Aggregate Effects of Public Health Insurance Expansion: The Role of Delayed Medical Care*

Mitchell VanVuren

This Version: Jan 2022

Please click here for the latest version

Abstract

A substantial body of evidence suggests that many U.S. adults delay medical care until after age 65 when they become eligible for Medicare. In this paper, I study the aggregate consequences of expanding public health insurance access for younger individuals, accounting for the subsequent reduction in delayed care. I focus on two main channels. First, expanding public health insurance can reduce delayed care, resulting in long-run cost savings, since early treatment tends to be less expensive than later treatment. Second, expanding public insurance can raise the total number of people over age 65, raising long-run costs, since earlier care tends to reduce mortality. Both channels raise welfare from an ex-ante perspective, but the second leads to larger increases in distortionary taxation. To study these channels, I construct a heterogeneous-agent overlapping generations general-equilibrium model featuring health investment, endogenous mortality, and public and private health insurance. I estimate the model to match quasi-experimental evidence on the extent of delayed medical care in older U.S. adults and on the effects of the 2014 ACA Medicaid expansion on mortality. Both channels are quantitatively important in determining the long-run costs of expansion; however, the cost savings of the first outweigh the cost increases of the second, reducing long-run costs and the need for distortionary taxes.

*I thank my advisers Valerie Ramey and David Lagakos for their comments and help on this project. I also thank Karen Kopecky, Roozbeh Hosseini, Mariachristina De Nardi, Johannes Weiland, Titan Alon, Jeff Clemens, Itzik Fadlon, Katherine Meckel, Rebecca Lian, Jacob Orchard, Tyler Paul, Minki Kim, and various seminar participants at UCSD for their helpful comments. Any errors, of course, remain my own. Mitchell VanVuren: Department of Economics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093

1. Introduction

A substantial body of evidence suggests that a large fraction of U.S. adults delay medical care until after age 65 when they become eligible for Medicare. For example, Card, Dobkin and Maestas (2008) document that a whole host of medical procedures – from doctor's visits to heart surgery to gall bladder removals – jumps discretely at age 65. Furthermore, McWilliams et al. (2003) show that the use of testing services – such as cholesterol, mammography, and prostate examination – rise substantially for uninsured individuals right after they turn 65, and Patel et al. (2021) show that cancer diagnoses rise substantially at age 65, particularly for early-stage cancers.

Delayed medical care carries potentially large financial costs. According to the Center for Disease Control, nearly 80 percent of adults approaching Medicare eligibility (ages 55-64) have been diagnosed with at least one chronic health condition and 37 percent have been diagnosed with at least two. Many medical studies (e.g. Gehi et al., 2007; Herkert et al., 2019; Fukuda and Mizobe, 2017) show that delaying treatment of these conditions risks both the individual's life and can lead to higher eventual treatment costs as the disease progresses. As an illustrative example, mild cases of coronary artery disease (i.e. the build-up of plaque in one's arteries) can be treated at relatively low cost through medications that help prevent or reduce the blockage of arteries. Atorvastatin, the generic version of popular anti-cholesterol medication Lipitor, costs roughly \$20 per month. However, if a mild case worsens due to delayed care, it can require surgical treatment, such as bypass surgery, which carries an average cost of \$169,000.[1] Some portion of those who delay care do so with deadly consequence. Miller, Johnson and Wherry (2021) show that receiving Medicaid coverage reduced all-cause mortality of low-income individuals aged 55-64 by 10 percent, consistent with the notion that these individuals are delaying important medical treatment when uninsured.

In this paper, I study the aggregate consequences of delayed medical care for public health insurance expansion in the spirit of the macro literature on health (e.g. De Nardi, French and Jones, 2016). I focus on two channels. First, expanding public health insurance can reduce delayed care since early treatment tends to be less expensive than later treatment, resulting in long-run cost savings. Second, expanding public insurance can raise the total number of people over age 65 since earlier care tends to reduce mortality, but this raises long-run costs. Both effects are ex-ante welfare increasing, but the second requires increases in distortionary taxation while the first can reduce necessary taxes.

To study these channels, I construct a heterogeneous-agent overlapping generations general equilibrium model featuring health investment and endogenous mortality. Following much of the literature on macroeconomics and health, individuals build and maintain health capital through medical spending each period (see Fang and Krueger, 2021, for an overview). Health capital reduces an individual's mortality risk, as in Ozkan (2014), as well as their chance of experiencing a costly health emergency. Individuals face a choice between purchasing health insurance or not which leaves some low-income individuals uninsured until they receive Medicare at age 65. For uninsured individuals approaching age 65, the optimal strategy is indeed one of delaying healthcare spending; they treat their health as an asset and substitute consumption for medical care, running down their health capital. After turning 65 and gaining health insurance, these individuals compensate for their period of low spending through higher use of medical care; however, some individuals die or end up needing more expensive medical care as a result of delaying care.

Expansion of public health insurance reduces incentives to delay care but requires increases in distortionary taxation, reducing output. Additionally, the model features two production sectors, one for consumption goods and the other for medical goods, with upwards sloping supply curves. The increase in demand for medical goods induced by public health insurance expansion increases the price of healthcare goods. In addition to paying higher taxes, individuals who

[1]The cost of Atorvastatin is taken from the following website: https://www.drugs.com/price-guide/atorvastatin#oral-tablet-20-mg. The average cost of bypass surgery comes from Benjamin et al. (2018).

are not beneficiaries of expansion must also pay higher prices for medical goods.

I estimate the model using the simulated method of moments to match two key quasi-experimental microeconomic studies from the health literature. Card, Dobkin and Maestas (2008) use administrative data from hospitals in California, New York, and Florida and a regression discontinuity framework to show that the number of medical procedures that individuals receive jumps discretely by 48 percent at age 65. I replicate this experiment by taking the percent change in average medical spending between individuals aged 55 and 64 in a steady-state of the model corresponding to the United States before the Affordable Care Act (ACA) and use this change as a target in model estimation. I also leverage the experiment of Miller, Johnson and Wherry (2021) which uses a diff-in-diff framework across states to study the change in mortality for low-income individuals aged 55 to 64 due to the ACA Medicaid expansion. They find that mortality declined by 10 percent for their sample. I reproduce this experiment in the model by beginning in a pre-ACA steady-state and simulating two alternative paths forward. In the first path, I expand Medicaid as in the ACA while in the second, I change nothing and the model remains in steady-state. The difference in mortality rates for low-income individuals between ages 55 and 64 corresponds to the diff-in-diff estimator of mortality and is used as a target in model estimation. The remaining parameters related to health are estimated, often directly, from individual health and healthcare expenditure data from the Medical Expenditure Panel Survey (MEPS).

To validate the model, I examine how it performs in replicating features of the data that were not targeted in calibration. The model successfully reproduces the strong left-skew of the cross-sectional distribution of health as measured by Hosseini, Kopecky and Zhao (2021) as well as broadly replicating the relationship between healthcare spending and age.

To test the model's core mechanism, I use data from before and after the 2014 ACA Medicaid expansion to validate the model's prediction that public health insurance expansion should decrease the amount of delayed care. In particular, the model predicts that an expansion similar in size to the ACA Medicaid expansion should decrease the jump in healthcare expenditure observed at age 65 by 17.6 percent. Using MEPS data from 2010 to 2012 and a regression discontinuity design I find that average medical expenditure jumps by 25.3 percent ($p<0.01$) at age 65 before the ACA. I then repeat this procedure on data from 2017 to 2019, after the expansion, and find that expenditure only jumps by 10.1 percent (statistically insignificant). In the data, the reduction in the jump is 15.2 percent, very close to the 17.6 percent predicted by the model. Although this result carries the caveat that standard errors are large and the difference between the pre- and post-expansion point estimates is insignificant ($p=0.17$), I interpret this as suggestive evidence for the model's core mechanism.

I use the estimated model to evaluate the impact of an expansion in Medicaid similar to the 2014 ACA expansion funded by an increase in distortionary taxes. I focus on both the immediate costs of expansion as well as the long-run cost after the dynamics of the delayed spending channel have converged to steady-state. I also focus on the speed of transition and, in particular, on how many of the long-term costs and savings manifest within the first 10 years after the policy change (consistent with the Congressional Budget Office evaluation window of 10 years).

Medicaid expansion successfully reduces delayed care, particularly for individuals between ages 60 and 64 who are approaching the Medicare qualification threshold of 65. For these individuals, average annual medical investment increases by 13 percent. This reduction in delayed care has two effects. First, these individuals are healthier and, as a result, spend less on medical care as they age, leading to less Medicare expenses. Their lifetime medical spending after age 65 decreases by 2.7 percent. Second, these individuals live longer; they are roughly 0.5 percentage points more likely to survive to age 65. This increases Medicare expenses as the program now covers individuals who would've died before receiving coverage before expansion.

To separate the impact of these two channels on total Medicare expenses, I use two counterfactual modifications of the baseline model. In the first model, I replace the endogenous mortality process with an exogenous one. In other words, an individual's mortality risk no longer depends on their health status and is simply a deterministic function of their age. In this model, expanding Medicaid no longer reduces mortality and thus no longer increases Medicare costs by increasing the number of individuals who survive to receive Medicare coverage. As a result, the difference in post-expansion Medicare costs between this first counterfactual model and the baseline model tells us about the total increase in Medicare costs due to lower mortality.

In the second counterfactual model, I also replace the endogenous choice of health investment with an exogenous health expenditure function. In this model, expanding Medicaid no longer reduces delayed care and no longer leads to reductions in medical spending by individuals older than 65 as such spending is determined by the same exogenous process both before and after expansion. The difference in post-expansion Medicare costs between this model and the first counterfactual model thus tells us about the total decrease in Medicare costs stemming from reductions in late-in-life medical spending due to more efficient early care.

Overall, I find that cost savings due to reductions in late-in-life spending substantially outweigh the increase in costs due to lower mortality. For every \$100 spent on Medicaid expansion, there is a net decrease in Medicare costs of \$49 (undiscounted) resulting in a spending-to-savings ratio of 0.49. This decrease in costs is the combined result of an increase in costs due to lower mortality (and thus an increase in the size of the Medicare population) and a decrease due to more efficient, earlier care. The mortality channel results in an increase in Medicare expenditure of \$7.24 for each \$100 spent on Medicaid expansion while the delayed care channel results in a decrease in Medicare expenditure of \$56.52 for each \$100 spent. Both channels are quantitatively important, but the reduction in costs due to less delayed care outweighs the increase in costs due to lower mortality, reducing the net cost of Medicaid expansion.

Expansion decreases average welfare by 0.4 percent of consumption, but a small portion of the population gains substantially. Welfare gains are concentrated entirely among new insurance recipients; those who gain new access to Medicaid experience welfare gains as large as 6 percent of consumption. Mortality reduction has a substantial impact on welfare with gains in life expectancy accounting for roughly one-third of the welfare gains. Non-recipients, including the very poor who qualify for Medicaid before expansion, lose about 1 percent of consumption through higher taxes and higher healthcare prices. Unsurprisingly, welfare gains are ex-post heterogeneous. Individuals who gain coverage and subsequently experience a bad series of lifetime health shocks receive up to 10 percent in consumption equivalent welfare while those who gain coverage but experience a good series of shocks only gain 3.7 percent.

This paper is inspired by and builds on a growing macroeconomic literature evaluating the impact of public health insurance expansion using macroeconomic models. De Nardi, French and Jones (2016) evaluate Medicaid in the context of late-in-life insurance and find that it is approximately the correct size. Aizawa and Fu (2020) examine the interaction between risk-pool cross subsidization and Medicaid expansion and find that expansion leads to higher welfare gains. Pashchenko and Porapakkarm (2013) evaluate whether the welfare gains from expansion come from primarily regulatory changes or primarily redistribution and find that the welfare gains overwhelming come from the latter. Jung and Tran (2016) examine this same question in a more complex model with endogenous health expenditure and find a similar answer.

This paper also contributes to the literature on macroeconomics and insurance such as Kaplan and Violante (2010). I add to a large and growing literature on self-insurance in two-asset heterogeneous agent models such as Kaplan, Moll and Violante (2018) since health functions as an asset in my model. In a similar vein, health in my model can also be thought of as a durable good as in McKay and Wieland (2019).

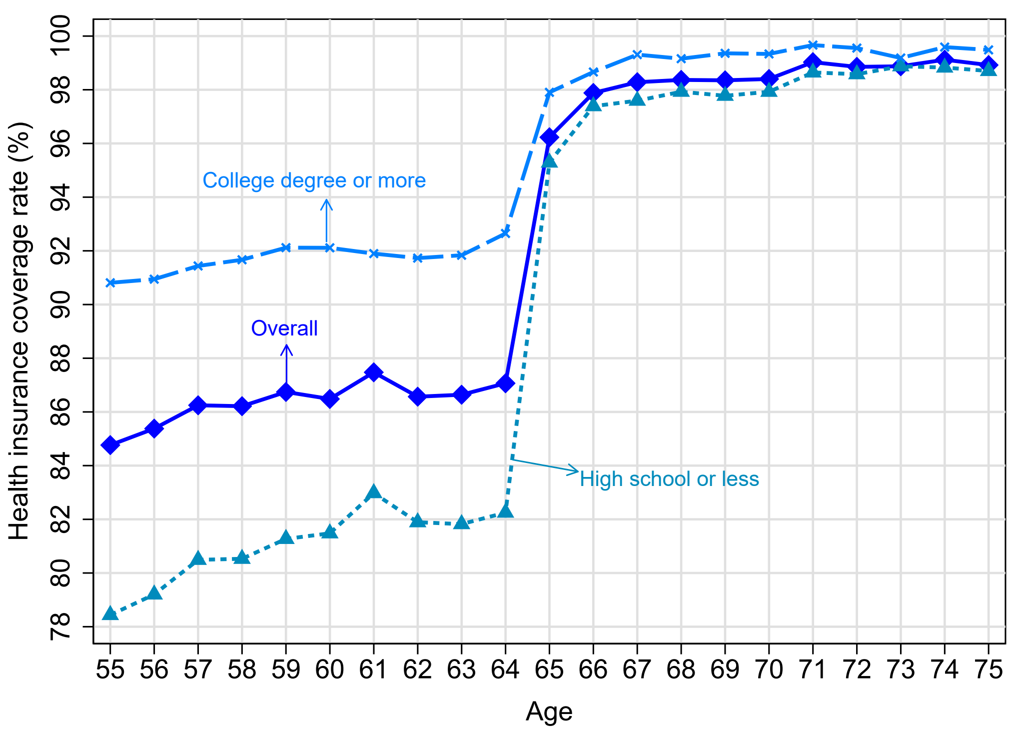
This paper is most closely related to the work of Ozkan (2014) which estimates a macroeconomic model of health spending and argues that shorter optimal lifespans for poorer individuals cause these individuals to under-spend on preventative care early in life, face a more costly distribution of late-in-life health shocks, and spend more on healthcare overall. My paper, instead of focusing on early-in-life preventative care, focuses on the incentives to delay the treatment of already-developed conditions induced by the age threshold of Medicare.

This paper also contributes to the broader literature on health and healthcare spending in macroeconomic models. De Nardi, French and Jones (2010) examine the role that late-in-life medical expenses play in individuals' optimal savings behavior. Cole, Kim and Krueger (2019) estimate optimal insurance policy in a model where health and labor market risks are intertwined and health insurance induces a moral hazard inefficiency.

Finally, this paper adds to a literature on modeling the trade-off between consumption and mortality beginning with Rosen (1988). Hall and Jones (2007) and Murphy and Topel (2006) expand on this analysis and use it to calculate the portion of health spending that can be attributed to higher incomes and the total welfare increase due to increased life expectancy within the US respectively. Jones and Klenow (2016) follow a similar approach to value the global increase in life expectancy. Finally, Córdoba and Ripoll (2017) expand on the standard modeling assumptions and suggest a more general preference specification that more closely matches observed behavior regarding mortality.

2. Some Facts on Healthcare Near Age 65

Figure 1: Health Insurance Coverage by Age and Education

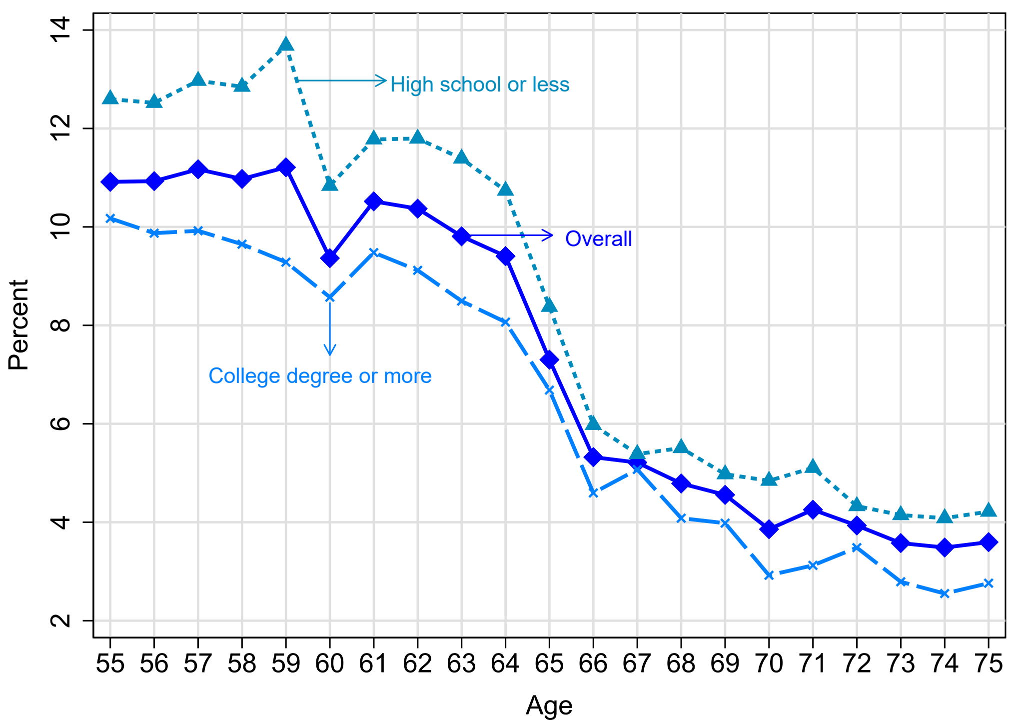


Displays the percentage of individual who self-report having health insurance coverage as a function of age and educational attainment. Calculated from NHIS data from 2002 to 2012.

A remarkable feature of the US healthcare system is the discrete and sudden increase in health insurance coverage that occurs at age 65. Before age 65, there is no universal government-provided health insurance or system, but after age 65, the government provides nearly universal healthcare through Medicare. Figure 1 displays the rate of health insurance coverage as a function of age and education calculated from the National Health Interview Survey. Before age 65, there is a substantial gap in coverage between educational groups of roughly 10 percentage points; however, at age 65 there is a jump in coverage for both education groups and a large convergence in coverage rates due to the sudden availability of Medicare.

The increase in insurance coverage is both quantitatively large – overall insurance coverage increases by about 10 percentage points – and extremely salient. It is a well-known fact among US individuals that Medicare eligibility begins at age 65. As a result, we might expect to see large changes in behavior around the age 65 threshold, particularly for individuals who have no health insurance or who are on cheaper high-deductible health plans and anticipate experiencing large declines in the marginal cost of receiving healthcare upon turning 65.

Figure 2: Delayed Medical Care by Age and Education



Displays the percentage of individual who self-report having delayed medical care in the last year for cost-related reasons. Calculated from NHIS data from 2002 to 2012.

Figure 2 displays the percent of individuals in the NHIS who reported delaying healthcare in the last year for cost-related reasons as a function of both age and education level. Unsurprisingly, highly educated individuals report delaying healthcare less often. The percentage of individuals who report delaying healthcare drops substantially from ages 64 to 66 as individuals become eligible for Medicare. Similar to Figure 1, the gap between education levels also shrinks substantially at age 65, consistent with the idea that the increase in insurance coverage (which is larger for the low-educated group) is driving the decline. Education provides a stand-in for lifetime income that is less subject to concerns of selection around the age threshold of 65. While it is intuitive that the decrease in delayed care may be different between low-income and high-income individuals, it is unclear that income at age 65 provides a good proxy of lifetime income. For example, wealthy individuals may choose to retire earlier than poor individuals and appears to have lower income at age 65. Education is less subject to these concerns of selection and is highly correlated with an individual's income.

One concern with interpreting these results is that changes in Medicare coverage at age 65 may be confounded with a jump in retirement. Figure A.1 displays both health insurance coverage rates and employment rates as a function of age. While health insurance coverage increases substantially at age 65, the employment rate declines smoothly with no sudden changes, assuaging any concerns that changes in working habits or leisure time may be driving the results.

Although it is difficult to make strong conclusions based on responses to survey questions about whether or not individuals delayed care, I interpret these figures as strong suggestive evidence that a fair number of individuals delay healthcare and that public health insurance can reduce the extent to which individuals delay. These facts motivate a model of endogenous health expenditure and credit-constrained individuals who face substantial incentives to delay care. In the rest of the paper, I lay out such a model and examine the implications for public health insurance expansion.

3. Model

I now present a macroeconomic model with endogenous health spending. The goal of the model is to allow the evaluation of the key tradeoffs in public health insurance expansion. Each period, individuals face a trade-off between consumption and investment into health. Insurance coverage, either purchased or provided by the government, reduces the marginal cost of health investment, increases individual health, and reduces mortality. Expansion must be funded by increases in the income tax rate which distorts individuals' labor supply decisions and reduces output. Additionally, the two-sector economy features an upwards-sloping relative supply curve for healthcare goods. As a result, the increase in demand for healthcare goods due to public health insurance expansion leads the relative price of healthcare to appreciate.

Because health and mortality occur at the individual level, I model the problem of individuals rather than of households. Although it lacks interesting behavior such as intra-household risk sharing, abstracting from household structure keeps the model tractable and creates a clear link between the model and health data, which are measured for individuals. Time is discrete and runs infinitely. Individuals are heterogeneous in their income y, savings b, health h, and age a. Exogenous measure n individuals are born at age 18 each period and age by 1 every period thereafter. At the end of each period, an individual faces an age and health dependent probability of dying π which will be discussed further in a later subsection. As a result of exogenous birth and endogenous death probability, there is endogenous measure N of individuals alive in any given period.

3.1. Preferences

Individuals have preferences over lifetime streams of consumption $\{c_a\}_{a=18}^{100}$, labor supply $\{l_a\}_{a=18}^{100}$, and mortality risk $\{\pi_a\}_{a=18}^{100}$. I've implicitly assumed that individuals die with certainty at age 100 (i.e. $\pi_{100} = 1$) and thus consumption, labor, and mortality beyond this age are irrelevant.

In the baseline model, I follow the sparse literature on modeling preferences over mortality by having period felicity be equal to felicity from consumption and labor $u(c,l)$ plus a "joy-of-life" parameter $\bar{u}$ which represents the additional utility an individual receives simply for being alive for the period. Thus period felicity is given by

$$\bar{u} + u(c,l)$$

As $\bar{u}$ is only a shift in the level of utility, it is meaningless without also specifying the level of utility realized when an individual dies which I normalize to zero for all agents. In a technical sense, this utility is an impulse that occurs in the moment of death, after which the individual ceases to exist.

Individuals discount the future exponentially. Death, if it occurs, occurs at the end of each period so that π_a denotes the probability that an individual dies at the end of period in which they are age a and doesn't live to see age $a+1$[2]. From the perspective of today, an individual of age a sees their future felicity at age $a+1$ as $\beta((1-\pi_a)(\bar{u}+u(c_a,l_a))+\pi_a \cdot 0)$ where $\pi_a \cdot 0$ is the normalized utility from death multiplied by the probability that the agent dies at the end of period t. Altogether, an individual's present-discounted lifetime utility is given by

$$U(c,l,\pi) = \sum_{a=18}^{100} [\prod_{j=17}^{a-1} (1-\pi_j)]\beta^a[\bar{u} + u(c_a,l_a)]$$

Under the normalization that the utility from being dead is equal to 0, the period mortality risk π_a acts as a time-varying discount factor by inducing individuals to put less weight on utility from periods that they are less likely to live to see. This results in some intuitive properties. For example, a young individual with low future mortality will have higher marginal utility from a reduction in current mortality risk than an old individual with higher future mortality as the younger individual has more "expected years" of life remaining (assuming, of course, that utility is parameterized such that life is a good and not a bad.). Similarly, an individual who expects higher future consumption will have higher marginal utility from a reduction in mortality than one who expects lower future consumption. Thus the marginal utility from a reduction in mortality risk is decreasing in future mortality risk and increasing in consumption.

3.1.1. The Value of Statistical Life

After specifying preferences over consumption and mortality risk, it is natural to think about the marginal rate of substitution between the two. In empirical studies and policy making, this is often referred to as the Value of Statistical Life (VSL) and is measured as an individual's willingness to pay to avoid a single expected death. That is, if an individual is willing to pay $10,000 to avoid a 0.01 percent increase in mortality risk, that individual exhibits a VSL of $1 million. The VSL can easily be expressed as the ratio of the marginal utility from a reduction in mortality risk to the marginal utility of consumption.

$$\text{VSL} = -\frac{\partial U}{\partial \pi_0} / \frac{\partial U}{\partial c_0}$$

Although the VSL is often used by policymakers evaluating trade-offs between mortality and dollar, I conceptualize the VSL as measuring a particular property of preferences rather than as a guide to normative policymaking. In the appendix, I extend the model to incorporate utility from health, allowing for agents to prefer healthiness over unhealthiness even beyond the degree to which being healthy saves their lives, as well as a more general preference specification, suggested by Córdoba and Ripoll (2017), which allows greater flexibility in calibrating the VSL.

[2]Under this timing convention, π_{17} is always equal to 0 as there is zero chance that the individual fails to live to age 18.

3.2. Health and Healthcare Expenditure

An individual's health status is written as a health index h with high values of h representing healthy individuals and lower values of h representing unhealthy individuals. In this way, h can be thought of as a sort of "health capital". When I bring the model to the data, h will correspond to a measured health index that lies in $[0,1]$ with $h = 1$ and $h = 0$ representing maximally and minimally healthy individuals respectively.

3.2.1. Medical Care Expenditure

Individuals can increase their health status using medical care, labeled i, through the health accumulation equation

$$h_{t+1} = (1 - \delta_a - \delta^x)h_t + \phi_a i_t^{\psi}$$

where δ_a is the natural rate of health depreciation which may depend on age, δ^x is the (possibly zero) depreciation from an acute emergency shock (discussed in detail in two paragraphs), and ϕ_a is a productivity parameter governing how effectively dollars of healthcare spending translate into units of health and also depends on age. The age dependence of δ_a and ϕ_a are key reasons why earlier treatment of disease is more cost-effective than later treatment of disease.

Non-emergency medical care expenditure represents all spending that is non-urgent and is done largely for the purpose of curing (or potentially preventing) disease. Spending on prescription or non-prescription drugs and spending on (non-emergency) bypass surgery to reduce arterial blockage are both examples of medical spending. For example, an individual with chronically high cholesterol may regularly purchase and take a statin aimed at reducing their cholesterol levels. This spending is non-urgent in the sense that there are no immediate consequences for an individual who chooses to forgo the spending; however, doing so may cause the individual's health to worsen. One may also consider non-monetary investments into health (e.g. exercise). In the appendix, I show how to extend the model to allow for such inputs and that, under certain conditions, this extended model is isomorphic to my baseline model.

The return to scale parameter $\psi < 1$ acts as an adjustment cost of jumping from unhealthy to healthy by spending heavily in a single period. Instead, health is acquired most efficiently through small investments made each year. Such a notion is intuitive and consistent with data; Hosseini, Kopecky and Zhao (2021) estimate an AR(1) persistence parameter of 0.99 for their measure of health status[3] indicating that health is highly autocorrelated.

Individuals invest in health for three reasons. First, being healthy decreases the probability that an individual experiences an emergency shock that leads to health depreciation and requires expensive emergency care. Second, if they do experience an emergency shock, a healthy individual incurs lower emergency costs on average. Intuitively, a healthy individual who regularly visits their doctor may catch an acute medical condition sooner and be able to treat the condition earlier and more cost-effectively. Finally, healthy individuals are simply less likely to die.

3.2.2. Emergency Healthcare Expenditure

In contrast to medical spending i which is an individual choice variable, emergency healthcare expenditure is compulsory. Each period, an individual faces an age- and health-dependent risk of experiencing a health emergency denoted $\pi_x(h,a)$. A health emergency carries two consequences. First, a health emergency results in an additional one-time depreciation of an individual's health stock, described by δ^x in the health accumulation equation above. Second, the individual must pay the costs of their emergency healthcare denoted x. These costs are stochastic and, conditional on experiencing an emergency, follow a log-normal distribution with a mean and variance that depend on the individual's

[3] It is worth noting that their persistence estimate is statistically different from 1 suggesting that there is no unit root in health

age and health.

$$x(h,a) \sim \begin{cases} 0 & \text{with prob. } 1-\pi_x(h,a) \\ \log N(\mu(h,a), \sigma(h,a)) & \text{with prob. } \pi_x(h,a) \end{cases}$$

The dependence of the mean and variance of the cost distribution on health and age can be thought of as representing the interaction between chronic and emergent health issues. For example, an individual with coronary artery disease is much more likely to have an emergent heart attack than one without. In the data section, I show that this intuitive relationship appears to hold for medical expenditure data; conditional on incurring positive emergency expenditures, expenditure is negatively correlated with an individual's measured health status. However, I find no statistically significant relationship between health status and the variance of emergency expenditure (conditional on positive expenditure). Still, I write the model to allow for such a relationship in order to keep with existing literature (e.g. De Nardi, French and Jones, 2016).

While medical expenditure includes all spending that could be deferred without immediate consequence, emergency expenditure represents urgent spending that cannot be delayed. A straightforward example would be angioplasty administered at the ER to stop a heart attack or emergency surgery for the victim of a severe car crash. Although not technically compulsory, most patients aside from the few who leave the ER, ICU, or otherwise act Against Medical Advice (AMA) treat this spending as effectively compulsory; a doctor prescribes care and the patient receives the treatment and later pays for it (or discharges the medical debt through bankruptcy).

3.2.3. Health and Mortality

As mentioned in 3.2.1, preventing mortality is a primary reason for an individual health investment. At the end of each period, every individual faces mortality risk denoted $\pi(h,a)$ which is allowed to depend on the individual's health and age. The function π is taken as exogenous.

3.2.4. Imperfect Information about Health

I also allow for a subset of individuals to be poorly informed about the benefits and importance of health and health investment. This assumption of imperfect information ends up being necessary for the model to match patterns of expenditure on preventative care; without some agents who underestimate the value of investment into their health, the model would predict counterfactually high levels of preventative care spending.

In particular, I assume that an individual with bad information perceives the risks of bad health to be smaller than they actually are. That is, a poorly informed individual with health status h perceives themself as facing the emergency expenditure and mortality risks of an individual with health status $h^* > h$. When I bring the model to the data, I measure h using a health index falling in the interval $[0,1]$ so a natural choice for h^* is

$$h^* = (1-\chi)h + \chi$$

where the parameter $\chi \in [0,1]$ governs the extent of misinformation. When $\chi = 0$, we have $h^* = h$ and the individual remains perfectly informed. When $\chi > 0$, h^* falls somewhere between the individual's true h and the maximum h of 1, and the individual perceives both their mortality and emergency risk to be lower than they actually are. Additionally, the agent perceives the marginal benefit of h to be exactly $(1-\chi)$ smaller than it actually is. In the limit of $\lambda = 1$, the individual is completely oblivious to the benefits of health.

Because emergency health events and mortality occur so rarely, poorly-informed individuals are rarely confronted with information that would cause them to substantially update their beliefs. I abstract from Bayesian updating and assume that an individual's information status follows a binary Markov process I and switches between "well-informed" and "poorly-informed" stochastically. Additionally, agents are naive about their present or future misinformation. They never suspect that they might be wrong in the current period or that they might be wrong in the future, even if they are correct today.

3.3. Health Insurance

Individuals can purchase health insurance to help pay for medical expenses and reduce the riskiness of emergency expenditure. Each period t, individuals choose to purchase (or not purchase) exactly one insurance plan from the set of plans for which they qualify and are offered. The individual is then covered by that plan in the next period $t+1$. In this way, individuals may reoptimize their choice of plan each period but must commit to buying a plan before they realize their exact draw of stochastic shocks for the period of coverage.

Every insurance plan p is indexed by a tuple (λ, ν, d, P) representing the plan's copay rate λ, coinsurance rate ν, deductible d, and premium P. These four plan parameters correspond more or less exactly to their real-life components. Emergency expenditure is covered through a standard deductible-coinsurance system; an individual facing emergency costs m in a single period must first pay up to their deductible d before any insurance coverage kicks in. Then the individual's insurance pays proportion $(1-\nu)$ of any costs beyond the deductible within the period leaving the individual responsible for paying the remaining fraction ν where ν is the plan's coinsurance rate. For simplicity, I abstract from out-of-pocket maximums, although these could be incorporated in the insurance scheme without much technical difficulty. In total, the individual's share of emergency costs m is given by $\min(d,m) + \nu \max(m-d, 0)$.

The insurance plan also subsidizes non-emergent care through the copay rate λ. Operating similarly to the coinsurance rate, an individual must pay proportion λ of their medical expenditure i while insurance pays the remaining $(1-\lambda)$. In essence, insurance subsidizes some portion of the costs of prescription drugs and other non-emergent expenditures. In addition to being a realistic feature of the model, such a subsidy makes sense from the perspective of the insurance company. The coverage of emergency expenditures introduces a moral hazard problem; individuals no longer face the full cost of their emergency expenditures and thus no longer receive the full benefit of investing in their health and experiencing a reduction in expected emergency costs. By reducing the marginal cost of preventative care through the copay rate, the insurance company is able to mitigate this distortion.

Finally, the premium P is the flat per-period cost of the individual's insurance plan which, in theory, may vary based on individual characteristics such as health. In practice, insurance companies will be mandated to charge premiums according to adjusted community rating which will be discussed in detail in a few paragraphs. Altogether, an individual with insurance plan p, indexed by $(\lambda_p, \nu_p, d_p, P_p)$, who spends i on medical care and faces emergency expenditure m must pay out-of-pocket costs given by

$$\chi_p(i,m) = \underbrace{\lambda_p i}_{\text{Preventative}} + \underbrace{\min(d,m) + \nu \max(m-d,0)}_{\text{Emergency}} + \underbrace{P_p}_{\text{Premium}}$$

3.3.1. Health Insurance Plans and Availability

There are four available health insurance plans and an option to be uninsured. The copay rate, coinsurance rate, and deductible (λ, ν, d) for each plan are taken to be exogenous and identical across individuals while the premium P is taken as endogenous for market-provided plans and exogenous for government-provided plans and, as mentioned above, can be individual-specific according to adjusted community rating.

All individuals are eligible to purchase an individual marketplace plan each period, denoted by $p = \text{IND}$. In addition, some individuals are eligible to purchase employer-provided insurance $p = \text{EMP}$. Although there is no *a priori* reason to prefer employer-provided insurance over marketplace insurance, when turning to the data, it is clear that employer-provided insurance plans offer lower deductibles and coinsurance/copay rates, on average, than marketplace plans. Additionally, government subsidies ensure that, despite better coverage, employer-based plans charge lower premiums than marketplace plans. Thus in the quantitative model, employer-provided insurance is strictly preferred to individual insurance.

Access to employer-provided insurance is not universal however. In reality, only individuals working for an employer who chooses to provide insurance or who previously worked for such an employer and remain covered through COBRA requirements have access to employer-provided programs. Replicating such a process in the model is difficult as the model lacks well-defined notions of job-switching or unemployment and tracking COBRA eligibility would involve many new state variables. Instead, I model eligibility as a simple binary Markov process M. The probability of transitioning from eligible ($e = 1$) to non-eligible ($e = 0$) is denoted by $\pi_{\text{EMP-IND}}$ while the probability of transitioning from non-eligible to eligible is denoted $\pi_{\text{IND-EMP}}$.

In addition to the employer-provided and individual marketplace plans, the government administers Medicare ($p = \text{MCR}$) and Medicaid ($p = \text{MCD}$). Medicaid is available to all individuals age 65 or older. Although in reality there are many different coverage and plan decisions an individual must make *within* Medicare, such as choices between different Medicare Advantage plans and optional prescription drug coverage, I abstract from these and model Medicare as a single insurance plan. Medicaid is made available to all individuals below a productivity threshold. Like with Medicare, I condense the complex reality of multiple Medicaid plans into a single representative plan. As part of this simplification, I model Medicaid availability as a function of productivity rather than earnings, simplifying away any labor market distortions.

Finally, individuals have the option to forgo insurance and enter the next period uninsured. For the sake of symmetry, I model this as a fifth insurance plan with a deductible, copay, and coinsurance all equal to zero. The premium for the uninsurance "plan" is also set by the government and may not be zero, reflecting the presence of an individual mandate which charges individuals for failing to purchase insurance. Such a mandate may make sense and might even be welfare-improving due to the problem of adverse selection, discussed in detail in the next section.

3.3.2. Insurance Firms and Pricing

While the government provides Medicare and Medicaid at exogenous (possibly zero) premiums, employer-provided and individual marketplace insurance are each provided by their own representative insurance firm. These firms take as given the exogenous plan parameters ($\lambda_{\text{EMP}}, \nu_{\text{EMP}}, d_{\text{EMP}}$ and $\lambda_{\text{IND}}, \nu_{\text{IND}}, d_{\text{IND}}$ respectively) and well as an exogenous load parameters $\kappa_{\text{EMP}}, \kappa_{\text{IND}} > 1$ which summarize the overhead costs of administration. For example, a firm that pays out x in total coverage must collect κx in premiums in order to break even for the period. Firms then set prices subject to a zero profit condition. Similar to wholesalers in New-Keynesian models which take intermediate goods and transform them into a single aggregate good, insurance firms operate without labor or capital and produce using purely intermediate goods collected through premiums. Thus κ represents the efficiency (or inefficiency) of this technology; κx goods are taken in by the firm and x goods are paid out. The remaining $(\kappa - 1)x$ goods are burnt up in the process.

Firms would like to charge different prices to different consumers of insurance. Although they may not be able to perfectly observe health status, easily observable characteristics such as age serve as strong proxies for expected health costs. However, firms are required to set prices following adjusted community rating rules which limit the extent of price discrimination. In particular, insurers are only allowed to price discriminate based on age and must follow strict age-by-age guidelines dictating the extent of price variation. These types of restrictions are stark features of the US health insurance market and have been in place since the implementation of the Afford Care Act. For example, insurance companies are restricted to charging a 40 year old individual no more than 1.278 times the amount they charge a 21 year old individual for the same coverage (1.786 for a 50 year old individual, etc).[4] In addition to being realistic, this restriction also dramatically simplifies the price-setting problem of the firms, effectively reducing a highly multi-dimensional problem across the ages and health status of different consumers to problem in a single variable.

Let the function $G(a)$ denote the exogenously enforced ratio between premiums for an individual of age a and for an individual of age 21. Then the zero profit condition for the insurance firms are given by

$$\int 1\{p = \text{EMP}\} G(a) P_{\text{EMP}} d\Omega = (1 - s_{\text{EMP}}) \qquad \kappa_{\text{EMP}} \int 1\{p = \text{EMP}\}[i^* + x - \chi_{\text{EMP}}(i^*, x)] f(x; h, a) dx d\Omega \quad (1)$$

$$\int 1\{p = \text{IND}\} G(a) P_{\text{IND}} d\Omega = \qquad \kappa_{\text{IND}} \int 1\{p = \text{IND}\}[i^* + x - \chi_{\text{IND}}(i^*, x)] f(x; h, a) dx d\Omega \quad (2)$$

where Ω is the distribution of individuals across states (b, h, a, p, e). The LHS of each equation is the firm revenue given by the baseline premium chosen by the firm P_{EMP} and P_{IND} multiplied by the age-specific premium schedule $G(a)$ and only taken for individuals who chose to purchase the plan last period. The RHS is the total outlays of the firm multiplied by the loading factor. The outlays are given by taking the individual's policy function for preventative spending i^* and (stochastic) emergency expenditure x and subtracting their share of out-of-pocket costs. This is aggregated across the health- and age-dependent distribution of emergency shocks, described by the pdf $f(x; h, a)$ and across all agents who purchased the plan last period. Finally, s_{EMP} is a proportional government subsidy for employer-based insurance where the government pays for proportion s_{EMP} of the healthcare costs and leaves the remaining fraction $1 - s_{\text{EMP}}$ for the company to pay. In the US, this subsidy occurs through the tax-exemption of employer-paid insurance premiums.

The assumption that health insurance is a zero-profit industry may be a contentious one but is of little consequence for this particular model. As an alternative, one could write a model where firms face a zero-profit loading factor of $\hat{\kappa} > 1$ and charge a constant markup given by $\sigma > 1$ due to market power. The observed loading factor would become $\hat{k}\sigma$. Because I observe loading factors directly from expenditure data as the ratio of total premiums paid by individuals and total dollars paid out by insurance, $\hat{\kappa}$ and σ are not separately identified, and I only measure their product. It is clear from the zero profit conditions that replacing κ with $\hat{k}\sigma$ when $\kappa = \hat{\kappa}\sigma$ changes nothing about the pricing decision of the firm and thus nothing about the insurance decision of the household. The only difference is that instead of $\kappa - 1$ percent of resources being lost in the production of insurance, $\hat{\kappa} - 1$ percent are lost and $\hat{\kappa}(\sigma - 1)$ percent are paid out to stakeholders in the insurance company. Given such a minor difference, I opt for a simpler model of perfectly competitive insurance firms.

3.4. Income and Labor Supply

Individuals younger than 65 years participate in the labor market and supply labor to the healthcare and consumption sectors, earning wage w_h and w_c per efficiency unit of labor supplied to each sector respectively. Individuals have a single measure of labor productivity z which summarizes their efficiency units of labor per hour of labor supplied in both sectors. Labor productivity is given by the following stochastic process (with time subscripts suppressed where

[4] A few states deviate from the national schedule by small amounts.

possible):

$$
\begin{aligned}
z(z^p, z^s, a) =& e^{g(a)+z^p+z^s} \\
z^p_{t+1} =& z^p_t \\
z^s_{t+1} =& \rho z^s_{t+1} + \varepsilon_t \\
\varepsilon_t \sim& N(0, \sigma)
\end{aligned}
$$

Here z^p is the individual's permanent productivity component which is invariant over their life-cycle while z^s represents a stochastic AR(1) component that leads to short-term fluctuations in income. Finally, $g(a)$ is a life-cycle component that depends on age a, allowing for deterministic life-cycle trends in productivity.

An individual who supplies labor l_m to the medical sector and labor l_c to the consumption goods sector has a pretax income given by

$$y_{\text{pre-tax}} = (w_m l_m + w_c l_c) z(z^p, z^s, a)$$

Individuals face disutility from their aggregate labor supply l which is given by a CES-style aggregator of labor supply to the healthcare and consumption sectors

$$l = \omega((1-\alpha_m) l_m^{\frac{\xi+1}{\xi}} + \alpha_m l_c^{\frac{\xi+1}{\xi}})^{\frac{\xi}{\xi+1}}$$

where $\xi > 0$ and ω is a level-adjustment constant that depends only on α_m and ξ. While abstract, this reduced-form description of labor supply captures an upwards-sloping relative supply curve for healthcare labor in a tractable way by allowing the relatively labor supply decision to be solved analytically. Conditional on aggregate labor supply l, the labor allocation problem of the individual is

$$
\begin{aligned}
&\max \; w_m l_m + w_c l_c \\
&s.t. \; l = \omega((1-\alpha_m) l_m^{\frac{\xi+1}{\xi}} + \alpha_m l_c^{\frac{\xi+1}{\xi}})^{\frac{\xi}{\xi+1}}
\end{aligned}
$$

which yields relative labor supply to healthcare given by

$$\frac{l_m}{l_c} = (\frac{1-\alpha_m}{\alpha_m})^{\xi} (\frac{w_m}{w_c})^{\xi}$$

Thus the relative supply of labor to healthcare $\frac{l_m}{l_c}$ exhibits a constant elasticity of ξ with respect to the relative wage $\frac{w_m}{w_c}$.

Despite being somewhat reduced-form, this upwards sloping relative labor supply curve captures an intuitive economic mechanism. The curve slopes upwards because the disutility of the marginal hour supplied to the healthcare sector is increasing in the (relative) number of hours supplied to the healthcare sector. In the short run where new doctors and healthcare workers cannot be trained, this is a straightforward implication of increasing marginal disutility of labor. In the long run where new doctors can be trained, the upwards-sloping labor supply curve can be thought of as a product of variation in preferences; the individuals with a strong taste for healthcare work are already healthcare workers and moving additional workers, who exhibit less of a taste for healthcare work, into the healthcare sector requires increasing (relative) wages. For the purpose of this paper, which is concerned with the long-term implications of public health insurance, I view ξ as a long-run elasticity.

I assume that ξ is common across all individuals so that the aggregate relative healthcare labor supply curve has constant elasticity ξ; however, I allow the share parameter α_h to vary across individuals. For tractability, I assume that α_h is a deterministic function of permanent productivity z_p. This captures the notion that healthcare workers are not evenly distributed across the income distribution; doctors and nurses tend to be high-paying professions. This variation is important when considering how mortality reduction due to Medicaid expansion interacts with the general equilibrium of the model. Because the reduction in mortality occurs largely for low-income individuals who supply little healthcare labor, it shifts demand and supply of healthcare goods differentially leading to price impacts.

3.4.1. Taxes and Retirement

Working individuals pay progressive income taxes that are used to fund government-provided health insurance (Medicare and Medicaid) as well as social security payments. Let $T(\cdot)$ denote an individual's after-tax income as a function of their before-tax income. After-tax period earnings for an individual younger than 65 are given by

$$y_{a<65}(z^p, z^s, a, l_h, l_c) = T((w_h l_h + w_c l_c) z(z^p, z^s, a))$$

where l_h and l_c are the individual's (endogenous) labor supply. It is important to note that as long as the tax function T is monotonically increasing, the solution to the labor allocation problem above does not change.

At age 65, individuals retire exogenously and fix their labor supply l to 0 for the remaining periods of their life. Retired individuals receive social security income which depends on their permanent productivity and is given by the exogenous function $y_{a\geq 65}(z^p)$. In reality, US social security income is determined by one's entire earnings history, taking the average earnings from the 35 years in which one's earnings were the highest, but faithfully modeling such a process would require keeping track of an individual's entire earnings history. Fortunately, permanent income provides a good approximation of this average[5] for everyone except the ex-post luckiest and unluckiest individuals who earned substantially more or less than their permanent income. Given the large increase in tractability for relatively little loss in accuracy, I opt for a simple model of retirement income.

3.5. Consumption, Savings, and the Budget Constraint

Individuals split their income between consumption c (the numeraire), medical expenditure i, emergency expenditure x, and assets b. The savings technology takes the form of a risk-free asset which pays an interest rate of r_t each period. Markets are incomplete and individuals cannot borrow, requiring so that assets b_t cannot be negative. The budget constraint of an individual with assets b_t, insurance plan p_t, and productivity $z(z_t^p, z_t^s, a)$ is given by

$$\begin{aligned} c_t + b_{t+1} + p_h \chi_p(i_t, m_t) &= (1+r_t)b_t + T((w_{h,t} l_{h,t} + w_{c,t} l_{c,t}) z(z_t^p, z_t^s, a)) && \text{if } a < 65 \\ c_t + b_{t+1} + p_h \chi_{\text{MCR}}(i_t, m_t) &= (1+r_t)b_t + y_{a\geq 65}(z_t^p) && \text{if } a \geq 65 \end{aligned}$$

where p_h is the price of healthcare goods.

3.6. The Individual Optimization Problem

Having specified the individual's preferences, budgets constraints, and the process for health, we can finally write their optimization problem. The individual faces eight individual-level state variables. They are

[5] The approximation must be adjusted for the fact that $\mathbb{E}(e^{z_t^s}) > 1$

1. Assets b

2. Health h

3. Age a

4. Permanent productivity z^p

5. Temporary productivity z^s

6. Insurance plan p

7. Access to employer-provided insurance e

8. Information status χ

They also face an aggregate state variable Ω describing the cross-sectional distribution of all individuals across the 8 individual-level states. The individual problem for a well-informed individual can be written recursively as in 3. $G(\Omega)$ is the perception function used by the individual to forecast the future aggregate state. The problem for a poorly-informed individual is similar but replaces the actual health-related stochastic processes $\pi(h,a)$, $\pi_x(h,a)$, $\mu(h,a)$, and $\sigma(h,a)$ with their perceived counterparts $\pi(h^*,a)$, $\pi_x(h^*,a)$, $\mu(h^*,a)$, and $\sigma(h^*,a)$for $h^* = (1-\chi)h + \chi$. The full problem for a bad-information individual can be found in the appendix.

It is worth noting that, although this problem looks complex, much of the complexity comes from the battery of exogenous stochastic processes and is absorbed by the expectation. The problem becomes even simpler once the labor allocation choice of $\frac{l_m}{l_c}$ is eliminated analytically. Broadly speaking, the model fits within a standard two-asset heterogeneous agent framework and can leverage the variety of algorithms aimed at efficiently computing these models. The only non-standard component of the model is the discrete choice induced by the decision of which insurance plan to purchase.

$$
\begin{aligned}
V(b,h,a,z^p,z^s,p,e;\Omega) =& \max\ \bar{u}+u(c,l)+\beta(1-\pi(h,a))\mathbb{E}[V(b',h',a+1,z^p,z^{s'},p',e';\Omega')] \\
s.t.\ c+b'+p_h\chi_p(i,m) =&(1+r(\Omega))b+T((w_m(\Omega)l_m+w_c(\Omega)l_c)z(z^p,z^s,a)) \ \text{ if } a<65 \\
c+b'+p_h\chi_{\text{MCR}}(i,m) =&(1+r(\Omega))b+y_{a\geq 65}(z^p) \ \text{ if } a\geq 65 \\
h' =&(1-\delta_a-\delta_x)h+\phi i^{\psi} \\
l =&\nu((1-\alpha_m)l_m^{\frac{\xi+1}{\xi}}+\alpha_m l_c^{\frac{\xi+1}{\xi}})^{\frac{\xi}{\xi+1}} \\
p' \in&\{\text{EMP},\text{IND},\text{UN},\text{MCD}\} \text{ according to eligibility} \\
b' \geq& 0 \\
i \geq& 0 \\
z^{s'} \sim& \rho z^s+\varepsilon,\ \varepsilon\sim N(0,\sigma) \\
x \sim& \begin{cases} 0 & \text{with prob. } 1-\pi_x(h,a) \\ \log N(\mu(h,a),\sigma(h,a)) & \text{with prob. } \pi_x(h,a) \end{cases} \\
e' \sim& M(e) \\
\Omega' =& G(\Omega)
\end{aligned}
\tag{3}
$$

3.7. Delayed Medical Care

Delaying medical care until age 65 emerges naturally from optimal behavior. Like all consumption-savings models, individuals use their assets b to smooth their consumption according to their Euler equation. Within-period optimization between consumption and health spending dictates equality between the marginal benefit of each. The first-order condition is

$$u_c(c^*,l^*)=\frac{\phi_a\psi i^{\psi-1}}{p_h\lambda_p}\beta(1-\pi(h,a))V_{h'}(h'^*) \tag{4}$$

where λ_p is the copay rate of the individual's insurance plan and I have suppressed most of the inputs into the value function for brevity. The marginal utility of consumption on the LHS of the equation is, as a result of the consumption-savings problem, roughly constant for individuals not near the borrowing constraint.

From equation 4, it is clear that as long as V exhibits diminishing marginal returns to health (as is the case in the estimated model), the marginal return to additional health $V_{h'}(h'^*)$ is inversely related to the individual's copay rate λ_p. The intuition is simple: under a lower copay rate, an individual will spend more on medical care resulting in higher health and a lower marginal return to any additional health.

The envelope condition for the marginal value of health

$$V_h(h)=\underbrace{-\beta\pi_h(h,a)\mathbb{E}(V(h'))}_{\text{Reduction in mortality}}+\underbrace{\beta(1-\pi(h,a))\frac{\partial}{\partial h}\mathbb{E}(V(h'))}_{\text{Reduction in Emg. Risk}}+\underbrace{(1-\delta_a-\delta_e)\beta(1-\pi(a,h))V_{h'}(h')}_{\text{"Health Tomorrow"}}$$

reveals that health is a forward-looking asset; part of the benefit of being healthy today is that one will continue to be healthy tomorrow and the marginal value of health today depends on the marginal value of health tomorrow discounted by the depreciation rate of health and the individual's subjective discount rate. Iterating this relationship forward through time, it is clear that the value of health today depends on the value of health t periods in the future discounted by $(1-\delta_a-\delta_e)^t\beta^t\prod_{i=0}^{t-1}(1-\pi(a+i,h_i))$.

Combining these two relationships, it is clear how delayed care arises. The individual expects to receive health insurance in the future, lowering λ_p, and thus the future marginal value of health. Because the value of health today depends on the value of health in the future, especially as the individual approaches the period they will receive insurance, this lowers the value of health today, resulting in under-spending.

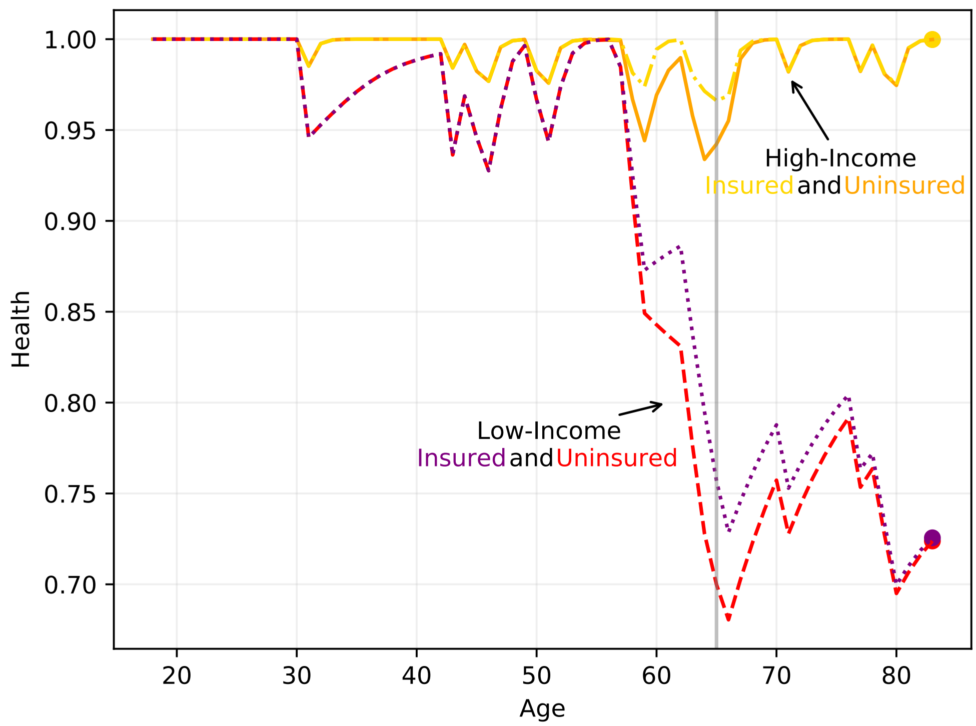
The economic intuition is straightforward. The individual anticipates their future insurance coverage and, in particular, their low copay rate. But if health is going to be so much cheaper tomorrow (or in two or three periods) and a large portion of the value of health comes from its continuation value rather than the immediate benefits, why bother investing in health today? The most effective strategy is to treat health as an asset to be run down and then replenished once covered by insurance. This incentive is mitigated by the decreasing returns to medical care each period which ensures that medical care tomorrow is not too good of a substitute for medical care today, but it is still strong enough to generate quantitatively important behavior.

3.7.1. A Quantitative Example

Figure 3 shows a quantitative example of individual health investment behavior and provides insight into delayed care. The figure displays individual health (on the y-axis) over the lifecycle of the individual as measured by their age (on the x-axis) for four different individuals. These individuals all receive an identical series of income and health shocks over the course of their life but differ by their permanent income and their access to employer-provided health insurance. In particular, the individual represented by the yellow line has high permanent income and maintains access to employer-provided health insurance for their entire life. The orange line plots the health of a high permanent income individual who never gains access to employer provided insurance. I label this individual as "Uninsured" since as they approach the age 65 threshold, they opt not to purchase marketplace insurance and go uninsured. The purple and red lines plot the experience of low permanent income individuals with and without access to employer-provided insurance respectively. Like the high income individual, the low-income individual without access opts not to purchase insurance.

The dynamics of health investment can be seen clearly in the figure. Early in life, the four individuals receive the same shocks but, because they spend more on healthcare, the high-income individuals recover from shocks faster and maintain higher average health. Insurance status seems to make minimal difference in health dynamics at this point. Later in life, around age 58, the four individuals experience a series of bad health shocks before retirement. Here the dynamics begin to diverge. As was the case with shocks early in life, the high-income individuals spend more on healthcare, recover from shocks faster, and keep their health at a higher level than the low-income individuals. However, as the individuals are approaching the age 65 threshold after which they will receive Medicare, they begin to respond to incentives to delay care. This is most noticeable in the differences between insured and uninsured low-income individuals. The insured individual, displayed in purple, continues to invest in health and recovers from the shock to the extent they can given their limited wealth, as exhibited by increase in health at ages 60 and 61. However, the situation is more dire for the uninsured individual who invests fewer resources and does not recover from the shocks as well as the insured individual. After these individuals turn 65 and receive coverage through Medicare, their health slowly converges and the gap disappears around age 78, evidence of the uninsured individuals higher post-65 spending.

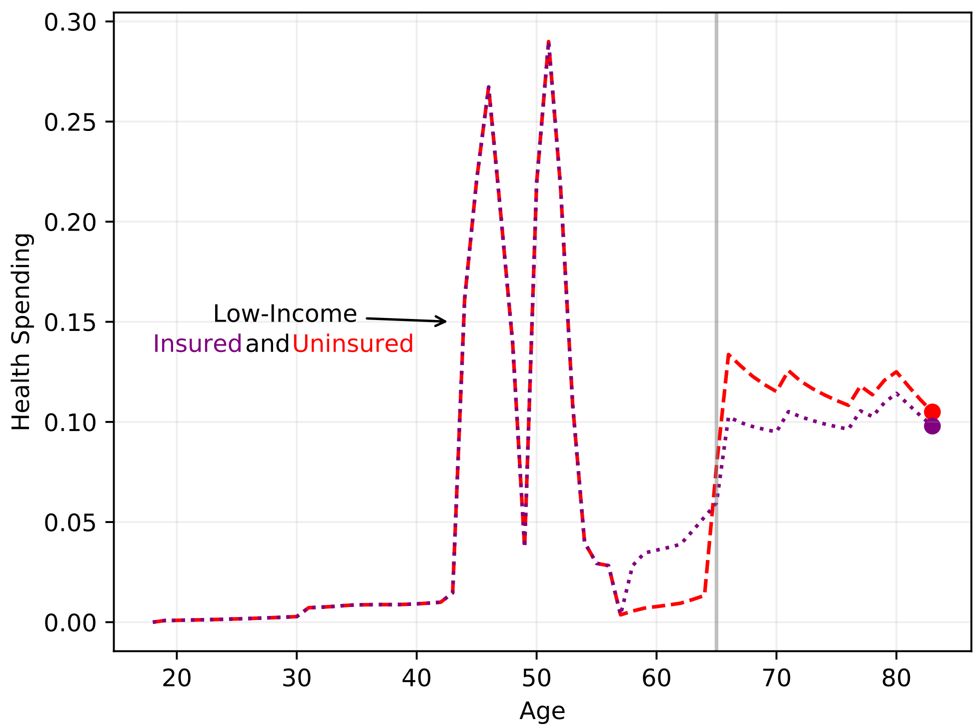
Figure 3: Life-cycle Health Dynamics for Four Individuals



Displays life-cycle health dynamics for four individuals who receive an identical series of health and income shocks. The yellow line displays the health of an individual with high permanent income and access to employer-based insurance. The orange line displays the same for an individual with high permanent income and no health insurance. The purple and red lines display the health of low permanent income individuals with and without access to employer based insurance respectively.

Figure 4 reinforces this interpretation by displaying annual healthcare spending for each of these two individuals over their life-cycle. From the figure, it is clear that, while both individuals delay healthcare as evidenced by the jump in expenditure at age 65, the uninsured individual spends less on healthcare before age 65 and more afterwards.

Figure 4: Life-cycle Health Spending by Low-Income Insured and Uninsured Individuals



Displays life-cycle health spending for two individuals. The purple line dispalys the spending for an individual with low permanent income and access to employer based insurance. The red line displays spending for an individual with low permanent income and no access to employer-based insurance.

3.8. Production

The production side of the economy is comparatively simple. There exist two representative firms producing healthcare and consumption goods respectively and setting output and input prices according to perfect competition. The production technology for healthcare and consumption goods takes the form of standard Cobb-Douglas production functions with a common share parameter α. Total output of medical goods Y_m and consumption goods Y_c are given by

$$p_m Y_m = p_m A_m L_m^{\alpha} K_m^{1-\alpha}$$
$$Y_c = A_c L_c^{\alpha} K_c^{1-\alpha}$$

Capital can flow freely between sectors so that aggregate demand for capital is simply given by $K = K_m + K_c$. As a result, the rate of return on capital is equalized between the two sectors $r_m = r_c$, justifying the modeling decision that households have only a single asset in which to invest.

The relative price of healthcare adjusts to equalize supply and demand for medical care. As the competitive firm retains no profits, increases in p_m translate to increases in wages w_m and returns r_m with unit elasticity, *ceteris paribus*. These increases lead to a relative increase in labor L_m and capital K_m directed towards healthcare, yielding an upwards-sloping aggregate supply curve for healthcare goods.

3.9. Recursive Competitive Equilibrium

The recursive competitive equilibrium of the model consists of

a) Individual value and policy functions for both good-information and bad-information individuals given by $(V, c, i, b', l_h, l_c, p')$ and $(V_\chi, c_\chi, i_\chi, b'_\chi, l_{h,\chi}, l_{c,\chi}, p'_\chi)$

b) Firm policy functions (L_m, K_m, L_c, K_c)

c) Price functions $(r, w_m, w_c, P_{\text{EMP}}, P_{\text{IND}})$

d) Perception function G

such that

1) The value and policy functions in a) solve the individual optimization problem (3)

2) The firm policy functions solve the firm optimization problem:

$$\max \; p_m A_m L_m^\alpha K_m^{1-\alpha} - rK_m - w_m L_m$$
$$\max \; A_c L_c^\alpha K_c^{1-\alpha} - rK_c - w_c L_c$$

3) Markets clear: $\int b d\Omega = K_m + K_c$ and (1) and (2) hold

4) Perceptions are correct: $\Omega' = G(\Omega)$

Details on the computation of the recursive competitive equilibrium can be found in the appendix.

4. Data, Calibration, and Estimation

The parameters of the model fall into three broad categories. The first are the macro parameters, such as the discount rate, which are simply calibrated to be equal to common values within the literature. Second, many of the less-common parameters relating to health, such as the function for mortality risk, are directly estimated from healthcare data. Finally, some health parameters are estimated using the simulated method of moments (SMM) to match important moments of aggregate or quasi-experimental data. In this section, I discuss each of these categories in turn, but first I will start with a discussion of the healthcare data used.

4.1. The Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS) data serve as the primary source of data for quantification of the model. These nationally-representative data contain detailed information on individual healthcare expenditure, insurance coverage and plan details, and health status. The data are collected as an overlapping panel; households are selected into the survey and complete an initial interview as well as four follow-up interviews. The final interview occurs roughly two years after the initial interview so that each individual has completed five interviews covering a two year span. A major advantage of the MEPS is that the panel structure allows me to observe an individual's actual health outcomes, conditional on their characteristics such as health or age, over time which allows direct estimation of the relationship between health status and outcomes such as mortality, and healthcare expenditures.

4.1.1. Measuring Healthcare Expenditure

True to its name, the MEPS contains detailed and, importantly, accurate measures of healthcare expenditure at the individual level. Given the complicated nature of medical billing, including how consumers are often largely removed from the true cost of their healthcare, it is worth taking some time to discuss how the MEPS is designed to collect accurate information even when survey households may be unaware of or misremember their expenditure. See Cohen (2003) and Zuvekas and Olin (2009) for elaboration on the discussion below.

In the MEPS, individual-reported data on healthcare expenditure is supplemented and often largely replaced with data collected from the Medical Provider component of the survey. These data are collected from the doctor, hospital, or other healthcare providers from which the surveyed individual received a healthcare service and include actual payments made to the provider, as opposed to charges which may or not may accurately reflect the payments made. The data also distinguish between sources of payment, allowing direct measurement of out-of-pocket costs vs costs covered by insurance. Using data directly from the medical provider also sidesteps any issues of recall or rounding that often plague survey-based financial data.

Unfortunately, the Medical Provider component does not cover all spending included in the MEPS. Table A.1 describes the MP component coverage for two key categories of provider: office-based physicians (including physicians assistants and nurse practitioners) and hospitals. The table details, for each category of provider, what percentage of households have their reported provider included in the MP component. For example, there is complete (100 percent) coverage of hospitals; any hospital reported by an individual as a healthcare provider will be included. In contrast, there is only 75 percent coverage for office-based physicians providing care to individuals covered by HMOs. In essence, 75 percent of HMO-covered individuals are chosen and all office-based physicians reported by these individuals are included in the MP component. The physicians reported by the remaining 25 percent are not included in the MP component.

The MEPS attempts to use information from the MP component to imputed survey-reported spending that is not covered by the MP component directly. Unfortunately, the details on this process are sparse and the public-use data even lack imputation flags, making it impossible to compare imputed spending to spending included in the MP component. Still, I take the reported spending data at face value.

4.1.2. Separating Health Investment and Emergency Spending

After measuring healthcare expenditure, the next step is to separate total expenditure out into spending on health investment and spending on emergency health events. While some spending, such as an emergency room visit, clear falls into one category, other forms of spending might be less clear. An emergency bypass surgery is clearly the result of a health emergency but also results in a long-term improvement in an individual's health by cleaning arteries. For lack of a clear defining line between types of spending and because it is measured clearly and unambiguously in the data, I use the presence of an emergency room as the distinguishing feature of emergency spending. If spending occurs in an emergency room, the spending is categorized as emergency spending. Otherwise, it is categorized as health investment.

Based on this definition, Table 1 reports some basic summary statistics for emergency and investment spending. The overall pattern is not surprising; non-zero investment spending is much more common than non-zero emergency spending and both types of spending are larger for older individuals. In the appendix, I explore how these patterns change for different definitions of emergency spending. Overall, they seem quite stable.

Table 1: Some Summary Statistics of Investment and Emergency Spending

	Mean	Median	% >0	Mean	Median	% >0
		All			65 or Older	
Investment	$3,847	$924	81.3%	$6,105	$2,892	95.9%
Emergency	$2501	$0	33.8%	$4041	$0	48.9%

This table provides some summary statistics for healthcare investment and emergency spending. Emergency spending is defined as any spending that takes place in the emergency department and investment spending is all other spending. Calculated from the Medical Expenditure Panel Survey

4.1.3. Measuring Individual Health

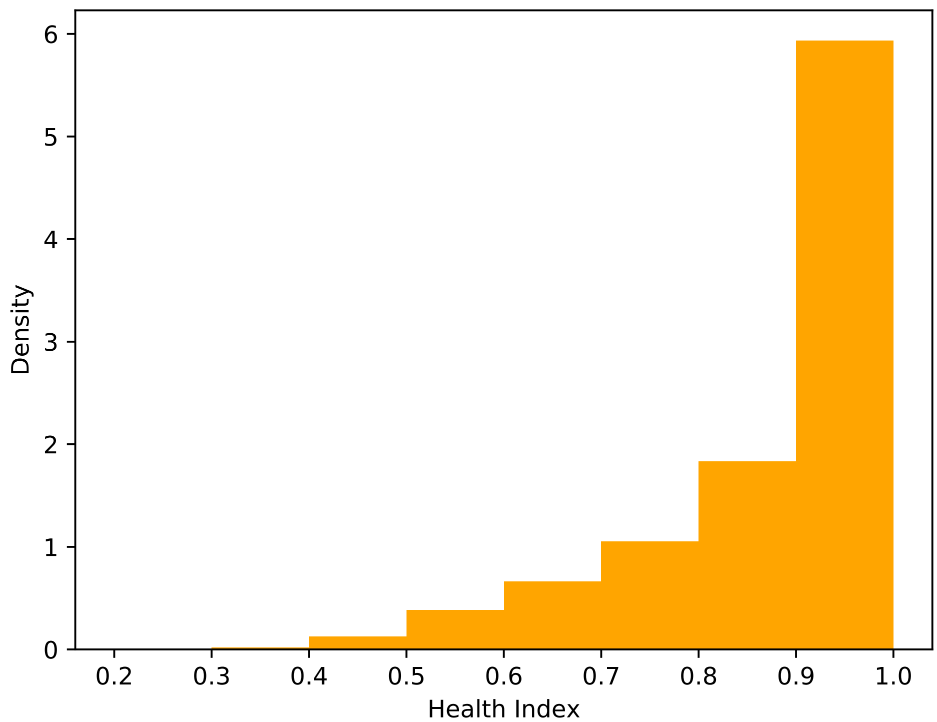
In addition to providing data on healthcare expenditure, the MEPS also provides crucial information on individual health status and outcomes. Health is an inherently high-dimensional, complex, and hard to quantify object and mapping the complex reality of health to a simplified, abstract concept amenable to economic modeling has long been a difficulty in the macro-health literature. Often the solution has been to restrict health to a small number of discrete categories such as "Good and "Bad" or ranging from "Excellent" to "Poor" (e.g. De Nardi, French and Jones, 2016; Yogo, 2016). Particularly in the case of the latter, these categories are often self-reported subjective measures of health that may or may not be related to an individual's actual health (see Spitzer and Weber, 2019, for an example).

To overcome these issues, I base my measure of health on the frailty index of Hosseini, Kopecky and Zhao (2021). This index has the advantage of being largely objective and close-to-continuous, allowing it to be a natural stand-in for health h in the model. I construct the frailty index from a wide variety of yes or no questions about an individual's health ranging from diagnoses (Have you ever been diagnosed with diabetes?) to cognitive limitations (Do you experience confusion or memory loss?) to common metrics known as Activities of Daily Living (Do you have difficulty getting dressed by yourself?). I also supplement these yes or no questions with some other objective measures of health obtainable from the MEPS such as an indicator for if the individual's BMI is greater than 30 and the individual's K6 score (a common measure of mental health). The frailty index is then constructed by summing up the number of yeses, referred to as the total number of health deficits, and rescaling by the number of possible deficits so that the minimum value of the index is 0, corresponding to an individual who reported zero health deficits, and the maximum value of the index is 1, corresponding to an individual who reported having every deficit.

I convert from an index of frailty to an index of health by simply subtracting the frailty index from 1 so that $h_i = 1 - f_i$ where h_i is an individual's health index and f_i is their frailty index. Thus a $h_i = 1$ represents a maximally healthy individual and $h_i = 0$ represents a minimally healthy individual. The distribution of health is shown in Figure 5. The data reveal a high concentration of healthy individuals possessing health indices between 0.9 and 1.0 with a thin tail on the left-hand side. Very few individuals accumulate an extensive number of health deficiencies.

Hosseini, Kopecky and Zhao (2021) discuss at length the usefulness of frailty as a measure of health and show

Figure 5: The Distribution of Health in MEPS Data



Displays the distribution of health in the Medical Expenditure Panel Survey, as measured by the health index based on Hosseini et al. (2021).

that it is a strong predictor of a variety of health outcomes, including medical expenditure and mortality, and that it outperforms self-reported measures of health. I corroborate these findings. Table 4 in subsection 4.2.2 below shows that the health index strongly predicts individual mortality, the probability of positive emergency expenditure, and the total amount of emergency expenditure conditional on positive expenditure. The index remains predictive even when age and the square of age are included in the regressions, demonstrating that its predictive power is orthogonal to the predictive power of age. In addition, the coefficient on the index is robust to the inclusion of a wide variety of controls including family income, race, sex, and geographic region, providing suggestive evidence that the index is not picking up variation in some non-health-related latent variables.

4.2. Calibration and Estimation

Having described the primary source of health data used, I can now discuss the quantification of the model. As mentioned before, model parameters fall into three broad categories: those calibrated to common values found in the literature, those estimated directly from data, and those estimated indirectly through the simulated method of moments. I discuss each in turn.

4.2.1. Parameters Taken from the Literature

Table 2 lists the parameters that are calibrated externally and their values. The parameters and their values are mostly typical or are normalizations. The temporary income process is taken from Floden and Lindé (2001) while the life-cycle component is chosen to match Lagakos et al. (2018) (plotted in Figure A.2). The functional form for post-tax income as well as the parameter values are chosen following Heathcote, Storesletten and Violante (2017). Post-retirement social security income is calculated using the actual social security scheme assuming that an individual with permanent productivity z_p earned exactly their ex-ante average lifetime earnings each period.

I choose to model period felicity from consumption and labor using the functional form for King-Plosser-Rebelo preferences laid out in Trabandt and Uhlig (2011). These preferences exhibit a constant Frisch elasticity of labor supply and, as a result, are commonly used in analyses of general equilibrium responses to taxation. As labor supply responses to taxation are an important channel in my model, KPR preferences are a natural choice.

It is worth commenting briefly on the parameter value for the coefficient of relative risk aversion (CRRA) which, due to well-known properties of expected utility, is also equal to the inverse intertemporal elasticity of substitution (IES). The value of 2 is commonly used in both the consumption-savings literature as well as the taxation literature; however, as both Murphy and Topel (2006) and Hall and Jones (2007) show, this parameter is also crucial in determining how an agent's willingness to trade-off consumption and mortality varies with respect to their income. Intuitively, this occurs because the IES governs how much the agent values living to see future consumption. If the IES is high, consumption today is a good substitute for consumption tomorrow, and the agent can consume heavily today and not worry about whether or not they will be alive to see tomorrow's consumption. Conversely, a low IES agent values living to see tomorrow very highly as there is no good substitute for tomorrow's consumption.

Thus the CRRA/IES parameter is doing "triple duty"; it governs the agent's risk tolerance (as in the consumption-savings literature), their labor supply response to income shocks (as in the taxation literature), and the income elasticity of their VSL (as in the sparse macro-health literature). In the baseline model, I accommodate this by choosing a value that seems to get all three behaviors roughly correct. In the appendix, I examine an extension using a variant of Epstein-Zihn-Weil preferences suggested by Córdoba and Ripoll (2017) that separates the single CRRA parameter into three parameters separately governing each behavior.

Table 2: Externally Calibrated Parameters

Description	Parameter	Value
Discount Factor	β	0.97
Utility from (c,l)	$u(c,l)$	$\frac{1}{1-\sigma}c^{1-\sigma}\left(1-\kappa(1-\sigma)l^{1+\frac{1}{\nu}}\right)^{\sigma}$
Coefficient of Relative Risk Aversion	σ	2
Frisch Elasticity of Labor	ν	1
Disutility of Labor	κ	0.15
Income Persistence	ρ	.91
Income SD	σ	.04
Life-cycle Income	$g(a)$	See Figure A.2
Labor Share	α	0.66
Healthcare Labor Supply Elasticity	ξ	2.22
Tax Function	$T(y)$	$\lambda_{\tau}y^{1-\tau}$
Tax Progressivity	τ	0.181
Tax Level	λ_{τ}	0.90
Social Security Function	$y_{a\geq 65}(z_p)$	See discussion

Displays model parameters calibrated to standard literature values. See discussion for details on each parameter.

The only other non-standard parameter is ξ, the elasticity of relative healthcare labor supply with respect to relative wage. I calculate this parameter from the quasi-experimental results of Finkelstein (2007) which uses a difference-in-difference framework exploiting pre-existing differences in elderly insurance coverage and the national implementation of Medicare to estimate the effect of Medicare on aggregate healthcare outcomes, including hospital employment and hospital payroll. I use the results from 5 to 10 years after the implementation of Medicare in an attempt to recover the longest-run elasticity possible. She finds that Medicare increased employment by 25.6 percent and payroll by 40.1 percent in the 5-10 years following implementation. Together, these results suggest that earnings per worker increased by 11.5 percent. I treat both the increase in employment and the increase in earnings per worker as increases in the relative employment share and relative wage for healthcare yielding an elasticity of $\frac{.256}{.115} \approx 2.22$.

How does this elasticity estimated using micro-data and quasi-experimental techniques compare to an aggregate elasticity? In particular, do its implications about the relative price of healthcare goods seem to hold in aggregate data extending beyond 1975 (the final year in Finkelstein (2007)). To answer this, I use a property of the firm maximizing problem; if the rate of return to capital is equalized between the healthcare and consumption sectors, then we have

$$(\frac{w_h}{w_c})^\alpha = p_h$$

That is, the relative healthcare wage raised to α is equal to the relative price of healthcare. Combining with the relative healthcare labor supply curve, taking logs, and differencing yields

$$\Delta \log(\frac{l_h}{l_c}) = \frac{\xi}{\alpha} \Delta \log(p_h)$$

which is a simple aggregate relationship between relative employment in healthcare and the relative price of healthcare. I test this relationship by comparing data on healthcare employment from the BLS and data on the relative price of healthcare from FRED between 1968 and 2008[6]. The BLS reports that total employment in healthcare increased from 4.7 percent to 11.6 percent from 1968 to 2008 while the relative price of healthcare increased from 0.85 to 1.70. Plugging these numbers along with $\alpha = 0.66$ into the above equation yields a value for ξ of 0.93.

Although the estimate of the elasticity from aggregate data is somewhat different than the elasticity implied by Finkelstein (2007), I still opt to use the elasticity of 2.22 implied by the microeconomic study. Because the variation in healthcare demand is quasi-experimental, the microeconomic study is not confounded by other secular trends that may confound estimation of the aggregate elasticity. In particular, the aggregate calculation may be confounded by differential productivity growth in the health and non-health sectors.

4.2.2. Parameters Estimated Directly

Table 3 lists the parameters of the model that are estimated directly or close-to-directly from data, their values, and the data source on which they are estimated. The distribution of individual-level permanent income is chosen to be log-normal with the mean and variance parameters calibrated to match US GDP per capita and median personal income. The effective loading factors $(1 - s_{\text{EMP}})\kappa_{\text{EMP}}$ and κ_{IND} are calculated as the ratio of total premiums paid over total covered costs for all individuals in the MEPS covered by employer-provided and marketplace insurance respectively. The Markov process for the availability of employer-provided insurance is chosen to match a ratio of employer-covered individuals over marketplace-covered and uninsured individuals of 3.6 as well as an annual hazard rate of losing employer-provided insurance of 7.8 percent for working-aged individuals. Both of these values are

[6] The data on healthcare employment can be found here while the data from FRED can be found here and here. Both were accessed on August 19, 2021.

calculated directly from MEPS data. The Medicaid productivity cutoff $\bar{z}$ is chosen so that Medicaid is offered to all individuals earning less than 138 percent of the federal poverty level for a single adult, the level prescribed by the ACA expansion of Medicaid.

Table 3: Directly Estimated Parameters

Description	Parameter	Value	Data
Mortality Function	$\pi(h,a)$	Table 4 Col. 1	MEPS
Emergency Prob. Function	$\pi_x(h,a)$	Table 4 Col. 4	
Emergency Mean Function	$\mu(h,a)$	Table 4 Col. 7	
Emergency Var Function	$\sigma(h,a)$	Table 4 Col. 10	
(Inverse) Weight on Healthcare Labor	See discussion	ACS	
Medicaid Prod. Cutoff	$\bar{z}$	0.68	Statutory
EMP availability	M	$\begin{bmatrix} .922 & .078 \\ .281 & .719 \end{bmatrix}$	MEPS
Effective Loading Factor for EMP insurance	$(1-s_{\text{EMP}})\kappa_{\text{EMP}}$	0.67	MEPS
Loading Factor for IND insurance	κ_{IND}	1.30	MEPS
Insurance Plans		See Table 5	
Permanent Income Distribution	z_p	$\log N(\mu_z,\sigma_z)$	US GDP and Median Income

Displays model parameters estimates directly or close-to-directly from data as well as the data source or aggregate target. See discussion for details on each parameter.

The mortality and emergency expenditure functions are all estimated directly from individual-level data on age, health, mortality, and emergency expenditure in the MEPS. Mortality and the probability of positive emergency expenditure are estimated using logit regression while the mean and variance of emergency expenditure, conditional on positive expenditure, are estimated with linear regression. In the case of the variance of emergency expenditure, I construct the individual-level variance for each observation as the squared residual from the regression used to estimate mean emergency expenditure. I then regress this individual-level variance on the predictors which recovers the best linear predictor of $\mathbb{E}([Y_i - \mathbb{E}(Y_i|X_i)]^2|X_i)$ where Y_i is emergency expenditure and X_i are the predictors which is exactly the definition of conditional variance. The procedure is very similar to performing a Breusch–Pagan test of heteroskedasticity. Table 4 columns (1), (4), (7), and (10) display the regression results for the outcomes of mortality, greater than 0 emergency expenditure, mean emergency expenditure, and the variance of emergency expenditure respectively. The remaining columns detail robustness checks described in further detail in subsection 4.1.3.

Table 4: Detailed Results of Mortality and Emergency Spending Regression

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Mortality	Mortality	Mortality	Emerg. > 0	Emerg. > 0	Emerg. > 0
Health	-5.756***	-5.785***	-3.604***	-4.932***	-4.887***	-4.276***
	(0.703)	(0.748)	(0.807)	(0.196)	(0.207)	(0.220)
Age	0.0562***	0.0529***	0.0641***	0.00736***	0.00746***	0.00831***
	(0.0104)	(0.0108)	(0.0104)	(0.00136)	(0.00140)	(0.00138)
Observations	11,158	9,808	11,158	11,158	10,877	11,158
Controls		YES			YES	
Self-Reported Health			YES			YES

	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	log(Emerg.)	log(Emerg.)	log(Emerg.)	Variance	Variance	Variance
Health	-2.315***	-2.595***	-1.537***	-0.213	-0.341	-0.0321
	(0.227)	(0.241)	(0.270)	(0.587)	(0.636)	(0.696)
Age	-0.00465	-0.0110	-0.00796	0.0158	0.0182	0.0151
	(0.00941)	(0.00974)	(0.00939)	(0.0217)	(0.0228)	(0.0218)
Age^2	4.81e-05	9.65e-05	9.70e-05	-0.000133	-0.000163	-0.000125
	(8.86e-05)	(9.15e-05)	(8.85e-05)	(0.000207)	(0.000216)	(0.000209)
Observations	3,752	3,648	3,752	3,752	3,648	3,752
R-squared	0.035	0.043	0.042	0.000	0.003	0.001
Controls		YES			YES	
Self-Reported Health			YES			YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Displays the results of regressions of various health outcomes on health, age, and controls. Columns 1 through 3 display the results of the regression of mortality on age and health with no controls, a battery of controls, and a control for self-reported measures of health respectively. Columns 4 through 6 display the same regression with the probability of positive emergency expenditure as the outcome. Columns 7 through 9 and 10 through 12 display the same results for the outcomes of log emergency spending of the variance of log emergency spending respectively. Calculated using MEPS data from 2018.

As mentioned in a previous section, I allow the labor disutility share of healthcare labor α_h to vary as a deterministic function of permanent income in order to capture the notion that healthcare workers are disproportionately high income. To discipline this with data, I turn to the American Community Survey (ACS). I limit my sample to employed adult individuals and estimate the probability that a given individual is classified as working in the healthcare industry as a function of the log of individual income using logit regression. The details of the regression can be found in the appendix. The predicted probabilities, denoted $P(\text{healthcare}|\log(\text{income}))$, range from about 3 percent at the bottom of the income distribution to over 15 percent at the top of the income distribution. Under the normalization that the baseline steady-state relative wage of healthcare is equal to one[7], the relative labor supply curve gives a straightforward relationship between relative labor supply towards healthcare and α_h, allowing me to choose α_h as a function of z_p to precisely match the pattern found in the data.

The estimated insurance plan parameters are listed in Table 5. The Medicare and uninsurance plans are straightforward as government-prescribed deductibles, copay rates, and coinsurance rates are easy to find. Although many Medicare plans are actually administered by private insurance companies, referred to as Medicare Advantage, I assume that these privately administered plans are competitive with the government-administered plan and provide roughly the same benefits. Although Medicaid plan parameters are similarly prescribed, they vary heavily from state to state. To avoid the complications of synthesizing this wide variety of plans, I simply estimate parameters for a representative Medicaid plan.

Estimation of the copay, deductible, and coinsurance parameters for the employer-provided, marketplace, and Medicaid insurance plans is more involved and necessitates its own discussion. The copay rate of the employer-provided plan is calculated by summing out-of-pocket costs for non-emergency care across all individuals listed as having healthcare through their employer and dividing by the sum of total costs, both out-of-pocket and covered by insurance, for these same individuals. Thus the copay rate for the employer-provided plan is the ratio of total out-of-pocket non-emergency care costs to total non-emergency care costs for all individuals covered by employer-provided insurance. The copay rates for marketplace insurance and Medicaid are calculated similarly.

The coinsurance rate for both plans can be calculated similarly with a small adjustment for the deductible. Instead of summing all out-of-pocket costs on emergency care for the numerator of the calculation, I could sum all out-of-pocket costs in excess of the individual's deductible. This is what I do; however, one complication arises from the fact that the public-use version of the MEPS data does not contain precise information about an individual's deductible. Instead, I only observe whether their plan falls into the categories of "zero deductible", "normal deductible", or "high deductible". I solve this problem by imputing an individual's deductible as the mean deductible of the category in which they fall[8] or with zero in the case of Medicaid. With this imputation, estimating deductibles and coinsurance rates becomes straightforward using the above method.

As mentioned in the model section, the premium for the employer-provided and marketplace plans is determined as part of the recursive competitive equilibrium. The premiums on the government-provided plans and uninsurance and determined exogenously and are calibrated to their statutory levels. Importantly, the price of uninsurance is set to \$0, representing the lack of any individual mandate.

[7]This normalization is possible because $\frac{w_h}{w_c}$ and $\frac{p_h A_h}{A_c}$ are not separately identified in steady-state

[8]The means of each category are listed separately in MEPS data tables available online. Details on this imputation can be found in the appendix.

Table 5: Estimated Insurance Plan Parameters

Plan	Copay Rate	Deductible	Coinsurance Rate	Premium
Employer-Provided	0.280	$2,400	0.103	Deter. in Eq.
Marketplace	0.383	$2,400	0.126	Deter. in Eq.
Medicaid	0.02	$0	0.02	$0
Medicare	0.20	$1,484	0.0	$180
Uninsured	1.0	$0	1.0	$0

Displays the calibrated and estimated insurance plan parameter. See discussion for details

4.2.3. Parameters Calibrated Internally

The remaining parameters are calibrated internally using the simulated method of moments. Table 6 lists the targeted moments, their values in the model and data as well as the source of the data value, and their rough correspondence to model parameters. The most straightforward correspondence is that between the parameter $\bar{u}$, which governs the utility an individual received for being alive each period, and individuals' value of statistical life within the model. There is a large literature attempting to estimate the average VSL within the US which finds answers ranging from $1 million (Ashenfelter and Greenstone, 2004) to $10 million (Viscusi and Aldy, 2003). This is further complicated by the fact that there is strong evidence that individual VSL varies with income (Hammitt and Robinson, 2011). The baseline model reproduces this property but, as discussed in subsection 4.2.1 above, lacks a parameter that can be used to independently target this elasticity. As a result, the model-implied elasticity may vary substantially from the data. To help compensate for this, I target the mean VSL for the marginal Medicaid recipient implied by a given elasticity rather than the mean VSL over the entire population. Using this approach, even if the elasticity is incorrect, the deviations in the model VSL from reality exist among the wealthy who are unaffected by the Medicaid expansion. I use a baseline VSL of $7.5 million and assume an income elasticity of 1. The marginal Medicaid expansion recipient has annual income equal to 138 percent of the federal poverty level which is roughly 30 percent of mean income with the US. Combining these numbers yields a VSL of the marginal recipient equal to $2.25 million which is my target used in estimation of the model.

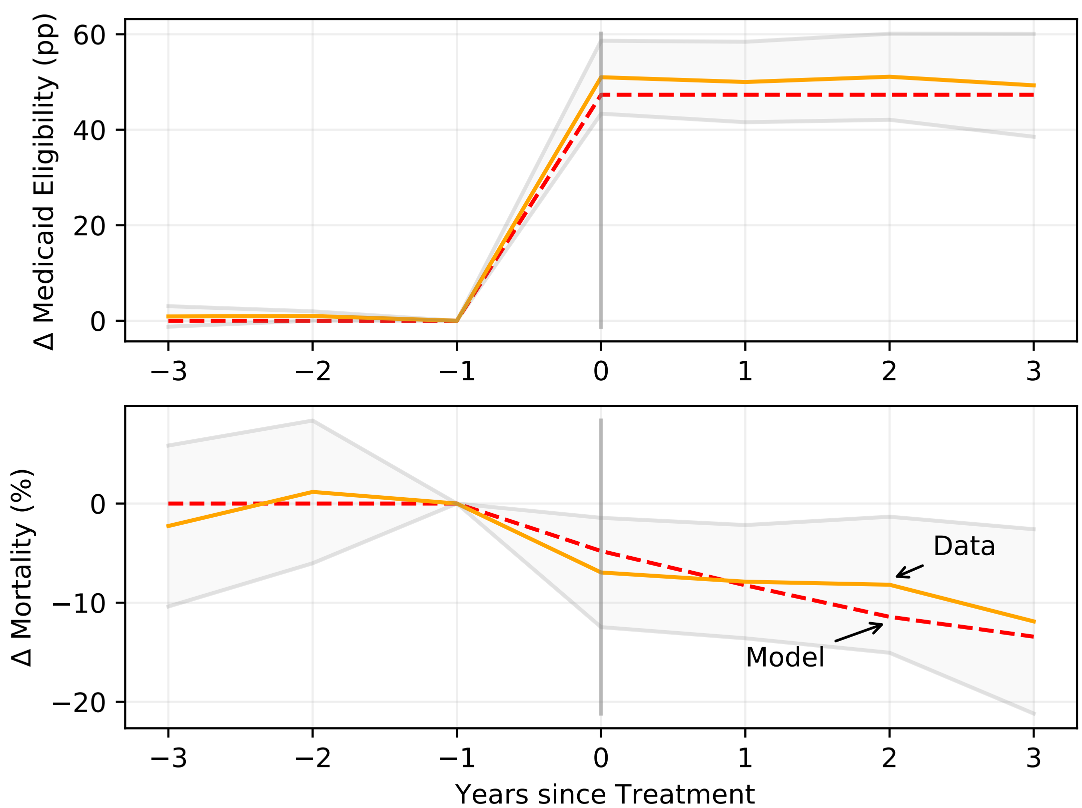
A substantial portion of the internally calibrated moments correspond to the parameters governing the accumulation and depreciation of health. As mentioned in the model section, the depreciation rate of health is allowed to be age-dependent. I choose to model it as a simple linear relationship so that $\delta_a = \beta_0 + \beta_1 a$. The level parameter β_0 is pinned down by average health while the slope parameter β_1 is pinned down by the standard deviation of healthcare expenditure. Finally, the portion of depreciation that occurs upon an emergency health event δ_e is pinned down by the average difference in health between those who experienced an emergency in a period and those who did not.

The age-dependent effectiveness of medical care at building health ϕ_a is also restricted to be linear. The slope is determined by the covariance between age and frailty. The level is largely pinned down by the mortality reduction due to Medicaid expansion measured in Miller et al. (2021), which estimates the effect of Medicaid using a diff-in-diff specification comparing states that expanded coverage in the 2014 ACA expansion to those that did not. I replicate this diff-in-diff in the model by first calculating the steady-state distribution under the pre-expansion Medicaid productivity cutoff $\bar{z}_{\text{PRE}}$. I then select a measure 0 subset of individuals with ages between 55 and 64 and who would qualify for

Medicaid post-expansion, following the sample selection procedure in Miller et al. (2021). I randomly assign selected individuals to treatment and control and increase the Medicaid qualification cutoff for the treatment group from $\bar{z}_{\text{PRE}}$ to $\bar{z}$. The diff-in-diff equivalent can then be measured by comparing outcomes between the treatment and control groups. I select the value of $\bar{z}_{\text{PRE}}$ so that the quasi-experiment in the model replicates the change in Medicaid eligibility estimated in the actual quasi-experiment, ensuring that the model experiment matches the reduction in mortality per newly eligible individual.

Figure 6 shows the results of the quasi-experiment in both the model and the data. I use the sum of squared differences between the reduction in mortality observed in the data and the model-implied reduction as the targeted moment for the SMM and set its target to 0.

Figure 6: Effect of Medicaid on Mortality: Model vs Data



This figure displays the impact of Medicaid expansion on Medicaid qualification and mortality of low-income 55-64 year old adults. The solid orange line displays the effects estimated in Miller et al. (2021) using a diff-in-diff design. The grey bands represent the estimated 95% CI. The dotted red line displays the change in mortality in the calibrated model. See the discussion for details on how the diff-in-diff design is replicated in the model.

The decision to select a measure 0 subset of individuals is in keeping with the growing literature on using experimental or quasi-experimental evidence to discipline macroeconomic models; however, in this particular case it is not clear if this is the correct decision. The choice is usually justified by the notion that the experimental group is a tiny subset of the overall population and thus treatment is unlikely to influence any general equilibrium variables. In this case, the quasi-experiment is performed "at scale" and is implemented by a substantial number of states. This is further complicated by the fact that it is not obvious how the insurance markets and labor markets between states are linked, especially in such a short time frame and by the fact that Medicaid eligibility requirements were highly heterogeneous across states before expansion. Despite these complications, I replicate the quasi-experiment on a measure 0 subset due to the dramatic increase in computational speed that such an assumption provides.

The returns to scale parameter in health production ψ is chosen to match quasi-experimental evidence on the extent of delayed care from Card et al. (2008). The authors show that the use of medical care jumps discretely when an individual turns 65 and becomes eligible for Medicare. As discussed in subsection 3.7, the model reproduces this property. The returns to scale parameter ψ is a large determinant of the size of the jump in the model for an intuitive reason; it represents the intertemporal elasticity of substitution for healthcare spending. If ψ is large, then medical care tomorrow is a good substitute for medical care today and the incentive introduced by the discrete jump in insurance at age 65 generates a large jump in expenditure at age 65 due to the delayed care effect. In contrast, if ψ is small, medical care tomorrow is a poor substitute for medical care today, reducing the incentive to delay care even if the agent anticipates a reduction in their copay rate in the future.

Although Card et al. (2008) measure large changes in the use of medical care, they do not measure the jump in overall health care expenditure. To translate their results to outcomes that can be compared to outcomes in the model, I select a handful of common medical procedures (Removal of arterial obstructions, heart bypass surgery, knee and hip replacements, and gall bladder removals) that they measure and calculate the simple average of the percentage increase in the frequency of these procedures at age 65. I use this average percentage increase as the target for the average increase in medical expenditure at age 65. Figure 7 displays these data, their average, and the average jump in medical expenditure at age 65 in the calibrated model.

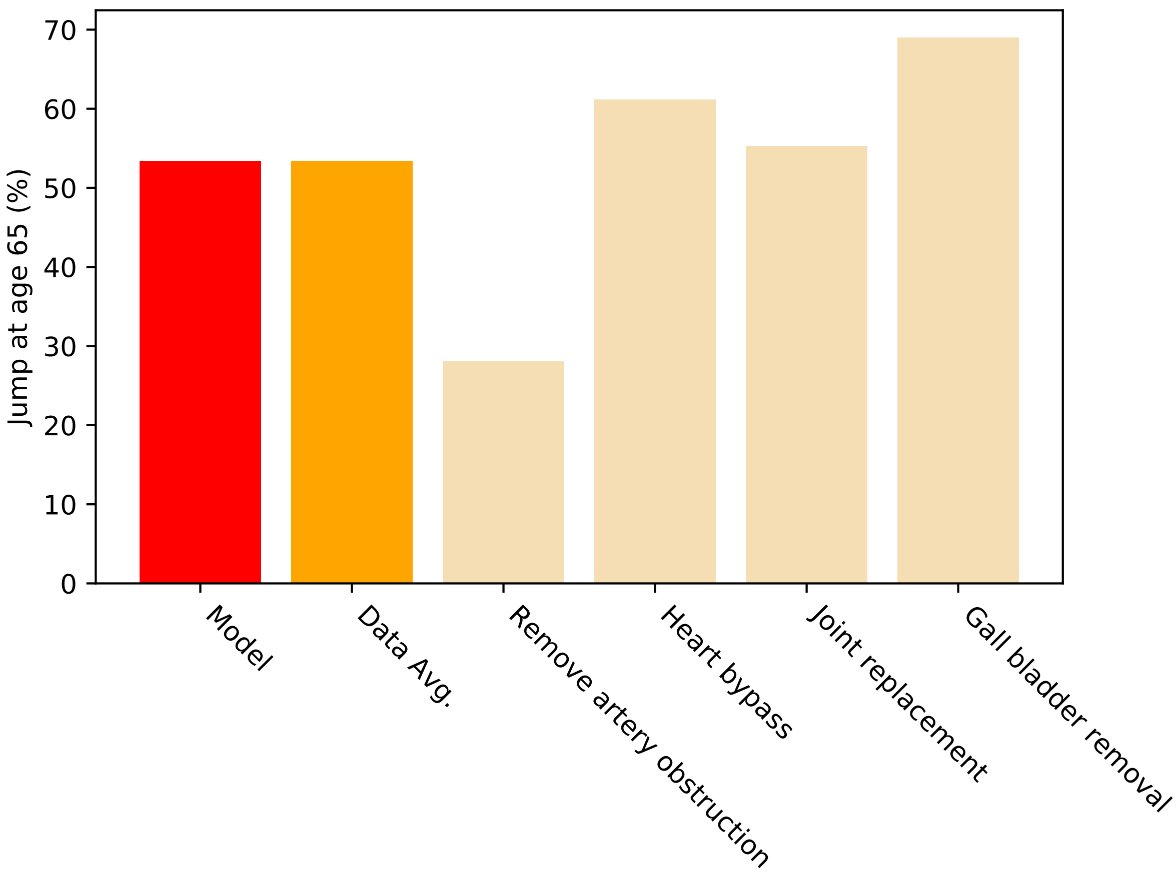
Finally, with the health production function pinned down, the parameter χ, which governs the magnitude of poorly-informed agents' misperception, pins down average health spending.

Table 6: Parameters Estimated by SMM

Moment	Model	Data	Source	Parameter
Avg. VSL of Medicaid Recipiant	\$2 million	See discussion		$\bar{u}$
Jump in Medical Exp. at 65	See Figure 7		Card et al. (2008)	ψ
Mortality Response to Medicaid	See Figure 6		Miller et al. (2021)	ϕ_a
Mean of Health Spending	\$6,220	\$6,086	MEPS	χ, I
SD of Health Spending	\$4,359	\$10,047	MEPS	δ
Avg. Health	0.886	0.877	MEPS	δ
cov(Health, Age)	-1.11	-1.21	MEPS	ϕ_a
mean(Health$\|x>0$) - mean(Health$\|x=0$)	-0.045	-0.090	MEPS	δ_e

Displays the moments targeted in the simulated method of moments estimation along with their value in both the estimated model and the data. Also displays a rough correspondence between targeted moments and model parameters. See discussion for more details.

Figure 7: Jump in Medical Expenditure at age 65: Model vs Data

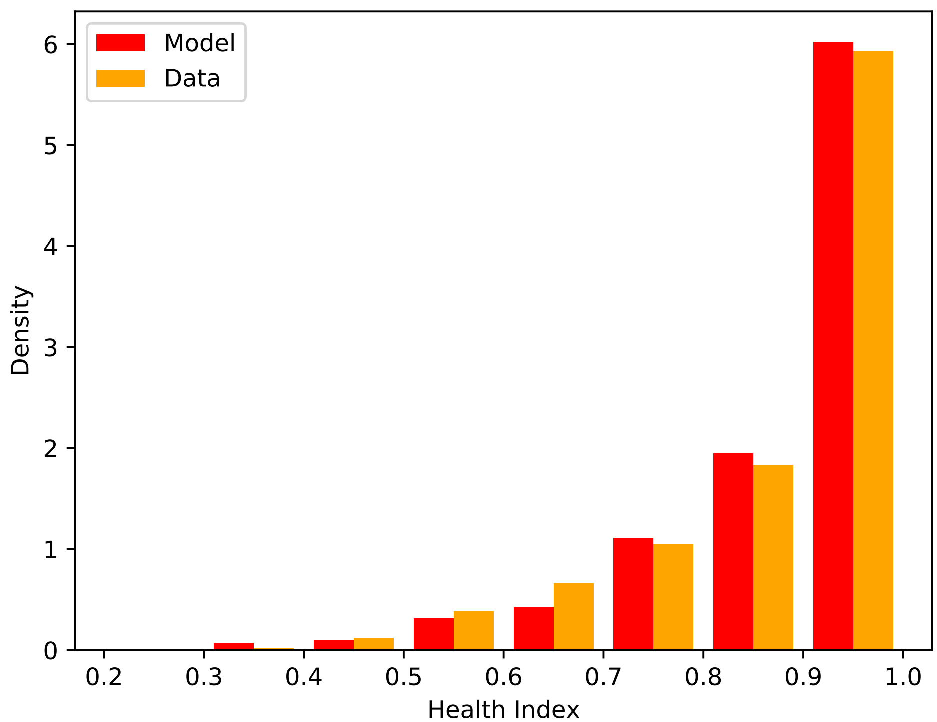


This figure displays the jump in various costly non-emergent medical procedures at age 65 estimated in Card et al. (2008) in tan. The orange bar represents the unweighted average of these four estimates. The red bar represents the jump in average medical expenditure in the pre-Medicaid expansion steady-state of the calibrated model.

4.3. Model Validation

To validate the model, I first begin by comparing the distribution of health in the model to the distribution of health as measured in MEPS data. This is displayed in Figure 8 as a pair of overlapping histograms where the red histogram displays the distribution of health in the model and the orange histogram displays the distribution of health measured by the MEPS. Although the mean of this distribution is the only moment targeted in model estimation, it is clear that the model successfully replicates the stark features of the data distribution, including bunching at the top and a thin left tail.

Figure 8: The Distribution of Health in the Data and Model

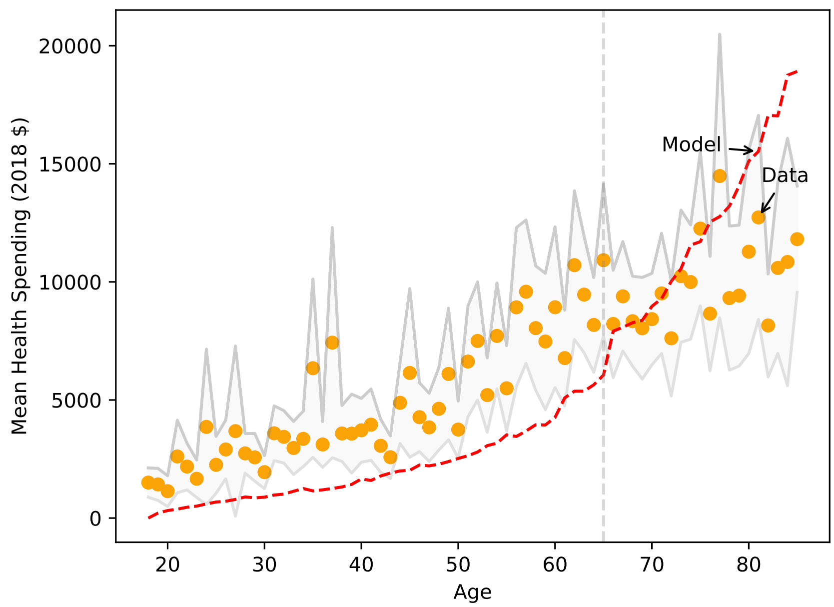


Displays the distribution of health in the Medical Expenditure Panel Survey, as measured by the health index based on Hosseini et al. (2021), in orange. See the discussion in section 4.1.3 for more details. The distribution of health in the calibrated model is displayed in red.

Figure 9 displays average health investment spending (in 2018 dollars) for each age between 18 and 85. The orange dots represent the data averages while the model averages are displayed by the red dotted line. The model replicates the basic relationship between healthcare spending and age well. Annual spending is low early in life and increases substantially as age increases, with a particularly fast increase after the age of 70. The model does miss somewhat on the level of spending, particularly for young individuals where the model predicts less spending than is observed in the data.

Figure A.3 displays average consumption, savings, and mortality for each age between 18 and 85 within the model. Although there is no data component to these observations, it is encouraging that all three exhibit standard life-cycle behavior with consumption increasing over the life-cycle and falling at retirement age, assets increasing during working years and then falling during retirement, and mortality increasing sharply in old age.

Figure 9: Average Health Spending by Age: Model and Data



Displays average preventative health spending for the year 2018 as measured in the MEPS as a function of age in orange. The 95 percent confidence interval is shaded. The red line displays average preventative health spending in the post-expansion calibration of the model.

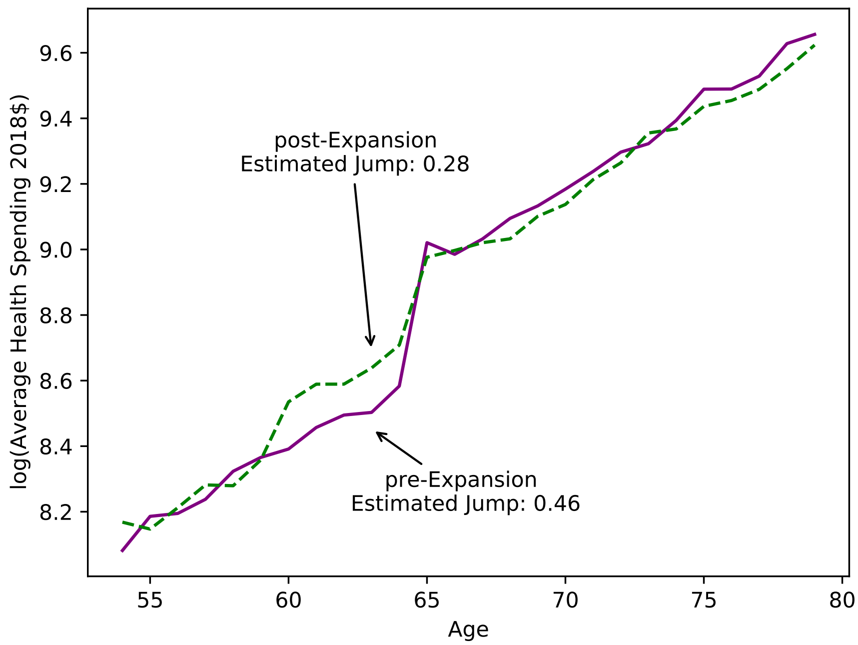
4.3.1. Validation from Pre- and Post- Medicaid Expansion Data

A stark prediction of the model is that the jump in healthcare consumption at age 65 should be significantly muted after public health insurance expansion as a result of reduced incentives to delay care. This prediction is displayed in Figure 10 and is discussed further in Section 5. In particular, the estimated increase in healthcare spending (using a regression discontinuity design on the model-generated data) at age 65 is 46 percent before public health insurance expansion and falls to 28 percent after expansion.

To test the validity of this model prediction, I perform the same exercise on data from before and after the 2014 ACA Medicaid expansion to test whether expansion actually decreased the observed discontinuity at age 65. To do this, I use pooled first-round MEPS data from 2010 to 2012 as my pre-expansion dataset and data from 2017 to 2019 as my post-expansion dataset. One issue with public-use MEPS data is that the exact date of medical expenditure is not reported and, instead, I only observe total expenditure for the entire year. As a result, for any individuals who start the year at age 64 and turn 65 at some point during the year, I will observe expenditure from both before and after they receive Medicare, effectively erasing any impact of the cutoff. To eliminate this issue, I drop all individuals who are reported to be age 65 as of December 31st of the year they are observed.

The results of the regression discontinuity are displayed in Table 7. Column (1) displays the results using the MSE-optimal bandwidth which is determined to be 5.88. While the point estimates support the broad prediction of the model that the estimated jump will be smaller post-expansion, the estimates are much too imprecise to be of any real use. One way to improve the precision of the estimates is to increase the bandwidth to allow more data to inform the estimation. Column (2) displays the results for a bandwidth of 25. The estimates here are much more precise with the standard errors for both the pre- and post- expansion results falling by more than half. The estimated jump in

Figure 10: Model Implied Delayed Care Pre- and Post- Expansion



Displays the log of average preventative health spending. The purple line shows spending from the pre-expansion steady-state of the calibrated model while the dotted green line shows spending the from post-expansion steady-state. The jump at age 65 is estimated using a regression discontinuity design with a cutoff at age 65, implemented using the Stata package **rdrobust**.

Table 7: Regression Discontinuity Results: Pre- and Post- Medicaid Expansion

	(1) log(Health Spending)	(2) log(Health Spending)	(3) Model
Pre and Post Difference	0.29784	**0.15222**	**0.17643****
pre-Expansion	0.21564	0.25294***	0.45637***
	(0.196)	(0.085)	(0.047)
Observations	37,719	37,719	500,000
post-Expansion	-0.08323	0.10072	0.27994***
	(0.174)	(0.073)	(0.057)
Observations	32,581	32,581	500,000
Bandwidth	5.88	25.0	5.414
Optimal Bandwidth?	YES	NO	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

This table displays the results of a regression discontinuity design to estimate the jump in medical expenditure that occurs at age 65 when individuals gain universal health coverage through Medicare. The pre-expansion panel uses MEPS data from the years 2010 to 2012 while the data in the post-expansion panel uses data from 2017 to 2019. Column (3) displays the results of applying the same empirical procedure to model-generated data. All estimates were calculated using the Stata package **rdrobust**.

healthcare spending at age 65 in the pre-expansion data is 25 percent and is significant at the 1% level. This is higher than the jump of 46 percent in the model which, recall, is estimated to match the results from Card et al. (2008). In the post-expansion data, the estimated jump falls by roughly 15.2 percentage points to a statistically insignificant 10 percent. This is remarkably close to the 17.6 percentage point fall (from 46 percent to 27 percent) predicted by the model. Although the large standard errors preclude any strong statements about these results (for example, the 95% confidence interval for the pre-expansion point estimate contains the post-expansion point estimate), I interpret this as strong suggestive evidence for the model's primary mechanism, namely that public health insurance expansion can reduce delayed care.

5. Quantitative Results

In this section, I use the calibrated model to evaluate the results of an expansion in public health insurance similar to the 2014 ACA Medicaid expansion. In the model, this takes the form of an increase in the productivity cutoff for Medicaid eligibility from $\bar{z}_{\text{PRE}}$ to $\bar{z}_{\text{POST}}$. The post-expansion cutoff $\bar{z}_{\text{POST}}$ is chosen so that the highest-earning Medicaid-qualifying individual has income equal to 138 percent of the federal poverty level, as is the case in states that expanded Medicaid. The pre-expansion cutoff $\bar{z}_{\text{PRE}}$ is chosen so that the increase in Medicaid eligibility from expansion matches that measured in Miller et al. (2021) (see the previous section for details). Because Medicaid qualification criteria were highly heterogeneous across states before 2014, matching the aggregate change in eligibility, rather than pre-expansion statutory qualification criteria, is the most natural way to represent the expansion at a national level.

5.1. Reduction in Delayed Care

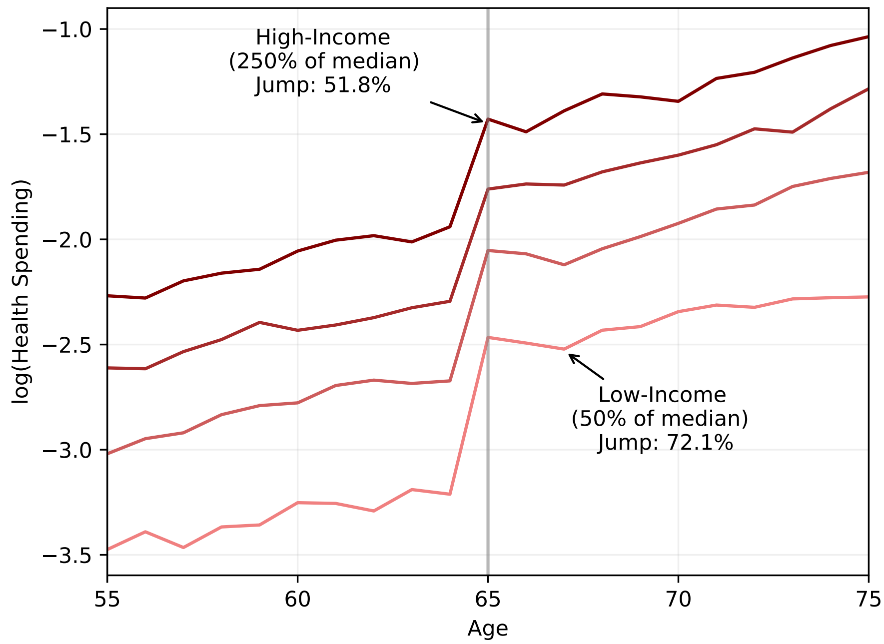
A natural first question is whether or not Medicaid expansion actually works to reduced delayed care. Figure 10 displays the log of average healthcare spending for individuals aged 55 to 80. The purple line displays the mean for the pre-expansion steady state while the dotted green line displays the mean for the post-expansion steady state. From the figure, it is clear that there is less delay of care in the post-expansion steady state. The reduction in delayed care is substantial; the RDD-estimated jump in expenditure at age 65 falls by about 18 percentage points from 46 percent to 28 percent. The reduction occurs almost entirely between the ages of 60 and 64. Spending between these ages is 13 percent higher after expansion. For individuals younger than age 60, average spending only increases by 2.9 percent. Given the relatively smaller share of total healthcare spending by individuals younger than 60, this increase is fairly small.

Although the decrease in delayed care is substantial, the reduction in expenditure for individuals older than age 65 is, at least in relative terms, small. Aggregate healthcare spending on individuals aged 65 or older is 2.7 percent lower in the post-expansion steady state than in the pre-expansion steady state; however, because late-in-life expenditures make up a substantial portion of healthcare costs, a 2.7 percent reduction is still large in absolute terms.

Figure 11 provides some insight into which individuals are most responsible for delaying healthcare. It displays the log mean of health spending as a function of individual age and permanent income in the pre-expansion steady-state of the model. The dark maroon line plots the spending of high-income individuals with average income equal to 250 percent of the median while the light pink line displays the spending of low-income individuals with average income equal to 50 percent of the median. It is important to note that the cutoff of 50 percent of the median is high enough that these individuals are very unlikely to receive Medicaid in the pre-expansion steady-state.

Two patterns are apparent in the figure. The first is that delayed care is substantial at all income levels. Even high-

Figure 11: Pre-Expansion Health Spending by Age and Income



Displays the log of average health spending as a function of age and permanent income in the pre-expansion steady-state of the model.

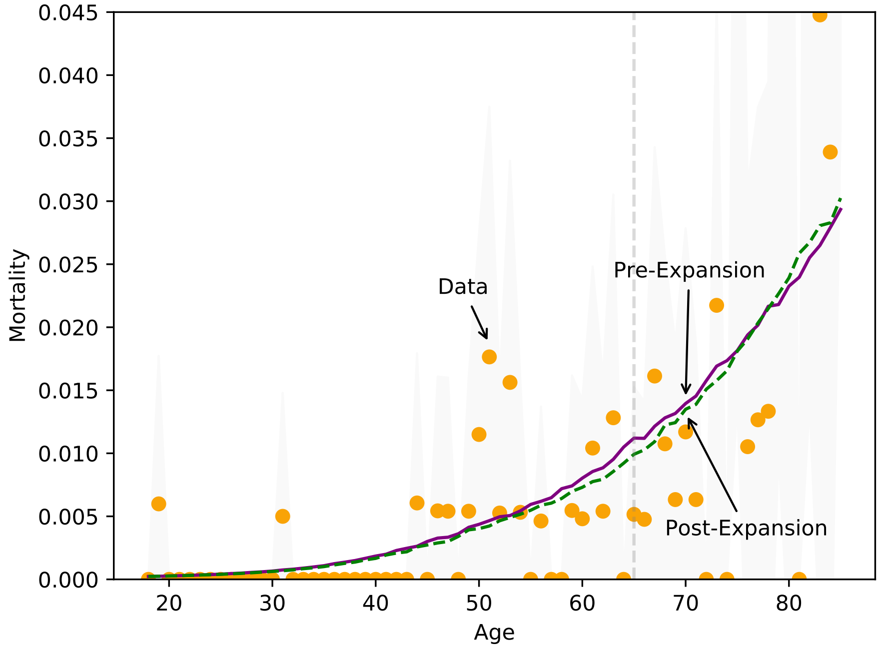
income earners who make more than 250 percent of median earnings exhibit a 51.8 percent jump in healthcare spending at age 65. The reason for this can be seen by looking at the estimated insurance parameters displayed in Table 5. Although high-income earners are more likely to be insured, access to Medicare still represents, at least on average, an improvement in insurance. The average copay rates for employer-provided and marketplace insurance are 0.28 and 0.38 respectively while the copay rate under Medicare is only 0.2. Thus even high-income individuals are incentivized to delay care. The second pattern is less surprising; low-income individuals delay care more than high-income individuals. Low-income individuals exhibit a much larger jump in health spending of 72.1 percent at age 65.

5.2. Reduction in Mortality

How substantial is the reduction in mortality due to less delayed care? Figure 12 displays mortality as a function of age in the pre- and post- expansion steady states. Mortality in the pre-expansion steady state is displayed by the purple line while the dotted green line displays mortality from the post-expansion steady state. Although the decline in mortality is small (recall the model is calibrated to match the decline in mortality measured in Miller et al. (2021)), it is clear that mortality declines for all individuals younger than age 75. The decline is largest for individuals between ages 60 and 64, consistent with the idea that this decline in mortality is driven by the decline in delayed care. For these individuals, mortality declines by 0.085 percentage points, a 9.6 percent decline in total mortality.

Total mortality for individuals older than age 75 increases as a result of expansion. This is the result of selection effects. Expansion reduces mortality, and individuals with low health are more likely to survive to old age. As a result, the average health of individuals older than age 75 counter-intuitively drops by 0.04 from 0.835 to 0.831. This decline in average health leads to an increase in average mortality despite the fact that life expectancy for every individual has

Figure 12: Model Implied Mortality Pre- and Post- Expansion



Displays average mortality as a function of age. The purple line shows mortality from the pre-expansion steady-state of the calibrated model while the dotted green line shows mortality the from post-expansion steady-state. The orange dots display mortality as measured in 2018 MEPS data with the 95 percent confidence interval shaded.

increased.

5.3. Cost of Expansion

After establishing that Medicaid expansion is effective in reducing delayed care, I turn to examining how this reduction impacts the long-run cost of expansion. Panel A Column 1 of Table 8 displays the increase in Medicaid coverage and the associated long-run Medicare savings per \$100 spent on expansion for the estimated model. In the baseline model, Medicare costs are reduced by \$49.63 for every \$100 spent on Medicaid expansion. This sizable reduction is a reflection of the fact that late-in-life medical care makes up the bulk of medical expenses.

The reduction of \$49.63 is the net result of an increase in costs due to higher mortality and a decrease in costs due to earlier care being more efficient. To separately measure the contribution of each channel, I use two counterfactual models. In the first counterfactual, the mortality function $\pi(a,h)$ is replaced with an alternative function $\tilde{\pi}(a,h)$ that does not depend on health so that $\frac{\partial \tilde{\pi}}{\partial h} = 0$. In this model, the increase in costs due to lower mortality is eliminated, as increases in health due to public health insurance coverage will no longer reduce mortality. However, it also eliminates the primary incentive, reduced mortality, for agents to invest in their health. To maintain an incentive, I also add a dependence on health to the agent's utility function so that they value being healthy. I discuss this change, along with other details about the counterfactual model, further in the appendix; however, it is worth noting that I choose the utility function for health so that the optimal policy function for medical spending is identical to that of the full model, ensuring that any differences in aggregate outcomes between the two models are due to the impact of $\tilde{\pi}$ on the distribution of households rather than on individual decisions. I choose $\tilde{\pi}$ so that average mortality for each age a in the counterfactual model is identical to average mortality in the full model in the pre-expansion steady-state.

The results of Medicaid expansion in this counterfactual model with exogenous mortality are displayed in Panel A Column 2 of Table 8. In this model, Medicare costs decrease by \$56.93 for each \$100 spent on Medicaid. The fact that this reduction is \$7.30 larger than in the full model is a result of the fact, in the full model, expanding Medicaid saves lives and results in a larger pool of individuals older than 65 for whom Medicaid must provide coverage, increasing total costs. When mortality is exogenous, however, expansion no longer carries this additional cost, leading to a larger reduction in total Medicare spending.

The second counterfactual model takes health spending as exogenous, turning off the delayed care channel and, implicitly, the mortality channel. Instead of endogenously choosing health investment i, individuals must consume an exogenous amount of healthcare goods each period $i(h,z_p,z_t,m)$ which is allowed to depend on an individual's current health status, productivity, and emergency shock m. Health follows the same accumulation process as in the full model. Under exogenous health expenditure, individuals no longer adjust their healthcare spending as a result of expansion and thus there is no avenue for expansion to reduce the extent of delayed medical care. Additionally, because health investment and thus an individual's level of health is now entirely a product of exogenous processes, this change also implicitly eliminates the mortality reduction channel as expansion will no longer save lives. Similar to the previous counterfactual model, $i(h,z_p,z_t,m)$ is chosen to match the average healthcare spending for each health-productivity-shock bin in the steady state of the full model. Details can be found in the appendix.

The results of expansion in the model with exogenous health expenditure are displayed in Panel A Column 3. By construction, Medicaid expansion results in no net change in Medicare spending as spending and health are determined entirely endogenously. Thus, going from the first counterfactual model with exogenous mortality to the second counterfactual model where both health spending and mortality are exogenous, effectively turning off the delayed care channel, decreases the Medicare savings per \$100 spent on expansion by \$56.93.

Taken together, these results suggest that both delayed medical care and mortality reductions play an important role in determining the long-run costs of Medicaid expansion; however, the delayed medical care channel dominates quantitatively. As shown by the counterfactual models, the mortality reduction channel increases Medicare purchases by $7.30 for each $100 spent on Medicaid while the delayed care channel reduces them by $56.93 for each $100. Taken together, these two channels result in a reduction in total Medicare purchases of $49.63 for each $100 spent on Medicaid which results in a cost-savings ratio of 0.4963 and helps offset the cost of Medicaid expansion by almost exactly half.

Panel B of Table 8 displays the total aggregate impact of expansion for the baseline and two counterfactual models. Column (1) displays the long-run implications of Medicaid expansion for annual Medicaid and Medicaid spending as well as the increase in tax receipts necessary to fund the expansion and the change in the relative price of healthcare goods due to the increase in relative demand. Expansion increases in total Medicaid spending by 1.37 percent of pre-expansion GDP while Medicare spending drops by 0.68 percent of GDP, an indicator that the delayed care channel is quite strong. Total taxes increase by 0.4 percent of pre-expansion GDP.

Table 8: Long-Run Effects of Medicaid Expansion

Panel A: Relative Effects

Variable	(1) Post-Expansion	(2) Exo. Mortality	(3) Exo. Medical Spending
Medicaid Coverage (% Population)	+15.7%	+12.3%	+15.7%
Medicare Savings per $100 Spent	$49.63	$56.93	$0

Panel B: Absolute Effects

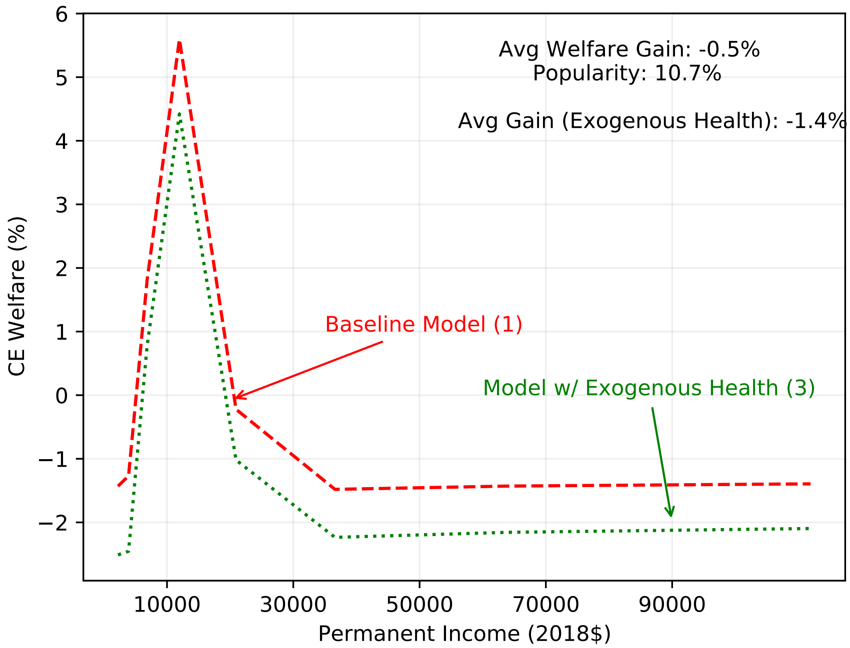
Variable	(1) Post-Expansion	(2) Exo. Mortality	(3) Exo. Medical Spending
Medicaid Coverage (% Population)	+15.7%	+12.3%	+15.7%
Total Medicaid Spending	+1.37% of GDP	+1.37% of GDP	+1.29% of GDP
Total Medicare Spending	−0.68% of GDP	−0.78% of GDP	−0.00% of GDP
Total Tax Receipts	+0.40% of GDP	+1.04% of GDP	+1.13% of GDP

Column 1 displays the changes in coverage, public healthcare spending, tax receipts, and the price of healthcare goods between the pre-expansion and post-expansion steady-states in the full estimated model. Column 2 displays these same changes for an alternative model with exogenous morality while column 3 displays the same changes for a model with exogenous medical expenditure. See discussion for details.

The results of Medicaid expansion in this counterfactual model with exogenous mortality are displayed in Column 2 of Table 8. Most notable is the reduction in total Medicare spending in the post-expansion steady-state which is equal to 0.78 percent of GDP. Column 3 of Table 8 displays the results for the counterfactual model with exogenous health spending. Most notable is that there is no change in Medicare spending as a result of expansion; this is a straightforward consequence of making healthcare spending exogenous. Medicaid spending increases by 1.29 percent of GDP, slightly less than in the full model as individuals no longer increase their medical spending in response to receiving coverage. The increase in the price of healthcare goods is dramatically larger than both previous models at 6.41 percent as taking consumption of medical goods as exogenous fixes the price elasticity of demand to 0. This necessitates larger price increases to reach healthcare goods market equilibrium than in the full model where demand responds to price changes.

5.4. Welfare

Figure 13: Consumption-Equivalent Welfare for Newborn Individuals by Permanent Income

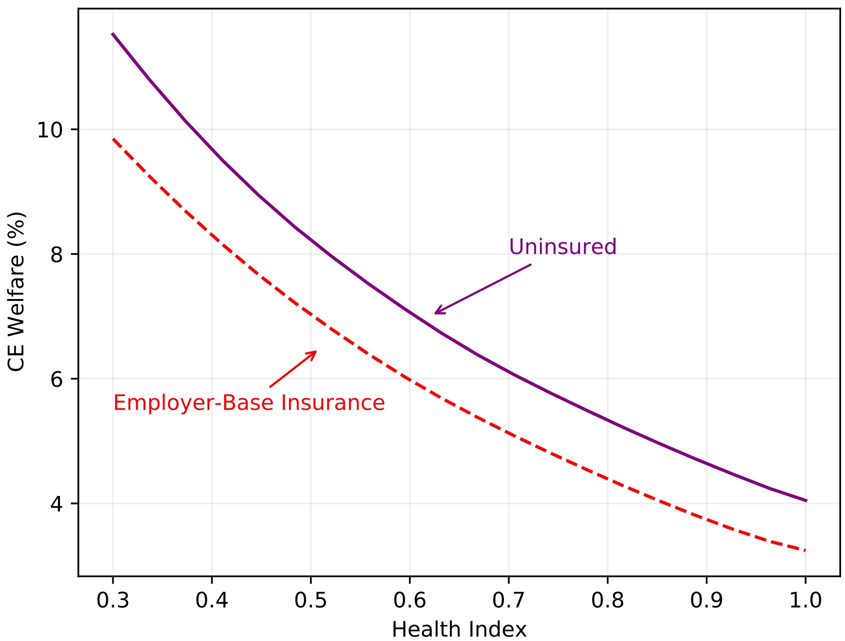


Displays average consumption equivalent welfare gains from Medicaid expansion for newborn individuals as a function of individual permanent income. Welfare for individuals born with access to employer-based insurance is displayed by the red dotted line while welfare for individuals born without coverage is displayed by the purple line. Popularity refers to the percentage of individuals who experience positive welfare gains from expansion.

Figure 13 displays the consumption equivalent welfare change from Medicaid expansion for a newly born individual as a function of individual permanent income. The red dashed line displayed the welfare change for individuals in the baseline model. Welfare is calculated as perceived by each individual ex-ante. In particular, low-information individuals who under-perceive the value of investing in health are asked how much better off they perceive themselves to be, given their information set.

It is evident from the figure that the welfare gains from Medicaid expansion are extremely concentrated. Individuals making around $10,000 per year experience gains as large as 5 percent of lifetime consumption with gains quickly tapering off as income increases and reaching 0 at roughly $20,000. Individuals expecting to earn more than $30,000 each year experience a loss in welfare equivalent to 1.25 percent of lifetime consumption due to higher taxes and distortion in the relative price of healthcare goods. Interestingly, individuals at the bottom of the income distribution also experience substantial welfare losses as they qualified for Medicaid even before expansion and now must pay slightly higher taxes. Despite large gains for low-income households, the average welfare effect of Medicaid expansion is negative. In aggregate, welfare drops by 0.5 percent of consumption. Unsurprisingly, the policy is also very unpopular. Only 10.7 percent of individuals are made better off by expansion and would vote for the policy absent altruistic motivations.

Figure 14: Consumption-Equivalent Welfare for Newly-Covered Age 40 Individuals by Individual Health Status



Displays average consumption equivalent welfare gains from Medicaid expansion for an aged 40 individual who becomes newly covered by expansion as a function of individual health. Welfare for individuals born with access to employer-based insurance is displayed by the red dotted line while welfare for individuals born without coverage is displayed by the purple line.

Examining welfare gains for only newly-born individuals masks important ex-post heterogeneity. Some individuals will receive a good series of emergency shocks of their life and will benefit less from expansion than individuals who receive a bad series of shocks. To examine this heterogeneity, Figure 14 displays the welfare gains for an age 40 individual with a permanent income of $12,000 (i.e. will be newly covered by expansion) as a function of individual health and insurance status. Gains vary substantially with both health and insurance. Healthy insured individuals gain 3.7 percent of consumption while healthy uninsured individuals gain 4.0 percent. These gains climb substantially as health decreases. At a health index of 0.8, an uninsured individual gains 5.3 percent of consumption. Although the health status is strongly left-skewed, there are still a fair enough of low-health individuals who benefit substantially as 20.4 percent of age 40 individuals have a health index of 0.8 or lower.

6. Conclusion

Evidence suggests that a substantial number of U.S. individuals delay healthcare until they receive health insurance through Medicare at age 65. This paper provides a quantitative model to explain this fact and analyze the extent to which public health insurance expansion can reduce individual incentives to delay care. A key question is whether reductions in delayed care lead to cost savings, due to earlier care being more effective than late care, or lead to cost increases due to reductions in mortality increasing the population share of adults over the age of 65 who are covered by Medicare. To discipline these channels, I use quasi-experimental results from the health literature. In particular, the delayed care channel is disciplined using the jump in healthcare consumption at age 65 from the regression discontinuity of Card et al. (2008), and the decline in mortality is disciplined using the 2014 ACA Medicaid expansion diff-in-diff results from Miller et al. (2021).

My results suggest that the cost-saving delayed care channel is substantially larger than the cost-increasing mortality channel. The model predicts that, in the long run, an expansion of Medicaid equivalent to roughly the size of the 2014 expansion under the ACA increases Medicaid spending by 1.37 percent of GDP but decreases Medicare spending by 0.68 percent of GDP. That is, for each dollar spent on Medicaid, Medicare saves 50 cents, resulting in a cost-savings ratio of 0.50. The net decrease in Medicare spending of 0.68 percent of GDP is the result of an increase in spending of 0.10 percent of GDP due to the mortality reduction channel and a 0.78 percent of GDP decrease due to the delayed care channel, leading to the conclusion that the savings from reducing delayed care are roughly eight times larger than the increase in costs. Overall, my findings point towards fairly large spillovers between Medicaid and Medicare spending and indicate that Medicaid expansion may, in the long run, be substantially less costly than a short-term analysis would suggest.

One key limitation of my analysis is that it assumes that all medical spending is productive on the margin and, thus, that all of the observed increase in healthcare consumption at age 65 represents increased investment into individual health. If, instead, increases in consumption at age 65 are a mix of necessary, productive care and of physicians prescribing unnecessary, unproductive care to exploit Medicare reimbursement rules, my analysis will overestimate the extent to which individuals are delaying important healthcare and the cost savings from reducing this delay. Empirical work has argued that some prescribed treatments are overused and have very little or even negative value. Kowalski (2021) is one example for the case of mammograms. Further analysis that explicitly models the decision problem of physicians and leverages solid evidence on the extent of over-prescription would be valuable.

References

Aizawa, N. and Fu, C. (2020), Interaction of the labor market and the health insurance system: Employer-sponsored, individual, and public insurance, Technical report, National Bureau of Economic Research.

Ashenfelter, O. and Greenstone, M. (2004), 'Using mandated speed limits to measure the value of a statistical life', *Journal of political Economy* **112**(S1), S226–S267.

Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., Chiuve, S. E., Cushman, M., Delling, F. N., Deo, R. et al. (2018), 'Heart disease and stroke statistics - 2018 update: a report from the american heart association', *Circulation* **137**(12), e67–e492.

Card, D., Dobkin, C. and Maestas, N. (2008), 'The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare', *American Economic Review* **98**(5), 2242–58.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.98.5.2242*

Cohen, S. B. (2003), 'Design strategies and innovations in the medical expenditure panel survey', *Medical care* pp. III5–III12.

Cole, H. L., Kim, S. and Krueger, D. (2019), 'Analysing the effects of insuring health risks: On the trade-off between short-run insurance benefits versus long-run incentive costs', *The Review of Economic Studies* **86**(3), 1123–1169.

Córdoba, J. C. and Ripoll, M. (2017), 'Risk aversion and the value of life', *The Review of Economic Studies* **84**(4), 1472–1509.

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

De Nardi, M., French, E. and Jones, J. B. (2016), 'Medicaid insurance in old age', *American Economic Review* **106**(11), 3480–3520.

Fang, H. and Krueger, D. (2021), The affordable care act after a decade: Its impact on the labor market and the macro economy, Technical report, National Bureau of Economic Research.

Finkelstein, A. (2007), 'The aggregate effects of health insurance: Evidence from the introduction of medicare', *The quarterly journal of economics* **122**(1), 1–37.

Floden, M. and Lindé, J. (2001), 'Idiosyncratic risk in the united states and sweden: Is there a role for government insurance?', *Review of Economic dynamics* **4**(2), 406–437.

Fukuda, H. and Mizobe, M. (2017), 'Impact of nonadherence on complication risks and healthcare costs in patients newly-diagnosed with diabetes', *Diabetes research and clinical practice* **123**, 55–62.

Gehi, A. K., Ali, S., Na, B. and Whooley, M. A. (2007), 'Self-reported medication adherence and cardiovascular events in patients with stable coronary heart disease: the heart and soul study', *Archives of internal medicine* **167**(16), 1798–1803.

Hall, R. E. and Jones, C. I. (2007), 'The value of life and the rise in health spending', *The Quarterly Journal of Economics* **122**(1), 39–72.

Hammitt, J. K. and Robinson, L. A. (2011), 'The income elasticity of the value per statistical life: transferring estimates between high and low income populations', *Journal of Benefit-Cost Analysis* **2**(1), 1–29.

Heathcote, J., Storesletten, K. and Violante, G. L. (2017), 'Optimal tax progressivity: An analytical framework', *The Quarterly Journal of Economics* **132**(4), 1693–1754.

Herkert, D., Vijayakumar, P., Luo, J., Schwartz, J. I., Rabin, T. L., DeFilippo, E. and Lipska, K. J. (2019), 'Cost-related insulin underuse among patients with diabetes', *JAMA internal medicine* **179**(1), 112–114.

Hosseini, R., Kopecky, K. A. and Zhao, K. (2021), 'The evolution of health over the life cycle', *Review of Economic Dynamics* .

Jones, C. I. and Klenow, P. J. (2016), 'Beyond gdp? welfare across countries and time', *American Economic Review* **106**(9), 2426–57.

Jung, J. and Tran, C. (2016), 'Market inefficiency, insurance mandate and welfare: Us health care reform 2010', *Review of Economic Dynamics* **20**, 132–159.

Kaplan, G., Moll, B. and Violante, G. L. (2018), 'Monetary policy according to hank', *American Economic Review* **108**(3), 697–743.

Kaplan, G. and Violante, G. L. (2010), 'How much consumption insurance beyond self-insurance?', *American Economic Journal: Macroeconomics* **2**(4), 53–87.

Kowalski, A. E. (2021), 'Mammograms and mortality: How has the evidence evolved?', *Journal of Economic Perspectives* **35**(2), 119–40.

Lagakos, D., Moll, B., Porzio, T., Qian, N. and Schoellman, T. (2018), 'Life cycle wage growth across countries', *Journal of Political Economy* **126**(2), 797–849.

McKay, A. and Wieland, J. F. (2019), Lumpy durable consumption demand and the limited ammunition of monetary policy, Technical report, National Bureau of Economic Research.

McWilliams, J. M., Zaslavsky, A. M., Meara, E. and Ayanian, J. Z. (2003), 'Impact of medicare coverage on basic clinical services for previously uninsured adults', *Jama* **290**(6), 757–764.

Miller, S., Johnson, N. and Wherry, L. R. (2021), 'Medicaid and mortality: new evidence from linked survey and administrative data', *The Quarterly Journal of Economics* .

Murphy, K. M. and Topel, R. H. (2006), 'The value of health and longevity', *Journal of political Economy* **114**(5), 871–904.

Ozkan, S. (2014), 'Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle', *Unpublished. https://sites. google. com/site/serdarozkan/Ozkan_2014. pdf* .

Pashchenko, S. and Porapakkarm, P. (2013), 'Quantitative analysis of health insurance reform: Separating regulation from redistribution', *Review of Economic Dynamics* **16**(3), 383–404.

Patel, D. C., He, H., Berry, M. F., Yang, C.-F. J., Trope, W. L., Wang, Y., Lui, N. S., Liou, D. Z., Backhus, L. M. and Shrager, J. B. (2021), 'Cancer diagnoses and survival rise as 65-year-olds become medicare-eligible', *Cancer* .

Rosen, S. (1988), 'The value of changes in life expectancy', *Journal of Risk and uncertainty* **1**(3), 285–304.

Sommers, J. P. (2007), 'Methodology report #17: Additional imputations of employer information for the insurance component of the medical expenditure panel survey since 1996', https://meps.ahrq.gov/data_files/publications/mr17/mr17.shtml. Accessed: 2010-09-30.

Spitzer, S. and Weber, D. (2019), 'Reporting biases in self-assessed physical and cognitive health status of older europeans', *PLoS One* **14**(10), e0223526.

Trabandt, M. and Uhlig, H. (2011), 'The laffer curve revisited', *Journal of Monetary Economics* **58**(4), 305–327.

Viscusi, W. K. and Aldy, J. E. (2003), 'The value of a statistical life: a critical review of market estimates throughout the world', *Journal of risk and uncertainty* **27**(1), 5–76.

Yogo, M. (2016), 'Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets', *Journal of Monetary Economics* **80**, 17–34.

Zuvekas, S. H. and Olin, G. L. (2009), 'Accuracy of medicare expenditures in the medical expenditure panel survey', *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* **46**(1), 92–108.

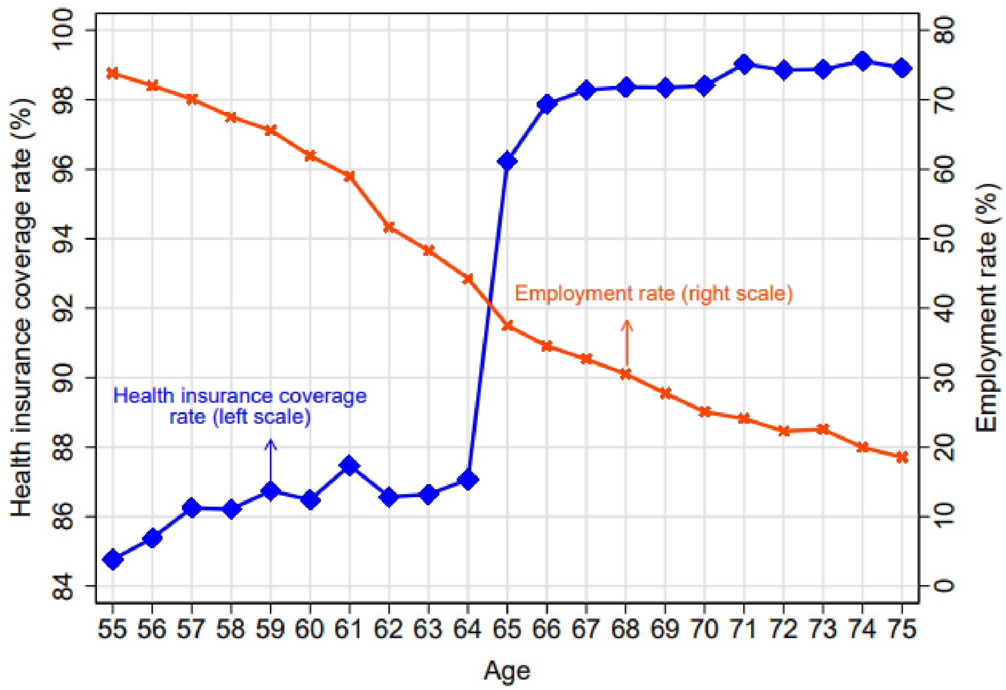
A. Additional Figures and Tables

Table A.1: MEPS Medical Provider Survey Coverage

Provider	Coverage
Hospitals	100%
Office-based Physicians	100% (Medicaid + Medicare covered individuals) 75% (HMO or managed care covered individuals) 25% (remaining individuals)

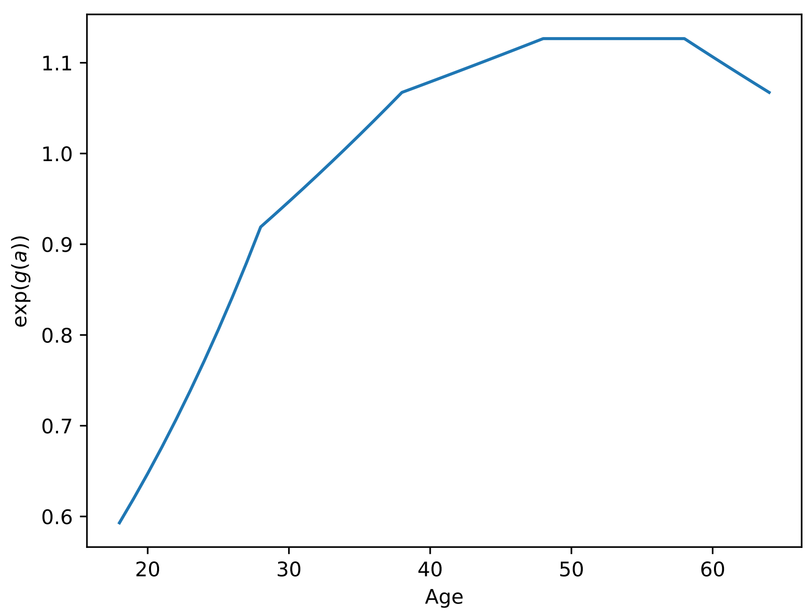
This table reports the percentage of medical spending covered by the Medical Provider component of the Medical Expenditure Panel Survey for 4 different categories of care. Information taken from Sommers (2007).

Figure A.1: Employment Rate as a Function of Age



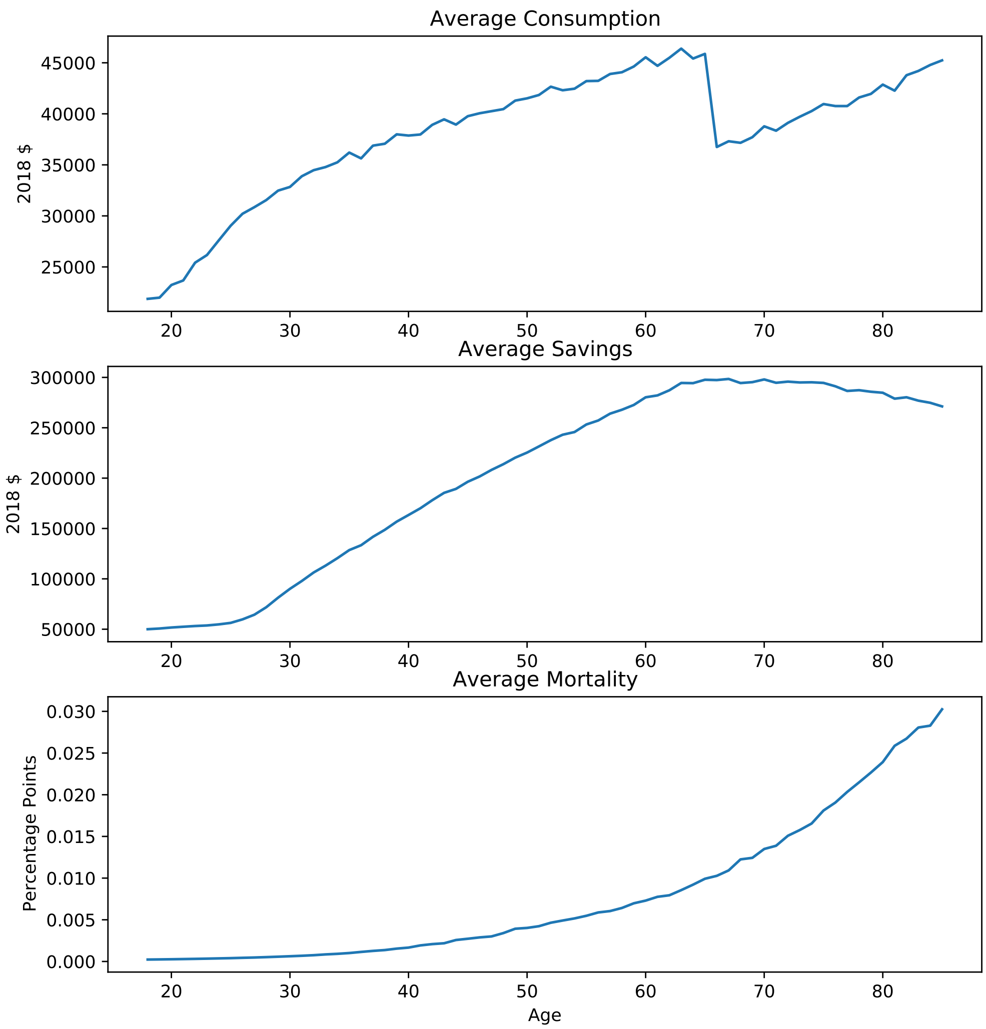
Displays health insurance coverage rates in blue and employment rates in red as a function of age. Calculated using NHIS data from 2002 to 2012

Figure A.2: Calibrated Lifecycle Component of Income



Displays the calibrated life-cycle component of income taken from Lagakos et al. (2018).

Figure A.3: Life-cycle Consumption, Savings, and Mortality



Average consumption, savings, and mortality across the life-cycle in the post-expansion steady state of the calibrated model.