Title
Essays on Macroeconomics with Imperfect Insurance

Author
VanVuren, Mitchell

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on Macroeconomics with Imperfect Insurance

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

in

Economics

by

Mitchell VanVuren

Committee in charge:

Professor Valerie Ramey, Chair
Professor David Lagakos, Co-Chair
Professor Titan Alon
Professor Juan Herreño
Professor Munseob Lee
Professor Karthik Muralidharan

2022

The dissertation of Mitchell VanVuren is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California, San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professors Valerie Ramey and David Lagakos for chairing my committee and providing research advice, career guidance, and support in the process of earning my PhD. I would also like to acknowledge Titan Alon, Tommaso Porzio, Natalia Ramondo and the many other members of the economics department, both present and past, for their myriad of suggestions, feedback, and comments which contribute to a fantastic intellectual environment.

I acknowledge my fellow student Minki Kim, a co-author on one of my chapters, and my office mates Tyler Paul, Jacob Orchard, and Pietro Spini for proving reliable sounding boards for new ideas.

Chapter 3, in full, has been submitted for publication of the material as it may appear in IMF Economic Review, 2022, Alon, Titan; Kim, Minki; Lagakos, David; VanVuren, Mitchell, Palgrave Macmillan, 2022. The dissertation author was a primary investigator and author of this paper.

VITA

2016	Bachelor of Science, Economics and Mathematics, University of Oregon
2016-2022	Teaching Assistant, University of California San Diego
2022	Doctor of Philosophy, University of California San Diego

FIELDS OF STUDY

Major Field: Economics

Essays on Macroeconomics with Imperfect Insurance
Professors Valerie Ramey and David Lagakos

ABSTRACT OF THE DISSERTATION

Essays on Macroeconomics
with Imperfect Insurance

by

Mitchell VanVuren

Doctor of Philosophy in Economics

University of California San Diego, 2022

Professor Valerie Ramey, Chair
Professor David Lagakos, Co-chair

This dissertation consists of three freestanding chapters, broadly linked by the theme of analyzing the macroeconomic effects of various policies using models of imperfect insurance. The term “imperfect insurance” refers to the notion that individuals and households cannot costlessly transfer resources between different states of the world and are subject to risk when facing uncertain outcomes. Chapter 1 examines the macroeconomic impact of public health insurance expansion in such a model where individuals face idiosyncratic health risk. Chapter 2 examines

the impact of publicly provided cash transfers for job searchers who face job finding risk. Finally, Chapter 3 uses an incomplete markets macroeconomic model to quantitatively explore the role of various cross-country differences in explaining cross-country COVID outcomes.

CHAPTER 1

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

by

Mitchell VanVuren

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.

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

[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 EJ et al. (2018).

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 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

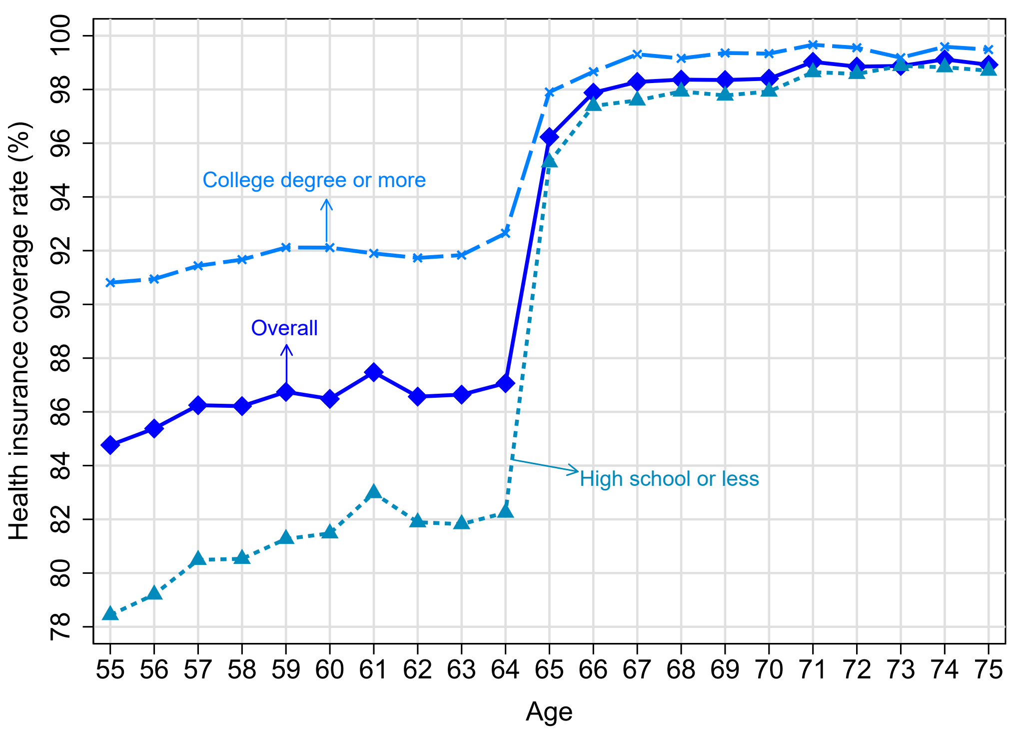


Figure 1.1: Health Insurance Coverage by Age and Education

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 1.1 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 1.2 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 1.1, 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.

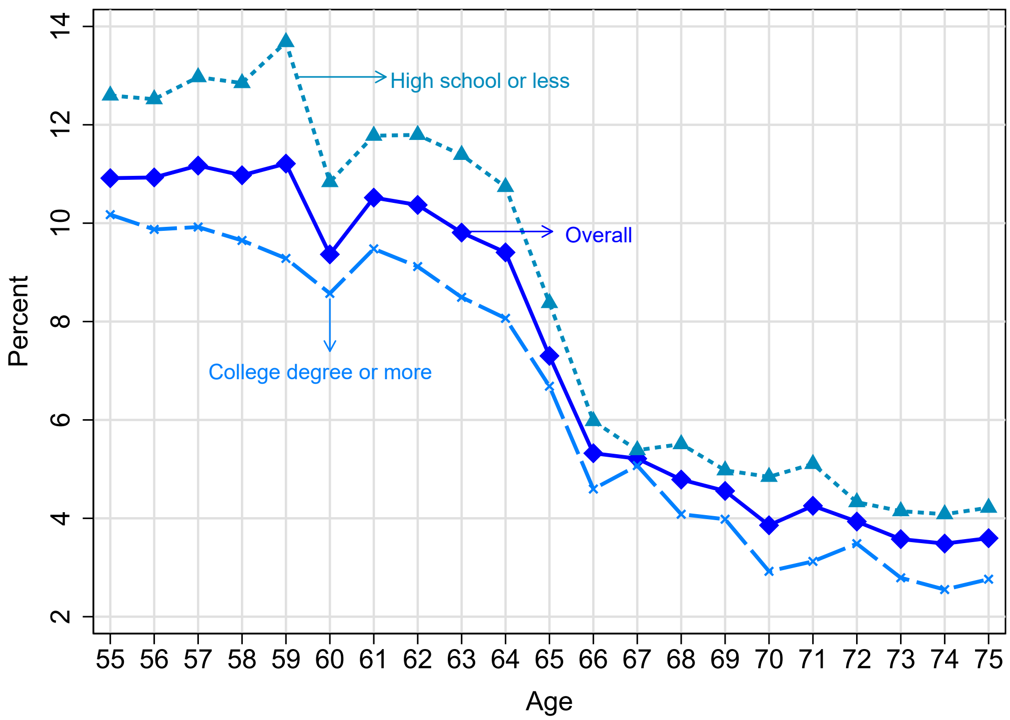


Figure 1.2: Delayed Medical Care by Age and Education

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.

One concern with interpreting these results is that changes in Medicare coverage at age 65 may be confounded with a jump in retirement. Figure 1.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

[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.

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.

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, con-

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

ditional on experiencing an emergency, follow a log-normal distribution with a mean and variance that depend on the individual's 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 (λ_{EMP}, ν_{EMP}, d_{EMP} and λ_{IND}, ν_{IND}, d_{IND} respectively) and well as an exogenous load parameters κ_{EMP} and κ_{IND} 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 a 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

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

firms are given by

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

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

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 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

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

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

1. Assets b

2. Health h

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

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.

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

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 1.3 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

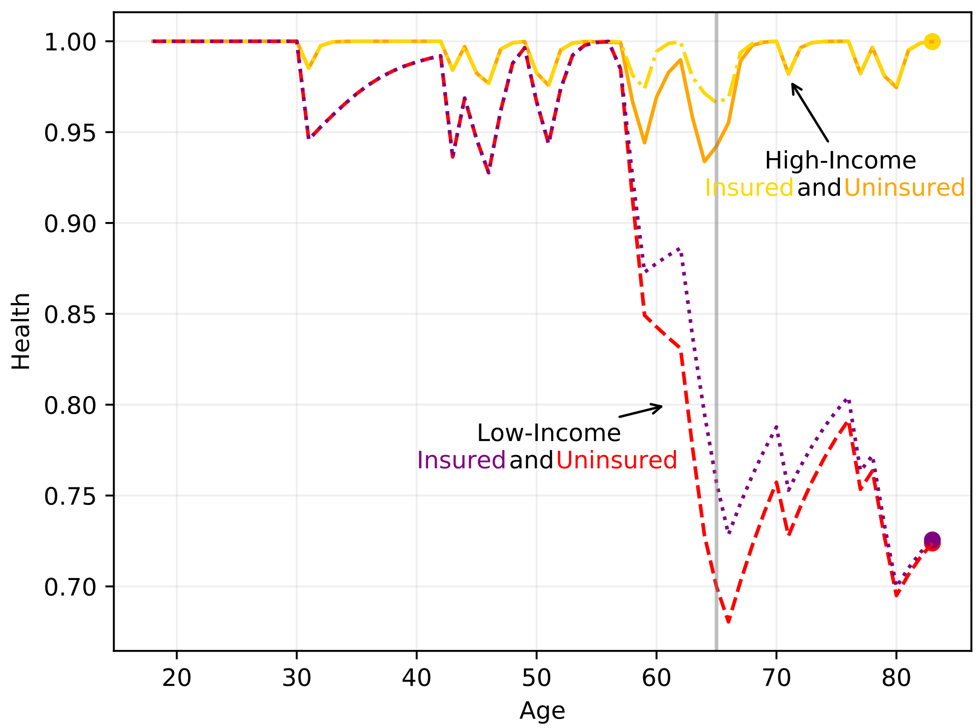


Figure 1.3: Life-cycle Health Dynamics for Four Individuals

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.

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 1.4 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.

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

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

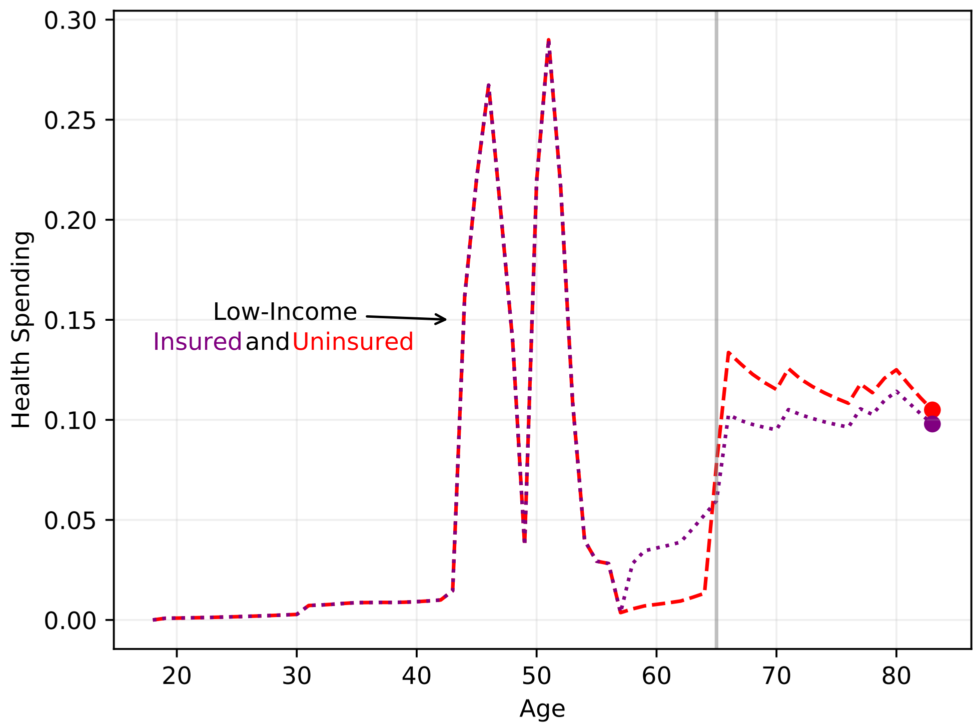


Figure 1.4: Life-cycle Health Spending by Low-Income Insured and Uninsured Individuals
Displays life-cycle health spending for two individuals. The purple line dispalys the spending for an individual with low permanent income and access to employer based insurance. The red line displays spending for an individual with low permanent income and no access to employer-based insurance.

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 1.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 1.1 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.

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

Table 1.1: 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

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 1.5. 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.

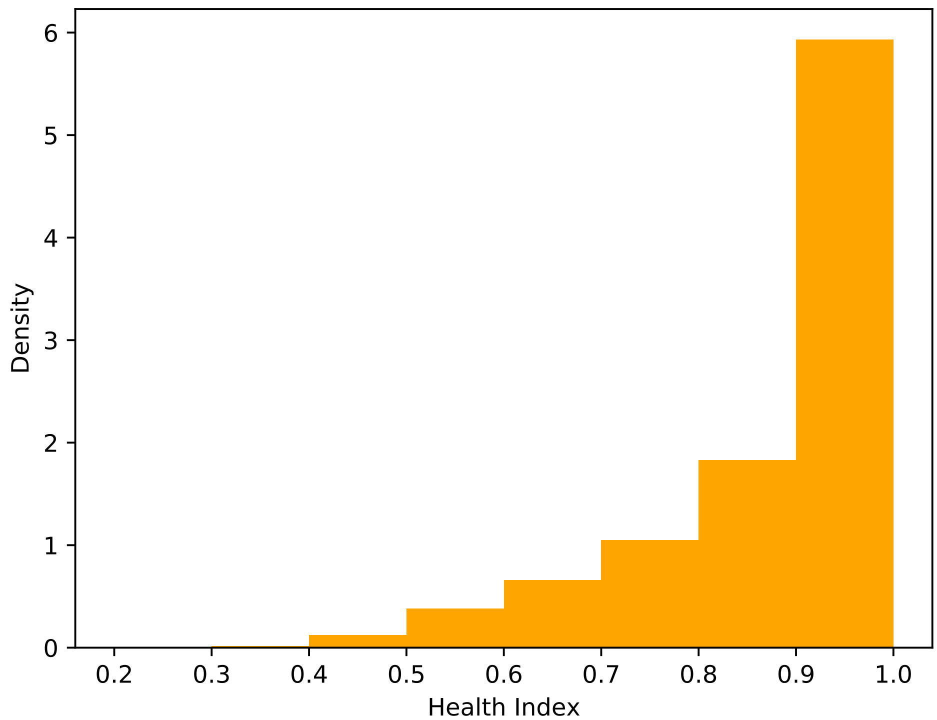


Figure 1.5: The Distribution of Health in MEPS Data

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

Hosseini, Kopecky and Zhao (2021) discuss at length the usefulness of frailty as a measure of health and show 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 1.4 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 1.2 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 1.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

Table 1.2: 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 1.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.

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.

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.

[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.

4.2.2. Parameters Estimated Directly

Table 1.3 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 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.

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 1.4 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.

Table 1.3: Directly Estimated Parameters

Description	Parameter	Value	Data
Mortality Function	$\pi(h,a)$	Table 1.4 Col. 1	MEPS
Emergency Prob. Function	$\pi_x(h,a)$	Table 1.4 Col. 4	
Emergency Mean Function	$\mu(h,a)$	Table 1.4 Col. 7	
Emergency Var Function	$\sigma(h,a)$	Table 1.4 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 1.5	
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 remaining columns detail robustness checks described in further detail in subsection 4.1.3.

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 straight-forward 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 1.5. 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

[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

Table 1.4: Detailed Results of Mortality and Emergency Spending Regression

VARIABLES	(1) Mortality	(2) Mortality	(3) Mortality	(4) Emerg. > 0	(5) Emerg. > 0	(6) 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

VARIABLES	(7) log(Emerg.)	(8) log(Emerg.)	(9) log(Emerg.)	(10) Variance	(11) Variance	(12) 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.

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.

4.2.3. Parameters Calibrated Internally

The remaining parameters are calibrated internally using the simulated method of moments. Table 1.6 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

[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 1.5: 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

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 1.6 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.

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

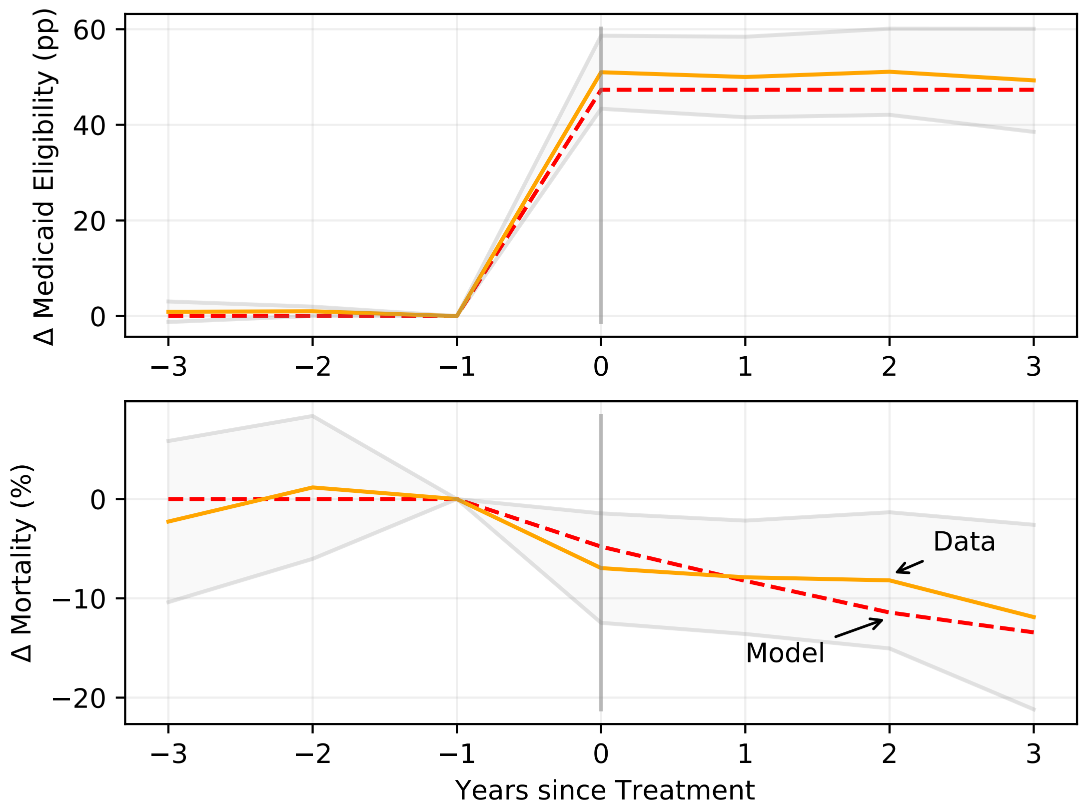


Figure 1.6: Effect of Medicaid on Mortality: Model vs Data

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.

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 1.7 displays these data, their average, and the average jump in medical expenditure at age 65 in the calibrated model.

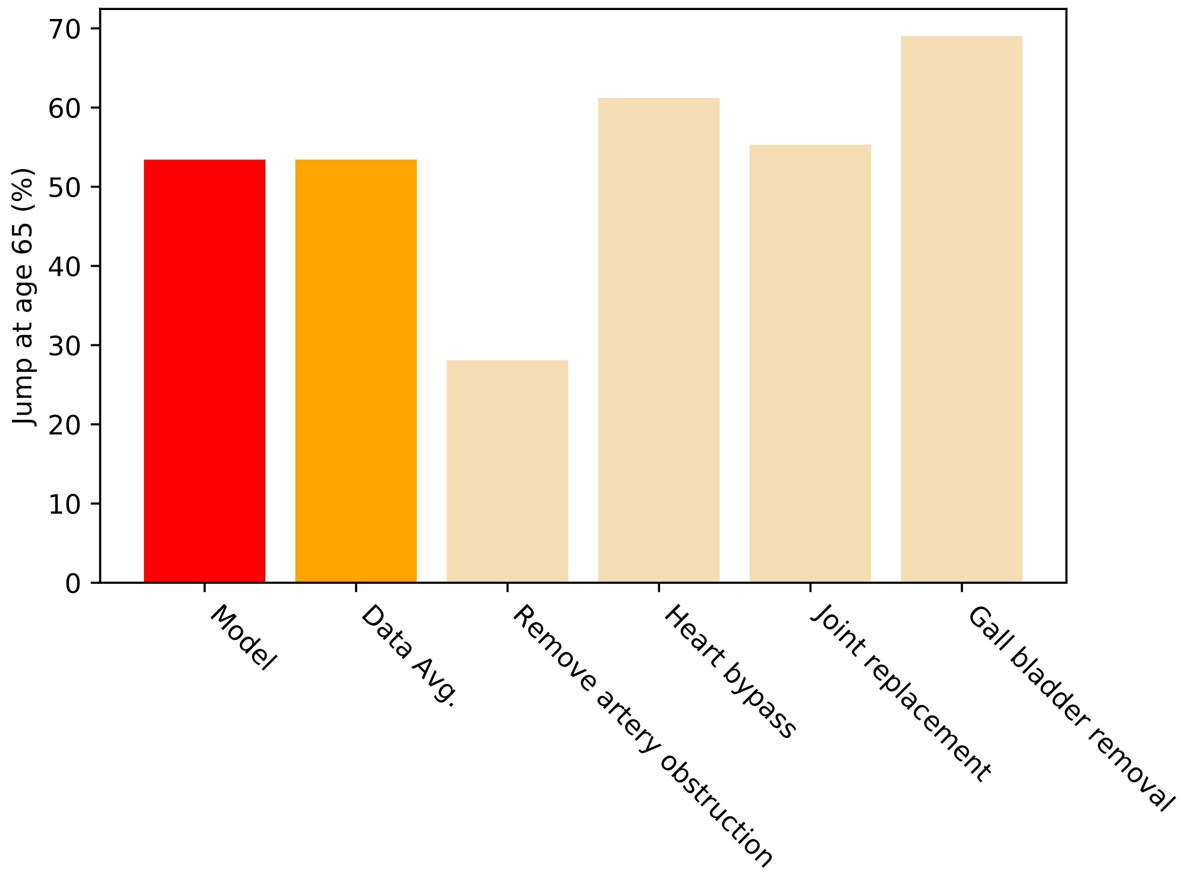


Figure 1.7: Jump in Medical Expenditure at age 65: Model vs Data

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.

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 1.6: Parameters Estimated by SMM

Moment	Model	Data	Source	Parameter
Avg. VSL of Medicaid Recipient	\$2 million	See discussion		$\bar{u}$
Jump in Medical Exp. at 65	See Figure 1.7		Card et al. (2008)	ψ
Mortality Response to Medicaid	See Figure 1.6		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$\mid x > 0$) - mean(Health$\mid x = 0$)	-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.

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 1.8 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 1.9 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 dis-

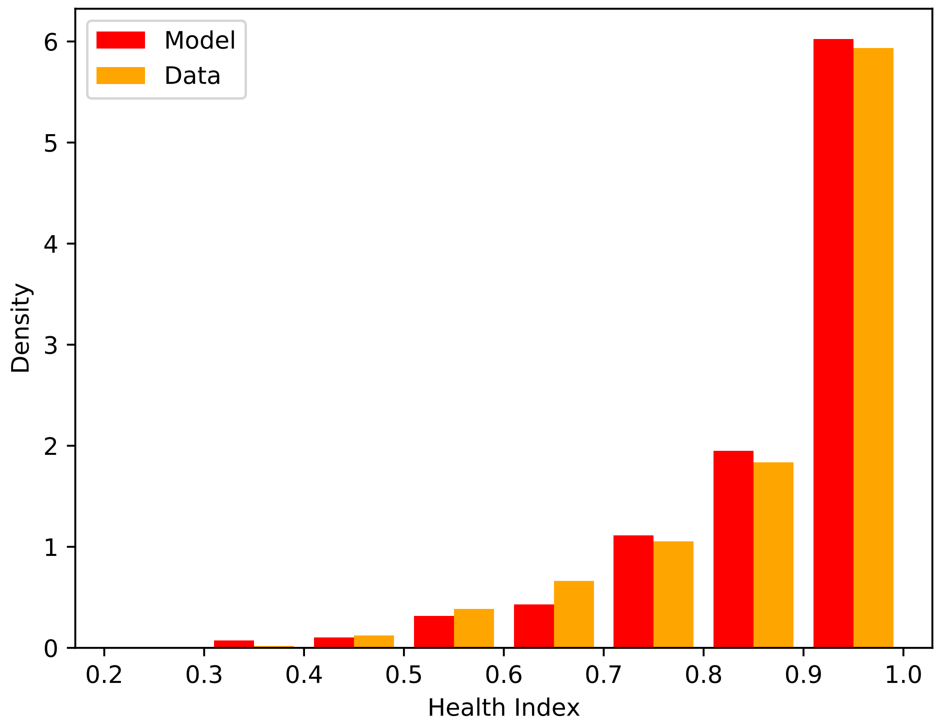


Figure 1.8: The Distribution of Health in the Data and Model

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.

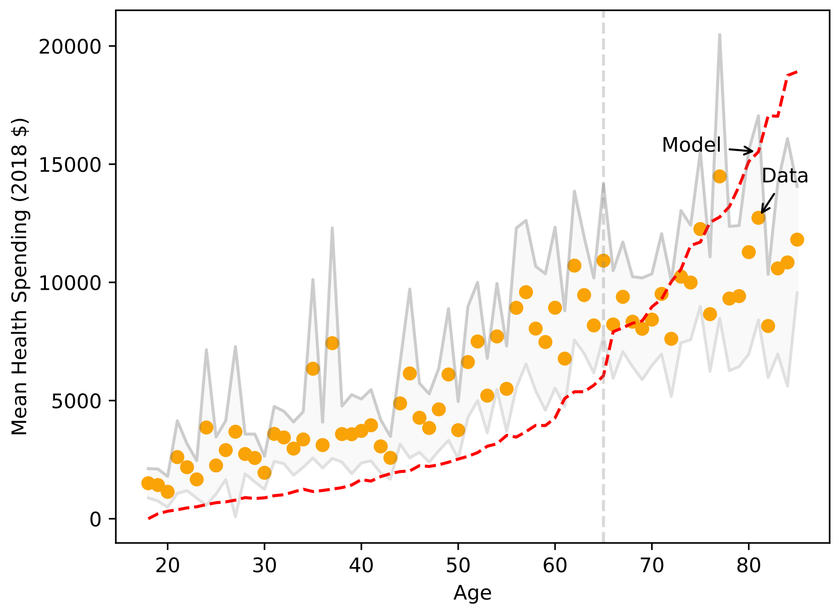


Figure 1.9: Average Health Spending by Age: Model and Data

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.

played 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 1.A.2 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.

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 1.10 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

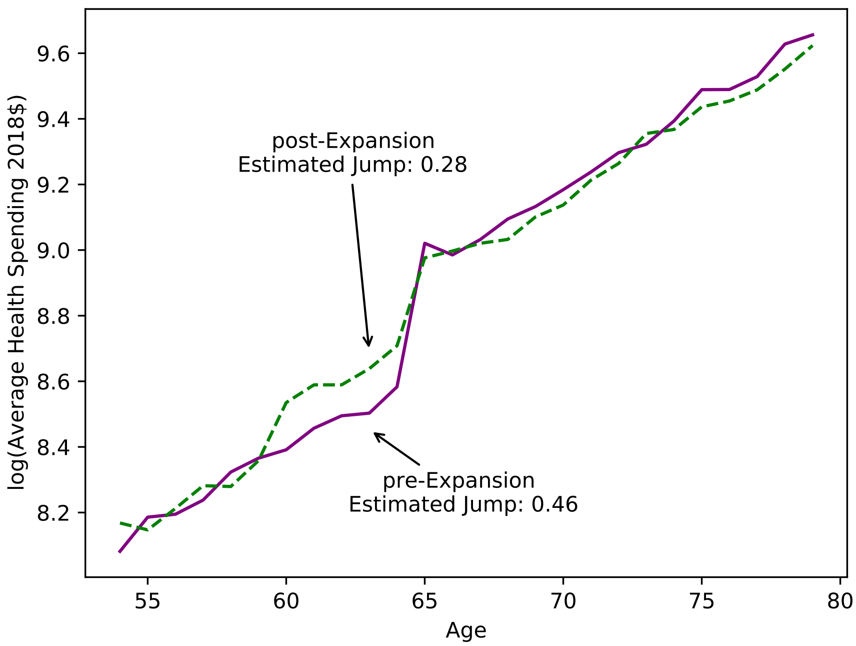


Figure 1.10: Model Implied Delayed Care Pre- and Post- Expansion

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**.

the year they are observed.

Table 1.7: 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**.

The results of the regression discontinuity are displayed in Table 1.7. 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 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 1.10 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 1.11 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-income earners who make more than 250 percent of median earnings

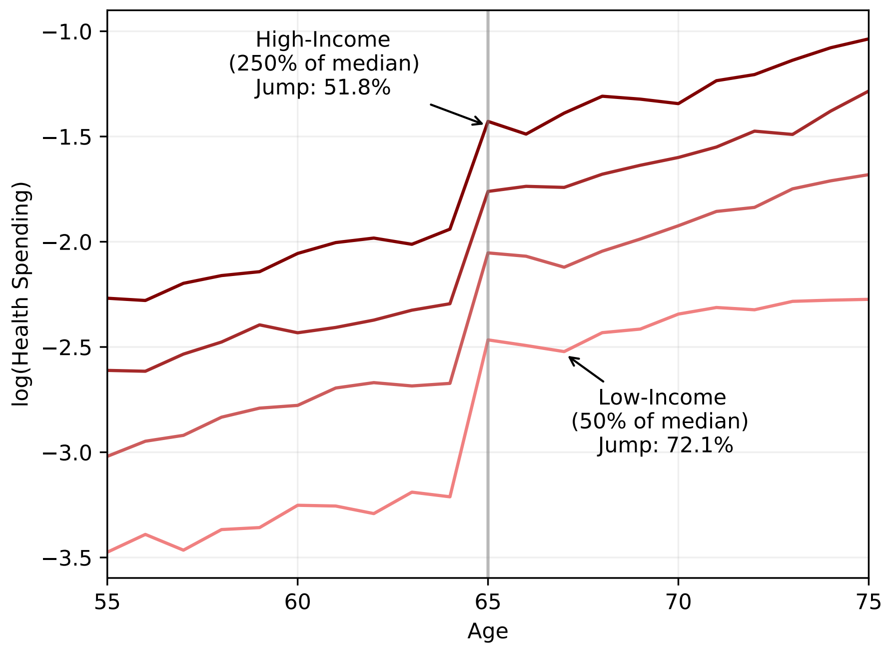


Figure 1.11: Pre-Expansion Health Spending by Age and Income

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

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 1.5. 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

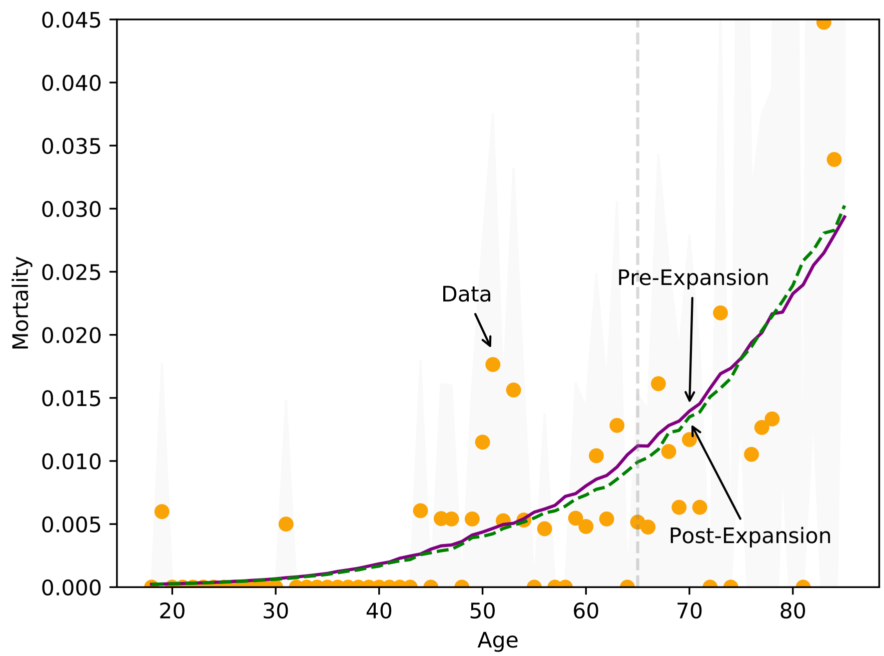


Figure 1.12: Model Implied Mortality Pre- and Post- Expansion

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.

How substantial is the reduction in mortality due to less delayed care? Figure 1.12 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 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 1.8 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 1.8. 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 1.8 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.

The results of Medicaid expansion in this counterfactual model with exogenous mortality are displayed in Column 2 of Table 1.8. 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 1.8 displays the results for the counterfactual model with exogenous health spending. Most notable

Table 1.8: Long-Run Effects of Medicaid Expansion

Panel A: Relative Effects

	(1)	(2)	(3)
Variable	Post-Expansion	Exo. Mortality	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

	(1)	(2)	(3)
Variable	Post-Expansion	Exo. Mortality	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.

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 1.13 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 per-

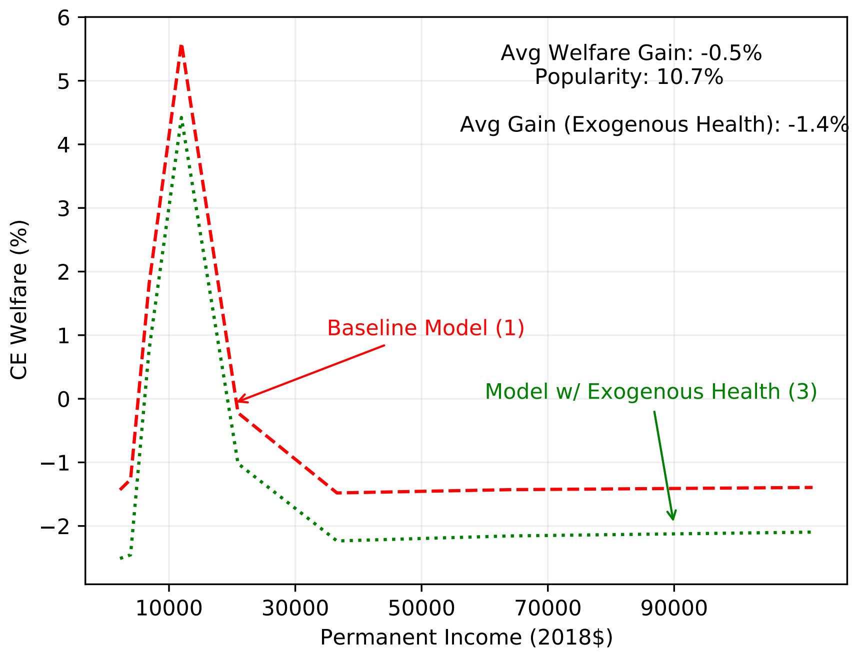


Figure 1.13: Consumption-Equivalent Welfare for Newborn Individuals by Permanent Income

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.

ceived 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.

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 1.14 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.

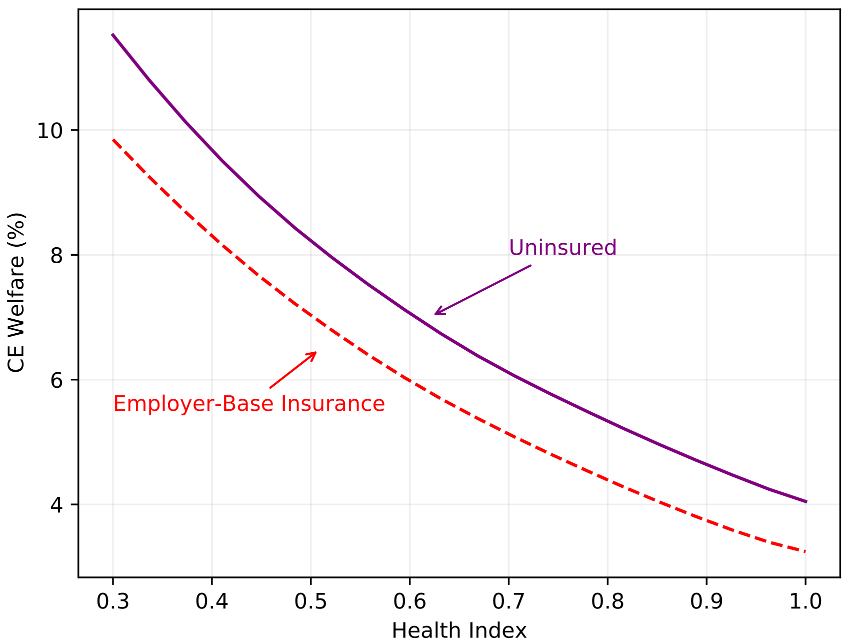


Figure 1.14: Consumption-Equivalent Welfare for Newly-Covered Age 40 Individuals by Individual Health Status

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.

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.

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 EJ, Virani SS, C. C. C. A. C. A. C. S. C. S. C. M. D. F. D. R. d. F. S. F. J. F. M. G. C. I. C. J. M. J. L. J. S. L. D. L. J. L. L. L. S. L. C. L. P. M. J. M. D. M. K. M. M. N. K. O. M. P. L. P. A. P. D. R. M. R. M. R. C. R. G. R. W. S. U. S. G. S. S. S. N. T. D. T. C. V. J. W. J. W. J. W. J. A. H. W. S. M. P. A. H. A. C. o. E., Committee, P. S. and Subcommittee, S. S. (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 health-care 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.

Appendix

A. Additional Figures and Tables

Table 1.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).

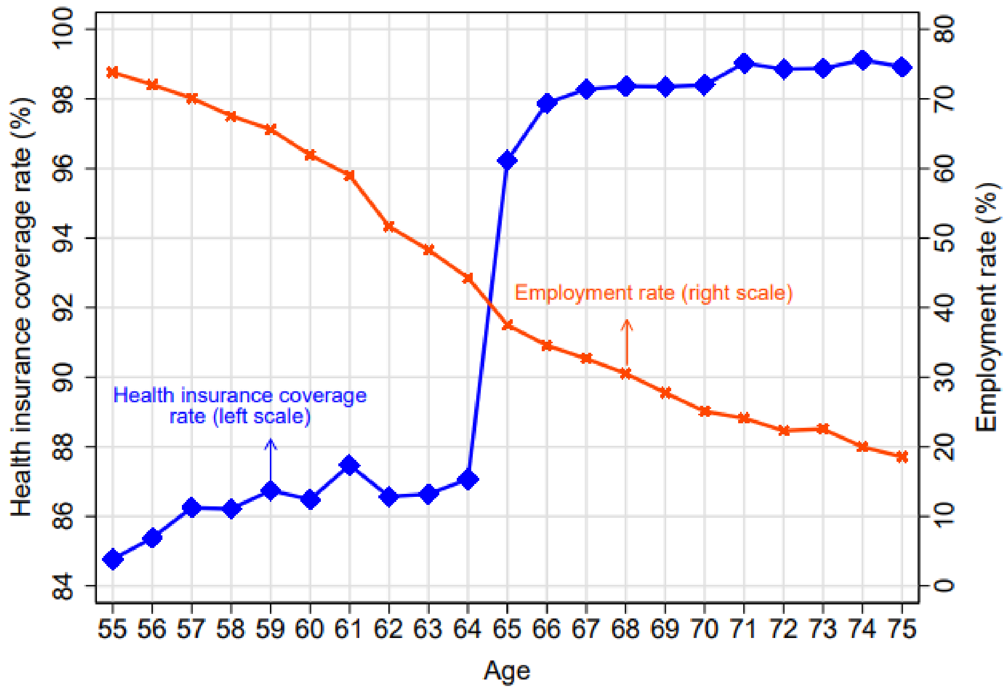


Figure 1.A.1: Employment Rate as a Function of Age

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

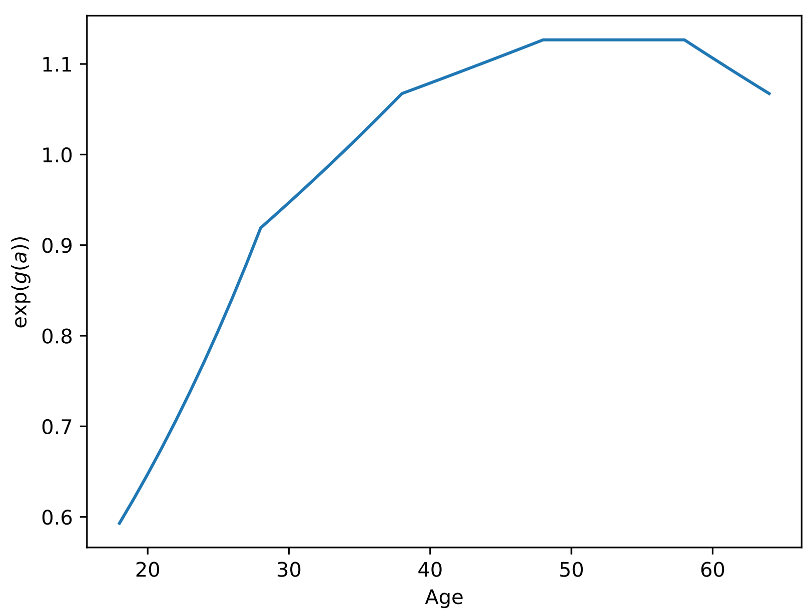


Figure 1.A.2: Calibrated Lifecycle Component of Income

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

CHAPTER 2

Active Labor Market Policies in General Equilibrium: Crowd-In or Crowd-Out?

by

Mitchell VanVuren

Abstract

Recent empirical work has shown that high search costs may contribute to the low levels of wage work in many developing countries, but the aggregate effects of job search assistance are unclear. Greatly increasing the number of searchers without an equivalent increase in the number of jobs could lead to substantial crowd-out effects and limit the effectiveness of such policies in promoting employment. Conversely, making it easier for firms to find qualified workers could reduce the cost of hiring and grow the wage sector, crowding in additional workers and accelerating the process of structural transformation. Which effect dominates is crucial in understanding the effectiveness of job search assistance at an aggregate level. I examine this question using a two-sector general equilibrium search model with a frictional wage sector and frictionless traditional sector. The model allows for both crowd-in and crowd-out effects, but neither effect dominates in general. I estimate the key model parameters using the simulated method of moments to match the results of an experiment that provided job search subsidies to job seekers in Ethiopia. Using the estimated model, I evaluate the impact of implementing a job search subsidy for the all households. I find that the crowd-out effect dominates. Ignoring equilibrium adjustment, the percent of households engaging in wage employment increases from 31 to 51 percent; however, after accounting for the adjustment in labor market tightness, wage employment only increases to 38 percent. The welfare gains follow a similar pattern. In partial equilibrium, the policy results in a gain of welfare equivalent to 1.2 percent of consumption which falls to only 0.6 in general equilibrium. These

results suggest that job search assistance alone is limited in its ability to move workers into the wage sector and may benefit from being accompanied by policies aimed at increasing the number of jobs posted by firms.

1. Introduction

It is well established in the development economics literature that the reallocation of labor from self-employment to wage work is a crucial aspect of structural change and economic development. For example, Gollin (2008) documents a substantial cross-country relationship between income and self-employment rates where low-income countries have substantially higher rates of self-employment. Poschke (2019) shows that this relationship continues to hold even when looking at only urban workers, suggesting that it is not merely the result of cross-country differences in agricultural share. Furthermore, while some self-employed individuals may be successful entrepreneurs, there is widespread belief that many of the self-employed in poor countries are unproductive entrepreneurs choosing self-employment out of necessity, sometimes referred to as "subsistence self-employment" (Schoar, 2010; Herreño and Ocampo, 2021). This notion is further reinforced by the fact that a large fraction of self-employed individuals in poor countries report turning to self-employment due to a lack of other employment opportunities (Poschke, 2013).

This observation that there are high levels of self-employment and little wage work has inspired policymakers in developing countries to implement Active Labor Market Policies (ALMPs) aimed at encouraging and assisting workers' participation in the labor market for wage jobs. These policies cover a wide variety of interventions including explicit subsidies for job seekers, free or subsidized vocational training or apprenticeship, job fairs, or algorithm-made matches between workers and jobs. As the popularity of such policies has grown, economists have begun to evaluate their effects experimentally. For example, Abebe, Caria, Fafchamps, Falco, Franklin and Quinn (2021); Caria, Gordon, Kasy, Quinn, Shami and Teytelboym (2020*b*); and Franklin (2018) evaluate the effects of providing explicit subsidies to job searchers through conditional cash transfers, unconditional cash transfers, and transportation subsidies respectively. These policies are primarily motivated by the observation that many individuals searching for jobs have little to no savings and face fairly high costs of searching for jobs (which is performed in-person in many developing countries). A search subsidy can help overcome this barrier and allow workers to find jobs in the wage sector. Other policies include various vocational training and apprenticeship programs (evaluated in Alfonsi, Bandiera, Bassi, Burgess, Rasul, Sulaiman and Vitali, 2020; Bratti, Ghirelli, Havari and Santangelo, 2018; Bandiera, Bassi, Burgess, Rasul, Sulaiman and Vitali, 2021; Crépon and Premand, 2018) and explicit hiring subsidies (as in Algan, Crépon and Glover, 2020) motivated by the fact that many firms self-report difficulties in finding reliable workers.

In this paper, I evaluate the effects of a particular ALMP, a subsidy for job searchers, in general equilibrium. I build a macroeconomic model of entrepreneurs, workers, and labor search in developing countries. Workers face a choice between engaging in self-employment or participating in the labor market for wage work (as in Herreño and Ocampo, 2021; Poschke, 2019). Workers experience idiosyncratic productivity shocks in self-employment and face incomplete markets in the spirit of Aiyagari (1994), creating an incentive for self-insurance. To participate in the wage sector, workers must first pay a search cost to find a job. Workers face idiosyncratic job-finding risk; paying the search cost may or may not lead to a job at the end of the period. Combined with incomplete markets, this job-finding risk induces workers to ensure that they are sufficiently self-insured before they attempt to search for a job and, if they are unlucky and fail to find one for a few periods, they will return to self-employment. Firms are run by heterogeneously productive entrepreneurs who face a financial friction in the form of a collateral constraint that restricts their choice of capital (as in Itskhoki and Moll, 2019; Buera, Kaboski and Shin, 2021). To overcome this friction, entrepreneurs accumulate collateral by reinvesting a portion of their profits each period, allowing them to continue to grow. Entrepreneurs hire workers in the frictional labor market by posting costly vacancies. Paying vacancy costs reduces entrepreneur profits and thus reduces their ability to grow through reinvestment each period.

The justification for job search subsidies is immediately apparent. Although wage work is, on average, more productive than self-employment, workers are unable to insure against job-finding risk and, as a result, poor workers will opt for the safety of self-employment, despite the fact that it is less productive. A job search subsidy can encourage participation in the labor market and provide insurance against the risk that an individual may not find a job. However, I identify two key general equilibrium channels that served to dampen and enhance the impact of the subsidy respectively. The first channel, which I refer to as the crowd-out effect, is that as workers move out of self-employment and into wage work, the labor market slackens, reducing the probability that any individual searcher will find a job at the end of the period, heightening the risk of participating in the market for wage labor. This mechanism is ubiquitous in labor search models and, in the context of my model, drives some individuals who would have searched for wage work absent this heightened risk to self-employment, shrinking the wage sector.

The second channel, which I refer to as the crowd-in effect, reflects the fact that the difficulty of matching with qualified workers may act as a substantial constraint to growth for entrepreneurs. As the subsidy moves workers into the wage sector, entrepreneurs spend less on hiring costs and grow faster. This reduction in hiring costs is more substantial for highly productive entrepreneurs who want to rapidly expand their firms and thus induces more productive entrepreneurs to grow faster than less productive entrepreneurs, increasing TFP and wages. As average earnings in the wage sector increase, additional workers are induced to move from self-employment to wage work.

I estimate the model using the simulated method of moments to match the results of an experiment in Addis Ababa, Ethiopia performed by Abebe et al. (2021). The experiment offered cash subsidies to job seekers in the city center and was designed to operate as a conditional cash transfer; that is, treated individuals only received the subsidy if they spent the day looking for work. I estimate the model to match the observed increase in search behavior and wage employment as a result of the subsidy as well as some data moments from the control arm of the experiment, such as the average savings held by job searchers and the average job-finding rate. In addition to data from this experiment, I also estimate the model to match typical macro aggregates for Ethiopia as well as firm-level moments calculated using the World Bank Enterprise Survey for Addis Ababa.

Using the estimated model, I evaluate the effects of implementing an economy-wide labor search subsidy funded by a tax on wage workers. Overall, I find that the policy is successful in moving workers out of the traditional sector; however, the crowd-out effect dominates and substantially limits the policy's effectiveness. Participation in the wage sector increases modestly from 34 percent to 39 percent as a result of the subsidy; however, the policy exhibits substantial crowding out. Fixing labor market tightness, and thus shutting down the crowd-in and crowd-out channels, suggests that the policy should increase wage sector participation to 50 percent in the absence of both effects. The large difference between these two results is evidence that the crowd-out effect dominates quantitatively, mostly arising from the fact that households' job search behavior changes strongly in response to changes in their probability of finding a job.

Despite substantial crowding out, the policy still increases average welfare by about 0.6 percent of consumption. When labor market tightness is fixed, welfare increases by 1.2 percent of consumption, suggesting that the net impact of the crowd-out and crowd-in effects is to reduce the welfare gains of the policy by about half. Rather than a large expansion of the wage sector, the welfare gains arise largely from the fact that the subsidy improves insurance by taking resources from the state of the world in which workers are employed and transferring them to the state in which workers are unemployed, which is highly valued by workers. The gains accrue entirely to the unemployed while the employed, who pay the tax required to fund the subsidy, suffer welfare losses of about 1 percent of consumption. Surprisingly, the gains and losses exhibit no substantial pattern with respect to household wealth; poor and rich households both gain (or lose) equally from the policy. Separating the welfare gains into the direct effects of the subsidy and the indirect effects due to higher taxes, crowd-in, and crowd-out does, however, reveal patterns in wealth. In particular, wealthier households benefit more from the direct effects of the subsidy as they are the most likely to be able to fund long periods of job search and collect the subsidy. But wealthier households also experience the largest welfare losses due to the dominance of the crowd-out effects and higher taxes, as they are the most likely to be engaged in wage work where they suffer from both effects. The net result is that changes in welfare are roughly equal for households of all levels of wealth.

1.1. Related Literature

Methodologically, this paper is closely related to the macroeconomic development literature studying the interactions of workers and entrepreneurs in developing countries. The model builds on Itskhoki and Moll (2019) who study optimal Ramsey policies in a model with credit-constrained entrepreneurs and households and find that optimal policy begins by subsidizing entrepreneurship at the expense of workers to encourage growth. Buera, Kaboski and Shin (2011) show that the allocation of capital across entrepreneurs is a key determinant of productivity, a channel also present in this paper and responsible for driving the crowd-in effect through higher wages. Buera, Kaboski and Shin (2021) study the macroeconomic effects of microloans in a model of heterogeneous agents and endogenous selection into entrepreneurship.

This paper also builds on work that distinguishes between subsistence self-employment and entrepreneurship in the developing world. Feng and Ren (2021) document stark differences between the self-employed with and without employees (referred to as own-account workers and employers respectively) and show that employers' labor share is increasing in GDP while own-account work declines as GDP rises, consistent with the ALMP's goal of moving own-account workers into the wage sector. The model is closely related to the model of Herreño and Ocampo (2021) who study the macroeconomics effects of microloans and cash transfers in a heterogeneous agent model in which poor agents use less productive self-employment to cope with the risks of wage employment (as in this paper). Donovan, Lu and Schoellman (2020) construct detailed measures of worker flows between employment, unemployment, and self-employment for countries of various incomes and show that, in developing countries, self-employment and unemployment exhibit similar flows to employment and that self-employment does not help workers climb the job ladder. These results are consistent with the idea that self-employment in developing countries largely exists as a subsistence activity.

This paper contributes to a recent literature documenting and examining the macroeconomic effects of labor search frictions across the cross-country income distribution. Feng, Lagakos and Rauch (2018) document that overall unemployment rates are increasing in GDP per capita and show that skill-biased productivity differences can explain a large fraction of the observed variation in a model with frictional labor markets and frictionless self-employment. I expand on their model by adding risk-averse households and financially-constrained entrepreneurs. Poschke (2019) shows that urban unemployment is substantially higher in developing countries and builds a model in which cross-country variation in search frictions can jointly explain cross-country variation in self-employment and urban unemployment rates, consistent with this paper's finding that individuals self-employment decisions respond strongly to changes in job-finding probabilities. In a similar vein, Banerjee, Basu and Keller (2021) find that skilled workers in developing countries exhibit higher unemployment rates, relative to unskilled workers, than in developed countries and show that this difference leads to differences in occupational choice. Finally, Porzio, Rossi and Santangelo (2021) use a model with frictional reallocation of labor from (self-employment dominated) agriculture to (wage work dominated) non-agriculture to quantify the importance of human capital in explaining the process of structural change.

This paper studies the effects of Active Labor Market Policies in general equilibrium and is thus closely related the empirical literature evaluating the effects of these policies. Abebe et al. (2021) and Franklin (2018) both study the effects of cash transfers to job searchers in extremely similar experiments and find that these subsidies increase search behavior and an individual's probability of being employed in a permanent, formal job after 16 weeks. Interesting, while they find substantial effects on job amenities and self-report job satisfaction, they find no significant effect on earnings. The results and data from these experiments play an important role in the quantitative discipline of this paper's model.

Algan et al. (2020) randomize a government program in France aimed at reducing recruitment and vacancy posting costs for firms and find that the program successfully increased vacancy posted and hirings. Similarly, De Mel, McKenzie and Woodruff (2019) find that wage subsidies effectively increase employment among microenterprises in Sri Lanka but that the impacts of the subsidy are fleeting and employment quickly returns to normal when the subsidy is removed. Alfonsi et al. (2020) evaluate the impact of free training programs, provided either directly to workers for free or provided through firms and subsidized by the experiment. Although this is less directly related to my results as there is no concept of training in the model, it is still an important experimental evaluation of ALMPs and sheds light on a main constraint preventing workers from finding wage sector employment, namely that they lack a credible mechanism through which to signal their abilities.

2. Model

The model features many properties that are characteristic of labor markets in the developing world while remaining computationally tractable. Time is discrete. Because my primary source of data is collected at a weekly frequency, I conceptualize one model period as one week. There is measure one of households and an endogenous measure of entrepreneurs. Households consume, save, and choose between working in self-employment or participating in the labor market while entrepreneurs operate firms, consume profits, and accumulate capital and labor for future periods.

Households are ex-ante homogeneous but face idiosyncratic shocks and incomplete markets. As a result, they accumulate assets for self-insurance (as in Aiyagari, 1994). In any period, an unemployed household must pick between self-employment, which they can participate in costlessly, or paying a cost to search for a permanent wage job. Job search is risky, and only households with sufficient self-insurance will opt to search. Employed households can choose between working at their job in the wage sector or self-employment (in equilibrium they will always choose to work at their wage job).

Entrepreneurs are ex-ante heterogeneous in ability but are financially constrained and must accumulate assets to use as collateral in renting capital. Each period, entrepreneurs earn profits and split these profits between consumption, hiring workers, and financing capital for the next period. Entrepreneurs face no idiosyncratic risk other than an exogenous death rate which ensures that the model has a steady-state in which collateral constraints are binding.

2.1. Labor Markets

The labor market for wage work exhibits typical search-and-matching frictions. Households must search for jobs and entrepreneurs must hire by posting vacancies. The cost of searching for a job and the cost of posting a vacancy are denoted by b and c respectively. Each period, the number of worker-firm matches is given by a homogeneous of degree 1 matching function $m(u,v)$ where u is the number of households searching for a job and v is the number of vacancies posted by firms. As is standard in search-and-matching models, I define $\theta = \frac{v}{u}$ to be labor market tightness. Then $p(\theta) \equiv m(\frac{1}{\theta}, 1) = \frac{m(u,v)}{v}$ is the probability that any vacancy is filled and $\theta p(\theta) = \frac{m(u,v)}{u}$ is the probability that any searcher finds a job. Matches between workers and firms are separated with exogenous probability λ at the end of every period.

2.2. Households

There exists a unit measure of infinitely-lived households. Households are ex-post heterogeneous due to their realizations of idiosyncratic shocks and are indexed by their wealth a, their employment status e, and their self-employment productivity z. Households are endowed with one unit of time each period which they supply inelastically and indivisibly to their activity of choice each period. Households gain utility from consumption according to a CRRA utility function and discount the future at rate β so that household lifetime utility is given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}}{1-\sigma}$$

where σ is a parameter governing the risk aversion and intertemporal elasticity of substitution of households.

Households can spend their time either working for a wage, searching for work, or engaging in self-employment. Importantly, each unit of time is indivisible; a household must commit its entire time endowment to a single activity each period.[9] Any household can engage in self-

[9]This assumption can be justified by the fact that this model is designed to be calibrated to a weekly frequency.

employment, but to work in the wage sector, a household must first search for and match with a job. A household's self-employment productivity y follows an exogenous Markov process described by transition matrix M. For expositional clarity, I assume that self-employment productivity takes a binary form with only a low and high value , y_l and y_h respectively, but more states can be easily accommodated. A household engaging in self-employment earns w_S per unit of productivity; I normalize w_S to one so that a household's earnings from self-employment are given by y.

Instead of engaging in self-employment, a household can choose to pay a search cost b and search for a wage job. A searching household earns nothing in the current period and finds a permanent job with probability $\theta p(\theta)$. Labor market tightness θ is an equilibrium object and depends on the aggregate state X. Wages in the wage sector are determined through bargaining and depend on the state variables of the entrepreneur employing the household as well as the household's earnings in the traditional sector which serves as the household's outside option. For notational simplicity, I suppress much of this dependence and write the permanent sector wage function as $w_t(z)$, depending only on the matched entrepreneurs productivity z, which the household takes as given each period. I will show in a later section that, although all entrepreneurs state variables appear in the bargaining problem, all entrepreneur-household pairs end up bargaining an identical wage conditional on entrepreneur productivity z, justifying this suppression of notation.

Insurance markets are incomplete, and households cannot insure themselves against idiosyncratic shocks; however, households can accumulate assets a as self-insurance. Each period, assets pay an exogenous rate of return $R \geq 1$ which does not vary over time. Households cannot borrow and must satisfy the restriction $a_t \geq 0 \forall t$. The budget constraint for the household can be written

$$a_{t+1} + c_t = Ra_t + (1 - s_t)y_t + s_t(e_t w_t(z_t) - (1 - e_t)b) \tag{5}$$

where $s_t \in \{0, 1\}$ is a choice variable for the household with $s_t = 1$ representing the decision to search in period t and $e_t \in \{0, 1\}$ is an indicator variable with $e_t = 1$ indicating that the household

Within a week, the returns to job search can exhibit increasing returns to scale. For example, the time and effort it takes to prepare a CV is a fixed cost regardless of how many jobs one applies for.

has a permanent job in period t.

A household employed in the wage sector can lose it's job for three reasons. First, the match between the household and firm separates with exogenous probability λ at the end of every period. Second, the entrepreneur employing the household can die at the end of the period with probability $(1-\xi)$. Finally, the entrepreneur can endogenously choose to downsize its labor force. The probability of a household being downsized μ is a function of the state variables of the entrepreneur employing the household as well as aggregate state variables. I will show later that entrepreneurs will never choose to downsize households in steady-state so that $\mu = 0$ in the steady-state equilibrium of the model (although entrepreneurs may choose to downsize along the transition path between two steady-states under certain conditions, to be described later). I define the total probability of a household keeping its job each period as $\lambda^* = (1-\lambda)\xi(1-\mu)$, suppress the dependence of μ on various state variables.

Taking all of the above, the household's problem can be written recursively as

$$\begin{aligned}
V(a,y,e,Z;X) = & \max_{c,a',s\in\{0,1\}} \frac{c^{1-\sigma}}{1-\sigma} + \beta E_{z',e'}[V(a',y',e',Z';X')|y,e,s] \\
s.t.\ a'+c = & Ra + (1-s)y + s(ew(z) - (1-e)b) \\
X' = & G(X) \\
y' \sim & M(y) \\
e',Z' \sim & \text{As described above}
\end{aligned} \tag{6}$$

where X is a vector of aggregate state variables (to be described later) and G is the household's perception function for the evolution of the aggregate state. Z is a vector containing the state variables of the entrepreneur that the household is matched with or a vector of zeros if the household has no match and is included for technical reasons. As mentioned above, I will show that although the full vector of state variables may be important in some contexts, for the purposes of this paper, the household problem will only depend on the matches entrepreneurs productivity z.

2.3. Entrepreneurs

There is an exogenous measure M of entrepreneurs born each period and, at the end of a period, entrepreneurs die with probability Δ. Entrepreneurs are born with idiosyncratic ability z drawn from a distribution described by pdf $h(z)$ and some starting level of financial wealth $\underline{f}$. They discount the future at rate β (the same rate as households), face an exogenous death probability Δ each period, and wish to maximize the following preferences over their consumption (labeled d_t for "dividends")

$$\sum_{t=0}^{\infty}(\beta\Delta)^t \log(d_t)$$

An entrepreneur with idiosyncratic ability z operates a production technology that takes capital k and labor n and produces output y according a Cobb-Douglas production function:

$$y_t = z k_t^{\alpha} n_t^{1-\alpha} \tag{7}$$

The entrepreneur rents capital from an international capital market at an exogenous rental cost $(r+\delta)$ that does not vary over time and pays workers at a wage w_t determined by bargaining. Similar to the previous section, I suppress the dependence of the wage on household and entrepreneur state variables, justified by the fact that the entrepreneur will end up bargaining the same wage with all paired households.

The entrepreneur is constrained in her choice of both k_t and n_t. In particular, the entrepreneur must provide her own assets f as collateral in order to rent capital. Thus k must satisfy the inequality

$$k_t \leq \gamma f_t \tag{8}$$

where $\gamma \geq 1$ is a parameter summarizing the degree of financial market frictions. The case where γ is equal to one corresponds to an economy where there are no financial markets and entrepreneurs must entirely self-finance. As γ goes to infinity, we approach the case with no financial frictions. I take this collateral constraint as exogenous, but it can be thought of as arising from unenforceability of contracts or other institutional features that make uncollateralized lending risky.

To hire labor and adjust n_t, the entrepreneur must post vacancies v_t. Each vacancy costs c units of output to post and is filled at the end of the period with probability $p(\theta)$. In addition, the exogenous separation rate means that the entrepreneur is separated from a proportion λ of her workforce each period. Thus the evolution of n_t is dictated by the equation

$$n_{t+1} = (1-\lambda)n_t + p(\theta)v_t \tag{9}$$

An entrepreneur's period profits are given by

$$\pi_t(z,k_t,n_t) = zk_t^{\alpha}n_t^{1-\alpha} - (r+\delta)k_t - w_t n_t \tag{10}$$

Due to the constraints on the choices of k_t and n_t, an entrepreneur will earn positive profits each period and splits her profits between consumption, posting vacancies, and accumulating additional collateral f_{t+1}. The entrepreneur's period budget constraint is

$$d_t + f_{t+1} = \pi_t(z,k_t,n_t) + f_t - cv_t \tag{11}$$

Taking the preferences and combining equations (7)-(11), the problem of the entrepreneur can be written recursively as

$$
\begin{aligned}
V(z,f,n;X) = \max_{f',n',k,v,d} \; & \log(d) + \beta \Delta V(z,f',n';X) \\
s.t.\; d + f' =& z k^{\alpha} n^{1-\alpha} - (r+\delta)k - wn + f - cv \\
n' =& (1-\lambda)n + p(\theta)v \\
k \leq& \gamma f \\
v \geq& 0 \\
X' =& H(X)
\end{aligned}
$$

where f is a vector of probability density functions summarizing the aggregate distributions of entrepreneurs and households in the economy and H is the perception function of entrepreneurs.

2.4. Wage Bargaining

Each period, entrepreneurs and workers bargain over wages. Because capital acts as a fixed factor of production (it is easy to show that an entrepreneur's collateral constraint will always be binding in equilibrium), an entrepreneur's output exhibit decreasing returns to scale in labor. I follow Smith (1999) and, more recently, Acemoglu and Hawkins (2014) and model production as a cooperative game between workers and entrepreneurs in which each agent is paid their Shapley value.

The entrepreneur enters the game with capital k and workforce n. Any worker that chooses not to cooperate will instead engage in self-employment for the period and then return to the bargaining table the next period. That is, the outside option for the worker takes the form of a temporary strike in which the match between worker and firm is preserved rather than the termination of the match. To simplify the problem, I assume that the period self-employment productivity of an uncooperative worker is drawn from an independent distribution and does not depend on the worker's productivity state. In particular, the worker has probability p of having high productivity and $(1-p)$ of having low productivity so that expected self-employment earnings are given by $py_l + (1-p)y_h$ denoted $\underline{w}$ for simplicity. This simplifies computation of the problem as it allows the productivity of every uncooperative worker to be known a priori, eliminating any potential dependence of the game on aggregate state variables such as the cross-sectional distribution of workers over productivity states and employment.

If the entrepreneur and x of the n workers form a coalition, they operate the entrepreneur's production technology and produce $zk^{\alpha}x^{1-\alpha}$. The remaining $(n-x)$ workers form their own coalition and produce $(n-x)\underline{w}$. Each agent is paid their Shapley value arising from this game, so that the wage per worker is given by

$$w = \chi z k^{\alpha} n^{-\alpha} + (1-\chi)\underline{w} \tag{12}$$

where χ is a parameter governing the bargaining power of the entrepreneur relative to workers. This wage determination equation is intuitive. The workers are simply paid some linear combination of their marginal labor product and their outside option, where the weight on each is determined by bargaining power.

2.5. Characterizing Equilibrium

The model has many moving parts and is somewhat complicated to write down, but households and entrepreneur behavior exhibits some intuitive properties given the environment they are facing. Here I give a brief characterization of the equilibrium behavior of both agents. Because very little analytical progress can be made on the household problem, I provide an intuitive description of household behavior and some quantitative simulations demonstrating the household dynamics. I also provide some analytic results from the entrepreneur problem that make clear the role of entrepreneurs in contributing to crowd-in and crowd-out effects.

2.5.1. Household Equilibrium Behavior

Because of credit constraints, households in the model face a stark trade-off between the lower risk of self-employment and the higher earnings of participating in the wage sector. As a result, only households who are sufficiently self-insured will opt to search for wage work while households without much self-insurance will enjoy the safety of self-employment. But households looking for wage work must pay a search cost which quickly diminishes their savings and reduces their self-insurance, driving them to self-employment to recoup the lost search costs. The result is that households near the threshold of self-insurance spend a few periods working in self-employment and accumulating assets, then switch to searching for a wage job for a few periods, and return to self-employment once their savings have been depleted.

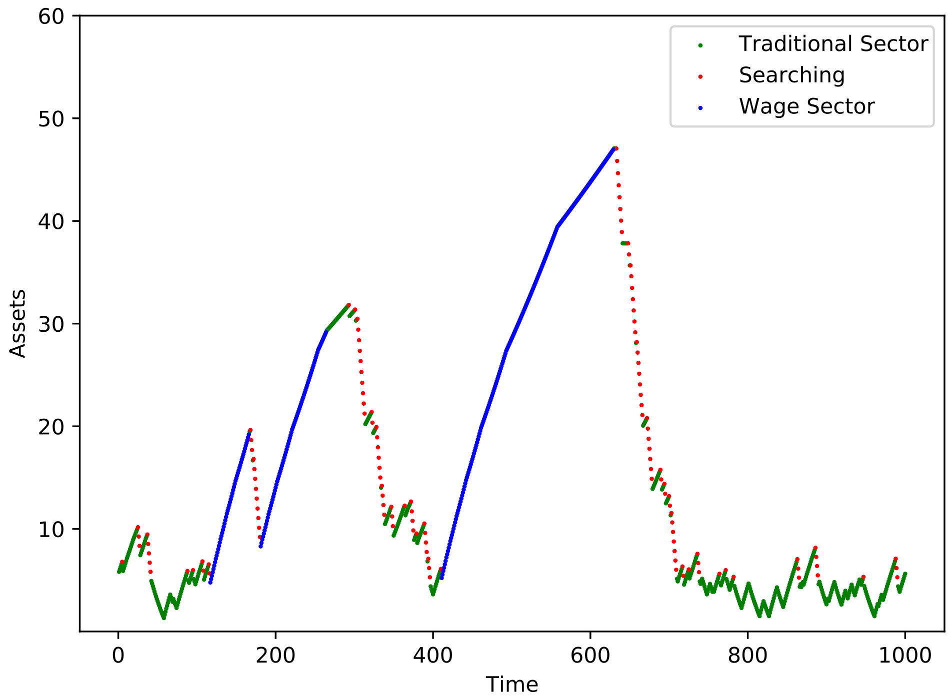


Figure 2.1: Household Self-Employment and Wage Sector Behavior over Time

This figure plots a simulated household's search, wage work, and self-employment behavior as well as assets over 1000 periods of the household's life. This simulation is performed using the full quantitative calibration of the model described in Section 3.

Figure 2.1 displays an example of this behavior for a single household simulated for 1000 periods (about 20 years). Time, denoted in model periods which correspond to one week, is on the x-axis while the household's stock of assets in any period is displayed on the y-axis. The color of each period shows the household's behavior in that period; green points represent periods where the household is engaging in self-employment, red points represent periods where the household is searching for wage work, and blue points represent periods where the household is matched with an entrepreneur and working a wage job.

The figure demonstrates the household behavior described above. At period 0, the household is near the threshold of self-insurance and alternates between working in self-employment and searching for wage work depending on their particular level of assets and self-employment productivity. Around period 150 the household's search is successful, and they acquire a high-earning wage job and quickly accumulate assets. They eventually separate from their employer but use their stock of assets to fund extensive search and remain in the wage sector. This behavior continues until around period 600 where the household has exhausted its savings without finding a wage job and returns to self-employment with occasional wage search. There is even a long period around period 800 where the household's assets fall so low that they stop searching altogether and only engage in self-employment.

With this behavior in mind, it is clear how implementing a subsidy for labor search will lead to crowd-out and crowd-in effects. When the subsidy is implemented, this directly encourages household participation in the wage sector and leads to a slackening of the labor market. This slackening has two primary effects. First, it decreases the probability that a searching household will be matched with a wage job. This decreases the expected earnings of participating in the wage sector, as a larger fraction of a household's time is spent searching rather than earning a wage, and also increases the level of self-insurance that households require before they will choose to search. These channels serve to reduce the number of households searching for wage jobs, shrinking the size of the wage sector and leading to crowd-out. Second, as I will explain below, a decrease in labor market tightness increases average wages in the wage sector. Higher average wages increase participation in the wage sector both by directly increasing the return to searching for a wage job and by increasing the amount that wage workers are able to save each period. With higher wealth, workers are able to sustain a longer period of job search after being separated from their job, increasing the likelihood that they remain in the wage sector rather than falling out. These effects serve to crowd-in additional workers.

2.5.2. Entrepreneur Equilibrium Behavior

While the entrepreneur's problem is complex, substantial progress can be made by reworking the problem analytically. In particular, it can be shown from the first-order conditions that an entrepreneur choosing to post positive vacancies will choose their collateral tomorrow f' and labor force tomorrow n' such that their ratio depends only on the entrepreneur's productivity and aggregate state variables (see the Appendix for the derivation of this result). In particular, this ratio does not depend on an entrepreneur's size as measured by either their current collateral f or current labor force n. Thus I define

$$\eta(z;X) = \frac{\gamma f'^*}{n'^*} \tag{13}$$

so that η denotes an entrepreneur's optimally-chosen capital-labor ratio which depends only on productivity z and the aggregate state X and thus, in steady-state, is constant over the lifecycle of the entrepreneur (as z is fixed for the entrepreneur's lifetime) and for all entrepreneurs born with the same productivity. The intuition for this result is fairly straightforward. Because entrepreneurs face a constant marginal cost of both financing capital and hiring workers and production is Cobb-Douglas with constant returns to scale in capital and labor, entrepreneurs will pursue a fixed capital-labor ratio.

Because bargained wages are a linear combination of an entrepreneur's marginal product of labor (which depends only on their productive and capital-labor ratio) and a worker's outside option (which is the same for all workers), this result ensures that the wage bargained with employed households depends only on the entrepreneur's productivity and not on any other entrepreneur or household state variables. This result, mentioned above in the discussion of the household problem, is extremely useful and makes the model substantially more tractable; it collapses the entrepreneur-level state variables that are relevant for the decision problem of a matched household from three (z, f, and n) to one (just z), reducing the size of the state-space for households. Additionally, it means that the steady-state equilibrium wage function w is one-dimensional (in z) rather than depending on all entrepreneur and household state variables.

A second useful result of the entrepreneur problem, stemming from the constant capital-labor ratio, is that entrepreneurs will pursue a growth rate that depends only on their productivity and aggregate state variables or, mathematically, f'^* will satisfy

$$\begin{aligned} f'^* &= g(z;X)f \\ \frac{\partial g}{\partial z} &> 0 \end{aligned} \tag{14}$$

for some function g (derivation of this result can be found in the Appendix). As with the capital-labor ratio η, this result also ensures that, in steady-state, an entrepreneur's growth rate g will be fixed for their entire lifetime. Importantly, g is increasing in z so that more productive entrepreneurs will choose to grow more quickly. Together, these two functions η and g are sufficient to fully characterize entrepreneur behavior as a function of their productivity z and the aggregate state X.

To shed light on crowd-out and crowd-in effects, which operate through labor market tightness θ, define functions $\hat{\eta}(z,\theta)$ and $\hat{g}(z,\theta)$ to be equal to the capital-labor ratio and growth rate chosen by an optimizing entrepreneur who faces constant exogenous labor market tightness θ. Because these alternative policy functions take labor market tightness as exogenous, they make it possible to perform comparative statics and examine how outcomes of the entrepreneur's problem depend on θ. In particular, aggregate vacancies V and the expected earnings in the wage sector $\bar{w}$

for a newly employed household can be written in the following forms using these functions

$$V = \int \frac{1}{p(\theta)} \Big(\frac{1-\Delta(1-\lambda)}{1-\Delta\hat{g}}\Big) \frac{\gamma}{\hat{\eta}} \underline{f} h(z) dz$$

$$\bar{w} = \frac{1}{p(\theta)V} \int \big(\chi z \hat{\eta}^{\alpha} + (1-\chi)\underline{w}\big) \Big(\frac{1-\Delta(1-\lambda)}{1-\Delta\hat{g}}\Big) \frac{\gamma}{\hat{\eta}} \underline{f} h(z) dz \tag{15}$$

where I've suppressed the dependence on z and θ. Intuitively, the equation for V arise simply by counting vacancy posting across all entrepreneurs and the equation for $\bar{w}$ is the vacancy-weighted average wage.

Proposition 1. *Let $\hat{g}$ and $\hat{\eta}$ be defined as above. Then*

$$\frac{d\hat{\eta}}{d\theta} > 0$$

$$\frac{d\hat{g}}{d\theta} < 0 \text{ and } \frac{\partial^2 \hat{g}}{\partial z \partial \theta} < 0$$

where partial derivatives denoted by ∂ are taken while holding other endogenous outcomes (i.e. $\hat{\eta}$) constant.

Proposition 1 provides a basic characterization of how entrepreneur growth rates and capital-labor ratios respond to changes in labor market tightness. In words, the proposition makes three claims. The first, expressed mathematically as $\frac{d\hat{\eta}}{d\theta} > 0$, says that an entrepreneur's capital-labor ratio is increasing in labor market tightness. This result is intuitive; a tighter labor market leads to higher hiring costs and thus increases the cost of labor relative to capital. Subsequently, the entrepreneur's optimal capital-labor ratio increases. The second claim, $\frac{d\hat{g}}{d\theta} < 0$, is similarly intuitive. It says that an entrepreneur's optimal growth rate is decreasing in labor market tightness. As labor market tightness increases and hiring costs rise, the entrepreneur must spend more on hiring, reducing the profit per unit of consumption good invested. In response, the entrepreneur chooses to consume a higher proportion of their wealth today and invest less for tomorrow, reducing their growth rate. The final statement $\frac{\partial^2 \hat{g}}{\partial z \partial \theta} < 0$ is somewhat more complex. This statement is best interpreted as a statement about how $\frac{\partial \hat{g}}{\partial \theta}$ changes with productivity z. In essence, it says that the reduction in the growth rate due to an increase in labor market tightness is larger (i.e. more negative) for more productive entrepreneurs. In other words, more productive entrepreneurs are more responsive to changes in θ. This result arises from the fact that faster-growing entrepreneurs must hire more workers and thus post more vacancies. When the labor market tightens, hiring costs for these entrepreneurs increase by more than less productive entrepreneurs, resulting in a large decrease in growth.

Combining these results with the formula for average expected wage sector earnings 15 yields two channels through which entrepreneurs' reactions to the decline in labor market tightness induced by a search subsidy lead to crowd-out and crowd-in. First, Proposition 1 notes that the capital-labor ratio $\hat{\eta}$ will decline as the labor market slackens. As capital per worker declines, so do average wage sector earnings $\bar{w}$ which are a function of the marginal product of labor, as exhibited by term $\chi z\hat{\eta}^{\alpha}$ in equation 15. Second, Proposition 1 states that declining labor market tightness will lead to higher growth rates for entrepreneurs and that this increase in the growth rate will be larger for more productive entrepreneurs (as $\frac{\partial^2 \hat{g}}{\partial z \partial \theta} < 0$). As a result, the relative share of the labor force employed by more productive entrepreneurs increases which, because more productive entrepreneurs pay their workers more, increase average earnings, crowding in additional workers. Which effect dominates is a quantitative question and a key outcome in determining the overall impact of labor search subsidies.

3. Model Estimation and Quantification

In this section, I discus the estimation and quantification of the model as well as perform some model validation exercises. Broadly speaking, the parameters of the model fall into two categories. The first are parameters that can be estimated directly from data or are well-known macroeconomic parameters with standard values. These parameters I simply set equal to their estimated or standard value. The second set of parameters I estimate using the simulated method of moments to match key moments measured using weekly data on job searchers. In the subsection below, I describe these data as well as the experimental context in which they were collected.

3.1. Experimental Evaluation of Search Subsidies

I use data from an experiment evaluating the effect of providing search subsidies to potential wage workers in Addis Ababa, Ethiopia. The experiment was performed by Abebe et al. (2021) and began in 2014. In the context of Addis Ababa, the majority of openings for permanent wage jobs are posted on job boards located in the city center. To apply for a job, an individual must first travel to the city center, typically by bus, to view the job posting. In this context, the cost of buying a bus ticket serves as a cost to job search that is both large and salient.

The experiment sampled young individuals who were likely to desire a permanent wage job. In particular, individuals included in the sample "(i) were between 18 and 29 years of age; (ii) had completed high school; (iii) were available to start working in the next three months; and (iv) were not currently working in a permanent job or enrolled in full time education." (Abebe et al., 2021). Individuals in the sample were randomly offered cash that could be collected in person at the job boards in the city center up to three times per week. To minimize the incentive to travel to the job boards and collect the subsidy with no intention to actually search for work, the subsidy was designed to offset the cost of a bus ticket from each individual's home to the city center. As a result, each individual was offered a different subsidy amount. I abstract from this heterogeneity when estimating the model and simply use the average amount of the subsidy collected per person per week. Treated individuals were offered the subsidy for 16 weeks. Weekly data on the search behavior and labor market outcomes of both the treated and control groups were collected through phone surveys.[10]

After 16 weeks, the authors calculate the effect of being offered the search subsidy on a variety of labor market outcomes. Their results are replicated in Table 2.A.1. The search subsidy has a significant effect on the type of jobs that searchers have at the end of the 16 weeks. Indi-

[10] At the current moment, these data are not publicly available. Instead, I use data collected from an almost identical pilot experiment. These data are published in Franklin (2018). In the future, I plan to use data from the full experiment.

viduals offered the subsidy are 3.4 percentage points and 5.4 percentage points more like to be in permanent and formal jobs respectively. For the purpose of my model, I interpret this as evidence of an increase in wage employment and choose to treat temporary, informal employment as part of the model's self-employment sector. There is also some suggestive evidence that the subsidy increases wages and employment, but these estimates are very imprecisely estimated.

These data are useful in quantifying the model for two primary reasons. First, the data collected on the control group provides a high-frequency look at the search behavior of workers. This allows direct observation of many important model moments, such as the probability of finding a job conditional on searching ($\theta p(\theta)$ in the model) or the average level of savings among searchers. Direct observation of these micro moments allows for more direct estimation of model parameters instead of relying on aggregate moments. Second, the experimentally evaluated impact of job search subsidies on wage sector employment provides a valuable moment that directly speaks to the effectiveness of subsidies in encouraging workers to search. Because I have enough parameters to estimate the model, I reserve this moment for model validation, allowing me to check whether the model's predicted increase in wage employment aligns with reality.

3.2. Directly Estimated and Calibrated Parameters

Table 2.1 displays the model parameters that are either calibrated directly from external sources or are estimated directly using data along with their source. The discount rate β is calibrated to match an annual discount rate of 0.95. I choose to conceptualize a model period as one week resulting in a very small value for β. The rate of return on worker's savings R is calibrated to be less than one, meaning that workers are unable to save in productive assets. Instead, workers save in the form of cash which is subject to devaluation due to inflation. I choose R to match an annual inflation rate of 10 percent, consistent with World Bank estimates of the rate of inflation in Ethiopia over the last few years. Capital's share of income in production is set to 0.33 as is standard.

I calibrate the search costs b to match the average cost of a bus ticket to the city center calculated in Abebe et al. (2021). As mentioned in the previous section, this cost exhibits substantial heterogeneity across individuals; however, for simplicity I choose to use the average and treat individuals as homogeneous in their search costs. The cost that entrepreneurs face of financing capital, given by $(r+\delta)$, is calibrated to match that reported in Banerjee et al. (2015). It's important to note that the rental rate of capital faced by entrepreneurs r and the rate of return on worker assets $R-1$ are not equal, implying the existence of some sort of wedge between these two rates. This is possible, even in general equilibrium, due to the assumption of a small open market economy. Because of this, r and δ are not separately identified and I choose to calibrate them together as a single parameter.

The probability of entrepreneur death is taken from Abebe et al. (2017) which reports detailed data on firms in Addis Ababa. I also use their data on firm vacancies to calculate a vacancy filling rate of 3.76 percent, to which I calibrate the efficiency parameter of the matching function a. I estimate the collateral constraint for entrepreneurs using the World Bank Enterprise Survey, limited to enterprises based in Addis Ababa, and find that the average loan requires approximately 75 percent collateral, implying a γ of 1.33. Finally, the exogenous separation rate of workers from jobs λ is calibrated to match an unemployment rate of 18.5 percent, consistent with World Bank estimates for Addis Ababa.

The remaining two parameters, the distribution from which newborn entrepreneurs draw their productivity $F(z)$ and the initial level of assets for newborn entrepreneurs $\underline{f}$ are set somewhat arbitrarily. The distribution is chosen to be Pareto with tail parameter 2.1, which, consistent with evidence, results in a Pareto distribution in establishment size. The tail parameter is chosen to be 2.1 so that the distribution of establishments exhibits finite mean and variance. The initial level of assets for newborn entrepreneurs $\underline{f}$ is chosen entirely arbitrarily.

3.3. Parameters Estimated using the Simulated Method of Moments

Table 2.2 displays the model moments that are targeted in the simulated method of moments estimation as well as their values measured in the data and in the model. The final column of Table

Table 2.1: Directly Estimated Parameters

Parameter	Value	Description	Source
β	.997	Discount rate	.95 annual discount rate
R	.998	Return to savings	10% annual inflation (World Bank)
α	.33	Capital share	Standard value
b	.137	Search cost	Abebe et al. (2021)
$r+\delta$	.0041	Capital cost for entrepreneurs	Banerjee et al. (2015)
Δ	.998	Entrepreneur death prob.	Abebe et al. (2017)
$p(\theta)=1-e^{-a/\theta}$	.0376	Job filling rate	Abebe et al. (2017)
γ	1.33	Collateral constraint	World Bank ES
λ	.0071	Unemployment rate of 18.5%	Poschke (2019)
M	$(1-\Delta).25$	Entre. population share of 25%	Itskhoki and Moll (2019)
$h(z)$	$\frac{1}{z^{2.1}}$	CDF of entpreneur z	
$\underline{f}$	0.1	Initial firm assets	

This table displays the model parameters that are estimated directly as well as their values and sources and/or aggregate target. See the discussion for details on each parameter.

Table 2.2: Moments Targeted using the Simulated Method of Moments

Moment	Data	Model	Parameter
Wage Work as % of Total Work	30.0%	34.4%	$u(c)=\frac{c^{1-5.6}-1}{1-5.6}$
Median Savings while Self-Employed	25.1% of earnings	25.6% of earnings	$\frac{y_h}{y_l}=2.088$
Control Wage Employment after 16 Weeks	17.1%	16.8%	$c=18.0$
Productivity Transition Prob.	21%	21%	$p=.21$

This table displays the moments targeted in the simulated method of moments estimation and their values in both the data and model. The final column lists the model parameters estimated using SMM and provides a rough, intuitive correspondence indicating which moment is most responsible for disciplining each parameter. See the discussion for details.

2.2 reports the four estimated parameters corresponding to the four moments. Although all four moments are determined jointly by all four parameters, the correspondence between moments and parameters displayed in the table gives intuition for which moment is most important in estimating which parameter.

The first moment is the percentage of the population engaged in wage work which I calculate to be 30 percent for Addis Ababa based on the data of Abebe et al. (2021). It's worth noting that this is substantially lower than the rates of wage work calculated by the World Bank for Ethiopia which are 10 to 15 percent. Although this is not particularly surprising as it would be expected that urban Addis Ababa would have higher rates of wage employment. In the model, this moment is closely pinned down by the CRRA parameter of the utility function. This correspondence is intuitive as participating in the wage sector carries a higher expected return that participating in self-employment but is subject to idiosyncratic job-finding risk. As a result, conditional on other variables (savings, relative earnings, and job-finding probability), the decision to participate in the wage sector is determined fully by an individual's risk tolerance. I estimate the CRRA parameter to be 5.6, reflecting the fact that workers seem to be very risk-averse in choosing whether or not to engage in wage work.

For those engaged in self-employment, the relative productivity of the high productivity state versus the low productivity state is pinned down by observing savings held by the self-employed or casually employed measured in Abebe et al. (2021). If the gap between productivity states is larger, individuals will hold higher savings to be more self-insured. I estimate that the high productivity state is roughly twice as productive as the low productivity state. Having fixed most of the search parameters in the previous sector, the remaining search parameter, the cost of vacancy posting c, is estimated using the rate of wage work in the control group of Abebe et al. (2021) after the 16 week observation period. Because all other search parameters have been fixed, c directly determines the job finding probability $\theta p(\theta)$ and thus corresponds closely to this moment. The final moment, the probability of transitioning between productivity states when self-employed, is estimated to match the average transition probability between a week spent without work and a week spent casually working for individuals observed in Abebe et al. (2021)

3.4. Model Validation

As my primary model validation exercise, I replicate the experiment performed by Abebe et al. (2021) in the model and compare the model outcomes to the experimentally estimated outcomes. To emphasize the appropriateness of this exercise to validate the model, it is important to make one note. When using data from the experiment to estimate the model in the section above, I make sure to only use data from the control group of individuals in the experiment. In other words, data from the treatment group is used nowhere in the estimation process. Thus comparing the treatment effect estimated in the experiment, which boils down to a difference in means between the treatment and the control group, provides validation of the model that is independent of the data used to estimate it.

To replicate the experiment in the model, I begin by selecting a representative but small portion of workers. This "representative but small" assumption is important because it captures the idea that an experiment providing a treatment to a few thousand individuals in a city of millions will have essentially zero impact on equilibrium outcomes. When replicating the experiment in the model, I want to capture this notion and ensure that the model predicted experimental effect arises purely due to the treatment and not due to equilibrium adjustment. In a technical sense, I select a representative measure zero set of workers. Because the set is measure zero, outcomes for this group will have no impact on equilibrium objects.

I split the sample into treatment and control groups. The control group receives no changes while the cost of searching for wage work b is changed to be equal to zero for the treatment group. Setting this cost to zero reflects the fact that, in reality, the treatment was designed to exactly offset the cost of a bus ticket to the city center. I then simulate the economy forward for sixteen weeks (sixteen periods), as in the experiment, while tracking the behavior and outcomes of the control and treatment groups. After these sixteen periods are up, the model equivalents of the experimentally estimated treatment effects can be constructed by comparing the mean outcome between the control and treatment groups.

Overall, I find that the model does a very good job of predicting the experimentally estimated outcomes. The model predicts that wage sector employment will be 3.5 percentage points higher (from a baseline of 16.8 percent) in the treatment group after 16 weeks. In reality, the experiment finds that wage sector employment is 3.3 percentage points higher (from a baseline of 17.1 percent) in the treatment group. The fact that the model prediction is remarkably close to the experimentally estimated treatment effect, despite no data from the treatment group being used in estimation, is an encouraging signal of the model's ability to accurately capture the sectoral decision of workers.

4. Quantitative Exercise and Results

As the main quantitative experiment, I implement a cash transfer each period targeted at all individuals who are searching for wage work. I choose the size of the subsidy to be equal size used to validate the model in the previous section. In particular, this subsidy is equal to 13.7 percent of average weekly earnings (across both sectors). Recall that this subsidy size was designed to exactly offset the costs of search. As a result, the subsidy essentially sets the search cost b to zero. For the main exercise, I assume that the subsidy is funded by a flat tax levied on wage workers, rather than a tax on all workers. This is an important distinction as it means that the tax itself serves to distort workers' choice of sector towards self-employment and, as a result, the tax contributes to the crowd-out effect. In the future, I plan to evaluate an alternative scenario where the subsidy is funded by a flat tax on all workers, eliminating this distortion, and compare how the results differ between these two cases.

Table 2.3 displays the results of this policy. Column (1) displays the value of moments key aggregate moments in the benchmark steady-state of the estimated model while column (2) displays the values of these moments in the post-subsidy steady-state. The policy results in a substantial increase in both GDP and welfare. Welfare increases by 0.6 percent of consumption on average while GDP increases by a little over 2 percent. This increase in GDP is the result of a 5.4 percentage point increase in the size of the wage sector, which is more productive than the self-employment sector, and an increase in wage sector earnings of 1.88 percent. This increase in earnings is the direct result of higher average wage sector TFP in the post-subsidy steady-state of the model. As the subsidy encourages wage work and the labor market slackens, entrepreneurs now dedicate fewer resources towards hiring and more resources to growth. This increase in growth is disproportionately beneficial to higher productivity entrepreneurs, allowing them to increase their market share and increasing TFP. A portion of this higher TFP is shared with workers through higher wages due to bargaining. However, it is important to note that the increase in wages due to higher TFP is not enough to overcome the increase in taxes necessary to fund the policy; post-tax earnings in the wage sector decrease by 0.5 percent

Table 2.3: Results of Implementing Search Subsidies

Variable	(1) Benchmark	(2) After Subsidy	(3) Equil. Values Fixed
GDP (relative to benchmark)		+2.06%	+4.10%
CE Welfare (relative to benchmark)		+0.60%	+1.25%
Size of Wage Sector	34.4%	39.8%	50.5%
Wage Sector Earnings (relative to benchmark)		+1.88%	+0.00%
Wage Sector Earnings (includ. tax)		-0.50%	+0.00%
Labor Market Tightness	0.094	0.074	0.094
Job-Finding Prob.	3.10%	2.95%	3.10%
Unemployment Rate	18.6%	19.4%	18.6%

This table displays the results of the primary quantitative exercise of subsidizing search for wage jobs. Column (1) reports key aggregate parameters in the steady-state of the model before implementation while Column (2) reports these same parameters in the new steady-state of the model once the policy has been implemented. Column (3) displays the results in a hypothetical steady-state where labor market tightness θ is fixed. See the discussion for details on how to interpret these results.

The search subsidy has only a modest impact on the size of the wage sector which increases from 34.4 percent to 39.8 percent. Labor market tightness decreases resulting in a small decrease in the job-finding probability from 3.1 percent to 2.95 percent and, consequently, an increase in the unemployment rate by 1.2 percentage points. The decrease in job-finding probability together with the decrease in post-tax earnings in the wage sector strongly suggests that the crowd-out effect dominates the crowd-in effect. To investigate this quantitatively, I perform an additional numerical experiment. Because the crowd-out and crowd-in effects operate through labor market tightness and earnings, both of which are equilibrium objects, I also compute the results of the subsidy if these equilibrium objects were fixed to their pre-subsidy values.

The results of this numerical experiment are displayed in column (3) of Table 2.3. I interpret these results (when compared to the pre-subsidy model) as revealing the direct impact of the subsidy on workers' decisions and outcomes while the difference between these results with fixed labor market tightness and wages then reveals the impact of the general equilibrium effects of the subsidy. The most striking difference between this numerical experiment and the post-subsidy steady state is the size of the wage sector. When equilibrium parameters are fixed, the subsidy increases wage sector participation by a remarkable 16.1 percent points to 50.5 percent. Nearly three times as much as the 5.4 percentage point increase induced by the policy in full equilibrium. This stark difference suggests that the direct impact of the search subsidy is large; search costs serve as a substantial constraint in preventing workers from participating in the wage sector.

The large difference in wage sector participation between the full equilibrium results and the results with equilibrium values fixed also suggests that the crowd-out effects play a substantially larger quantitative role than the crowd-in effects. As can be seen from column (3), when equilibrium adjustment is shut down, the crowd-in and crowd-out channels are shut down. Labor market tightness is fixed, there is no change in the job-finding probability or in taxes that may crowd out wage workers. Similarly, because wages are fixed, there is no increase in the wage due to higher TFP that could crowd-in additional workers. Once both these channels are introduced, the size of the wage sector falls substantially, consistent with the notion that the crowd-out channels dominate.

Interestingly, the crowd-out effect seems to be large despite a fairly small decrease in the job-finding probability in the new equilibrium. The probability falls by 0.15 percentage points from 3.10 percent to 2.95 percent, a small decline. This large change in the size of the wage sector despite a small decline in job-finding probability indicates that the semi-elasticity between an individual's search choice and their probability of finding a job must be fairly large, likely a direct result of high estimated risk aversion. This behavior seems consistent with experimental interventions such as Alfonsi et al. (2020) and Abebe et al. (2017) that find large impacts on search behavior of treatments that lead individuals to substantially revise their expectations of their job-finding likelihood.

4.1. Welfare

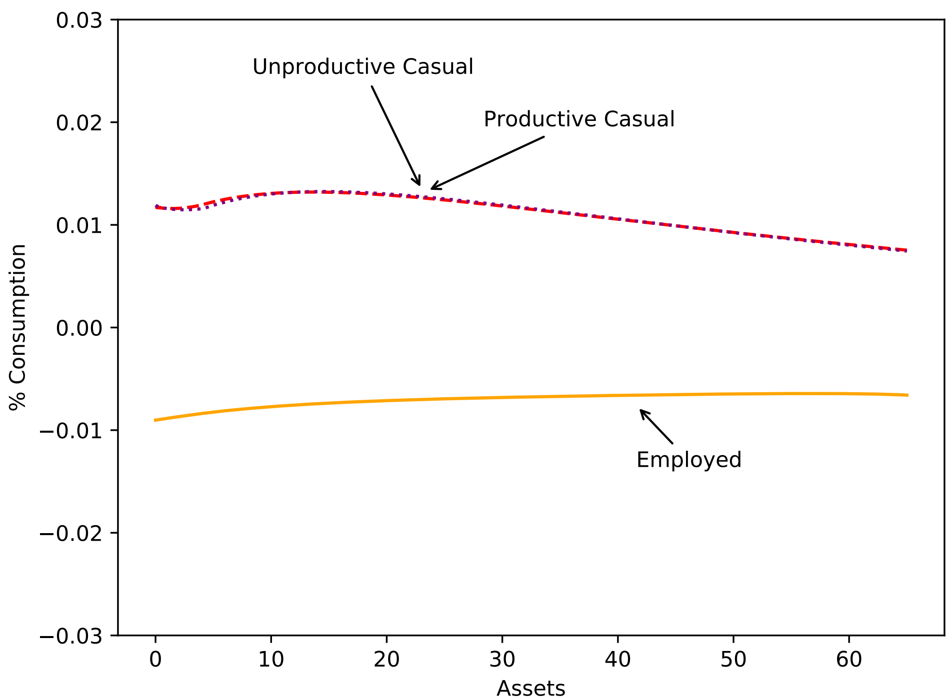


Figure 2.2: Welfare Effects of Search Subsidy as a Function of Household Assets
This figure displays the change in welfare, measured in consumption equivalent welfare, of the search subsidy policy as a function of a household's assets as well as their employment status and self-employment productivity.

Figure 2.2 displays the welfare impact of the search subsidy as a function of individual assets and employment status. For now, these numbers are calculated by comparing steady-states, although I plan to compute welfare along the transition path in the future. The red and purple lines display the welfare impact for workers without a wage sector job in the high productivity and lower productivity states respectively while the orange line displays the impact for workers matched with a wage job. Two aspects of the figure are striking. The first is that the welfare effects are highly dependent on an individual's employment state. The workers without a wage job, who switch between engaging in self-employment and searching for work, experience large welfare gains equal to around 1 percent of consumption while workers matched with an employer experience welfare loss of a little less than 1 percent. This gap is intuitive; workers without a wage job are either searching or anticipate to be searching in a few periods and thus are direct beneficiaries of the subsidy while workers already matched with a job pay a tax in order to fund the subsidy.

The second striking aspect of Figure 2.2 is that the welfare impacts exhibit very little heterogeneity with respect to an individual's level of wealth; individuals with zero assets experience welfare changes similar to the highest asset individuals. At first glance this result seems puzzling; however, splitting the welfare impact into the direct impact of the subsidy and the indirect impact through equilibrium objects reveals the intuition. Figure 2.3 displays the effect of the subsidy on welfare as a function of assets while fixing the equilibrium values of labor market tightness, wages and taxes (i.e. corresponding to column (3) of Table 2.3) while Figure 2.4 displays the difference between this counterfactual and the full results. In essence, Figure 2.3 displays the direct impact of the subsidy while Figure 2.4 displays the indirect impact.

In these figures, the impact of the policy is clearly heterogeneous with respect to individual wealth. The direct effect of the subsidy exhibits the largest welfare gains for the wealthiest individuals. Recall that households will participate in the wage sector until their self-insurance falls below a certain level, after which they will turn to self-employment until they have accumulated a buffer stock of savings. Because wealthy individuals can run down their assets for longer than poor individuals while searching for a job, they expect to collect the subsidy for more periods than poor households, who may only be able to search for a handful of periods before turning to self-employment. The welfare losses from the indirect effects of the policy are largest for wealthy households for a similar reason. Because wealthy households expect to participate in the wage sector the longest, they face the largest losses from a decline in the job-finding probability and an increase in taxes. Although the indirect effect and the direct effect individually exhibit substantial heterogeneity with respect to wealth, when they are combined the larger gains and larger losses for wealthy households serve to counteract each other and the overall welfare change doesn't vary much with wealth.

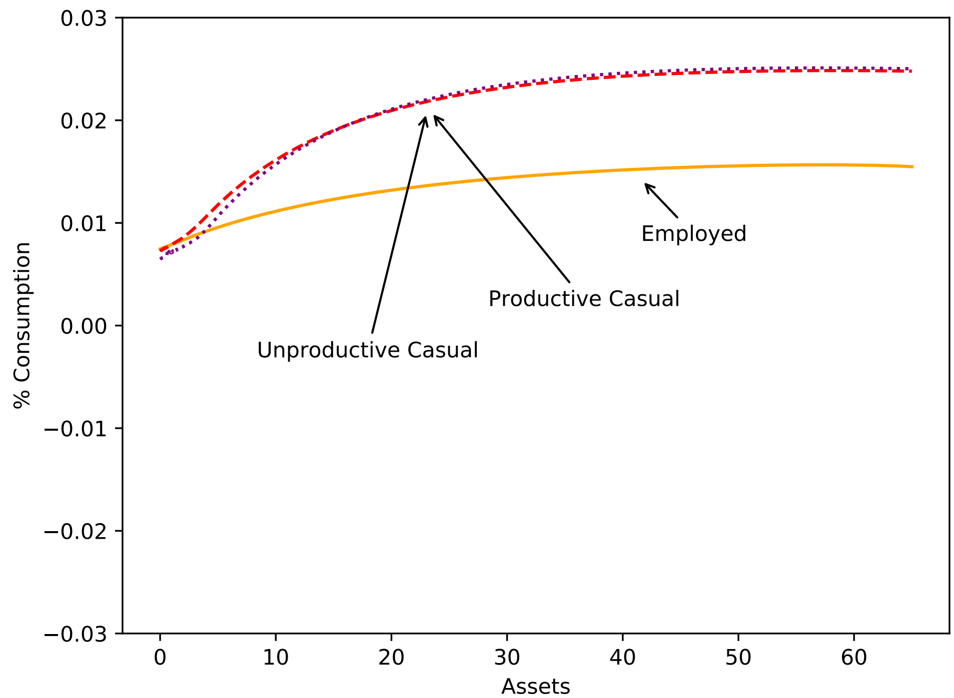


Figure 2.3: Welfare Effects of Search Subsidy as a Function of Household Assets (Fixed θ)
This figure displays the change in welfare, measured in consumption equivalent welfare, of the search subsidy policy as a function of a household's assets as well as their employment status and self-employment productivity in an alternative model where labor market tightness θ is fixed and does not change as a result of the policy. See the discussion for intuition on how to interpret these results.

5. Conclusion

Overall, my results suggest that the impact of subsidies for labor search is complex but, generally speaking, substantially smaller in general equilibrium than experimental results would suggest. As an Active Labor Market Policy designed to encourage participation in the market for

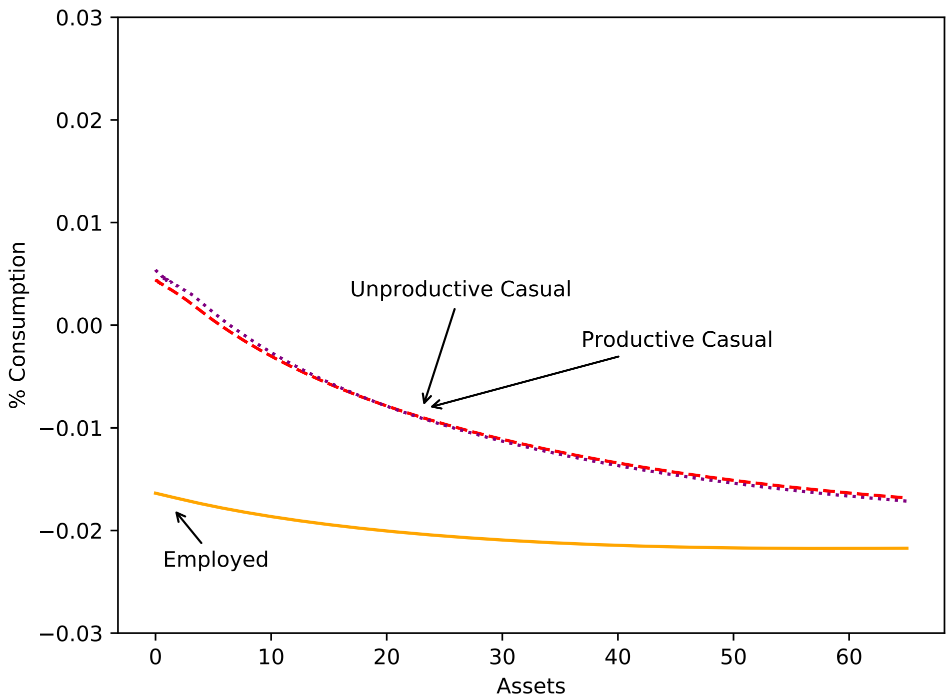


Figure 2.4: Difference Between Welfare Effects of Subsidy with and without Fixed θ

This figure the difference in the change in welfare as a function of household assets, employment status, and self-employed productivity between the full model and the alternative model with fixed θ. See the discussion for intuition on how to interpret this figure.

wage labor and reduce self-employment, the effects are substantially muted, largely due to households' high elasticity of labor search with respect to job-finding probability. Even the substantial subsidy evaluated in this paper only increases wage sector participation by 5.4 percentage points in general equilibrium. The subsidy results in a small increase in TFP in the wage sector as a large wage sector allows more productive entrepreneurs to increase their relative size. Although capital per worker declines in response, this increase in TFP is still enough to boost wage sector earnings by 1.88 percent.

Despite its muted effects in expanding the wage sector, the subsidy does substantially increase welfare by about 0.6 percent of consumption. This gain occurs almost entirely due to the increase in insurance that the subsidy provides. The subsidy transfers resources from a good state of the world (wage employment) to a bad state of the world (search) which is very valuable to households as they lack the means to do so effectively. These gains accrue entirely to unemployed households of all asset levels while employed households suffer welfare losses. The intuition is straightforward as unemployed households are the direct beneficiaries of the policies while employed households pay the taxes required to fund it.

One potentially important channel missing from this analysis is that of entry into entrepreneurship. It seems intuitive that subsidies expanding the market for wage labor would have an impact on business formation and entry into entrepreneurship; however, it is unclear, even in theory, which direction this effect will push. On one hand, it might be the case that the reduction in hiring costs lowers the cost of operating a business and encourages entrepreneurship. On the other hand, while I model entrepreneurs and workers as two completely different types of agents, it's possible that a subsidy for search would induce some entrepreneurs to close their businesses and pursue wage work, reducing the number of entrepreneurs. Additionally, in both these cases, the marginal entrepreneur choosing to close or open a business likely possess lower than average productivity, leading their entry decision to affect TFP and average earnings as well. Because of a lack of solid empirical evidence to discipline any of these channels and because of their theoretical ambiguity, I choose to abstract from them. However, future work could examine these channels more closely.

Future work could also examine the impact of Active Labor Market Policies aimed at firms such as hiring subsidies or subsidized apprenticeships. My results suggest that search subsidies alone are not sufficient to expand the wage sector, largely because the labor market slackens and the probability of finding a job decreases. These effects could be mitigated by policies aimed at increasing hiring by firms which would tighten the labor market. A combination of subsidies for job seekers and subsidies for firms may be the most effective tool for policymakers looking to expand wage sector employment.

Abebe, G., Caria, A. S., Fafchamps, M., Falco, P., Franklin, S. and Quinn, S. (2021), 'Anonymity or distance? job search and labour market exclusion in a growing african city', *The Review of Economic Studies* **88**(3), 1279–1310.

Abebe, G. T., Caria, S., Fafchamps, M., Falco, P., Franklin, S., Quinn, S. and Shilpi, F. (2017), 'Job fairs: Matching firms and workers in a field experiment in ethiopia', *World Bank Policy Research Working Paper* (8092).

Acemoglu, D. and Hawkins, W. B. (2014), 'Search with multi-worker firms', *Theoretical Economics* **9**(3), 583–628.

Aiyagari, S. R. (1994), 'Uninsured idiosyncratic risk and aggregate saving', *The Quarterly Journal of Economics* **109**(3), 659–684.

Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M. and Vitali, A. (2020), 'Tackling youth unemployment: Evidence from a labor market experiment in uganda', *Econometrica* **88**(6), 2369–2414.

Algan, Y., Crépon, B. and Glover, D. (2020), 'Are active labor market policies directed at firms effective? evidence from a randomized evaluation with local employment agencies', *J-PAL working paper* .

Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M. and Vitali, A. (2021), 'The search for good jobs: Evidence from a six-year field experiment in uganda', *Available at SSRN 3910330* .

Banerjee, A., Basu, P. and Keller, E. (2021), Cross-country disparities in skill premium and skill acquisition, Technical report.

Banerjee, A., Duflo, E., Glennerster, R. and Kinnan, C. (2015), 'The miracle of microfinance? evidence from a randomized evaluation', *American economic journal: Applied economics* **7**(1), 22–53.

Bratti, M., Ghirelli, C., Havari, E. and Santangelo, G. (2018), 'Vocational training for unemployed youth in latvia: Evidence from a regression discontinuity design'.

Buera, F. J., Kaboski, J. P. and Shin, Y. (2011), 'Finance and development: A tale of two sectors', *American economic review* **101**(5), 1964–2002.

Buera, F. J., Kaboski, J. P. and Shin, Y. (2021), 'The macroeconomics of microfinance', *The Review of Economic Studies* **88**(1), 126–161.

Caria, S., Gordon, G., Kasy, M., Quinn, S., Shami, S. and Teytelboym, A. (2020*b*), 'An adaptive targeted field experiment: Job search assistance for refugees in jordan'.

Crépon, B. and Premand, P. (2018), 'Creating new positions? direct and indirect effects of a subsidized apprenticeship program', *Direct and Indirect Effects of a Subsidized Apprenticeship Program (August 20, 2018). World Bank Policy Research Working Paper* (8561).

De Mel, S., McKenzie, D. and Woodruff, C. (2019), 'Labor drops: Experimental evidence on the return to additional labor in microenterprises', *American Economic Journal: Applied Economics* **11**(1), 202–35.

Donovan, K., Lu, W. J. and Schoellman, T. (2020), Labor market dynamics and development, Technical report, JSTOR.

Feng, Y., Lagakos, D. and Rauch, J. E. (2018), Unemployment and development, Technical report, National bureau of economic research.

Feng, Y. and Ren, J. (2021), Skill bias, financial frictions, and selection into entrepreneurship, Technical report.

Franklin, S. (2018), 'Location, search costs and youth unemployment: experimental evidence from transport subsidies', *The Economic Journal* **128**(614), 2353–2379.

Gollin, D. (2008), 'Nobody's business but my own: Self-employment and small enterprise in economic development', *Journal of Monetary Economics* **55**(2), 219–233.

Herreño, J. and Ocampo, S. (2021), 'The macroeconomic consequences of subsistence self-employment'.

Itskhoki, O. and Moll, B. (2019), 'Optimal development policies with financial frictions', *Econometrica* **87**(1), 139–173.

Porzio, T., Rossi, F. and Santangelo, G. V. (2021), The human side of structural transformation, Technical report, National Bureau of Economic Research.

Poschke, M. (2013), 'Entrepreneurs out of necessity: a snapshot', *Applied Economics Letters* **20**(7), 658–663.

Poschke, M. (2019), 'Wage employment, unemployment and self-employment across countries'.

Schoar, A. (2010), 'The divide between subsistence and transformational entrepreneurship', *Innovation policy and the economy* **10**(1), 57–81.

Smith, E. (1999), 'Search, concave production, and optimal firm size', *Review of Economic Dynamics* **2**(2), 456–471.

Appendix

A. Additional Tables and Figure

Table 2.A.1: Effect of Search Subsidy on Labor Market Outcomes (Abebe et al., 2021)

Outcome	Control Mean	Effect of Subsidy
Any Work	0.526	0.037
		(0.029)
Hours Worked	26.18	0.183
		(1.543)
Monthly Wages	857.9	65.88
		(63.86)
Permanent Job	0.171	0.033*
		(0.018)
Formal Job	0.224	0.054**
		(0.019)
Job Satisfaction	0.237	-0.001
		(0.027)

This table reproduces the primary results of Abebe et al. (2021) and displays the control mean for a variety of labor market outcomes as well as the experimentally estimated treatment effect of a conditional cash transfer to job seekers.

B. Derivations and Proofs from Section 2.5.2

The first result to show is that the entrepreneur's optimal choice of f' and n' satisfy $\eta(z;X) = \frac{\gamma f'^*}{n'^*}$ for some function η depending only on z and X. Substituting in the wage determination equation (which the entrepreneur takes as given) and the vacancy posting constraint, the first-order condition for f' and n' can be combined with the envelope condition for f and n to generate

$$\mu\left((1-\alpha)(1-\chi)z(\frac{\gamma f'}{n'})^{\alpha}-\left((1-\chi)\underline{w}-\frac{c}{p(\theta(X'))}(1-\lambda)\right)\right)=\frac{c}{p(\theta(X'))}\beta\Delta\mu'$$
$$\mu\left(\gamma\alpha(1-\chi)z(\frac{\gamma f'}{n'})^{\alpha-1}+1-\gamma(r+\delta)\right)=\mu'\beta\Delta$$

where μ is the Lagrange multiplier on the budget constraint, μ' is the Lagrange multiplier on the budget constraint in the following period, and $\theta(X')$ is a price function mapping aggregate states X to equilibrium values of θ. Combining these two equations, substituting in η, and defining A, $B(X')$, and $C(X')$ for clarity yields

$$Az\eta^{\alpha}+B(X')z\eta^{\alpha-1}+C(X')=0$$

which, for $0<\alpha<1$, can be shown to have a unique and positive solution for η for any value of z and X'. Call this solution $\tilde{\eta}(z;X')$. Finally, substituting $X'=H(X)$ and defining $\eta(z;X)=\tilde{\eta}(z;G(X))$ completes the derivation.

The next result to show is that entrepreneurs choose a growth rate that depends only on their z and aggregate state variables. This follows almost directly from the previous result. Substituting $n=\frac{\gamma}{\tilde{\eta}(z;X)}f$ in to the budget constraint of the entrepreneur problem reveals that the RHS of the budget constraint is now linear in f and can be written

$$d+\left(1+\frac{c}{p(\theta(X))}\frac{\gamma}{\eta(z;X)}\right)f'=\left((1-\chi)\gamma z\tilde{\eta}(z;X)^{\alpha-1}-\left((1-\chi)\underline{w}-\frac{c}{p(\theta(X))}(1-\lambda)\right)\frac{\gamma}{\tilde{\eta}(z;X)}+\left(1-\gamma(r+\delta)\right)\right)f$$
$$\Rightarrow\ d+E(z,X)f'=D(z,X)f$$

where $D(z,X)$ and $E(z,X)$ are defined such that the second line is equivalent to the first line. E functions as the price of collateral f relative to the price of consumption d while D functions as the return to collateral. Because entrepreneurs possess log utility, the entrepreneur problem has the well-known solution of a constant growth rate in f depending on the values of D and E which are given by z and X so that $f'=g(z;X)f$.

The final result to show is the proof of Proposition 1. By assumption, θ is assumed to be constant. Let $\hat{E}(z,\theta)$ and $\hat{D}(z,\theta)$ denote E and D respectively, but with $\theta(X)$ simply replaced by θ, the argument to the function. Note that this is possible because E and D only depend on X through θ. Then we have the explicit solution

$$\hat{g}(z,\theta) = \beta\Delta\frac{\hat{E}(z,\theta)}{\hat{D}(z,\theta)} = \beta\Delta\frac{\Big((1-\chi)\gamma z\hat{\eta}(z;\theta)^{\alpha-1} - \big((1-\chi)\underline{w} - \frac{c}{p(\theta)}(1-\lambda)\big)\frac{\gamma}{\hat{\eta}(z;\theta)} + \big(1-\gamma(r+\delta)\big)\Big)}{\big(1+\frac{c}{p(\theta)}\frac{\gamma}{\hat{\eta}(z;\theta)}\big)}$$

The chain rule yields $\frac{d\hat{g}}{d\theta} = \frac{\partial\hat{g}}{\partial c/p(\theta)}\frac{dc/p(\theta)}{d\theta} + \frac{\partial\hat{g}}{\partial\hat{\eta}}\frac{d\hat{\eta}}{dc/p(\theta)}\frac{dc/p(\theta)}{d\theta}$. Using either direct calculation of partial derivatives or implicit differentiation (in the case of $\frac{d\hat{\eta}}{dc/p(\theta)}$), we can express each individual piece as

$$\begin{aligned}
\frac{\partial\hat{g}}{\partial c/p(\theta)} &= -\frac{(\frac{\hat{g}}{\beta\Delta}-1)+\lambda}{\frac{\eta}{\gamma}+\frac{c}{p(\theta)}} \leq 0 \\
\frac{\partial\hat{g}}{\partial\hat{\eta}} &= \frac{\frac{\beta\Delta}{\hat{g}}-\frac{\hat{g}}{\beta\Delta}}{\frac{\eta}{\gamma}+\frac{c}{p(\theta)}} \leq 0 \\
\frac{d\hat{\eta}}{dc/p(\theta)} &= \frac{\gamma\big(\alpha(1-\chi)z\hat{\eta}^{\alpha-1}-(r+\delta)\big)+\lambda}{J(\theta)} > 0
\end{aligned}$$

where $J(\theta)$ is a placeholder for a complex but unambiguously positive expression and I have made use of the first-order condition for f' in the second expression. It is worth commenting briefly on why the claimed inequalities hold. Both the first and second expressions follow directly from the fact that an optimally acting entrepreneur will ensure that $g \geq \beta\Delta$. This is clearly true as an entrepreneur can always choose to select $k=0, n=0$ and simply eat their cake, yielding $g=\beta\Delta$. An entrepreneur will only choose to operate if they can be weakly better off by doing so. The third and final expression follows from the first-order condition for capital which ensures that the marginal product of capital $\alpha(1-\chi)z\hat{\eta}^{\alpha-1}$ is greater than the marginal cost of capital $r+\delta$ (the MPK is greater, rather than equal to, the marginal cost due to the presence of the financing constraint). Because $\frac{dc/p(\theta)}{d\theta} > 0$ by construction, combining these inequalities with the chain rule provides the result $\frac{d\hat{g}}{d\theta} < 0$ and along the way we have shown $\frac{d\hat{\eta}}{d\theta} > 0$.

The result for $\frac{\partial \hat{g}}{\partial \theta \partial z}$ is straightforward. We have $\frac{\partial \hat{g}}{\partial z} = \frac{(1-\chi)\hat{\eta}^{\alpha}}{\frac{\hat{\eta}}{\gamma}+\frac{c}{p(\theta)}}$ which is also clearly greater than zero and decreasing in θ. Although this result holds only for partial derivatives (i.e. with $\hat{\eta}$ being held constant), it can also be shown to hold for total derivatives in the case where $\hat{\eta} \geq \alpha(1+\frac{c}{p(\theta)}\gamma)$ by applying the chain rule as above and computing $\frac{d\hat{\eta}}{dz}$ using implicit differentiation.

CHAPTER 3

Macroeconomic Effects of COVID-19
Across the World Income Distribution

by

Titan Alon, Minki Kim, David Lagakos, and Mitchell VanVuren

Abstract

The macroeconomic effects of the COVID-19 pandemic were most severe for emerging market economies, representing the middle of the world income distribution. This paper provides a quantitative economic theory for why emerging markets fared worse, on average, relative to advanced economies and low-income countries. To do so we adapt a workhorse incomplete-markets macro model to include epidemiological dynamics alongside key economic and demographic characteristics that distinguish countries of different income levels. We focus in particular on differences in lockdown stringency, public insurance programs, age distributions, healthcare capacity, and the sectoral composition of employment. The calibrated model correctly predicts the larger output losses and greater fatalities in emerging market economies, matching the data. Quantitatively, differences in the size of public transfer programs, age demographics, and the sectoral composition of employment explain most of the cross-country variation. Emerging markets fared especially poorly due to their high employment share in occupations requiring social interactions and their low level of pubic transfers, which leads economically vulnerable households to continue working in the market rather than sheltering at home. Low income countries fared relatively better due mainly to their younger populations, whom are less susceptible to disease, and larger agricultural sectors, which require fewer social interactions.

1. Introduction

While every country has been adversely affected by the coronavirus pandemic, the damage it has wrought varied widely around the world. In this paper, we investigate how and why the pandemic's macroeconomic consequences have differed (so far) across the world income distribution. We focus in particular on variation in output and excess mortality across three broad groups of countries: low-income economies, emerging markets, and advanced economies, as classified by the International Monetary Fund (IMF). As we detail below, data from a variety of sources reveals that the pandemic's cost in terms of lives and livelihoods was roughly U-shaped in national income, with emerging markets experiencing the worst public health and macroeconomic consequences. On average, GDP per capita in emerging markets declined by 6.7 percent from 2019 to 2020, compared to 2.4 percent in advanced economies and 3.6 percent in low income countries. Excess mortality exhibits a similar pattern. According to estimates by *The Economist*, excess mortality was 75 percent higher in emerging markets than in advanced economies. While credible excess mortality data for low-income countries are still largely unavailable, the few existing estimates point to lower mortality rates.

We assess the extent to which variation in policy or preexisting economic and demographic characteristics can explain the cross-country GDP and mortality outcomes in the data. In part, these outcomes could stem from differences in government policy responses to combat the coronavirus pandemic. While most countries enacted similar "lockdown style" policies and expanded social insurance programs, the scope of such efforts varied substantially. According to the Oxford Coronavirus Government Response Tracker, the stringency of lockdown policies aiming to restrict individual behavior (such as school and workplace closures) were somewhat stricter emerging markets. The generosity of social insurance programs, in contrast, were substantially higher in richer countries. Accounting for these differences in policy is important because they can directly affect both fatalities and growth during the pandemic. The cross-country variation may also arise from stark underlying differences in economic and demographic characteristics that predate the pandemic. For instance, low-income countries may face very different public health risks than wealthier ones, as they have substantially younger populations but also less developed healthcare systems. Moreover, systematic differences in the sectoral composition of employment make some countries better able to preserve income while mitigating health risks through social distancing or lockdowns. Low-income countries may benefit from their large agricultural sectors and rural populations, which provide a resilient source of income that can be sustained while limiting social contacts. On the other hand, Gottlieb, Grobovsek, Poschke and Saltiel (2021*b*) show that in urban areas, the ability to work from home is far more limited in lower income countries. Combining their estimates with data on urbanization rates, we can measure the total share of labor in *social sector employment*, as in Kaplan, Moll and Violante (2020), to capture cross country differences in the ability to work from home or with limited in-person interactions. The composite measure shows that emerging markets have the highest share of workers in social employment, due to their large urban workforce concentrated in high-contact sectors such as manufacturing and retail trade. In contrast, low-income countries have the smallest social employment shares, due to the predominance of rural agricultural work.

To investigate the extent to which these factors can explain observed differences in mortality and output, this paper follows the newly emerged literature on the macroeconomics of pandemics by combining a variant of the SICR model standard in epidemiology with a workhorse macro model. In particularly, our framework builds on the heterogeneous-agent incomplete-markets model of Aiyagari (1994), Bewley (1977) and Huggett (1996), which allows us to capture the individual-level trade-offs between consumption and health that have been the focus of most economic analysis during the pandemic. The model distinguishes between social and non-social jobs, differentiating individuals by their ability to work from home or while socially distancing. We incorporate age heterogeneity following Glover, Heathcote, Krueger and Ríos-Rull (2020) and allow death rates to depend on a person's age, consistent with a vast medical literature. Our model also allows for a time-varying infection rate that captures, in a reduced-form way, the various other non-modeled determinants of disease progression, such as seasonal conditions, improved treatment, or virus mutation. Finally, we include constraints on peak healthcare capacity which represent limits on the ability of certain healthcare systems to treat many patients at once, due to factors like the availability of protective equipment, hospital beds, or supplemental oxygen.

In the model, the propagation of disease depends in large part on individual decisions to stay home during the pandemic or continue working in the market. The model therefore features a public health externality that creates space for welfare improving government interventions. We model lockdown policies in a simple way that is consistent with policy variation observed during the pandemic. Specifically, we feed in time-varying lockdown measures that replicate the changing stringency of government policies over the course of the pandemic, as measured by the Oxford Coronavirus Government Response Tracker (OxCGRT). In the model, lockdown stringency corresponds to the fraction of susceptible individuals who are confined to their home, where they are less likely to become infected but incur income losses depending on their job type. While we do not allow households to disobey lockdowns, individuals can voluntarily elect to work from home at any point in time. Households also receive time-varying public transfers to support or replace lost income. As with lockdowns, we set the level of public financial assistance to match the time-path reported in the OxCGRT financial support index.

To evaluate the quantitative importance of these channels in explaining the facts at hand, we parameterize the model to match key pre-pandemic economic and demographic characteristics of the United States. Parameters governing the epidemiological process are set using estimates from the relevant medical literature. We compute the model's equilibrium response to the COVID-19 pandemic as a surprise "MIT shock," where a small exogenous fraction of the population becomes infected with the virus, and then allow the disease to spread endogenously through the populous. We feed in the time-series of vaccination rates, as reported by OxCGRT, allowing a random fraction of the population to be vaccinated in each period, consistent with rates we observe in the data. We set the non-parametric component of the infection probability so that the model's endogenous disease path (nearly) exactly replicates the time-path of fatalities from COVID-19 in the United States during the pandemic. We calibrate the productivity penalty incurred during lockdowns to match the cumulative 2019-2020 year-on-year employment loss in the United States. We also allow for a one-off shock to aggregate total factor productivity (TFP), which is calibrated to match the cumulative 2019-2020 year-on-year decline in U.S. real GDP per capita.

We use the calibrated model to simulate how the United States would have fared during the pandemic if it counterfactually had the characteristics of emerging market or low-income economies. Comparing the model's predictions to the actual outcomes allows us to assess the importance of each characteristic in explaining cross-country differences in GDP declines and mortality rates. Including all characteristics, the model is able to generate the U-shaped pattern in output losses and mortality rates observed across the world income distribution. The model can fully account for the relatively larger GDP declines and higher fatalities in emerging markets compared with advanced economies. Similarly, the model correctly predicts the more modest output losses and mortality in low income countries, albeit to a quantitatively greater extent than what is observed in the data.

Simulating the contribution of each factor in isolation, we find that variation in the sectoral composition of employment is the most important factor in accounting for cross-country GDP declines. Emerging markets suffered the greatest output losses in large part because they had high employment shares in close-contact occupations. In contrast, output declines in low income countries were substantially moderated by their large agricultural sector. The sectoral composition of employment also plays an important role in explaining cross-country mortality outcomes, along with variation in age demographics and the size of social insurance programs. In both low income and emerging markets, low levels of public financial assistance during the pandemic substantially amplified fatalities by leading many economically vulnerable individuals to continue working in the market rather than sheltering at home during times of peak infection. Our counterfactuals predict that if the United States had implemented the more limited transfer programs in low income and emerging market economies, cumulative fatalities from the pandemic would have been 50 percent greater. In low income countries, these higher fatalities were avoided largely thanks to their substantially younger populations with greater natural immunity to infection and serious illness.

High agricultural employment shares, where transmission is lower while working, also reduces mortality in low income countries. In contrast, emerging market economies experienced much greater mortality because they do not benefit from favorable age demographics and also have high social sector employment shares.

Following the counterfactual simulations, we conclude the analysis by reporting multiple correlations between cross-country changes in GDP per capita during the pandemic and covariates representing the various channels embodied in our model. Consistent with our findings, the data show that the agricultural employment shares are highly correlated with GDP changes during the pandemic, while lockdown stringency exhibits a strong negative correlation. Median age and indices of government economic support show weaker correlations. Altogether, the covariates greatly reduces the observed U-shape pattern in GDP declines across the world income distribution. The result suggests that this parsimonious set of variables, and the economic mechanisms they represent, are empirically relevant in explaining cross-country macroeconomic outcomes during the pandemic.

Our work builds on the first generation of papers addressing the aggregate effects of COVID-19 in the developing world, which were largely written in the early months of the pandemic (Loayza and Pennings, 2020; Alon, Kim, Lagakos and VanVuren, 2020; Alfaro, Becerra and Eslava, 2020; von Carnap, Almås, Bold, Ghisolfi and Sandefur, 2020; Djankov and Panizza, 2020). The current paper differs in its efforts to explain observed macroeconomic outcomes through the first year and a half of the pandemic, in particular the larger declines in GDP and employment in emerging markets. Sanchez (2021) also notes the larger decline in GDP middle-income countries, but does not attempt to explain this finding. We also emphasize the inability of individuals in emerging market economies to work from home, following Gottlieb, Grobovsek, Poschke and Saltiel (2021*a*,*b*), though we argue that low-income developing countries, on account of their large agriculture sectors, are better able to work without social interactions.

On the modeling front, our study most closely follows the structural macro work on the pandemic using models of heterogeneity in income, age and occupation/sector of employment (e.g. Acemoglu, Chernozhukov, Werning and Whinston, 2020; Bairoliya and Imrohoroglu, 2020; Kaplan, Moll and Violante, 2020; Glover, Heathcote, Krueger and Ríos-Rull, 2020; Brotherhood, Kircher, Santos and Tertilt, 2021; Chopra, Devereux and Lahiri, 2021). Our model of disease dynamics features endogenous behavioral responses to changes in infection rates, even in the absence of government intervention, as in Greenwood, Kircher, Santos and Tertilt (2019); Alvarez, Argente and Lippi (2020); Krueger, Uhlig and Xie (2020) and other studies. To our knowledge ours is the first to evaluate the quantitative predictions of a model of this sort for how the experience of emerging markets differed from richer (or poorer) countries.

Our study abstracts from many important features of reality that may also be relevant for the effects of the pandemic outside of the world's advanced economies, such as negative impacts through shocks to global supply chains (Cakmakli, Demiralp and Ozcan, 2020; Bonadio, Huo, Levchenko and Pandalai-Nayar, 2021), the ability to issue sovereign debt (Arellano, Bai and Mihalache, 2020), or the ability to test and trace infections (Berger, Herkenhoff and Mongey, 2020). We also abstract from differences in the prevalence of co-morbidities, such as diabetes and cardiovascular disease, and differential ability or willingness or ability to mask or get vaccinated. These issues would be valuable to consider in future studies trying to explain cross-country differences in the macroeconomic effects of the pandemic.

2. Macroeconomic Effects of the Coronavirus Pandemic by Income Level

This section presents the main facts regarding excess mortality and output losses across the world income distribution resulting from the coronavirus pandemic. Following the IMF classification, we focus in particular on three major income groups: low-income economies, emerging markets, and advanced economies. In 2019, the median GDP per capita of these three country groups was \$1,124, \$6,700, and \$43,144, respectively, in constant 2010 USD. While there is interesting variation even with these group, we focus the main part of our analysis on just the three aggregate groups. Section 5 of the paper looks at empirical patterns in the full set of countries for which data are available. Here, drawing on various data sources, we show that both output losses and excess mortality exhibit hump-shaped outcomes with middle income countries experiencing the worst. We then present in a systematic way the important differences in policy and underlying economic and demographic conditions. For each, we briefly discuss their relevance for the pandemic's impact in order to help motivate the model and quantitative analysis which follows.

2.1. The Impact of COVID-19

The first fact we highlight is the differential impact of the pandemic on output losses and employment declines across the world income distribution. Figure 3.1 displays the data by plotting changes in output and employment for low-income, emerging, and advanced economies. While there is considerable variance even within groups, a clear U-shaped patterns emerges in which output losses were greatest in emerging economies. GDP per capita fell by 6.7 percent and employment by 5.4 percent in emerging economies, considerably worse than both wealthier countries where output and employment losses were 4.6 percent and 2.4 percent, respectively, and lower income countries where those losses stood at 3.6 and 3.1 percent.[11] Such outcomes are surprising given the tremendous resources and technology that wealthy countries brought to bare in combating COVID-19, resources that low-income countries had no ability to marshal or match in any comparable way.

The second important fact pertains to the fatalities caused by COVID-19. These deaths are commonly measured using excess mortality, the difference between total deaths in a given month of the pandemic and those that would be normally expected, measured as expected deaths during the same month over the previous (typically five) years. Figure 3.2 displays the data by comparing mortality outcomes in advanced and emerging economies. As with output losses, we find that the emerging economies experienced the worst outcomes. According to estimates *The Economist*, excess deaths in emerging economies stands at 112.9 per hundred thousand people, which is around

[11] Appendix Figures 3.A.1 and 3.A.2 illustrate that the relationship also holds in the un-binned data and Appendix Figure 3.A.3 displays similar trends in cross-country consumption data.

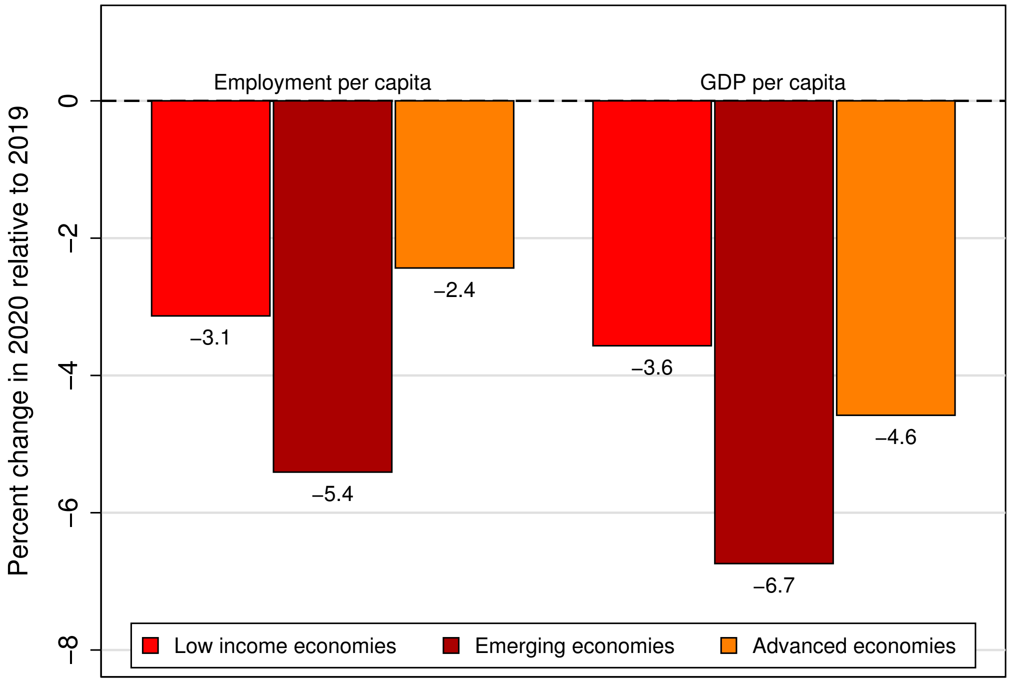


Figure 3.1: GDP and Employment Growth from 2019 to 2020 by National Income

Note: Employment data comes from the ILO Statistical Database and data on GDP per capita is taken from the World Bank World Development Indicators.

75 percent higher than the average estimate for advanced economies, which experienced 64.1 excess deaths per hundred thousand. Estimates from the World Mortality Database of Karlinsky and Kobak (2021) show 164.5 excess deaths per hundred thousand people, or 65 percent larger than the 99.5 deaths per hundred thousand of advanced ones. The gap is even wider in the New York Times mortality tracker which records 148.1 deaths per hundred thousand in emerging economies, compared to 63 in advanced ones.

Internationally comparably data on excess mortality in low-income countries are more difficult to find. The most comparable statistics of which we are aware contain very few observations from low-income countries (see Appendix Figures 3.A.4 and 3.A.5). These data, from *The Economist* and Karlinsky and Kobak (2021), have two and five observations from the low-income group respectively. Deaths for this small set of countries average around 100 excess deaths per hundred thousand people, putting them well below the level of the emerging markets. Official

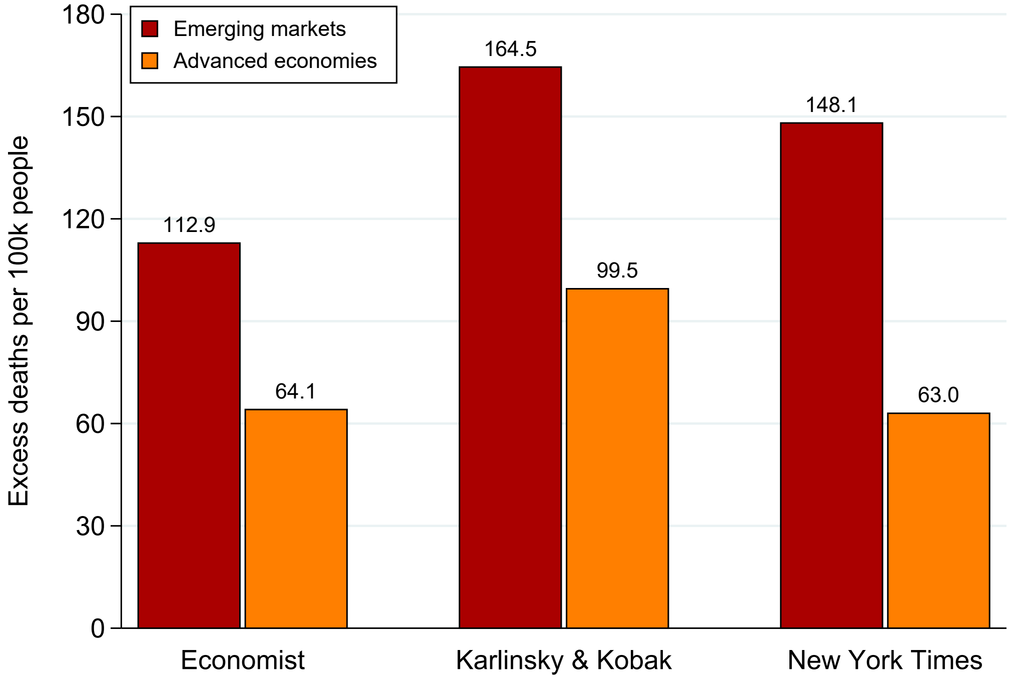


Figure 3.2: Excess Deaths from 2019 to 2020

Note: Data sourced from the New York Times and Economist excess mortality trackers, and Karlinsky and Kobak's (2021) World Mortality Database.

data on deaths from COVID-19 in low-income show remarkably low levels of fatalities (see e.g. Appendix Figure 3.A.6), though there is widespread belief that official statistics undercount deaths there. Our read of the literature is that there is still no clear consensus on what the true death rates have been in low-income countries, though it seems unlikely that they are worse than the high rates estimated in emerging markets such India (Deshmukh et al., 2021; Ramachandran and Malani, 2021), Mexico (Dahal et al., 2021) and Brazil (Yamall Orellana et al., 2021).

Taken together, the data reveal that the impact of the COVID-19 pandemic across the world income distribution has been highly non-linear. Emerging economies have been hit the hardest most in terms of output losses and likely in terms of excess mortality as well. Equally surprising is that the data suggest that low-income countries have fared better than advanced economies in terms of output losses, and possibly also in terms of mortality rates, despite the far greater economic and

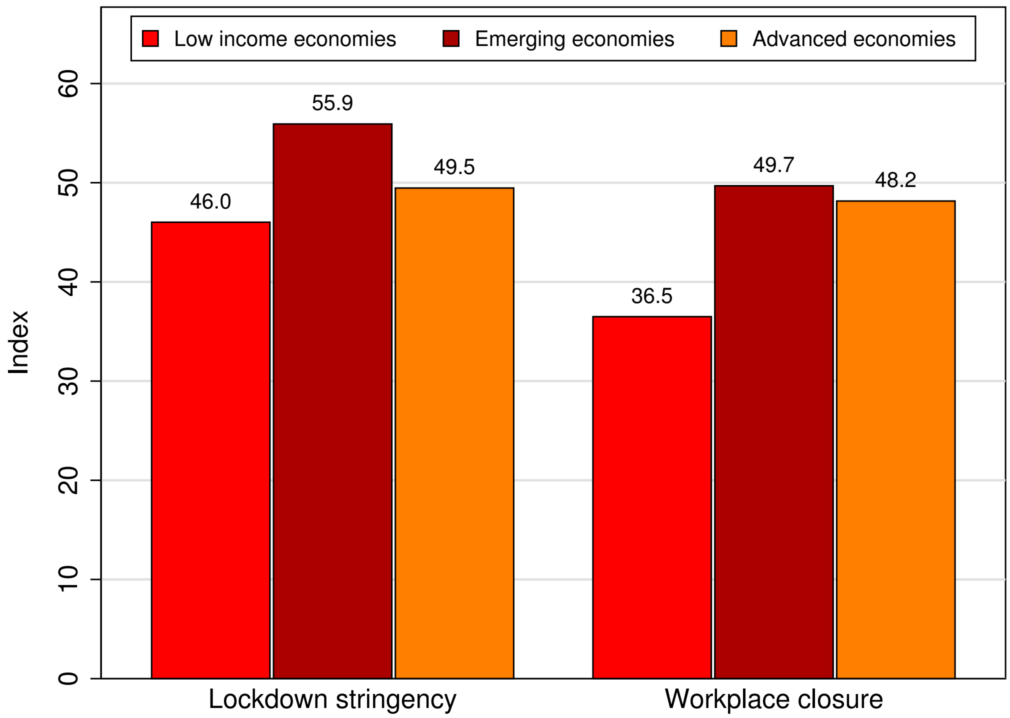


Figure 3.3: Oxford Lockdown Stringency Index

Note: The Government Stringency Index is taken from the Oxford Government Response Tracker (OxCGRT). GDP per capita is expressed at PPP and taken from Penn World Table 9.1 (Feenstra et al., 2015).

technological resources mustered by the latter to combat the crisis.

2.2. Differences in Policy Response

A natural candidate explanation for the cross-country variation is that they reflect differences in policy responses to the COVID-19 pandemic. While nearly all countries implemented some sort of lockdown and transfer programs, they varied widely both in the stringency of restrictions and in the generosity of transfers. The policy distinction matters for how well countries manage the endogenous path of infections through the public health externality and for the ability of households to protect themselves by staying home for prolonged periods without income.

By lockdown policies, we refer to those whose primary aim is to restrict individual behavior and social interactions to stem the spread of disease. These include school closures; workplace closures; public event cancellations; restrictions on public gatherings; closure of public transport;

stay-at-home requirements; public information campaigns; and domestic and international travel restrictions. The Oxford Coronavirus Government Response Tracker's (OxCGRT) *stringency index* provides a parsimonious quantifiable measure of how strict these policies were across countries. Figure 3.3 plots the index of each country group, and shows that the most stringent lockdown policies were implemented by emerging economies (the un-binned data are displayed in Appendix Figure 3.A.7). When we simulate lockdown policies, we implement them using the time-series of workplace closures reported by OxCGRT to be consistent with how such policies are represented in the model. As cross-country and time-series data (see Appendix Figure 3.A.8) show, the variation in workplace lockdowns is similar to the overall stringency of policies across countries. One concern is that these data only represent *de jure* differences in policies, and that *de facto* lockdowns actually varied markedly less. Google workplace mobility data suggests this is not the case, confirming that the largest gap in workplace mobility is between the richest and poorest countries (see Appendix Figure 3.A.9).

Another important dimension of the policy response in nearly all countries was the expansion of social insurance payments, such as unemployment benefits. These payments are viewed as critical to offsetting lost income and make isolating at home economically feasible for those with low savings or little income. However, as the crisis unfolded it quickly became clear that governments in many developing countries lacked the fiscal capacity to sustain substantial transfers to major segments of their population for very long. Consequently, we observe substantially more cross-country variation in the size and scope of social insurance programs than in lockdown policies.

Figure 3.4 provides two measures capturing the scope and generosity of transfer programs implemented in response to COVID-19 across the world income distribution. The left side histogram plots national pandemic spending as a share of GDP, which includes comprehensive measures of budgetary fiscal support to individuals and firms estimated by the IMF. While pandemic spending appears similar in low-income and emerging economies, they are only about one-third the spending undertaken by advanced economies which reached nearly 10 percent of GDP. The right

Table 3.1: Oxford Covid-19 Government Response Indices in 2020

	Country Income Group		
Index	Low-Income	Emerging Markets	Advanced Economies
Panel A: Included in both Stringency and Health & Containment Indices			
School closures	53.8	64.8	50.1
Workplace closures	34.6	47.0	45.1
Cancellation of public events	57.0	69.4	63.7
Restrictions on public gatherings	50.9	59.5	61.3
Closure of public transport	22.5	32.0	17.8
Stay at home requirements	25.0	35.7	24.9
Restrictions on internal movements	32.9	47.7	31.8
International travel controls	57.6	63.6	63.4
Public information campaigns	79.7	83.8	87.0
Panel B: Included only in Health & Containment Index			
Contact tracing policy	54.4	61.5	67.6
Facial coverings	43.8	46.4	37.3
Testing policy	37.9	52.2	58.8
Vaccination policy	22.8	31.3	35.3
Protection of the elderly	19.4	40.8	57.3
Panel C: Included only in Economic Support Index			
Income support	17.3	29.3	57.8
Contract/Debt relief	31.0	49.6	58.9
Observations	52	67	33

Note: Countries are grouped into low income, emerging markets, and advanced economies using the IMF's economic classification of countries. Data in the table is the average level of the Oxford Covid-19 government response tracker by country income group.

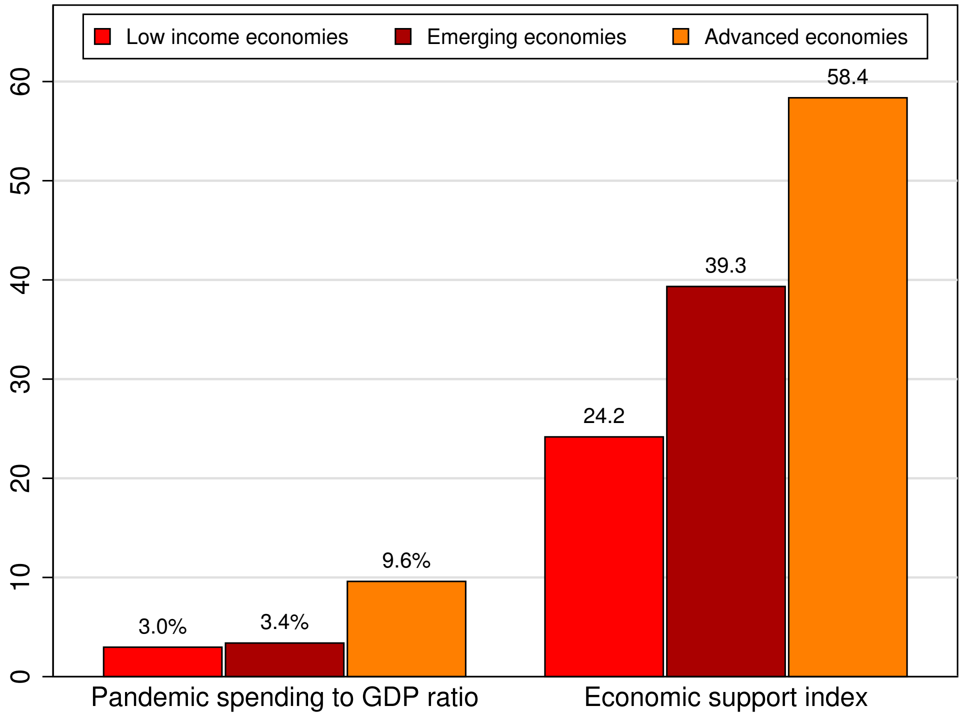


Figure 3.4: Pandemic Spending and Economic Support

Note: The left side histogram plots the ratio of pandemic spending to GDP, taken from the IMF. The right side histogram displays the Oxford Economic Support Index available through the Oxford Coronavirus Government Response Tracker's (OxCGRT).

side histogram displays the Oxford's Government *Economic Support Index* which records financial assistance programs such as income replacement and debt relief for individual citizens. The index should be interpreted as an ordinal measure of economic assistance for individual citizens in that it does not include support to firms or business and does not take into account the total fiscal value of economic support programs. Nevertheless, the data reveal a similar pattern with spending on economic support rising monotonically with national income.[12]

These cross-country differences in lockdown policies and public insurance programs are even more apparent when one examines the underlying components of the OxCGRT's indices which are displayed in Table 3.1. The first noticeable feature is that low-income countries have the least stringent policies in every lockdown category, and in all other categories except "Facial Cov-

[12]The greater cross-country variation in economic support policies, as compared to lockdown policies, is most apparent in these underlying data. See Appendix Figures 3.A.10 and 3.A.11.

erings." The near opposite is true for emerging economies which have the most stringent policies across all sub-categories of lockdown measures (Panel A) except "Public Information Campaigns." The largest deviations in emerging economy lockdowns pertain to the closure of public transport, stay at home orders, and restrictions on internal movements. This is notable since these measures likely imposed the largest restrictions on commercial activity, especially in emerging economies where the ability to work from home is not widespread (see Section 2.5) and substituting to e-commerce and delivery services is limited by infrastructure. Finally, it is interesting to note that the stringency of emerging economy policies does not extend beyond lockdowns; as Panels B and C show, direct public health interventions and economic support policies were generally less encompassing in emerging economies. Taken altogether, the scope of differences in the stringency and aim of policies across the world income distribution offer ample scope for them to drive the differences in outcomes we observe in the data.

2.3. Differences in Population Structure

It has been well known since the beginning of the pandemic that COVID-19 poses dramatically greater health risks to older individuals, in particular those over the age of 65 (Ferguson et al., 2020; Glynn, 2020). Early centers of infection in the west, such as Italy, experienced health impacts concentrated on those in this older age range, with particularly severe fatality rates for those in their 80s and 90s. At the same time, the number of deaths linked to COVID-19 for those under 20 has been negligible, though certainly not zero.

A basic demographic difference between advanced and developing economies is that populations are far younger in the developing world. Since fatality rates from COVID-19 are very low for young individuals but rise sharply with age, these demographic differences suggest much smaller populations of vulnerable individuals in the developing world. One can see these demographic differences starkly when looking at cross-country data on the median age. Figure 3.5 plots the median age against GDP per capita in a set of 158 countries using data from UN Population Division and Penn World Tables. Data from the UN Population Division show that countries in

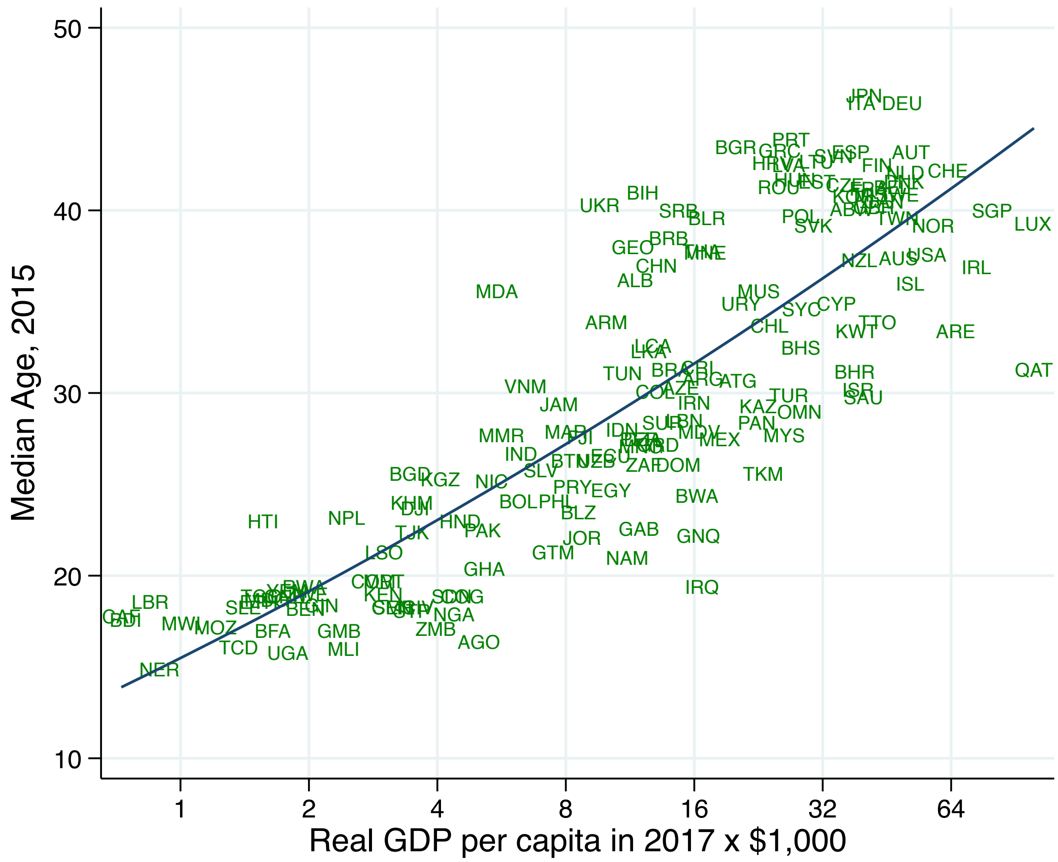


Figure 3.5: Median Age of the Population

Note: Median age data corresponds to 2015 and is from the UN Population Division. GDP per capita is expressed at PPP and taken from Penn World Table 9.1 (Feenstra et al., 2015).

the bottom quartile of the world income distribution have a median age of 19.1 years. Nigeria, Africa's most populous country, has a median age of 17.9, while countries like Angola and the Democratic Republic of the Congo have median ages of just 16.4 and 16.8 years old. By contrast richer countries like Italy, the United Kingdom and France have median ages of 45.9, 40.2 and 41.2, respectively.

Another statistic indicative of the much smaller vulnerable population in the developing world is the cross-country data on the population above 65. In the world's poorest countries the fraction of the population that is above age 65 is negligible, with an average of around 3 percent for countries in the bottom quartile of the world income distribution. The older population is much larger as a fraction of the total in richer economies, and reaches around one quarter of the population in Japan. Among countries in the topic quartile of the world, the average is about 15 percent of the population being above age 65 (see Appendix Figure 3.A.12).

It is hard to look at statistics like these and not see how different the impacts of COVID-19 will be in less developed countries. Concretely, while almost everything about COVID-19 suggests a more severe impact in less-developed countries, the far younger demographic is clearly in their favor.

2.4. Differences in Healthcare Capacity

Developing countries typically have substantially less ability to control disease than do richer countries. Sanitation and hygiene are more of an issue given the lack of widespread piped water and functioning sewage systems. Health infrastructure, especially hospital and health clinic capacity, is also less developed. For mild cases of COVID-19 infections, this may make little differences, as bed rest is likely to suffice in these mild cases. However, for critical cases, the lack of intensive-care capacity is a clear disadvantage for developing countries in their attempts to save lives during the pandemic.

Figure 3.A.13 plots the number of hospital beds per 10,000 people, as reported by the World Health Organization (WHO), against GDP per capita. The number of hospital beds is an imperfect measure of hospital capacity for many reasons, most importantly because it is not a bed per se that helps critical patients recover from COVID-19 but trained doctors, equipment like ventilators, and appropriate pharmaceuticals. Still, for lack of more comprehensive cross-country data, we take hospital beds as a proxy for medical care capacity.

By this metric there are stark differences in healthcare capacity across countries. Richer countries, which have quite some range amongst themselves, average around 49 hospital beds per 10,000 people. Countries like Japan and Korea have even more beds per capita, having 134 and 115 beds per 10,000 people, respectively. This is still far higher than the capacity in developing countries, which is a paltry 12 beds per 10,000 people on average in the bottom quartile of the income distribution. In Appendix Table 3.B.1, we report the availability of intensive care unit (ICU) beds and per capita healthcare costs across a limited set of countries. Consistent with the patterns observed from the number of hospital beds, it appears that low income countries possess significantly fewer ICU beds than high income countries.

2.5. Differences in Sectoral Composition of Employment

It is widely known that the sectoral composition of employment varies systematically with economic development. These differences are important because commercial disruptions brought on by COVID-19 and the resulting lockdowns differed substantially by occupation. Non-essential jobs that could not be performed remotely or while socially distancing experienced the largest and most sustained drops in employment throughout the recession; in contrast, occupations that were amenable to working from home experienced minimal disruption and some even flourished during the pandemic. In our model, we highlight two systematic differences in the composition of employment between advanced and developing economies which are relevant to the pandemic's macroeconomic outcomes across countries: the share of rural employment and the extent to which the urban workforce can work from home.

It is well known that the share of agricultural employment varies widely with economic development (see Figure 3.A.14). In the poorest countries, over 70 percent of the population is engaged in agricultural work on average, often subsistence farming on family plots; in advanced economies, that share is in the low single digits. The high agricultural share, while often considered a drag on economic modernization, offers a resilient source of income during pandemics. A good deal of agriculture in the developing world takes place on household-run farms, allowing it to continue during "stay-at-home" orders. Even in the absence of lockdowns, farming can often continue while socially distanced or with contact restricted to household members. Agricultural workers therefore do not face the same stark trade-offs in choosing between protecting their health or incomes since farming can often continue without substantially increasing the risk of infection.

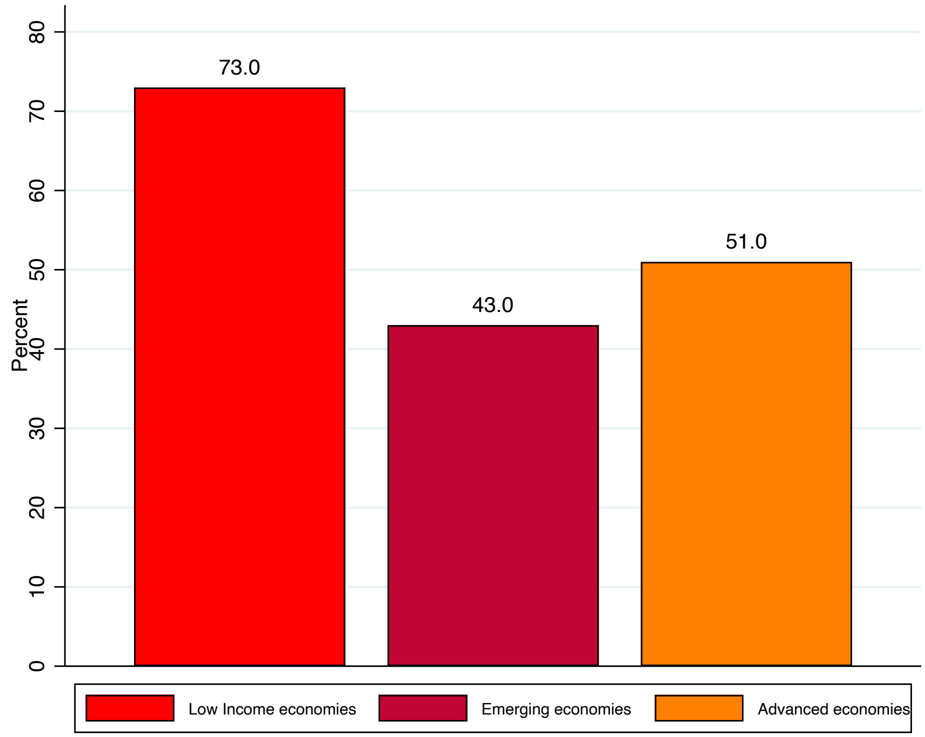


Figure 3.6: Non-Social Sector Employment Share

Note: The non-social sector includes rural employment and urban jobs that can be done from home, as estimated by Gottlieb et al. (2021*b*). See text for details. GDP per capita is expressed at PPP and is taken from the Penn World Table 9.1 (Feenstra et al., 2015).

Consequently, while agricultural workers may be vulnerable because of low wages, their employment is more resilient to large losses from lockdowns or voluntary self-isolation.

Outside of the agriculture sector, labor markets in lower income countries are characterized by widespread informality and employment concentrated in high-contact sectors. Large informal sectors will generally make economies more vulnerable to COVID-19 since, like agriculture, these jobs generally pay low wages while, unlike agriculture, most informal jobs cannot be performed from home or while socially distancing. To summarize these effects at the country level, we follow Kaplan et al. (2020) and aggregate employment into *social* and *non-social* sectors. Social sector workers have limited ability to work from home and suffer large income losses during lockdowns, while non-social sector workers can substitute more easily to remote work. We calculate the non-social sector share to include rural employment and all urban jobs that can be worked from home.

For the latter, we use the cross-country estimates of Gottlieb et al. (2021*a*) which are constructed using worker level data on the task-content of jobs in urban labor markets. Figure 3.6 displays the resulting estimates of non-social employment and illustrates that it varies substantially across countries. Emerging market economies have the lowest ability to work from home, with only 43 percent employed in non-social, low-contact jobs. In advanced economies, the non-social share is 60 percent, due to the greater number of high skill, professional jobs. However, the non-social share is largest in low-income countries, at 73 percent of aggregate employment, driven by the large agricultural labor force.

As a consequence of theses differences in the sectoral composition of employment, emerging market economies are more exposed to economic losses during the pandemic. Having less jobs that can be done from home or while socially distanced leads to greater economic losses during lockdowns and workplace closures. Moreover, in the absence of robust transfers, many social sector workers can become desperate and so voluntarily elect to continue working, rather than shelter at home, during times of peak infection. Such decisions will generally provide only marginal income gains, while amplifying the infection risk for the whole population through the public health externality. Large social sector employment can therefore be a liability for emerging market countries fighting COVID-19, as these workers are particularly vulnerable with limited options to avoid increasing their risk of becoming infected, or infecting others.

3. Model

Our analysis draws on a quantitative heterogenous-agent macroeconomic model with epidemiology as in the SICR model to analyze how policy responses to the COVID-19 pandemic should differ in developing countries. The model is equipped with several features that vary between advanced and developing economies that are relevant for the pandemic response, as motivated by the data presented in the previous section. These include uninsurable idiosyncratic health and income risks, age heterogeneity, fiscal capacity constraints, healthcare capacity, and availability to work from home across sectors. This section now presents these features in detail.

3.1. Households and Preferences

The economy is populated by a unit mass of heterogenous individuals who make consumption and savings decisions subject to idiosyncratic income and health risks. Individuals differ in their age $j \in \{\text{young adult}, \text{old adult}\}$ and permanent labor productivity $z \sim G$. Time is discrete and each period represents two weeks. Preferences are given by:

$$U = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta_j^t \left\{ \log(c_t) + \bar{u} \right\}\right], \tag{16}$$

where the discount factors β_j^t capture age heterogeneity in the population, and $\beta_{\text{young}} < \beta_{\text{old}}$. This specification follows the tractable formulation of Glover et al. (2020) that abstracts from explicitly modeling age, appealing to the logic that pandemics are sufficiently short-lived relative to entire lifetimes. It thus suffices to model only the expected number of years left to live, which is captured by the heterogeneity in discount factors. The term $\bar{u}$ represents the flow utility value of being alive, following the specification of Jones and Klenow (2016), and represents the reason that model households try to avoid fatality risk. Once an individual dies, they receive a fixed utility level that potentially depends on their individual characteristics, as we describe below.

There are two sectors, which we denote as social ($s = S$) and non-social ($s = N$). We assume that households are born with the sector they supply labor and cannot switch sectors. The social sector represents the workers with little availability of remote work. Examples of the occupations in the social sector includes waitresses, hair dressers, to name a few. The non-social sector represent the occupations that can be done with low level of social contacts. Such occupations include farmers in agricultural sector who can work while distancing from others, or college professors who can easily work remotely. Households in sector s supply labor to a representative firm where they can earn wage w_s per effective hour worked.

At the beginning of life, workers draw their permanent productivity, $z \sim G$. Incomes in both sectors are also subject to idiosyncratic productivity shocks as in Bewley (1977), Huggett (1993) and Aiyagari (1994). Specifically, we assume that individual labor productivity in each sector is composed of the sector-specific permanent component z and an idiosyncratic component v following the stochastic process:

$$\log v_{t+1} = \rho_v \log v_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim F(0, \sigma_v). \tag{17}$$

We include idiosyncratic income risk because developing countries are far from having full insurance, and so accounting for how people insure themselves in response to policies which may keep them away from work for prolonged periods of time is a first order consideration.

After observing their income realization, households make consumption and savings decisions given the interest rate, r, and subject to a no-borrowing condition, $a \geq 0$. Formally, the budget constraint of a household in sector s before the pandemic is given by:

$$U = \mathbb{E}\left[\sum_{t=0}^{\infty} \beta_j^t \left\{ \log(c_t) + \bar{u} \right\}\right], \tag{18}$$

where the discount factors β_j^t capture age heterogeneity in the population, and $\beta_{\text{young}} > \beta_{\text{old}}$. This specification follows the tractable formulation of Glover et al. (2020) that abstracts from explicitly modeling age, appealing to the logic that pandemics are sufficiently short-lived relative to entire lifetimes. It thus suffices to model only the expected number of years left to live, which is captured by the heterogeneity in discount factors. The term $\bar{u}$ represents the flow utility value of being alive, following the specification of Jones and Klenow (2016), and represents the reason that model households try to avoid fatality risk. Once an individual dies, they receive a fixed utility level that potentially depends on their individual characteristics, as we describe below.

There are two sectors, which we denote as social ($s = S$) and non-social ($s = N$). We assume that households are born with the sector they supply labor and cannot switch sectors. The social sector represents the workers with little availability of remote work. Examples of the occupations in the social sector includes waitresses, hair dressers, to name a few. The non-social sector represent the occupations that can be done with low level of social contacts. Such occupations include farmers in agricultural sector who can work while distancing from others, or college professors who can easily work remotely. Households in sector s supply labor to a representative firm where they can earn wage w_s per effective hour worked.

At the beginning of life, workers draw their permanent productivity, $z \sim G$. Incomes in both sectors are also subject to idiosyncratic productivity shocks as in Bewley (1977), Huggett (1993) and Aiyagari (1994). Specifically, we assume that individual labor productivity in each sector is composed of the sector-specific permanent component z and an idiosyncratic component v following the stochastic process:

$$\log v_{t+1} = \rho_v \log v_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim F(0, \sigma_v). \tag{19}$$

We include idiosyncratic income risk because developing countries are far from having full insurance, and so accounting for how people insure themselves in response to policies which may keep them away from work for prolonged periods of time is a first order consideration.

After observing their income realization, households make consumption and savings decisions given the interest rate, r, and subject to a no-borrowing condition, $a \geq 0$. Formally, the budget constraint of a household in sector s before the pandemic is given by:

$$c + a' \leq (1 - \tau) w_s z v n + (1 + r) a + T \tag{20}$$

where τ is the income tax rate and T is government transfers.

3.2. Aggregate Production Technology

The economy produces a single final good by combining capital with labor services supplied by the three sectors. The aggregate production technology is given by:

$$Y = AL^{\alpha}K^{1-\alpha},$$

where A is the total factor productivity and $0 < \alpha \leq 1$ is labor's share of value-added. We abstract from the domestic capital market. The aggregate capital stock is composed entirely of foreign sources, $K = K^F$, which can be rented at an exogenously given international rental rate r^F and which depreciates at rate δ. Aggregate labor depends on the total supply of labor services from the social and non-social sector,

$$L = L_S + L_N.$$

3.3. Credit and Capital Markets

Credit market incompleteness prevents households from borrowing against future earnings. As a result, individuals must maintain non-negative assets in formulating their consumption plans subject to (20), giving rise to hand-to-mouth consumers as well as a precautionary savings motive in response to idiosyncratic health and income risks. The precautionary motive is important for getting aggregate welfare measurements correct since it creates another feedback between the epidemiological and economic dynamics, as individuals withhold some consumption to increase precautionary savings in response to the pandemic's onset.

3.4. Public Health and Hospital Capacity

Households face idiosyncratic health risk which can reduce their labor productivity and increase the probability of dying. Susceptibility to infection is determined in part by economic decisions taken by households. Once infected, progression of the disease depends on an individual's age and the availability of public health infrastructure offering treatments.

Health risks are modeled using an SICR epidemiological model with five health states: susceptible ($\mathbb{S}$), infected ($\mathbb{I}$), critical ($\mathbb{C}$), recovered ($\mathbb{R}$), and deceased ($\mathbb{D}$). We denote by N_t^x the mass of individuals in each health state $x \in \{\mathbb{S}, \mathbb{I}, \mathbb{C}, \mathbb{R}, \mathbb{D}\}$ at time t and use $N_t = N_t^{\mathbb{S}} + N_t^{\mathbb{I}} + N_t^{\mathbb{C}} + N_t^{\mathbb{R}}$ to measure the non-deceased population. Figure 3.7 illustrates how these states evolve:

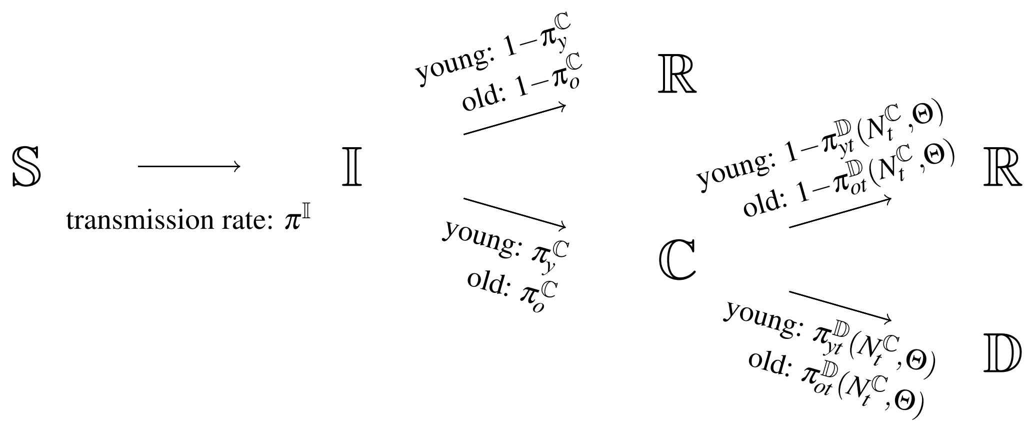


Figure 3.7: Dynamics of Health States and Transition Probabilities

The probability a susceptible person becomes infected is given as:

$$\pi_t^{\mathbb{I}} = \beta_t^{\mathbb{I}} \times \frac{N_t^{\mathbb{I}}}{N_t}$$

where $\beta_t^{\mathbb{I}}$ is the time-varying infection rate, reflecting the disease's natural progression (e.g. new variants), seasonal variations in infection rates, better medical treatments and other un-modeled factors that change infection rates over time.

Individuals who contract the virus experience a proportional drop in productivity of $1-\eta$ for one model period (two weeks), at which point they either recover or enter a critical health state. The probability of becoming critically ill depends on an individual's age and is given by π_j^C. Those in critical health are unable to work and require hospitalization. The likelihood of recovery in the hospital depends again on their age in addition to the availability of public health infrastructure, such as ICU beds and ventilators. In particular, the fatality rate of a critically ill patient of age j is given by:

$$\pi_{jt}^{\mathbb{D}}(N_t^{\mathbb{C}}, \Theta) = \begin{cases} \pi_j^{\mathbb{D}} & \text{if assigned ICU bed} \\ \kappa \times \pi_j^{\mathbb{D}} & \text{if not assigned} \end{cases}$$

where π_j^D is a baseline fatality rate for age j individuals in critical health and κ governs the impact on fatality rates of strained hospital resources. Whether or not a critically ill patient receives an ICU bed depends on overall hospital capacity and the number of other patients. Specifically, letting Θ denote hospital ICU capacity, the probability a new patient receives an ICU bed is given by $\min\{\Theta/N_t^C, 1\}$. In other words, all critically-ill patients receive an ICU bed if hospital capacity constraints are not binding, and beds are rationed amongst the critically-ill with probability Θ/N_t^C when constraints bind.

3.5. Voluntary Substitution Away From Workplace and Lockdowns

Voluntary Substitution While the disease's progression is exogenous, the probability a susceptible person becomes infected depends on endogenous economic decisions and the prevalence of infections in the population. To incorporate the feedback from economic behavior to infections, we allow individuals to lower the degree of exposure to the virus by voluntarily substituting away their labor supply to remote work. Specifically, we allow workers to choose between going to workplace and working remotely in each period. Remote work involves less social contacts, providing protection from being infected. Specifically, remote work lowers the probability of infection by ξ.

While it provides protection from being infected, working remotely is also less productive than going to the workplace. The productivity penalty of working remotely is parameterized by ϕ_s, where $s \in \{S, N\}$, by assuming that the effective labor supply of a worker in sector s can provide is given as $\phi_s n$, where $0 \leq \phi_s < 1$. We assume that $\phi_S < \phi_N < 1$, implying that the jobs in the non-social sector are more suited to be done remotely. Consequently, the probability a susceptible person becomes infected is given by:

$$\pi_t^{\mathbb{I}} = \begin{cases} \beta_t^{\mathbb{I}} \times N_t^{\mathbb{I}}/N_t & \text{if go to workplace} \\ \beta_t^{\mathbb{I}} \times N_t^{\mathbb{I}}/N_t \times \xi & \text{if work remotely} \end{cases}$$

Given the trade-offs between productivity penalties and lowered infection risks, individuals choose whether or not to go to workplace in each period. Specifically, the value for an individual at any period is:

$$V = \max\{V^w + \varepsilon_w, V^r + \varepsilon_r\}$$

where V^w and V^r each represents the value of going to the workplace and the value of working remotely. For each of the two options, we also introduce taste shock ε_w and ε_r, which are drawn from *i.i.d* Gumbel distribution with variance σ_g. The variance σ_g is calibrated to match the fraction of workforce already working remotely in the pre-pandemic steady state.

Lockdowns Infection rates can be further mitigated by containment policies, such as lockdowns. As in Kaplan et al. (2020), we model lockdowns as a certain fraction of workforce being chosen to work remotely through stay-at-home orders. Under a lockdown, households who would otherwise go to workplace hours are forced to switch to remote work. The stringency of lockdown varied across time and countries. Following Bick et al. (2020), we assume that 70 percent of the workers are forced to work at home under a full lockdown. Because remote work lowers the number of new infections, lockdowns mitigate the pandemic by exogenously decreasing the aggregate supply of workplace labor. We assume that lockdown policies are applied by group with the same intensity. For example, if the lockdown intensity if 70 percent in a period, then 70 percent of each group (young social, young non-social, old social, old non-social) are required to work remotely.

3.6. Vaccinations

Susceptible individuals can obtain immunity through vaccination as well. In each period, a susceptible individual draw a nonnegative probability of receiving vaccination. Once vaccinated, the individual obtains immunity and joins the recovered population. The exact probability of vaccination in each time period is taken from the actual path of vaccination in the United States. We will explain it in more details in the calibration section.

3.7. Government and Taxation

The government has power to tax, transfer, and impose economic lockdowns subject to the constraints imposed by limited fiscal capacity and labor market informality. We further require that the government run a balanced flow budget which satisfies,

$$B_t + \tau \int y(a,x,v) dQ = T$$

where $y(a,x,v)$ is pretax income for individual $(a,x,v) \sim Q$, τ is the prevailing tax rate, and T is aggregate transfers to households. In addition to tax revenue, we allow developing countries access to emergency bonds, B_t, which can be used to finance additional welfare transfers during government imposed lockdowns. The source of these funds is international donors and multinational institutions such as the IMF, World Bank, and World Health Organization. Funds borrowed for emergency transfers accrue interest at rate $1 + r^F$ until the pandemic ends, at which they are repaid through annual annuities. Formally, emergency transfers are given by:

$$B_t = \begin{cases} \bar{B} & \text{during the lockdown} \\ -\frac{r^F}{1+r^F} \times \sum_{t_l - t_s}^{t_l - t_e} \left(1 + r^F\right)^t \bar{B} & \text{after pandemic ends} \\ 0 & \text{otherwise} \end{cases}$$

where $\bar{B}$ is the size of per-period emergency transfers during lockdown, which we take parametrically, and t_s, t_e, and t_l index the lockdown's start, the lockdown's end, and the pandemic's end, respectively.

4. Quantitative Analysis

In this section, we discuss the calibration strategy, validate the model's fit, and present our counter-factual results. To evaluate the quantitative importance of each channel in explaining the cross-country variation in outcomes, we calibrate the model to match the U.S. economy and then vary key economic and demographic characteristics of the U.S. to match those of low-income and emerging economies. For each variation, we display the dynamic path of output and fatalities predicted by the model. To identify the most salient channels, we report the cumulative effects of each counterfactual on the U.S. economy compared to the calibrated benchmark.

4.1. Data Sources and Calibration

For expositional clarity, we divide the calibrated targets into three broad categories corresponding to those governing economic mechanisms, those controlling epidemiological dynamics, and those delineating differences between the advanced, emerging, and low-income countries.

Table 3.2: Calibration of Economic Parameters

Var	Description	Value	Source / Target
r^F	Exogenous interest rate	0.0006	Pre-COVID T-Bills rate 1.5%
ρ_v	Persistence of idiosyncratic income shock	0.91	Floden and Lindé (2001)
σ_v	St.Dev of idiosyncratic income shock	0.04	Floden and Lindé (2001)
α	Labor share	0.6	Gollin (2002)
β_y	Discount factor for the young	0.9984	Glover et al. (2020)
β_o	Discount factor for the old	0.9960	Glover et al. (2020))
σ_g	Variance of remote / non-remote work taste shock	0.0101	Pre-COVID Remote Workers 8.2%
ϕ_n	Productivity remote work, non-social sector	1	Barrero et al. (2021)
ϕ_s	Productivity remote work, social sector	0.62	COVID-19 Employment Declines - 6.4%
$A(P)$	Pandemic Total Factor Productivity	1.042	COVID-19 Output Declines -4.1%

Table 3.3: Calibration of Epidemiological Parameters

Var	Description	Value	Source or Target
η	Effect of infection on productivity	0.3	Alene et al. (2021)
ξ	Reduction of infection probability by working from home	0.6	Mossong et al. (2008)
κ	Impact of hospital overuse on fatality	2	Glover et al. (2020)
$\pi_y^{\mathbb{C}}$	Rate of young entering $\mathbb{C}$ from $\mathbb{I}$	6.7%	Ferguson et al. (2020)
$\pi_o^{\mathbb{C}}$	Rate of old entering $\mathbb{C}$ from $\mathbb{I}$	38.0%	Ferguson et al. (2020)
$\pi_y^{\mathbb{D}}$	Rate of young entering $\mathbb{D}$ from $\mathbb{C}$	2.7%	Glynn (2020)
$\pi_o^{\mathbb{D}}$	Rate of old entering $\mathbb{D}$ from $\mathbb{C}$	9.0%	Glynn (2020)

Table 3.2 reports the parameters that govern the core economic dynamics of the model. Population demographics are modeled using age dependent discount factors accounting for differences in the remaining years of life for young and old workers. The age specific discount factors are taken from Glover et al. (2020), and the stochastic income processes are taken from Floden and Lindé (2001), who estimate similar income processes in the United States and Sweden. The taste-shock for remote work σ_g is chosen so that 8.2 percent of the pre-pandemic laborforce works remotely, consistent with the estimates in Bick et al. (2020). Finally, labor's share of income comes from Gollin (2002), and the rental rate of capital is set to the two-week return on pre-COVID Treasury Bills. We set the productivity penalty for remote work in the nonsocial sector, ϕ_n, to unity, consistent with evidence of small productivity losses for these workers in most cases, and potentially even productivity gains in some cases (Barrero et al., 2021). Finally, the penalty for remote work in the social sector, ϕ_s, and the TFP shock accompanying the pandemic $A(P)$, are jointly calibrated to match aggregate 2019-2020 year-on-year employment and output declines in the United States.[13]

Table 3.3 reports parameters controlling the epidemiological transmission of disease and their interactions with public health infrastructure and lockdown policies. We take parameters governing the fatality infection rates from Glynn (2020) and the rates of infected cases becoming

[13] Appendix Table 3.B.2 summarizes the internally calibrated parameters and the model's fit to the data. Note that TFP in normal times, $A(N)$ is set to one, so that $A(P)$ should be interpreted as a relative TFP shock in effect during the Pandemic.

critical from Ferguson et al. (2020). The effect of hospital congestion on disease fatality rates, κ, is taken from Glover et al. (2020). The productivity penalty of becoming infected, η, is set to match a 30 percent share of asymptomatic infection cases, as estimate in the meta-analysis of Alene et al. (2021). Such a choice is motivated by the observation that those known to be infected cannot work, and so have productivity of zero, while those who are infected but asymptomatic may continue to work unhindered. Finally, we choose the time-varying behavioral-adjusted infection probability, $\beta_t^{\mathbb{I}}$, so that the model's endogenous path of fatalities precisely matches the experience of the United States. The simulated endogenous path of the virus also account the time path of vaccinations and lockdowns in the U.S.. Vaccination data is taken from the COVID-19 Data Repository by CSSE at John Hopkins University, and we assume vaccination rates continue to grow at 1% per period after the last available data point, until period 60. The time path of lockdown policies comes from the Oxford Coronavirus Government Response Tracker (see Appendix Figure 3.A.8). We assume lockdown policies are gradually lifted starting in the last period of available data until they are completely discontinued by period 60. Figure 3.8 plots the fitted results and validates the model's ability to replicate these dynamics exactly.

The mortality dynamics (and output losses) are shaped by both government lockdown policies and voluntary household substitution away from market work and consumption. Figure 3.9 illustrates that both margins play an important role in the calibrated model by plotting the equilibrium population share under lockdowns or voluntarily sheltering from home. The dashed purple line represents the strictness of prevailing lockdowns, reporting the share of the susceptible population forced to stay home. Any mass above this "Lockdown" curve represents voluntary substitution to working from home. The figure shows that there is a considerable amount of voluntary sheltering at home, above and beyond what is required by lockdowns, especially during times of peak infection (such as in Winter of 2020-2021). Voluntary substitution to working from home also varies substantially across age and sector of employment. Consistent with differences in health risks and economic costs, we see more voluntary working from home among the older population and in the social sector, where the health risks of the pandemic are most acute. The result also

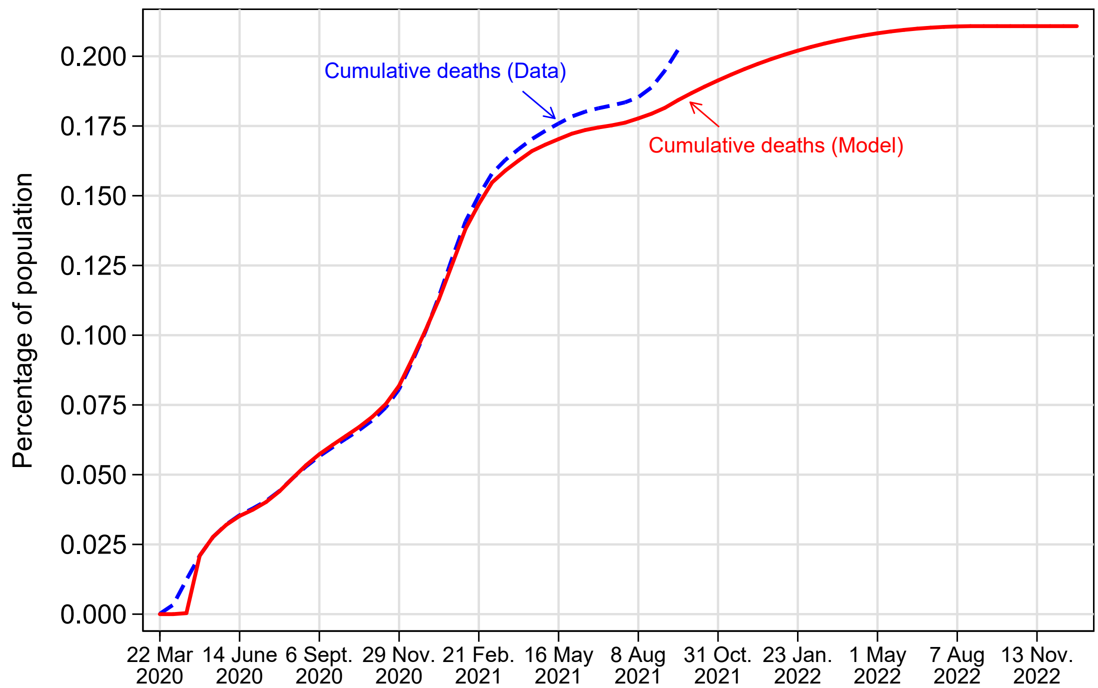


Figure 3.8: Predicted and Actual COVID-19 Mortality in the United States

Note: Time path of U.S. COVID-19 mortality taken from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at John Hopkins University.

highlights the importance of accounting for heterogeneity in age and the sectoral composition of work alongside government policies in evaluating output and mortality dynamics throughout the pandemic.

Table 3.4 summarizes parameters which vary across advanced and developing countries. The tax rates for the advanced and developing countries are taken from Besley and Persson (2013). Age demographics ω_y come from the World Bank and measure the share of the population under 65. The youth share in advanced economies corresponds to the U.S. economy, as it is our benchmark calibration, and we set the shares for emerging and low-income countries to their group averages. The share of workers in the social sector, ω_s, is constructed using estimates from Gottlieb et al. (2021*b*) on the share of urban labor that can work from home and adjusting the ratio to account for the rural population. Specifically, we take the shares of urban and rural labor from the UN Population Division and assuming the entire rural sector is non-social, calculate the ω_s as the

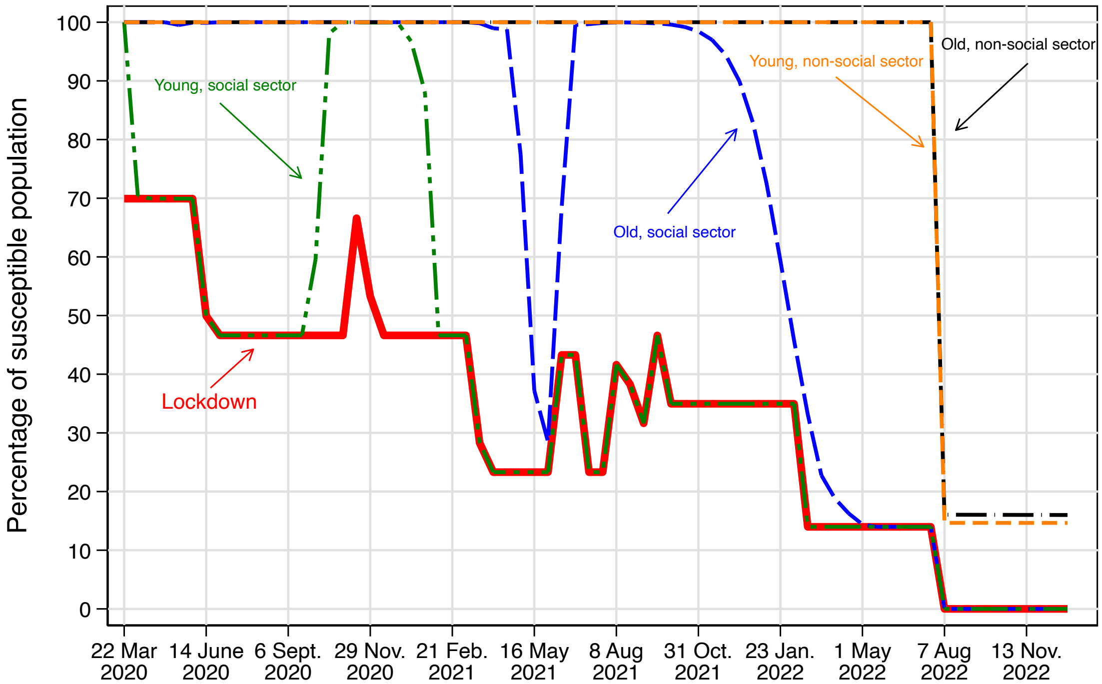


Figure 3.9: Lockdowns and Voluntary Working from Home

weighted average of the urban and rural populations.

The flow value of life, $\bar{u}$, is calibrated using the value of statistical life (VSL) approach. Following Glover et al. (2020), we set the per-period statistical value of life to \$515,000 for advanced economies, equal to 11.4 times average US consumption. The value for $\bar{u}$ is then computed so that the behavioral response to a marginal increase in the risk of death is consistent with the VSL. Specifically, we get $\bar{u}$ by solving,

$$\text{VSL} = \frac{dc}{d\rho}|_{E(u)=k,\rho=0} = ln(\bar{c}) - \bar{u}$$

where ρ is the risk of death and $\bar{c}$ is average consumption. Absent better evidence, we assume the VSL has unitary income elasticity and adjust $\bar{u}$ for developing countries accordingly.

The final cross-country parameter to be set govern the ICU hospital capacity in developing and developed countries. One challenge is that while many countries report hospital bed capacity,

Table 3.4: Calibration of Parameters Varying Across Advanced and Developing Economies

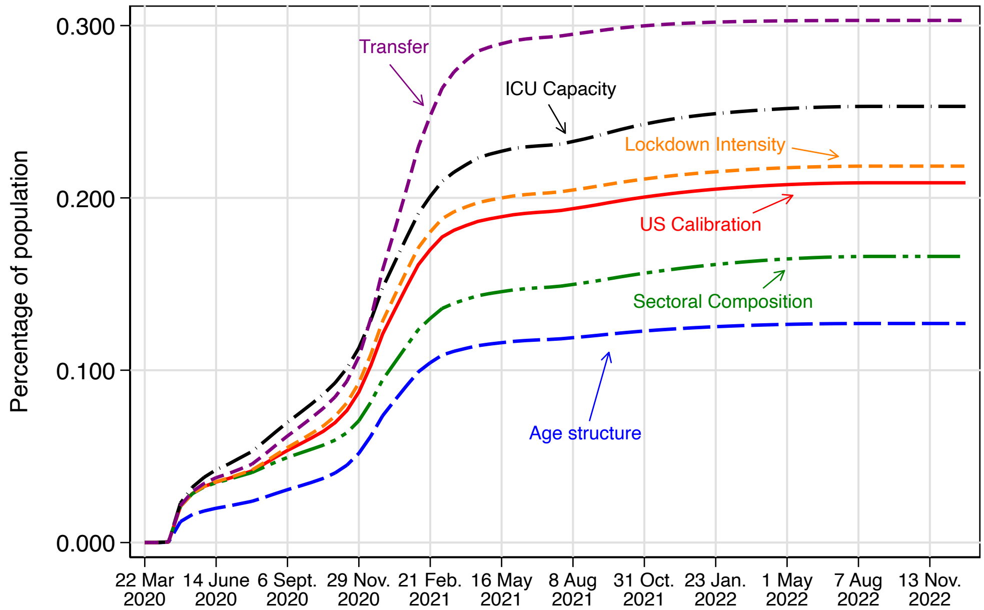
Var	Description	Advanced Economies	Emerging Economies	Low-Income Economies	Source or Target
$\bar{u}$	Flow value of being alive	$11.4\bar{c}^{US}$	$11.4\bar{c}^{MID}$	$11.4\bar{c}^{DEV}$	Glover et al. (2020)
τ	Marginal tax rate	0.25	0.20	0.15	Besley and Persson (2013)
ω_y	Share of young in population	83%	84%	92%	UN Population Division
ω_s	Share of social sector workforce	40%	57%	27%	Gottlieb et al. (2021*b*)/IPUMS
Θ	Hospital capacity per capita	0.00042	0.00025	0.00011	Glover et al. (2020) / WHO

few developing countries distinguish explicitly between general hospital capacity and ICU capacity in the data. To address this, we assume the ratio of hospital beds to ICU beds is constant across countries, and calibrate Θ by adjusting WHO data on the availability of hospital beds in the top and bottom quartiles of country income levels (as in Figure 3.A.13) by the ratio of hospital beds to ICU beds taken from Glover et al. (2020).

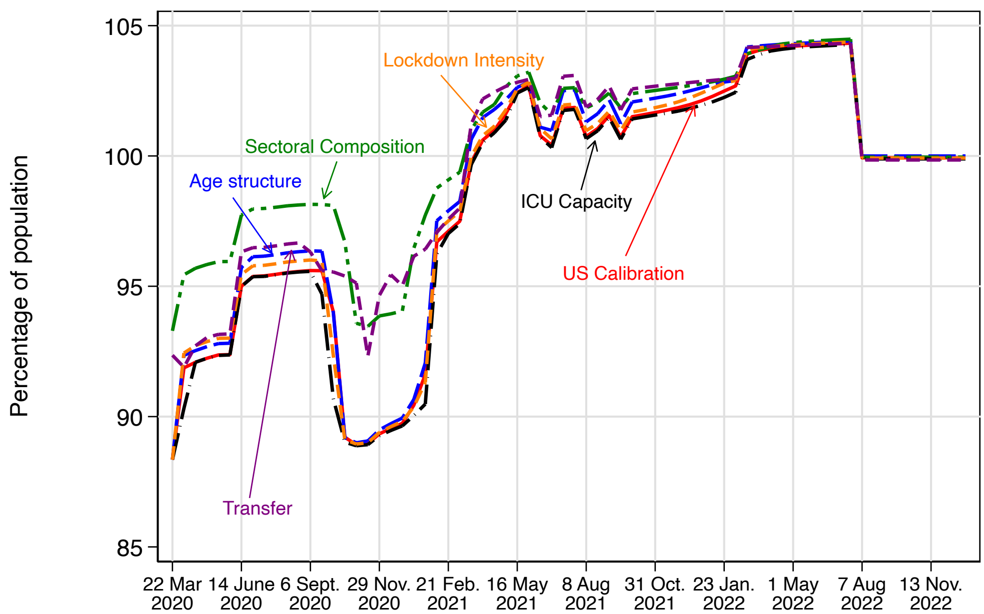
4.2. Economic and Demographic Sources of Cross-Country Differences

Figures 3.10a and 3.11a plot the dynamic path of GDP per capita and fatalities as a percentage of population during the COVID-19 pandemic in the United States in each of our counterfactual simulations. The panels on top display cumulative fatalities and those on the bottom plot GDP. Each figure provides six simulated paths: the benchmark U.S. calibration and the five counterfactual exercises which vary demographics, the sectoral composition of employment, public healthcare capacity, government transfer programs, and the stringency of lockdowns. Figure 3.10a reports counterfactuals that endow the U.S. economy with the characteristics of low-income countries; Figure 3.11a reports the results of endowing the U.S. with emerging market economy characteristics.

Looking across the panels, one can see that all five mechanisms play an important role to some degree, but differences in age demographics, the sectoral composition of employment, and the size of public transfer programs are the most quantitatively prominent. In both low income and emerging market economies, low levels of public financial assistance during the pandemic lead to much higher levels of fatalities. Without transfers to support or replace income lost in the pandemic, many households are not able to shelter at home during times of peak infection and instead must work outside the home, further propagating thee spread of disease which increases fatalities. Quantitatively, low levels of public transfers are the largest factor pushing fatalities higher outside advanced economies. The simulations show that if the United States scaled back its public transfer programs to the levels of low income and emerging market economies, that alone would lead to cumulative fatalities from the pandemic to grow by 50 percent.

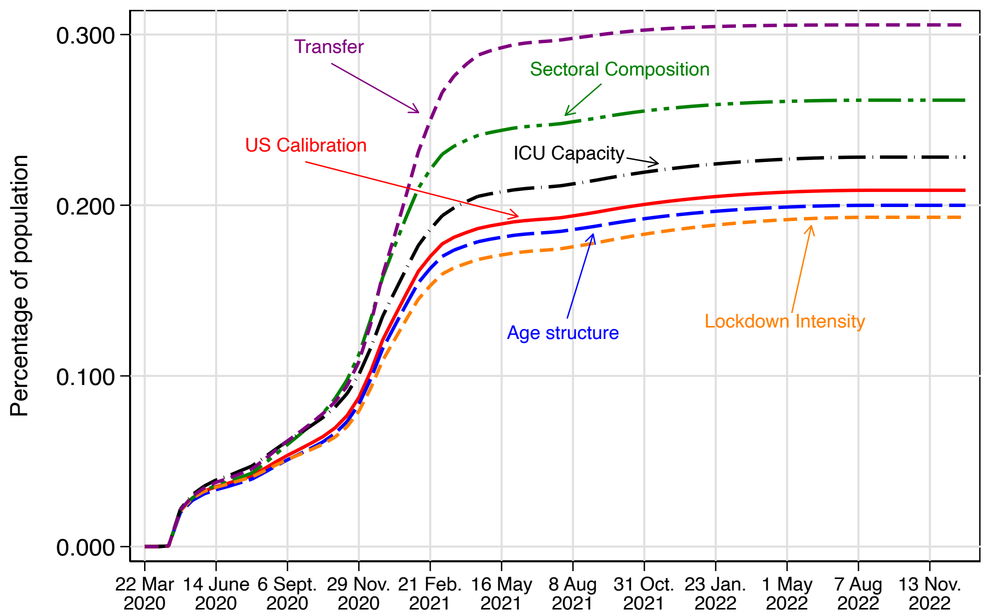


(a) Cumulative Death, US with Low Income Economies' Features

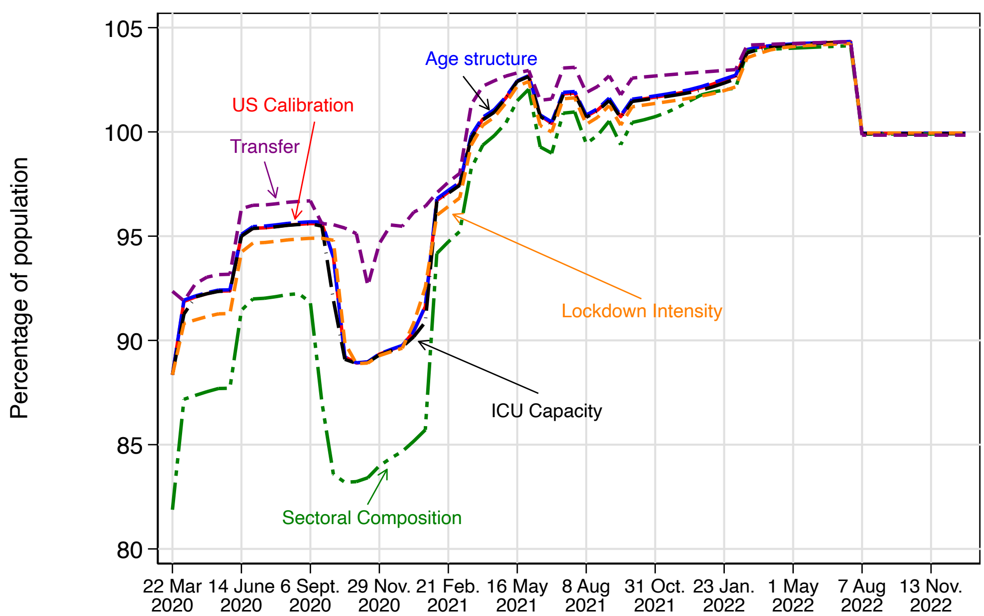


(b) GDP per capita, US with Low Income Economies' Features

Figure 3.10: Time Path of Cumulative Deaths and GDP: Low Income Economies



(a) Cumulative Death, US with Emerging Economies' Features



(b) GDP per capita, US with Emerging Economies' Features

Figure 3.11: Time Path of Cumulative Deaths and GDP: Emerging Economies

Despite the effect of low transfers, fatalities in low income countries remained modest because of the offsetting effect of its substantially younger population. The high agricultural employment share in low-income countries also substantially reduces fatalities. In contrast, emerging markets experienced far higher mortality rates because they do not benefit from the favorable demographics of low income countries and have a high social sector employment share, making it difficult to control the spread of disease while working. ICU constraints exacerbated fatalities in both low and emerging economies, though played a secondary role overall. Differences in lockdown intensity play the smallest role, suggesting the more important cross-country policy difference during the pandemic was in the size of public insurance programs.

The output counterfactuals exhibit less variation than what we see in fatalities, suggesting the mechanisms we study contribute more equally to observed economic declines. Among the channels, only the sectoral composition of employment and public transfers stand out as having an especially important quantitative role. In low-income countries, economic losses were moderated by a large agricultural sector that was minimally disrupted by lockdowns and social distancing requirements. In emerging markets, high levels of urban employment in jobs that cannot be done from home explains a substantial part of their larger economic losses. Somewhat perversely, the low levels of public transfers which amplify fatalities in low income and emerging market economies also serve to *reduce* output losses by causing financially vulnerable households to continue working outside the home. The impact is most pronounced at times of peak infection, as is visible in the transition paths during the winter months of 2020.

To assess what may be driving the especially bad outcomes observed in emerging markets, Table 3.5 reports the cumulative effect of our counterfactuals on 2019-2020 year-on-year changes in GDP and fatalities. For comparison, the first data column displays the data for advanced and emerging economies discussed in the introductory sections (see Appendix Figures 3.1 and 3.2). The second data column reports the simulation outcomes when all features are allowed to vary (i.e. demographics, sectoral employment, ICU capacity, and lockdown policies). The entry for advanced economies corresponds to our benchmark calibration to the United States data; the entry for emerging economies corresponds to the simulation which endows the United States with

Table 3.5: Cumulative Effect of the COVID-19 Pandemic in Emerging Markets

Panel (a): GDP Changes from 2019 to 2020			
	Data	Model	
		All Features	Age/Sector/ICU
Advanced Economies	-4.6	-4.7	-4.7
Emerging Economies	-6.7	-6.9	-8.3
Ratio	1.47	1.46	1.77
Panel (b): Excess Mortality			
	Data	Model	
		All Features	Age/Sector/ICU
Advanced Economies	64	208	208
Emerging Economies	113	367	265
Ratio	1.76	1.76	1.27

all the features of emerging economies. The third column reports results when we endow the United States with only the age demographics, sectoral employment, and ICU capacity of emerging economies. We distinguish these features since we view them as largely immutable throughout the pandemic's duration. To facilitate comparisons, the final row of each column reports the ratio of outcomes in emerging markets relative to advanced economies.

In panel (a) we see that the model does relatively well at replicating variation in GDP. In the data, GDP in advanced economies contracted by -4.6 percent while emerging economies shrank by -6.7 percent. The benchmark model nearly replicates these data, predicting GDP declines of -4.7 percent and -6.9 percent in advanced and emerging economies, respectively. While the model predicts contractions in GDP that are slightly larger than in the data, it accurately replicates the relative severity of the pandemic across countries. In the model, emerging markets experience contractions in GDP that are 46 percent larger, in line with the data, at 47 percent larger.

Panel (b) reports excess mortality per hundred thousand people in advanced economies and emerging markets, both in the data and full counterfactual. The model substantially over-predicts the total fatality rate since the benchmark advanced economy calibration is set to match the United States, which has been an outlier in terms of reported COVID-19 mortality amongst advanced economies. The level discrepancy suggests there exist other important public health differences even within country income groups – such as the prevalence of mask wearing and co-morbidities – which are missing from our model but may be important. Nevertheless, the model once again does a good job at replicating the relative severity of the pandemic in emerging markets. Endowing the United States with all the features of an emerging market economy leads to a 76 percent rise in excess mortality. In the data, emerging market economies registered excess mortality that was on average 76 percent greater than advanced economies. The model can therefore completely account for the relatively higher number of fatalities in emerging markets during the pandemic.

Finally, in light of the large differences in emerging economies, it is natural to ask if there is anything emerging market economies could have done differently to improve their outcomes. While we do not model the optimality of different policies, our framework allows us to study the extent to which outcome differences depend on features that are outside the control of policymakers throughout the pandemic's duration. In particular, we view a country's age demographics, sectoral composition of employment, and healthcare capacity to be largely fixed throughout the pandemic. That governments cannot choose the age of their population is obvious. Similarly, it's generally widely held that the industrial composition of the economy is rigid in the short-run. While public healthcare capacity can in principle be expanded (and was, rather rapidly in a few places like China), we believe that emerging market economies by and large only had limited ability to do so during the pandemic, especially given the concurrent global competition for medical equipment, oxygen, and protective gear.

The final column of Table 3.5 reports the cumulative counterfactual impact on output and fatalities if only these immutable characteristics varied between emerging markets and advanced economies. For output, these characteristics alone predict a -8.3 percent decline in GDP for emerging markets, over-accounting for the -6.7 percent decline observed in the data. The simulation suggests that the especially large GDP declines in emerging markets may have been largely outside the control of policymakers, depending instead on prevailing demographic and structural conditions that cannot be easily changed. In fact, comparing the full (column 2) and restricted (columns 3) counterfactuals in panel (a) shows that the more stringent lockdowns and public transfers policies in emerging markets actually reduced cumulative GDP losses from the pandemic by 1.4 percentage points.

For mortality, these fixed features lead to a 27 percent rise in fatalities, explaining about one-third of the 76 percent higher excess mortality in emerging markets during the pandemic. Much of the remaining two-thirds is accounted for by the lower level of social insurance in emerging markets. Limited public financial assistance results in many lower-income households continuing to outside the home during the pandemic, propagating infections that lead to higher mortality. The result suggests that constraints on the ability of emerging markets to support large scale public transfer programs, such as limited fiscal capacity and borrowing costs, were an important determinant of the greater excess mortality they experienced during the pandemic. Likewise, the result suggests that expansions in international emergency financial assistance during the pandemic, including debt relief and lending programs such as the IMF's Rapid Credit Facility (RCF) and Rapid Financing Instrument (RFI), contributed meaningfully to reducing mortality globally and particularly so in lower income countries.

Table 3.6 reports the cumulative counterfactual effects of the COVID-19 pandemic in low income countries. The model correctly predicts the more modest GDP declines and mortality rates in low income economies relative to advanced ones, as in the data, albeit with larger differences. In the counterfactual, low income countries experience GDP declines that are 25 percent the size of contractions in advanced economies, while the data show declines that were 78 percent the size of advanced economy losses. The result suggests that there may be other important factors driving GDP declines in low income countries that are not present in our current model, such as foreign demand shocks, supply chain disruptions, constraints on sovereign debt, and limited fiscal capacity. The model also predicts mortality rates in low income countries that are 80 percent the size of mortality rates in advanced economies, primarily due to their younger age demographic

Table 3.6: Cumulative Effect of the COVID-19 Pandemic in Low Income Economies

Panel (a): GDP Changes from 2019 to 2020			
	Data	Model	
		All Features	Age/Sector/ICU
Advanced Economies	-4.6	-4.7	-4.7
Low Income Economies	-3.6	-1.2	-2.2
Ratio	0.78	0.25	0.47

Panel (b): Excess Mortality			
	Data	Model	
		All Features	Age/Sector/ICU
Advanced Economies	64	209	209
Low Income Economies	-	167	128
Ratio	-	0.80	0.61

and higher agriculture employment share. While systematic data on COVID-19 fatalities in low income countries is not yet available, the result is consistent with the limited evidence on excess deaths available for some countries in Africa (see Appendix Figures 3.A.4 and 3.A.5).

As with emerging markets, we can assess the extent to which outcomes were driven by policy choices or fixed short-run characteristics of low income countries by comparing the cumulative counterfactuals in the last two columns of Table 3.6. Endowing advanced economies with only the immutable characteristics of low income countries generates a GDP declines of -2.2 percent, nearly double the benchmark level in the full counterfactual. For mortality, these fixed factors alone lead fatalities to fall by nearly quarter, declining from 80 percent to 61 percent of the mortality level in advanced economies. Taken together, the results suggest that younger populations and high agricultural employment shares predisposed low-income countries to have fewer fatalities from COVID-19, but public lockdowns and transfer policies played an important role in moderating the economic fallout accompanying the pandemic.

5. Empirical Correlates of GDP Declines During the Pandemic

In this section we explore the empirical correlates of changes in GDP per capita from 2019 to 2020, focusing on the same variables emphasized in the model. We make no claim at uncovering causal patterns in this section. Instead, we assess the extent to which correlations between aggregate income changes during the pandemic and a country's demographic, economic, and policy characteristics are consistent with the model's predictions and quantitative exercises.

Table 3.7: Correlates of GDP per Capita Change from 2019 to 2020

	Dependent variable: GDP per capita change from 2019 to 2020					
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
GDP per capita in 2019	-0.10**	0.037	-0.17*	-0.076*	-0.11	-0.052
	(0.046)	(0.068)	(0.094)	(0.044)	(0.068)	(0.11)
GDP per capita in 2019^2	0.0014**	0.00021	0.0020*	0.0011*	0.0014*	0.00084
	(0.00066)	(0.00071)	(0.0010)	(0.00063)	(0.00080)	(0.0011)
Agriculture emp. share		0.076***				0.062**
		(0.027)				(0.030)
Median age			0.083			0.074
			(0.079)			(0.082)
Lockdown stringency				-0.13***		-0.13**
				(0.043)		(0.053)
Economic support					0.0042	0.024
					(0.036)	(0.038)
Constant	-4.21***	-8.03***	-5.67***	2.38	-4.29***	-2.97
	(0.60)	(1.66)	(1.48)	(2.07)	(1.09)	(3.34)
Observations	144	144	144	140	140	140
R^2	0.031	0.071	0.037	0.129	0.030	0.163

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

We begin with the basic relationship between declines in GDP per capita and pre-pandemic level of GDP per-capita. The first column of Table 3.7 shows that this relationship is U-shaped, as we argued earlier. Both the level and quadratic coefficients on GDP per capita in 2019 are statistically significant at the five-percent level, with the former negative and latter positive. The second column includes controls for the agricultural employment share. The variable exhibits a significant positive correlation with changes in GDP, holding constant differences in national income, means that countries with larger percentages of their workforce in agriculture also experienced smaller declines in national income, all else equal. Interesting, the coefficients on GDP per capita and its square are now statistically indistinguishable from zero, with the former switching signs. The third column includes median age as a control which exhibits no significant correlation, somewhat puzzlingly. The fourth column controls for the stringency of lockdowns, which is positive and statistically significant. The fifth column adds controls for the generosity of economic support programs during the pandemic, which turns out to be statistically insignificant.

Column six of Table 3.7 adds all the covariates at once. This specification shows that agriculture's employment share remains a strong positive correlate of GDP changes, while lockdown stringency remains a strong negative correlate. Median age and the economic support index continue to be insignificant. Collectively, the inclusion of these controls eliminates the statistical significance of the original U-shape pattern in GDP per capita, and substantially reduce the magnitude of the original correlations. We take this as suggestive evidence that these variables are important empirical determinants of macroeconomic performance across the world income distribution, at least thus far, during the pandemic.

6. Conclusion

The macroeconomic impact of the COVID-19 pandemic was most severe in emerging market economies, which represent the middle of the world income distribution. This paper provides a quantitative economic theory to explain why these economies fared so poorly compared to both poorer and wealthier nations. Our model is motivated by key economic and demographic differences across the world income distribution, including variation in lockdown policies, public insurance, demographics, healthcare capacity, and the sectoral composition of employment.

Our quantitative model does well in predicting the greater output declines and higher mortality rates in emerging market economies, as in the data. The model also quantitatively predicts the more modest output losses and fatalities in low income countries, albeit to a greater extent than what is observed in the data. Among the factors we consider, the size of public transfer programs, age demographics, and the sectoral composition of employment are the most quantitatively important. Low levels of public financial assistance and a high share of jobs which require social interaction explains most of the greater GDP losses and higher fatalities in emerging markets. Low income countries also suffered from low levels of public transfers, but the negative consequences were largely blunted by their substantially younger populations, whom are naturally more resistant to illness, and large agricultural employment share, which provide a resilient source of income during lockdowns and while socially distancing. Quantitatively, the results suggest that cross-country differences are mostly driven by variation in public transfer programs and factors outside the short-term control of government officials. The out-sized role of public transfer programs in explaining cross-country differences highlights the importance of constraints which may limit the ability of governments to enact and sustain large scale social insurance programs during emergencies. A valuable avenue for future research is to better identify the sources of these policy differences and what impact they had on macroeconomic outcomes following the pandemic.

Overall, our findings suggest that much of the variation in aggregate outcomes across country income groups during the pandemic can be attributed to a small set of economic characteristics and broad policy choices. The model is stylized in many ways, however, and does not attempt to analyze the many more granular policy choices that surely mattered for the first year of the pandemic. Absent from this study are policy decisions regarding school closings (e.g. Fuchs-Schündeln, Krueger, Ludwig and Popova, 2020), mask use (e.g. Abaluck et al., 2021; Karaivanov et al., 2021), testing and tracing policies (e.g. Berger et al., 2020), and vaccine provision (e.g. Arellano, Bai and Mihalache, 2021). Future research could also fruitfully assess the quantitative importance of other policy choices for cross-country macroeconomic performance during the pandemic.

Another key limitation of our analysis is that it relies on a large exogenous time-varying component of the infection rate in order to match the observed path of excess deaths in the United States. In reality, however, much of the time variation in infection probabilities is likely due public policy choices that are not modeled here. These include policies that increase the prevalence of mask wearing, the development of better treatments for the infected, the rate of vaccination, or general knowledge about how COVID-19 can and cannot be transmitted. Future research should more explicitly consider the role these factors play in determining cross-country differences in the pandemic's consequences.

Chapter 3, in full, has been submitted for publication of the material as it may appear in IMF Economic Review, 2022, Alon, Titan; Kim, Minki; Lagakos, David; VanVuren, Mitchell, Palgrave Macmillan, 2022. The dissertation author was a primary investigator and author of this paper.

Abaluck, J., Kwong, L. H., Styczynski, A., Haque, A., Kabir, M. A., Bates-Jeffries, E., Crawford, E., Benjamin-Chung, J., Raihan, S., Rahman, S., Benhachmi, S., Zaman, N., Winch, P. J., Hossain, M. M., Reza, H. M., Jaber, A. A., Rahman, S. G. M. F. L. B. A., Huq, T. S., Luby, S. P. and Mobarak, A. M. (2021), The impact of community masking on covid-19: A cluster randomized trial in bangladesh. Unpublished Working Paper, Yale University.

Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M. D. (2020), A multi-risk sir model with optimally targeted lockdown. NBER Working Paper No. 27102.

Aiyagari, S. R. (1994), 'Uninsured Idiosyncratic Risk and Aggregate Saving', *The Quarterly Journal of Economics* **109**(3), 659–684.
URL: *https://doi.org/10.2307/2118417*

Aktay, A., Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M., Kamath, C., Kansal, M., Lange, A., Mandayam, C., Oplinger, A., Pluntke, C., Roessler, T., Schlosberg, A., Shekel, T., Vispute, S., Vu, M., Wellenius, G., Williams, B. and Wilson, R. J. (2020), 'Google covid-19 community mobility reports: Anonymization process description (version 1.0)', *arXiv preprint arXiv:2004.04145* .

Alene, M., Yismaw, L., Assemie, M. A., Ketema, D. B., Mengist, B., Kassie, B. and Birhan, T. Y. (2021), 'Magnitude of asymptomatic covid-19 cases throughout the course of infection: A systematic review and meta-analysis', *PloS One* **16**(3), e0249090.

Alfaro, L., Becerra, O. and Eslava, M. (2020), Emes and covid-19: Shutting down in a world of informal and tiny firms. NBER Working Paper No. 27360.

Alon, T. W., Kim, M., Lagakos, D. and VanVuren, M. (2020), How should policy responses to the covid-19 pandemic differ in the developing world? NBER Working Paper No. 27273.

Alvarez, F. E., Argente, D. and Lippi, F. (2020), A simple planning problem for covid-19 lockdown. NBER Working Paper No. 26981.

Arellano, C., Bai, Y. and Mihalache, G. P. (2020), Deadly debt crises: Covid-19 in emerging markets. NBER Working Paper No. 27275.

Arellano, C., Bai, Y. and Mihalache, G. P. (2021), Covid-19 vaccination in emerging markets. Unpublished Working Paper, University of Rochester.

Bairoliya, N. and Imrohoroglu, A. (2020), 'Macroeconomic consequences of stay-at-home policies during the covid-19 pandemic', *COVID Economics Vetted and Real-Time Papers* **13**, 71–90.

Barrero, J. M., Bloom, N. and Davis, S. J. (2021), Why working from home will stick, Working Paper 28731, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w28731*

Berger, D. W., Herkenhoff, K. F. and Mongey, S. (2020), An seir infectious disease model with testing and conditional quarantine. NBER Working Paper No. 2690.

Besley, T. and Persson, T. (2013), Taxation and development, *in* 'Handbook of Public Economics', Vol. 5, Elsevier, pp. 51–110.

Bewley, T. (1977), 'The permanent income hypothesis: A theoretical formulation', *Journal of Economic Theory* **16**(2), 252 – 292.
URL: *http://www.sciencedirect.com/science/article/pii/0022053177900096*

Bick, A., Blandin, A. and Mertens, K. (2020), Work from home after the covid-19 outbreak. Federal Reserve Bank of Dallas Working Paper 2017.

Bonadio, B., Huo, Z., Levchenko, A. A. and Pandalai-Nayar, N. (2021), 'Global supply chains in the pandemic', *Journal of International Economics* **133**(3).

Brotherhood, L., Kircher, P., Santos, C. and Tertilt, M. (2021), An economic model of the covid-19 pandemic with young and old agents: Behavior, testing and policies. Unpublished Working Paper, University of Mannheim.

Cakmakli, C., Demiralp, S. and Ozcan, S. K. (2020), Covid-19 and emerging markets: An epidemiological model with international production networks and capital flows. IMF Working Paper No. 20/133.

Chopra, A., Devereux, M. B. and Lahiri, A. (2021), 'Pandemics through the lens of occupations', *Canadian Journal of Economics* **54**.

Dahal, S., Banda, J. M., Bento, A. I., Mizumoto, K. and Chowell, G. (2021), 'Characterizing all-cause excess mortality patterns during covid-19 pandemic in mexico', *BMC Infectious Diseases* **21**.

Deshmukh, Y., Suraweera, W., Tumbe, C., Bhowmick, A., Sharma, S., Novosad, P., Fu, S. H., Newcombe, L., Gelband, H., Brown, P. and Jha, P. (2021), Excess mortality in india from june 2020 to june 2021 during the covid pandemic: Death registration, health facility deaths, and survey data. Unpublished Working Paper, University of Toronto.

Djankov, S. and Panizza, U., eds (2020), *COVID-19 in Developing Economies*, Centre for Economic Policy Research Press, London, UK.

Feenstra, R. C., Inklaar, R. and Timmer, M. P. (2015), 'The next generation of the penn world table', *American Economic Review* **105**(10), 3150–82.

Feng, Y., Lagakos, D. and Rauch, J. E. (2018), Unemployment and development. NBER Working Paper No. 25171.

Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L., van Elsland, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P., Walters, C., Winskill, P., Whittaker, C., Donnelly, C., Riley, S. and Ghani, A. (2020), Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. Unpublished Working Paper, Imperial College London.

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.

Fuchs-Schündeln, N., Krueger, D., Ludwig, A. and Popova, I. (2020), The long-term distributional and welfare effects of covid-19 school closures. Unpublished Working Paper, University of Pennsylvania.

Glover, A., Heathcote, J., Krueger, D. and Ríos-Rull, J.-V. (2020), Health versus wealth: On the distributional effects of controlling a pandemic. NBER Working Paper No. 27046.

Glynn, J. R. (2020), 'Protecting workers aged 60–69 years from covid-19', *The Lancet Infectious Diseases; (published online April 16)* .

Gollin, D. (2002), 'Getting income shares right', *Journal of Political Economy* **110**(2), 458–474.

Gottlieb, C., Grobovsek, J., Poschke, M. and Saltiel, F. (2021*a*), 'Lockdown accounting', *The B.E. Journal of Macroeconomics* .

Gottlieb, C., Grobovsek, J., Poschke, M. and Saltiel, F. (2021*b*), 'Working from home in developing countries', *European Economic Review* **133**.

Greenwood, J., Kircher, P., Santos, C. and Tertilt, M. (2019), 'An equilibrium model of the african hiv/aids epidemic', *Econometrica* **87**(4), 1081–1113.

Huggett, M. (1993), 'The risk-free rate in heterogeneous-agent incomplete-insurance economies', *Journal of Economic Dynamics and Control* **17**(5-6), 953–969.

Huggett, M. (1996), 'Wealth distribution in life-cycle economies', *Journal of Monetary Economics* **38**(3), 469–94.

Jones, C. I. and Klenow, P. J. (2016), 'Beyond gdp? welfare across countries and time', *American Economic Review* **106**(9), 2426–57.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20110236*

Kaplan, G., Moll, B. and Violante, G. L. (2020), The great lockdown: Macroeconomic and distributional effects of covid-19. Unpublished Manuscript, University of Chicago.

Karaivanov, A., Lu, S. E., Shigeoka, H., Chen, C. and Pamplona, S. (2021), 'Face masks, public policies and slowing the spread of covid-19: Evidence from canada', *Journal of Health Economics* **78**.

Karlinsky, A. and Kobak, D. (2021), The world mortality dataset: Tracking excess mortality across countries during the covid-19 pandemic. https://doi.org/10.1101/2021.01.27.21250604.

Krueger, D., Uhlig, H. and Xie, T. (2020), Macroeconomic dynamics and reallocation in an epidemic. NBER Working Paper No. 27047.

Loayza, N. and Pennings, S. M. (2020), Macroeconomic policy in the time of covid-19: A primer for developing countries. World Bank Research and Policy Brief 147291.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M. and Edmunds, W. J. (2008), 'Social contacts and mixing patterns relevant to the spread of infectious diseases', *PLoS Med* **5**(3), e74.

Prin, M. and Wunsch, H. (2012), 'International comparisons of intensive care: Informing outcomes and improving standards', *Current Opinion in Critical Care* **18**(6), 700.

Ramachandran, S. and Malani, A. (2021), All-cause mortality during sars-cov-2 pandemic in india: Nationally-representative estimates independent of official death registry. Unpublished Working Paper, University of Toronto.

Sanchez, J. M. (2021), Covid-19's economic impact around the world. Federal Reserve Bank of St. Louis Regional Economist.

von Carnap, T., Almås, I., Bold, T., Ghisolfi, S. and Sandefur, J. (2020), 'The macroeconomics of pandemics in developing countries: an application to uganda', *COVID Economics Vetted and Real-Time Papers Issue 27, CEPR* pp. 104–122.

Yamall Orellana, J. D., da Cunha, G. M., Marrero, L., Ismerio Moreira, R., da Costa Leite, L. and Lessa Horta, B. (2021), 'Excess deaths during the covid-19 pandemic: Underreporting and regional inequalities in brazil', *Cadernos de SaÃºde PÃºblica* **37**(1).

Appendix

A. Appendix Figures

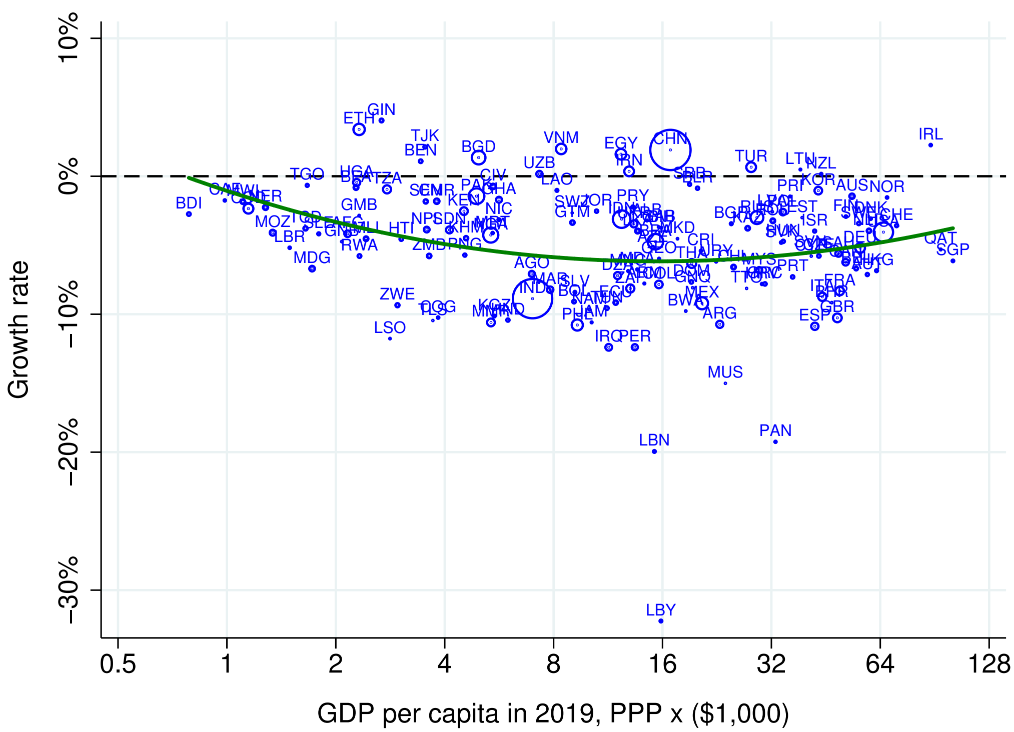


Figure 3.A.1: GDP per capita Growth from 2019 to 2020

Note: GDP per capita data comes from the World Bank World Development Indicators. GDP per capita is expressed at PPP and is taken from the Penn World Table 9.1 (Feenstra et al., 2015).

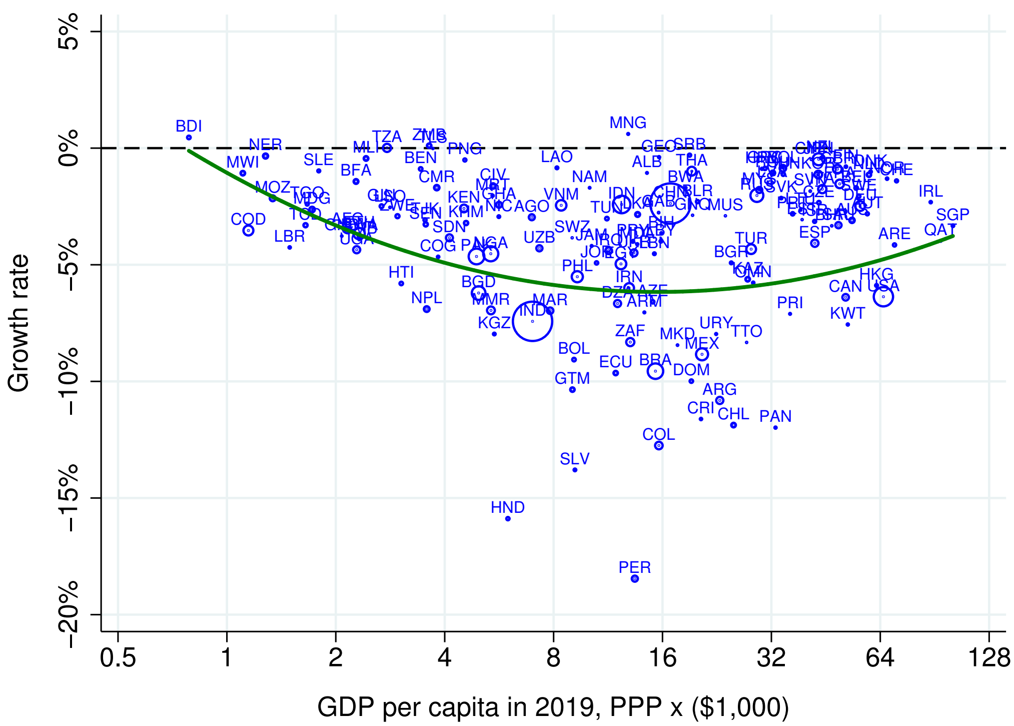


Figure 3.A.2: Employment Growth from 2019 to 2020

Note: Employment data comes from the ILO Statistical Database. GDP per capita is expressed at PPP and is taken from the Penn World Table 9.1 (Feenstra et al., 2015).

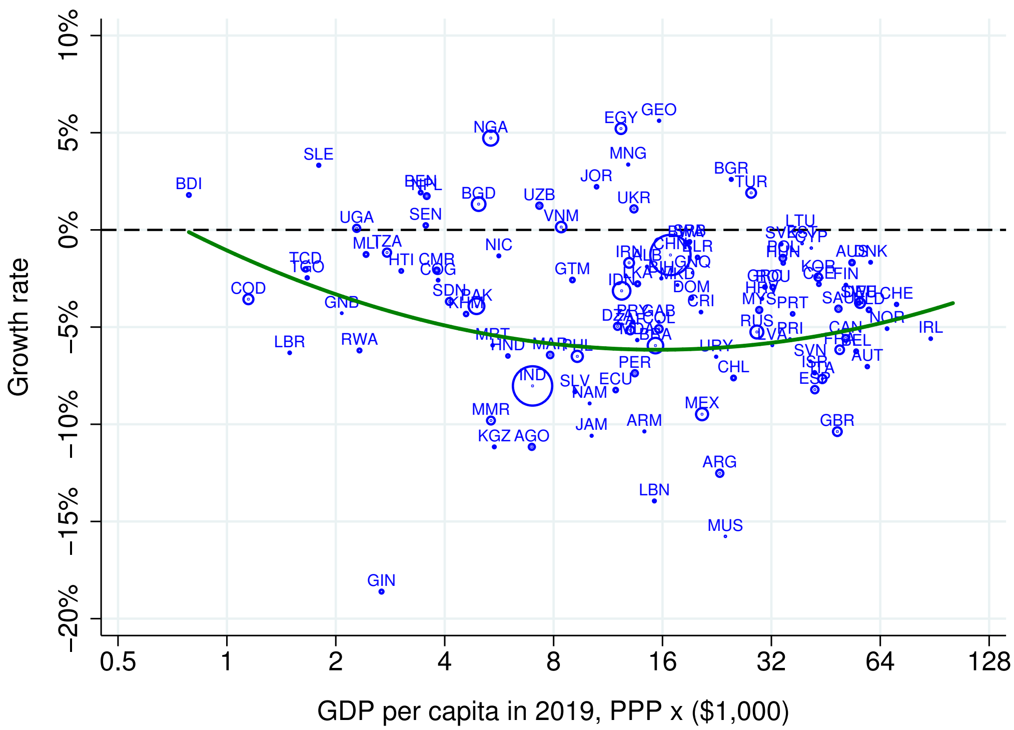


Figure 3.A.3: Consumption-per-capita Growth from 2019 to 2020

Note: Consumption data comes from the World Bank World Development Indicators. GDP per capita is expressed at PPP and is taken from the Penn World Table 9.1 (Feenstra et al., 2015).

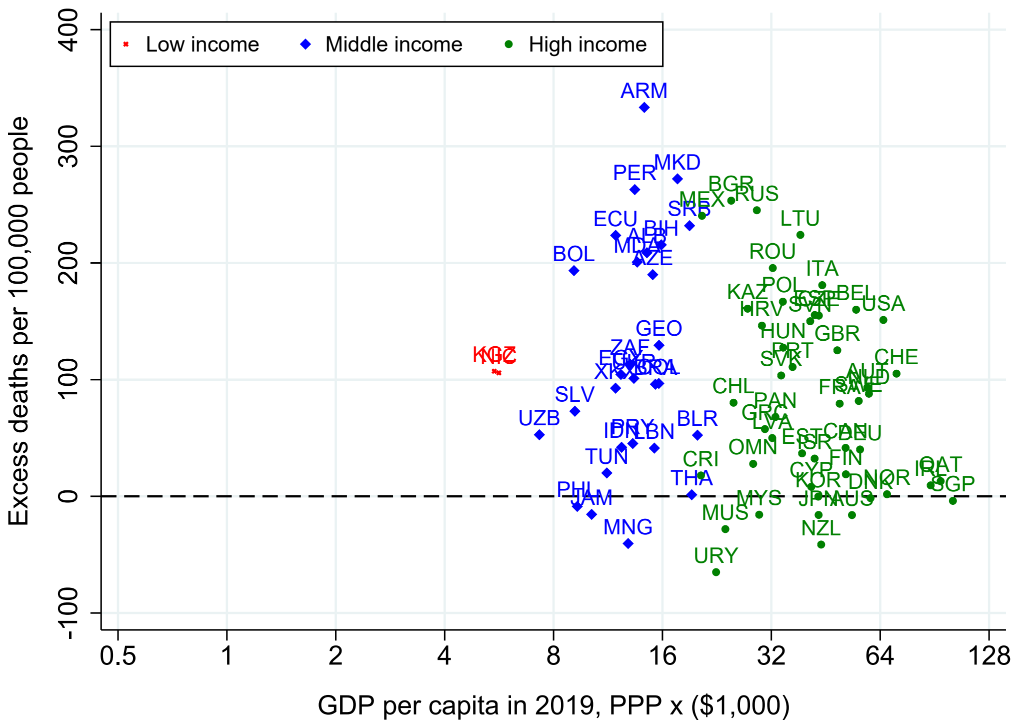


Figure 3.A.4: Excess Deaths Estimated by *The Economist*

Note: Data sourced from the Economist excess mortality tracker.

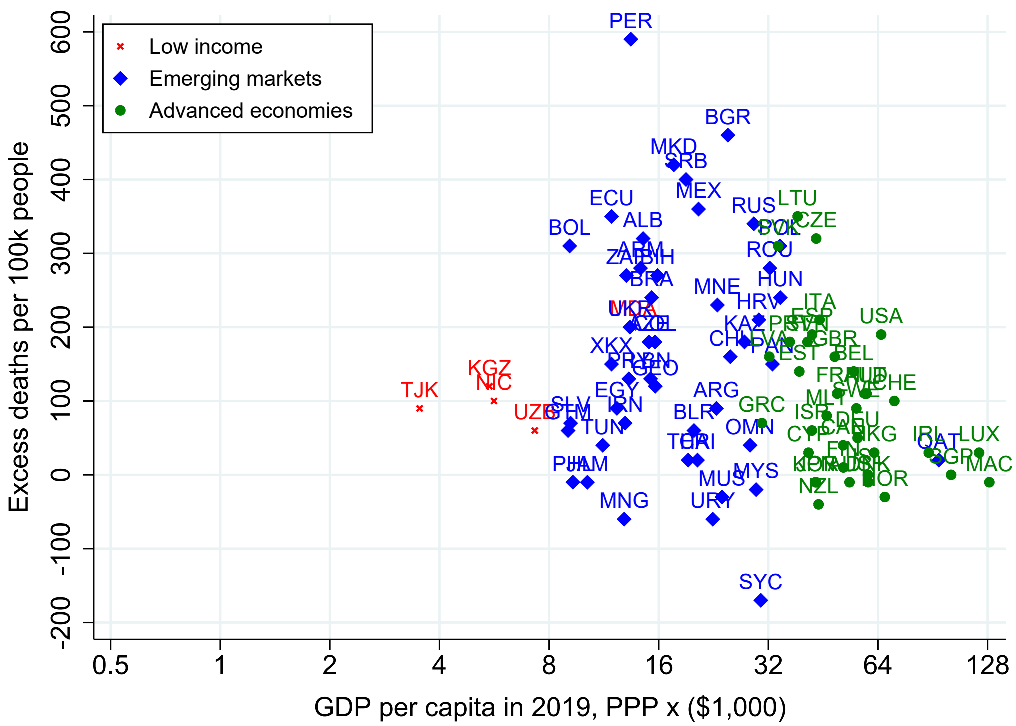


Figure 3.A.5: Excess Deaths Estimated by Karlinsky & Kobak (2021)

Note: Data sourced from Karlinsky & Kobak (2021)'s World Mortality Database.

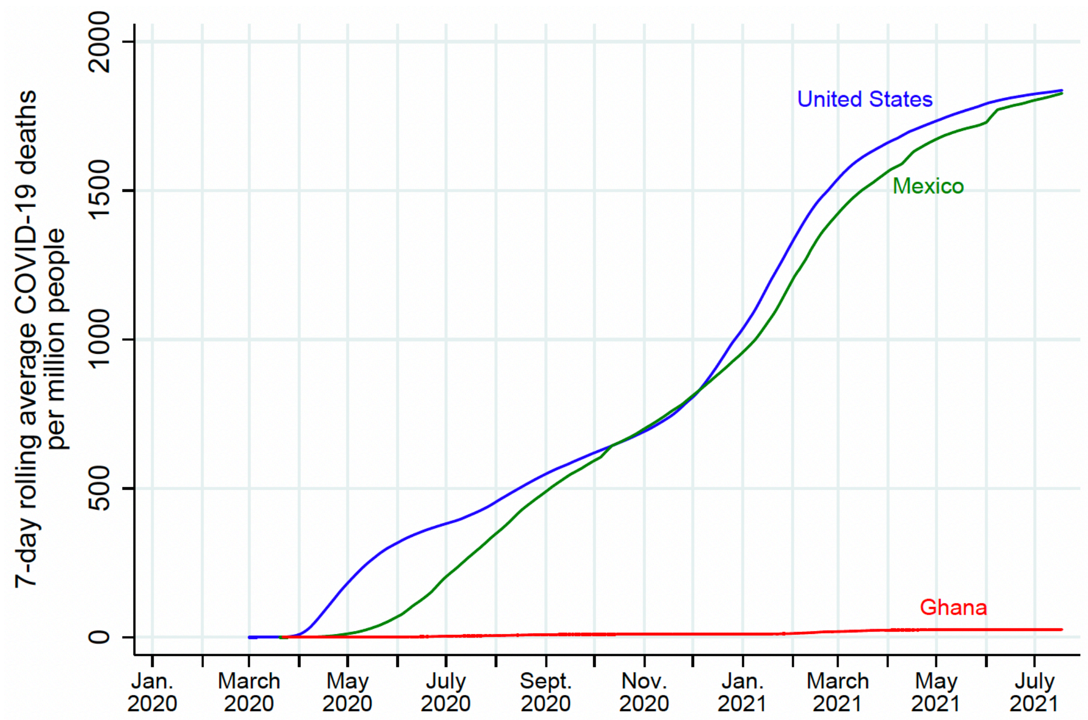


Figure 3.A.6: Official COVID-19 Deaths in the United States, Mexico and Ghana

Note: This figure plots cumulative official deaths from COVID-19, according to Our World in Data, in the three focus countries: the United States, Mexico and Ghana.

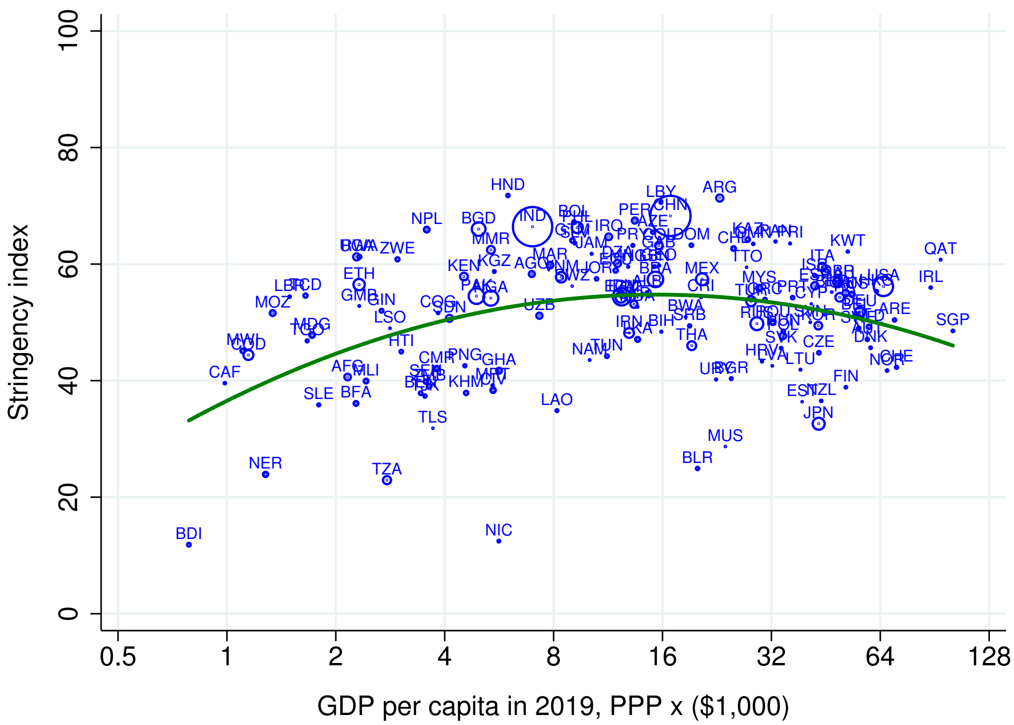


Figure 3.A.7: Oxford Lockdown Stringency Index

Note: The Government Stringency Index is taken from the Oxford Government Response Tracker (OxCGRT). GDP per capita is expressed at PPP and taken from Penn World Table 9.1 (Feenstra et al., 2015).

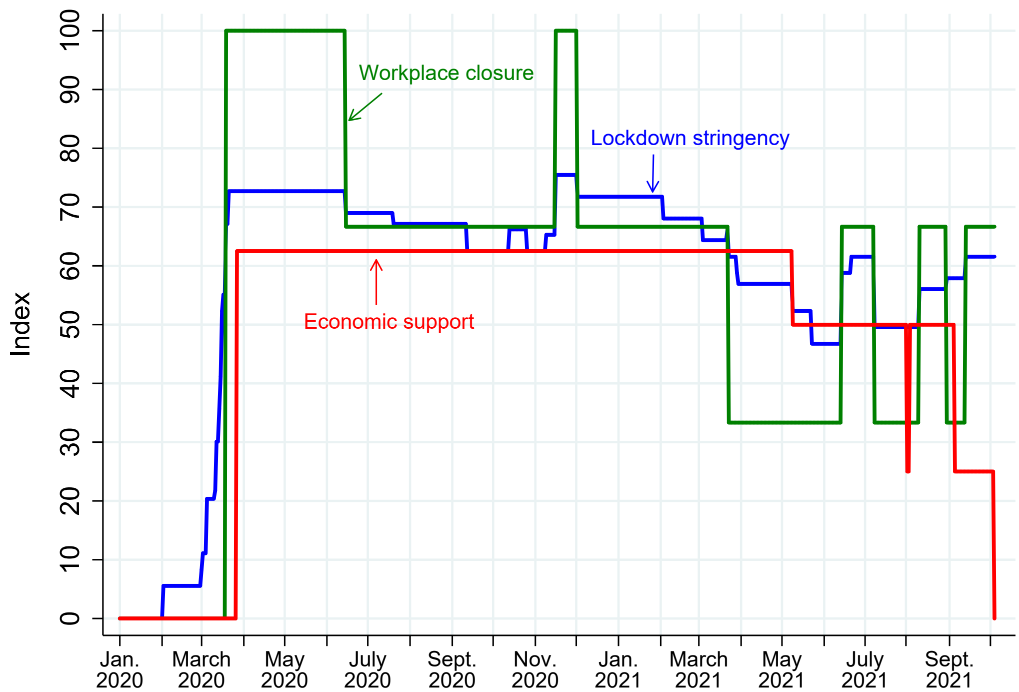


Figure 3.A.8: Time-Series of Lockdown Policies and Economic Support in the United States

Note: This figure displays the time-series of Oxford Lockdown Stringency Index, Economic Support Index, and Workplace Closures for the United States.

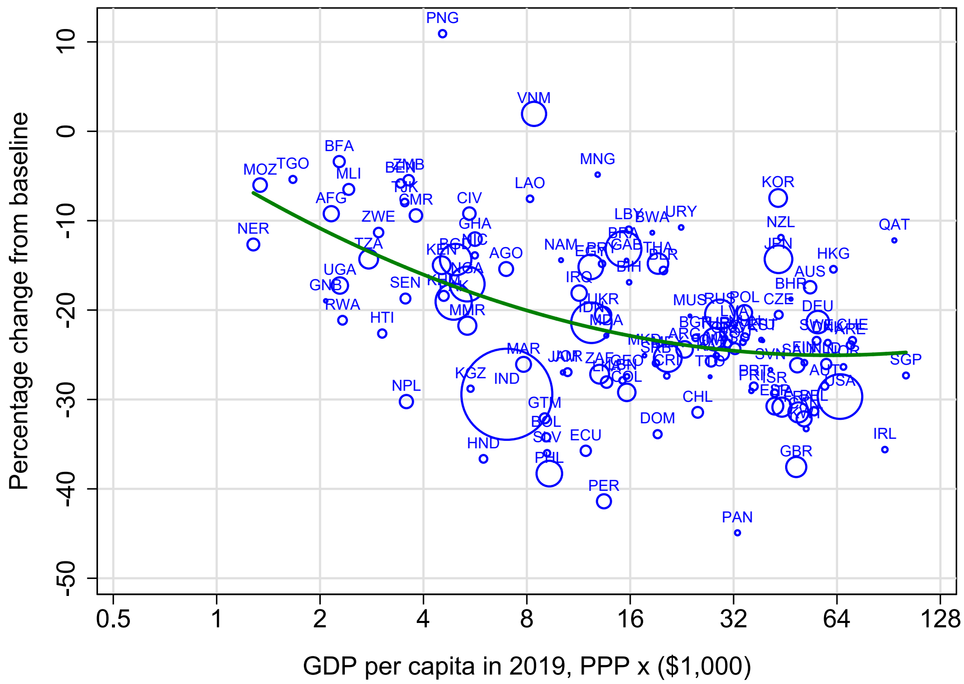


Figure 3.A.9: Changes in Workplace Mobility in 2020

Note: This figure plots the average percent change in visits and time spent at workplaces from baseline in 2020 against GDP per capita in 2019. The baseline is the median value, for the corresponding day of the week, during the 5-week period Jan 3 - Feb 6, 2020. GDP per capita is expressed at PPP and taken from Penn World Table 9.1 (Feenstra et al., 2015). Percent change in visits and time spent at home and workplace in 2020 comes from the Google COVID-19 Community Mobility Reports.

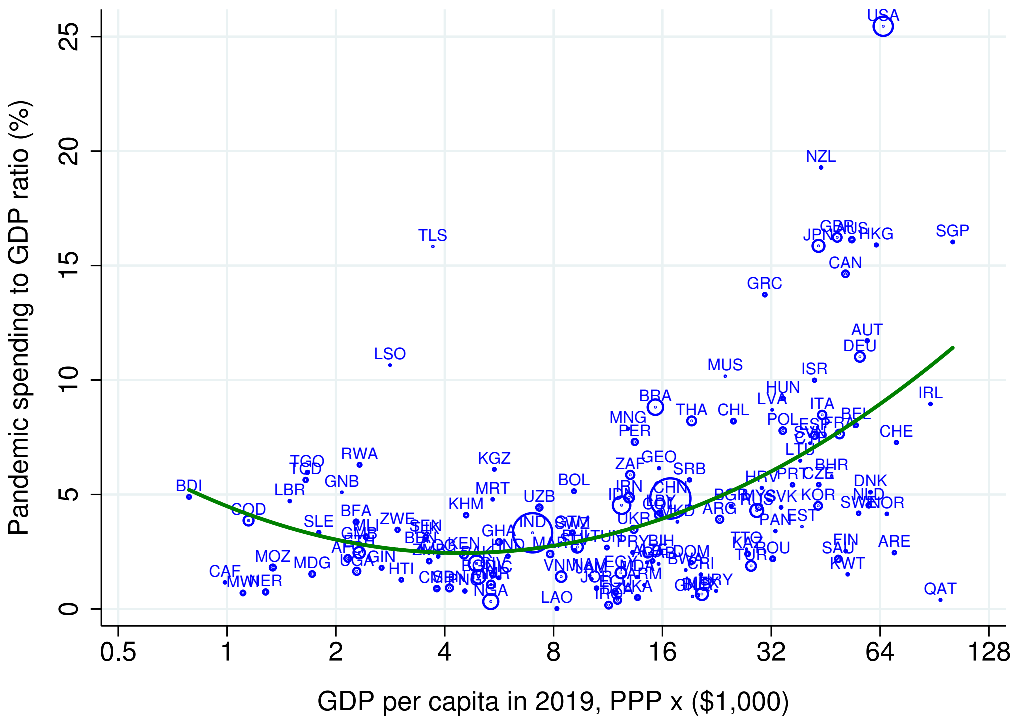


Figure 3.A.10: Pandemic Spending as Share of GDP

Note: Data on pandemic spending come from the IMF Fiscal Monitor Database. GDP per capita is expressed at PPP and taken from Penn World Table 9.1 (Feenstra et al., 2015).

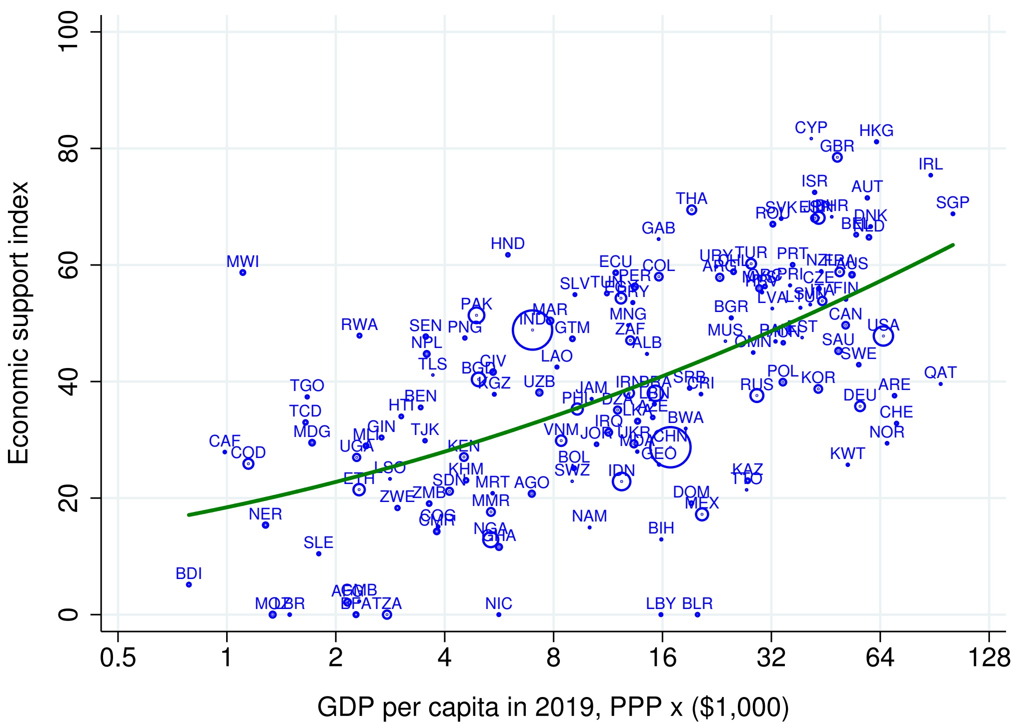


Figure 3.A.11: Economic Support Index

Note: Oxford Coronavirus Government Response Tracker's *Economic Support Index*.

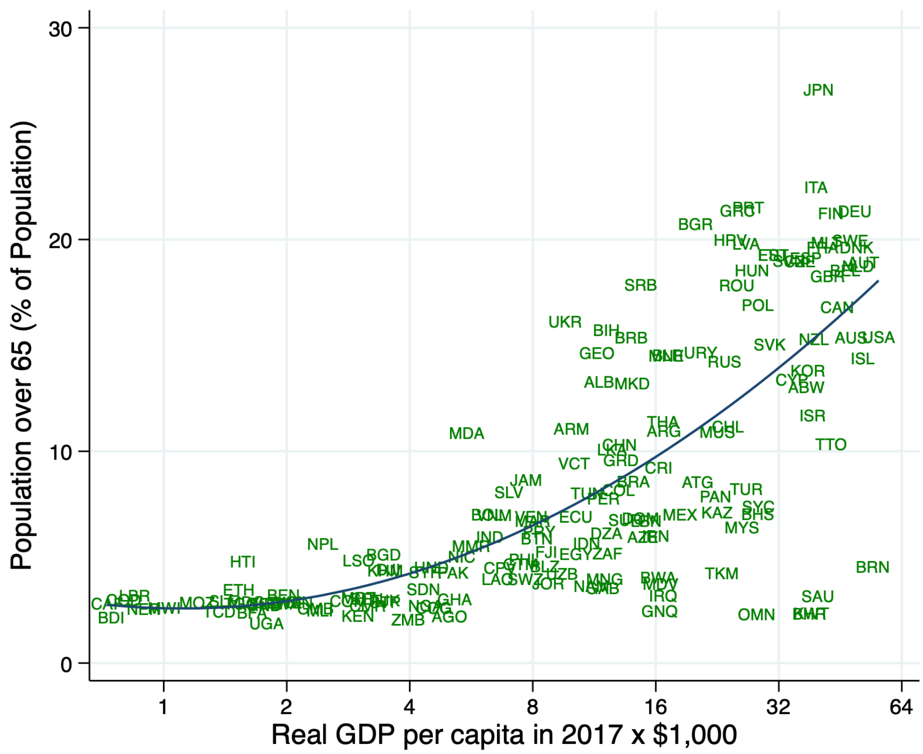


Figure 3.A.12: Fraction of the Population Older than Age 65

Note: This figure plots the proportion of population ages over 65 and above as a percentage of total population across 162 countries. GDP per capita is from Penn World Table 9.1 (Feenstra et al., 2015). Population data is World Bank staff estimates using the World Bank's total population and age/sex distributions of the United Nations Population Division's World Population Prospects: 2019 Revision.

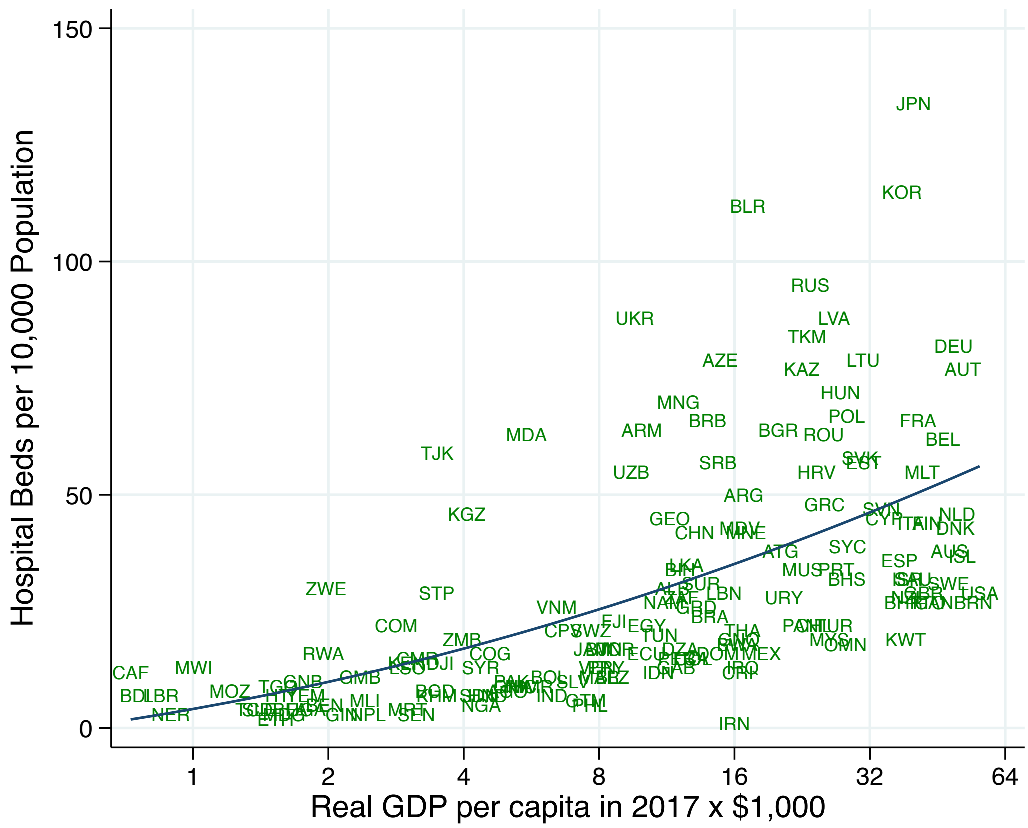


Figure 3.A.13: Hospital Beds per 10,000 People

Note: This figure plots the number of hospital beds available per 10,000 inhabitants in 153 countries. GDP per capita is at PPP and taken from the Penn World Table 9.1 (Feenstra et al., 2015). The hospital bed data are from the World Health Organization's Global Health Observatory.

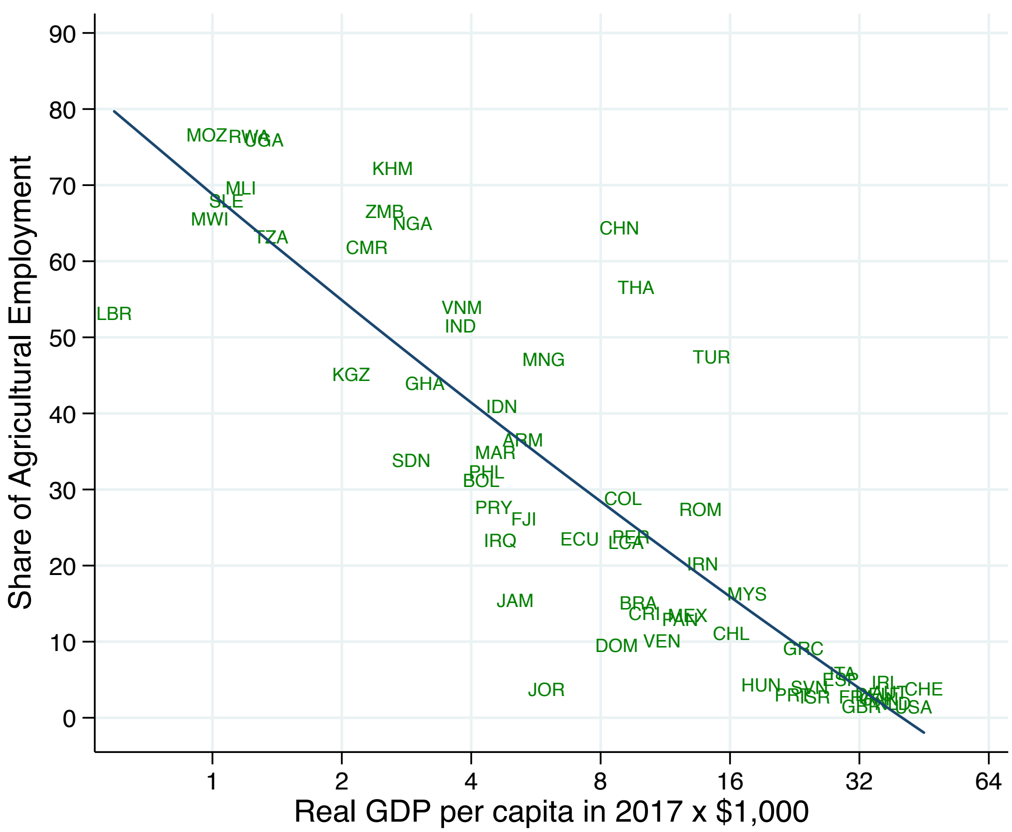


Figure 3.A.14: Size of the Agricultural Sector

Note: Agriculture employment data is taken from the IPUMS database. GDP per capita is expressed at PPP and is taken from the Penn World Table 9.1 (Feenstra et al., 2015).

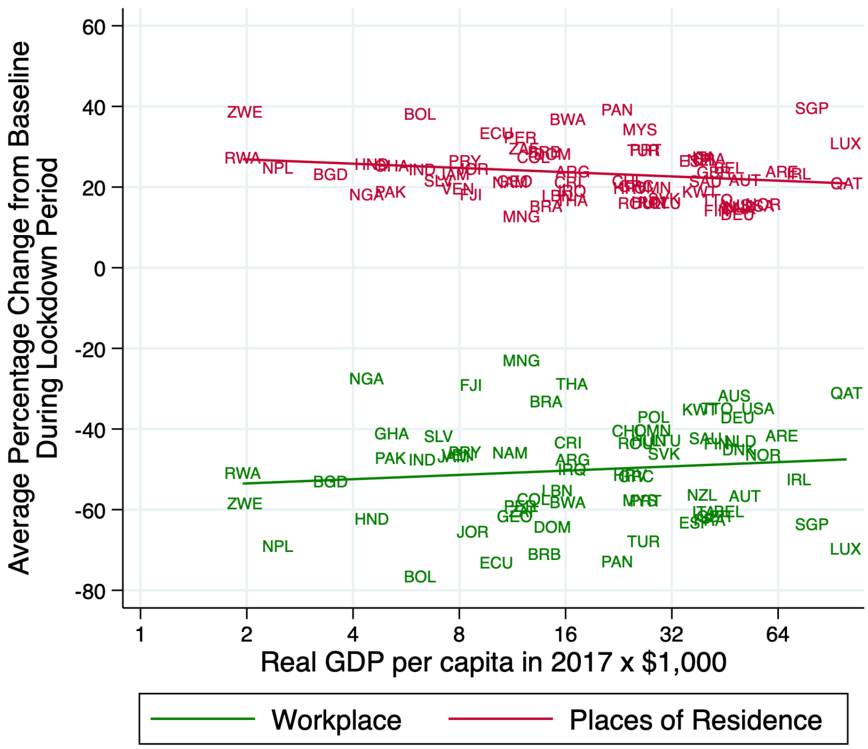


Figure 3.A.15: Changes in Mobility Across Countries During Lockdown Periods

Note: This figure plots the average percentage changes of the mobility metric in the 'Places of Residence' and 'Workplace' categories in the Google Community Mobility Report (Aktay et al., 2020), during the lockdown periods for the 65 countries which had implemented or are implementing lockdown. GDP per capita is from Penn World Table 9.1 (Feenstra et al., 2015). The average across all 65 countries is 23.44 percent. The slope of the fitted line is 1.52, with *p*-value of 0.354 for the 'Workplace' category. For the 'Places of Residence' category, the slope of the fitted line is -1.52, with *p*-value of 0.083.

B. Appendix Tables

Table 3.B.1: ICU Bed Availability Across Countries

Country	ICU beds per 100,000 population	Per capita healthcare cost
United States	20.0-31.7	$7,164
Canada	13.5	$3,867
Denmark	6.7-8.9	$3,814
Australia	8.0-8.9	$3,365
South Africa	8.9	$843
Sweden	5.8-8.7	$3,622
Spain	8.2-9.7	$2,941
Japan	7.9	$2,817
UK	3.5-7.4	$3,222
New Zealand	4.8-5.5	$2,655
China	2.8-4.6	$265
Trinidad and Tobago	2.1	$1,237
Sri Lanka	1.6	$187
Zambia	0	$80

Source: Table 1 in Prin and Wunsch (2012). Healthcare cost includes all public and private expenditures.

Table 3.B.2: Internally Calibrated Parameters and Model Fit

	Data	Model	Parameters	Description
U.S. GDP Decline, '19-'20	-4.10%	-4.01%	$A(P)$	Pandemic TFP
U.S. Employment Decline, '19-'20	-6.40%	-6.36%	ϕ_s	Productivity of remote work, social sector
Fraction Remote Workers pre COVID	8.20%	8.14%	σ_g	Variance of remote work taste shock