). Leisure $l_{t,j}$ is defined as $l_{t,j} = 1 - h_{t,j} - \theta_j^h \mathbb{I}[h_{t,j} > 0]$, which is the total time endowment minus work hours $h_{t,j}$ and an age dependent fixed cost of work θ_j^h. The term ξ_t grows at an exogenous constant rate and multiplying the disutility of work by $\xi_t^{1-\sigma}$ allows for a balanced growth path. Bequests in this specification are accidental. The production function is defined by

$$Y_t = F(K_t, \xi_t L_t; z_t) = z_t K_t^{\alpha} (\xi_t L_t)^{1-\alpha}$$

where ξ_t is the level of labour augmenting technology (King et al., 1988). We set the technological growth rate λ_ξ to 1.8%. We adjust the model's variables to account for technological growth and re-estimate the model subject to the constraint $\tilde{\beta} = \beta \lambda_\xi^{1-\sigma} < 1$. We find that there are small changes in the transition optimum. However, the political feasibility constraint again binds before any movements away from the current Social Security system are permitted.

Higher Frisch elasticity The estimates of Section 4.3 suggest that the Frisch elasticity of a 40 year old household is 0.38 which is in line with the micro literature. To generate realistic volatility in aggregate annual hours of work however, a larger Frisch elasticity is required. To address this, we adjust the coefficient that controls the utility function's curvature with respect to leisure (γ) and set it to set γ=6, so that the model generates an aggregate annual hours of work volatility that is approximately equal to the volatility of output, as the stylized business cycle facts suggest. Using the new value, the life-cycle profiles for hours of work and participation rates are affected, but results for welfare and voting have similar implications[29].

Lower population growth rate The model has a population growth rate of 1.1%, the U.S. population growth however is projected to decrease to 0.5% by 2050. Reducing the population growth rates, while keeping everything else constant does not affect the results or their interpretation significantly. Similar results are reported in Jones and Li (2020).

4.8 Conclusions

In this article we assess the optimal structure of the U.S. Social Security system, taking into account the current system's unfunded liabilities, transition dynamics and political feasibility constraints. To perform this assessment, we introduce an estimated overlapping generations general equilibrium model that features both aggregate and idiosyncratic risks. The model reflects the value of insurance, along with the distortions imposed by the existence and reform of a realistic Social Security system.

Results suggest that transition costs restrict the government's ability to move away from the current system. However, the government can implement a new Social Security system that generates welfare gains in both the

[29]See Imrohoroglu and Kitao (2009) and Bagchi (2015) for further analysis on the effects of labour supply elasticity and Social Security.

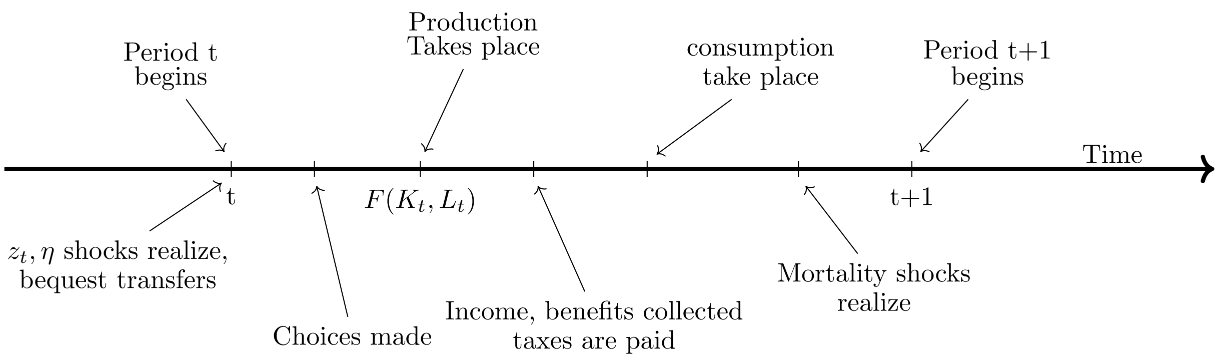
transition and the longer-term. The new system is smaller and more progressive than the current U.S. Social Security system. Finally, given the model's results, we create the sets of welfare increasing and politically feasible Social Security reforms. We find that the intersection between the two sets is equal to the empty set, a result that could explain the hesitance of multiple countries to implement Social Security reforms.

Although the model in this paper features a rich environment, it follows a macroscopic approach that comes at the cost of omitting factors on the household level and the government's degrees of freedom. On the household level, interesting avenues to extend this work include modeling i) myopia as in Moser and Olea de Souza e Silva (2019) ii) household joint decisions as in Nishiyama (2019), ii) mortality's dependence on household characteristics as in Jones and Li (2020) and iv) in addition to endogenous labour supply, the inclusion of safe and risky assets as much of the previous literature has done Krueger and Kubler (2006); Olovsson (2010); Harenberg and Ludwig (2019). In terms of the government's degrees of freedom, this work could be extended to include the role of government debt and smooth reform transitions, although Conesa and Krueger (1999) find that these reforms gather fewer votes as compared to once-and-for-all reforms, and could greatly increase the solution's computational cost.

4.9 Appendix A - Model details

4.9.1 Model Timing

The timing of the model is the following



Once period t begins, the shock z_t, η are realised and households receive bequests. Given the state of the economy and their characteristics vector, households make choices and supply labour and their capital to the firms. After production, they receive income, benefits and pay taxes. Then consumption / savings take place. Finally, households die stochastically before the end of the period.

4.9.2 Private Pension Payments

The private pension benefits of a household are calculated based on its remaining pension wealth and age. Denote by $1 + r$ the returns in the unconditional steady state of the economy. The benefits of a household with age $T - s$ and pension wealth n_{T-s} can be derived as the solution to the system of equations

$$\begin{aligned}
&n_{T-s}(1+r) - k = n_{T-s+1} \\
&\quad\vdots \\
&n_{T-2}(1+r) - k = n_{T-1} \\
&n_{T-1}(1+r) - k = n_T \\
&n_T(1+r) = k
\end{aligned}$$

The solution of (4.9.2) is a flat annuity payment that will exhaust pension wealth at age T, taking into account that after each payment, the remaining amount will compound in future periods. The value of the payment as a function of the household's current pension wealth and age is

$$k = \frac{n_{T-s}}{\sum_{j=1}^{s+1}(1+r)^{-j}}$$

As the economy is subject to aggregate shocks and returns fluctuate, the pension fund performs the computation (4.9.2) for each household, at the start of every period t and benefits are adjusted to reflect changes in pension wealth due to these fluctuations.

4.9.3 Social Security Payments

After the eligibility age, Social Security annual benefits (ssb_t) are calculated as a piece-wise linear function of Average Indexed Monthly Earnings (η).

$$ssb_t = \begin{cases} 0.9\eta, & \text{if } \eta \leq 9,132\$ \\ 9,132 + 0.35(\eta - 9,132), & \text{if } \eta > 9,132\$ \text{ \& } \eta \leq 55,032\$ \\ 55,032 + 0.15(\eta - 55,032), & \text{if } \eta > 55,032\$ \end{cases}$$

In Section 4.6, we approximate this piece-wise linear function using a flexible two-parameter function.

4.10 Appendix B - Computational details

Below we present the technical details of the computational problem. A general version of the code that allows for additional aggregate shocks and alternative solution methods is available online.

4.10.1 Household problem

Household problem solution algorithm: As the problem we consider is non-convex, we use the G^2EGM, where the solution algorithm of Carroll (2006) is supplemented with the upper envelope of Druedahl and Jørgensen (2017). The idea that non-convex models can be solved using first-order condition methods stems from early results of Clausen and Strub (2020).

Discretization: Due to the high persistence of the idiosyncratic productivity process, the corresponding state variable and its law of motion are discretized following Rouwenhorst (1995). See Kopecky and Suen (2010) for performance comparisons with alternative discretization methods.

Interpolation method: Throughout the solution and simulation of the model, multidimensional linear interpolation is used. To reduce interpolation errors, the marginal utility of consumption is transformed using $f(x) = 1/x$.

Grid definition: The baseline grid sizes (discrete variables included) for the steady state model are $\#m \times \#n \times \#\eta \times \#ss \times \#J = 70 \times 12 \times 8 \times 12 \times 81$ and for the aggregate uncertainty model they are $\#m \times \#n \times \#\eta \times \#K_1 \times \#K_2 \times \#z \times \#ss \times \#J = 70 \times 12 \times 8 \times 10 \times 10 \times 2 \times 12 \times 81$.

Machine: The algorithm was written and executed using JuliaPro-1.5.4-1.

Parallelism: We use Julia's multi-threading. We solve the household's problem in parallel conditional on age (j) and simulate households in parallel conditional on year (t).

Simulation: Households are simulated following the non-stochastic method of Young (2010).

4.10.2 Algorithm accuracy checks

Aggregate law of motion accuracy: Following Den Haan (2010) instead of using R^2 that looks at the one period ahead forecast, we condition on an initial value and generate the entire 10,000 period history[30] of i) aggregate capital (the 1st moment), ii) aggregate average index monthly earnings (the 2nd moment), iii) aggregate labour supply and iv) labour taxes, using their forecast laws of motion. We compare the forecasts with the history generated by the model's non-stochastic simulation. Below we provide three statistics on their difference and a graph of the comparison. The statistics are the maximum absolute error, the mean absolute error and the root mean squared error.

max abs-K1 error	mean abs-K1 error	rmse-K1 error
0.60%	0.08%	0.49%
max abs-K2 error	mean abs-K2 error	rmse-K2 error
0.81%	0.17%	0.23%
max abs-L error	mean abs-L error	rmse-L error
0.64%	0.08%	0.51%
max abs-τ^l error	mean abs-τ^l error	rmse-τ^l error
0.90%	0.23%	0.34%

Table 4.6: The table shows three statistics on the household forecast errors.

Relative Euler-residuals: To measure accuracy, following Judd (1992) and Santos (2000) we consider the common logarithm of the relative absolute Euler-errors ($\mathfrak{E}$). As the Euler-Equation does not have to be true on the constrained region, a lower bound on savings, $\underline{a} = 0.002$ is imposed.

$$\mathfrak{E} = \frac{\sum_{i=1}^{N}\sum_{1}^{T}\mathbb{I}[a_{i,t+1} > \underline{a}]error_{i,t}}{\sum_{i=1}^{N}\sum_{1}^{T}\mathbb{I}[a_{i,t+1} > \underline{a}]}$$

$$error_{i,t} = log_{10}\left(\left|c_{i,t} - \left(\beta(1+r)\,\mathbb{E}[c_{i,t+1}^{-\sigma}]\right)^{-\frac{1}{\sigma}}\right| \Big/ c_{it}\right)$$

We find that throughout the estimation and optimal design the errors are in the range of $\mathfrak{E} \in (-5, -3)$.

[30]Note that in life-cycle models weaker versions of the three displayed statistics can be used. The households have finite lives and only need to forecast prices for the remaining of their lives. Accumulation of errors beyond their remaining years does not affect their behaviour.

Comparisons to VFI: The G^2EGM solution (given the estimated coefficients) is compared to the solution obtained by standard Value Function Iteration (VFI). As VFI is inherently less accurate, the grid sizes used were $\#m \times \#n \times \#\eta \times \#K_y \times \#K_2 \times \#z \times \#ss \times \#J = 400 \times 30 \times 10 \times 10 \times 10 \times 2 \times 20 \times 81$ and $\#a' \times h \times b = 4000 \times 3 \times 3$. The two solutions are nearly identical, with the VFI producing larger relative Euler errors.

Increasing the Grid size: The problem was re-solved (given the estimated coefficients) using G^2EGM and larger grid sizes. Increasing the size by a factor of two does not significantly affect the relative Euler errors or the equilibrium statistics.

Longer simulation horizon: For the model's estimation we simulate 10,000 overlapping generations. Increasing the number of generations does not affect the conditional means (lifecycle profiles). For the welfare analysis we simulate 25,000 overlapping generations. Increasing the number of generations does not affect results.

Alternative minimization algorithm: We verified the estimation results obtained using the BOBYQUA algorithm of Powell (2009) with the Simplex algorithm of Nelder and Mead (1965). The difference in the estimated coefficients (and their standard errors) is small.

Deterministic Steady State algorithm

In most cases the deterministic steady state algorithm requires 50 to 90 iterations to converge. The algorithm proceeds as follows.

Step 1. We start with an initial guess for factor prices and the bequest vector

Step 2. We solve the household problem taking the factor prices and bequest vector as given. Then we simulate the model

Step 3. Given the model's simulation we evaluate the new factor prices and bequest vector

Step 4. We take a convex combination of the old and new prices and go back to *Step* 2. This is repeated until the factor prices and bequest vector converge

Aggergate Uncertainty algorithm

Step 1. We start with an initial guess for the law of motion coefficients. We obtain these guesses from the deterministic steady state algorithm

Step 2. We solve the household problem taking the laws of motion as given. Then we simulate the model for a large number of periods. The distribution of households and bequests in the initial simulation period are equal to the deterministic steady state distributions. We discard the first 300 periods.

Step 3. Given the model's simulation we evaluate the new law of motion coefficients. First we use OLS on the simulated data and then take a convex combination between the old and new coefficients. For every simulation we evaluate every relevant accuracy tests.

Step 4. We back to *Step* 2. This is repeated until the law of motion coefficients converge

Transition path algorithm

Step 1. We solve the benchmark economy with the current U.S. Social Security system in place and store the distribution of households over their characteristics state vector. We do this for a large number of periods all 10 years apart from each other.

Step 2. We solve for the new Social Security system described by (ϕ_0, ϕ_1) and a value for the minimum pension contribution floor. Then we store the new law of motion coefficients.

Step 3. We commence a simulation for every distribution stored in *Step* 1 and the new law of motion coefficients obtained in *Step* 2. We simulate forward for a large number of periods. For every simulation we evaluate every relevant accuracy tests.

Step 4. We evaluate the social welfare function for each of the simulations and then evaluate the average.

4.11 Appendix C - Estimation details

4.11.1 First step

Idiosyncratic Productivity GMM estimation This subsection presents the GMM estimation procedure for the parameters that govern idiosyncratic productivity. Empirically, we specify the model as

$$\begin{aligned} \eta_{i,t} &= \alpha_{i,t} + m_{i,t} \\ \alpha_{i,t} &= \rho\alpha_{i,t-1}, + \zeta_{i,t}, \qquad && \zeta_{i,t} \sim \mathcal{N}(0, \sigma_\zeta^2) \\ m_{i,t} &= \mu_{i,t} + \phi\mu_{i,t-1} \ , && \mu_{i,t} \sim \mathcal{N}(0, \sigma_\mu^2) \end{aligned}$$

where by assumption the errors $\mu_{i,t}$ and $\zeta_{i,t}$ are serially uncorrelated and mutually orthogonal. We interpret $\alpha_{i,t}$ as the "true" process that evolves according to an $AR(1)$ and $m_{i,t}$ as a $MA(1)$ measurement error. We estimate the four-element parameter vector $\theta = (\rho, \phi, \sigma_\zeta^2, \sigma_\mu^2)$ by minimizing the weighted distance between the theoretical and empirical variance-covariance matrices of the random variable[31] $\eta_{i,t}$. As the dataset contains $T = \#\{1999, 2001, ..., 2019\} = 11$ waves, the variance-covariance matrices have $T(T+1)/2 = 66$ unique elements. Each

[31]To be precise it is a random vector $(\eta_{i,1999}, \eta_{i,2001}, ...\eta_{i,2019})'$.

difference between these unique element will be used to form a moment condition. As the panel is unbalanced, we follow French and Jones (2004) and denote by n the total number of observations present in any wave and by $n_{t,t+s} = \sum_{i=1}^{n} \mathbb{I}[\eta_{i,t}, \eta_{i,t+s} \neq missing]$ the number of observations present in both waves t and $t+s$. Using $\sigma_\alpha^2 = \sigma_\zeta^2/(1-\rho^2)$, the sample moment conditions are defined as

$$
\begin{aligned}
m(\theta)_{t,t} &= \frac{n_{t,t}}{n}\left[\frac{1}{n_{t,t}} \sum_{i:\eta_{i,t}\neq missing} \left(\eta_{i,t}^2 - \frac{1}{1-\rho^2}\sigma_\zeta^2 + \sigma_\mu^2 + \phi^2\sigma_\mu^2 \right)\right] \mathbb{I}[\eta_{i,t} \neq missing] \\
m(\theta)_{t,t+1} &= \frac{n_{t,t+1}}{n}\left[\frac{1}{n_{t,t+1}} \sum_{\substack{i:\ \eta_{i,t},\ \neq missing \\ \eta_{i,t+1}}} \left(\eta_{i,t}\eta_{i,t+1} - \frac{\rho}{1-\rho^2}\sigma_\zeta^2 + \phi\sigma_\mu^2 \right)\right] \mathbb{I}[\eta_{i,t}, \eta_{i,t+1} \neq missing] \\
m(\theta)_{t,t+\kappa} &= \frac{n_{t,t+\kappa}}{n}\left[\frac{1}{n_{t,t+\kappa}} \sum_{\substack{i:\ \eta_{i,t},\ \neq missing \\ \eta_{i,t+\kappa}}} \left(\eta_{i,t}\eta_{i,t+\kappa} - \frac{\rho^\kappa}{1-\rho^2}\sigma_\zeta^2 \right)\right] \mathbb{I}[\eta_{i,t}, \eta_{i,t+\kappa} \neq missing]
\end{aligned}
$$

where $\kappa > 1$. The GMM estimate $\hat{\theta}$ is defined as

$$\hat{\theta} = \underset{\theta\in\Theta}{argmin}\Big\{\ n\ g(w_i,\theta)'\ \mathbb{W}\ g(w_i,\theta)\ \Big\}$$

where $g(w_i, \theta)$ is a $T \times 1$ vector of the stacked sample moment conditions and $\mathbb{W}$ is a weighting matrix. Using the optimal weighting matrix $\mathbb{W} = \mathbb{E}[g(w_i,\theta)'g(w_i,\theta)]^{-1}$ the estimator is asymptotically normal with the limiting distribution

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(\,0, (G'\mathbb{W}G)^{-1}\,)$$

where $G = \nabla_\theta\, \mathbb{E}[g(w_i,\theta)]$ denotes the Jacobian matrix of the sample moment condition vector. In practise, we perform the estimation using the two-step feasible GMM procedure and locate the global minimum with the Nelder-Mead (Simplex) implementation of the NLopt package. As a robustness check, we re-estimate the model using the diagonal weighting matrix suggested by Pischke (1995), the identity matrix and a bootstrapped empirical diagonal weighting matrix, where each moment is weighted by the inverse of its variance (resampling from the panel's cross-section dimension as suggested by Kapetanios (2008)). We find that the choice of the weighting matrix does not have a significant effect on the transition probabilities or the grid-vector for η, once the process is discretized.

In addition, measurement error's specification does not seem to have a significant effect on the transition probabilities or the grid-vector for η. For example, although heteroscedastic measurement error ($\sigma_{\mu,t}^2\ \forall t$) improves the model's fit and has a lower Sargan-Hansen $\mathfrak{J}$-statistic, it leaves the $AR(1)$ parameters unchanged.

4.11.2 Second step

Selection correction For this subsection, define $W_{i,t} = log(w_{i,t}) - \alpha^h log(h_{i,t})$. The objects of interest in the second part of Section 4.3.1 are $\mathbb{E}[W_{i,t}|j]$ and the residuals $\eta_{i,t} = W_{i,t} - \mathbb{E}[W_{i,t}|j]$. The FE estimator however, instead of $\mathbb{E}[W_{i,t}|j]$, identifies $\mathbb{E}[W_{i,t}|j, h_{i,t} > 0] = \mathbb{E}[W_{i,t}|j] + \mathbb{E}[\eta_{i,t}|j, h_{i,t} > 0]$ and the FE residuals are $\eta_{i,t}^{FE} =$

$W_{i,t} - \mathbb{E}[W_{i,t}|j, h_{i,t} > 0] = \eta_{i,t} - \mathbb{E}[\eta_{i,t}|j, h_{i,t} > 0]$ and thus $\eta_{i,t}^{FE} \neq \eta_{i,t}$. According to this inequality, in the general case, in addition to the life-cycle wage profile, the persistence and variance parameters that govern idiosyncratic productivity will be affected by the selection bias as well. In this article we extend the wage profile selection-correction method of French (2005), so that it corrects the persistence and variance parameters as well. The modified algorithm is the following

1. Estimate the model using the biased FE estimates.
2. Using the model simulated data where both. workers and non-workers are observed, the wage profiles are iteratively adjusted until the simulated wage profile, conditional on being employed in the model world, is equal to the estimated FE profile.
3. Once the fixed point of 2. is reached, use the new wage profiles and the PSID data to obtain new residuals and estimate the AR(1)+MA(1) model.
4. Estimate the model using the fixed-point wage profile of 2. and the AR(1)+MA(1) estimates of 3.
5. Repeat steps 2. to 4. until the life-cycle wage profile and AR(1)+MA(1) estimates converge.

At the end of this iterative process, the life-cycle profile of wages is equal between the FE estimates and the model simulated data, conditionally on being employed in the model world and the parameters that govern idiosyncratic productivity are estimated using selection-corrected regression residuals.

Bibliography

Aaronson, D. and E. French (2004). The effect of part-time work on wages: Evidence from the social security rules. *Journal of Labor Economics 22*(2), 329–352.

Altonji, J. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy 94*(3), S176–S215.

Altonji, J. G. and L. M. Segal (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics 14*(3), 353–366.

Auerbach, A., R. Auerbach, and L. Kotlikoff (1987). *Dynamic Fiscal Policy.* Cambridge University Press.

Bagchi, S. (2015). Labor supply and the optimality of social security. *Journal of Economic Dynamics and Control 58*, 167–185.

Ball, L. and N. G. Mankiw (2007, August). Intergenerational Risk Sharing in the Spirit of Arrow, Debreu, and Rawls, with Applications to Social Security Design. *Journal of Political Economy 115*(4), 523–547.

Bell, F. C. and M. L. Miller (2005). *Life tables for the United States Social Security Area 1900-2100.*

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). Female labor supply, human capital, and welfare reform. *Econometrica 84*(5), 1705–1753.

Blundell, R., E. French, and G. Tetlow (2016). Retirement Incentives and Labor Supply. In J. Piggott and A. Woodland (Eds.), *Handbook of the Economics of Population Aging*, Volume 1 of *Handbook of the Economics of Population Aging*, Chapter 0, pp. 457–566. Elsevier.

Blundell, R. and T. Macurdy (1999). Chapter 27 - labor supply: A review of alternative approaches. Volume 3 of *Handbook of Labor Economics*, pp. 1559–1695. Elsevier.

Caliendo, F. N., N. L. Guo, and R. Hosseini (2014). Social security is not a substitute for annuity markets. *Review of Economic Dynamics 17*(4), 739–755.

Carroll, C. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics Letters 91*(3), 312–320.

Clausen, A. and C. Strub (2020). Reverse calculus and nested optimization. *Journal of Economic Theory 187*, 105019.

Conesa, J. C., S. Kitao, and D. Krueger (2009, March). Taxing capital? not a bad idea after all! *American Economic Review 99*(1), 25–48.

Conesa, J. C. and D. Krueger (1999). Social security reform with heterogeneous agents. *Review of Economic Dynamics 2*(4), 757–795.

Den Haan, W. J. (2010). Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents. *Journal of Economic Dynamics and Control 34*(1), 79–99. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty.

Druedahl, J. and T. H. Jørgensen (2017). A general endogenous grid method for multi-dimensional models with non-convexities and constraints. *Journal of Economic Dynamics and Control 74*, 87 – 107.

Fehr, H. and F. Kindermann (2015). Taxing capital along the transition-not a bad idea after all? *Journal of Economic Dynamics and Control 51*, 64–77.

Feldstein, M. (1985). The optimal level of social security benefits. *The Quarterly Journal of Economics 100*(2), 303–320.

Feldstein, M. and J. Liebman (2002). Social security. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics* (1 ed.), Volume 4, Chapter 32, pp. 2245–2324. Elsevier.

Fella, G. (2014, April). A generalized endogenous grid method for non-smooth and non-concave problems. *Review of Economic Dynamics 17*(2), 329–344.

French, E. (2005, 04). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *The Review of Economic Studies 72*(2), 395–427.

French, E. and J. Jones (2012, April). Public pensions and labor supply over the life cycle. *International Tax and Public Finance 19*(2), 268–287.

French, E. and J. B. Jones (2004). On the distribution and dynamics of health care costs. *Journal of Applied Econometrics 19*(6), 705–721.

Goda, G. S., J. B. Shoven, and S. N. Slavov (2011). Implicit Taxes on Work from Social Security and Medicare. *Tax Policy and the Economy 25*(1), 69–88.

Golosov, M., A. Shourideh, M. Troshkin, and A. Tsyvinski (2013, May). Optimal pension systems with simple instruments. *American Economic Review 103*(3), 502–07.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Gourinchas, P.-O. and J. A. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

Harenberg, D. and A. Ludwig (2019). Idiosyncratic risk, aggregate risk, and the welfare effects of social security. *International Economic Review 60*(2), 661–692.

Hong, J. H. and J.-V. Rios-Rull (2007, January). Social security, life insurance and annuities for families. *Journal of Monetary Economics 54*(1), 118–140.

Hosseini, R. and A. Shourideh (2019). Retirement financing: An optimal reform approach. *Econometrica 87*(4), 1205–1265.

Huang, H., S. Imrohoroglu, and T. J. Sargent (1997). Two computations to fund social security. *Macroeconomic Dynamics 1*(1), 7–44.

Huggett, M. and J. Parra (2010). How well does the u.s. social insurance system provide social insurance? *Journal of Political Economy 118*(1), 76–112.

Imrohoroglu, A., S. Imrohoroglu, and D. H. Joines (1995, June). A Life Cycle Analysis of Social Security. *Economic Theory 6*(1), 83–114.

Imrohoroglu, A., S. Imrohoroglu, and D. H. Joines (1998, September). The Effect of Tax-Favored Retirement Accounts on Capital Accumulation. *American Economic Review 88*(4), 749–768.

Imrohoroglu, S. and S. Kitao (2009, August). Labor supply elasticity and social security reform. *Journal of Public Economics 93*(7-8), 867–878.

Jones, J. and Y. Li (2020, 07). Social security reform with heterogeneous mortality. *Federal Reserve Bank of Richmond Working Papers 20*, 1–65.

Judd, K. L. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory 58*(2), 410 – 452.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal 11*(2), 377–395.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988). Production, growth and business cycles : I. The basic neoclassical model. *Journal of Monetary Economics 21*(2-3), 195–232.

Kitao, S. (2010). Individual retirement accounts, saving and labor supply. *Economics Letters 108*(2), 197–200.

Kitao, S. (2014). Sustainable social security: Four options. *Review of Economic Dynamics 17*(4), 756–779.

Kopecky, K. and R. M. H. Suen (2010). Finite state markov-chain approximations to highly persistent processes. *Review of Economic Dynamics 13*(3), 701–714.

Krueger, D. and F. Kubler (2006, June). Pareto-improving social security reform when financial markets are incomplete!? *American Economic Review 96*(3), 737–755.

Krusell, P. and A. A. J. Smith (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Ludwig, A. and N. Grevenbrock (2021). The optimal joint design of unemployment and pension insurance.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy 89*(6), 1059–1085.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*(5), 995–1026.

Moser, C. and P. Olea de Souza e Silva (2019). Optimal Paternalistic Savings Policies. Technical report.

Ndiaye, A. (2018). Flexible Retirement and Optimal Taxation. Technical report.

Nelder, J. A. and R. Mead (1965, 01). A Simplex Method for Function Minimization. *The Computer Journal 7*(4), 308–313.

Nishiyama, S. (2011). The budgetary and welfare effects of tax-deferred retirement saving accounts. *Journal of Public Economics 95*(11), 1561–1578. Special Issue: International Seminar for Public Economics on Normative Tax Theory.

Nishiyama, S. (2015). Fiscal policy effects in a heterogeneous-agent olg economy with an aging population. *Journal of Economic Dynamics and Control 61*, 114–132.

Nishiyama, S. (2019). The joint labor supply decision of married couples and the u.s. social security pension system. *Review of Economic Dynamics 31*, 277–304.

Nishiyama, S. and K. Smetters (2007). Does social security privatization produce efficiency gains? *The Quarterly Journal of Economics 122*(4), 1677–1719.

Nishiyama, S. and K. Smetters (2008, September). The Optimal Design of Social Security Benefits. Working Papers wp197, University of Michigan, Michigan Retirement Research Center.

O'Dea, C. (2018). Insurance, efficiency and the design of public pensions.

Olovsson, C. (2010). Quantifying the risk-sharing welfare gains of social security. *Journal of Monetary Economics 57*(3), 364 – 375.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Peterman, W. B. and K. Sommer (2019a). A historical welfare analysis of social security: Whom did the program benefit? *Quantitative Economics 10*(4), 1357–1399.

Peterman, W. B. and K. Sommer (2019b). How well did social security mitigate the effects of the great recession? *International Economic Review 60*(3), 1433–1466.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business & Economic Statistics 13*(3), 305–314.

Powell, M. (2009, 01). The bobyqa algorithm for bound constrained optimization without derivatives. *Technical Report, Department of Applied Mathematics and Theoretical Physics*.

Rouwenhorst, G. (1995). Asset pricing implications of equilibrium business cycle models. *T.F. Cooley (Ed.), Frontiers of Business Cycle Research, Princeton University Press*.

Santos, M. S. (1999). Chapter 5 numerical solution of dynamic economic models. Volume 1 of *Handbook of Macroeconomics*, pp. 311–386. Elsevier.

Santos, M. S. (2000). Accuracy of numerical solutions using the euler equation residuals. *Econometrica 68*(6), 1377–1402.

Shiller, R. J. (1999). Social security and institutions for intergenerational, intragenerational, and international risk-sharing. *Carnegie-Rochester Conference Series on Public Policy 50*, 165 – 204.

Young, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control 34*(1), 36–41.