The Causal Factors Driving the Rise in U.S. Health-services Prices

Feldman, Maria and Pretnar, Nick

6 July 2023

The Causal Factors Driving the Rise in U.S. Health-services Prices

Maria Feldman[1*] Nick Pretnar[2†]

[1] *University of Wuerzburg*
[2] *University of California, Santa Barbara*

July 6, 2023[‡]

Abstract

We explore several possible avenues which have driven the rise in aggregate U.S. health-services prices since the mid-twentieth century. Our multi-sector general equilibrium model is structural change meets health macro, featuring endogenous population aging, market concentration in the health sector, and differential rates of sectoral technological change. The rise in the relative price of health services is almost exclusively a result of increasing market concentration in the health services sector, as well as slow health-sector TFP growth. Rising health prices have had no impact on life expectancy. Further, our results partially challenge the idea that population aging is responsible for dampening GDP growth rates. While health-sector TFP grows slowly, this is offset by gains to the efficiency of converting health investment to healthy outcomes which leads to increases in expenditure and higher rates of GFP growth.

Keywords: health services, market concentration, health prices, growth, structural change, aging
JEL Classification: I1, L1, O3, O4

* maria.feldman@uni-wuerzburg.de
† npretnar@ucsb.edu
‡ We are grateful for comments from Daniel Carroll, Dirk Krueger, and Karen Kopecky, as well as seminar participants at the Federal Reserve Bank of Cleveland and the UCSB, LAEF 2nd Annual Labor Markets and Macroeconomic Outcomes Workshop.

1 Introduction

In the United States (U.S.) the aggregate price level of health services relative to all other consumption has increased by a factor of three since 1948. The rise has been attributed to everything from increasing market concentration, slow sectoral productivity growth, and even population aging. But, which of these phenomena is most responsible for rising prices?

In this paper we build a model of structural transformation with features unique to the health-services sector in order to explore the forces that are causally responsible for the rising relative price of health services. We are motivated by the corollary of a question explored in Zhao (2014), as to why health expenditure shares have risen: why have relative prices risen? In this sense, our work also complements the analyses in Fonseca et al. (2021), who explore the factors which have led to rising relative health expenditure, and that of Horenstein and Santos (2019) and Fonseca et al. (2020), who attribute the increase in U.S. relative prices to mark-ups and price wedges.

Borrowing on evidence from Horenstein and Santos (2019) that the health-services sector has become increasingly concentrated, we allow for health-services production to be (possibly) monopolistically competitive in the spirit of Dixit and Stiglitz (1977). Given evidence that consumers at different ages have different demand elasticities for health-services consumption, we deploy a Grossman (1972) model like that of Hall and Jones (2007), allowing for health-services consumption to affect survival probabilities differently depending on a consumer's age. The model thus features endogenous aging acting on aggregate demand and, by extension, market-clearing prices. In the spirit of structural change models, we allow for unbalanced relative total-factor productivity (TFP) growth between the health-services and non-health-services sectors following evidence in Triplett and Bosworth (2004), Blumenthal, Stremikis, and Cutler (2013), and Shatto and Clemens (2022) that health-sector TFP is relatively slow growing. Finally, in order for changing demand patterns to affect prices in general equilibrium, we require the capital intensity of the health sector to be different than the capital intensity of the non-health sector, following theory in Alonso-Carrera, Caballé, and Raurich (2015).

This paper is thus structural change meets health and aging. We allow for the health sector to use a different technological structure than the broader economy. This is uncommon in the macro-health literature and somewhat uncommon even in the structural change literature, but critical in order to generate testable implications from theory. Indeed, we show that if the changing composition of demand is contributing to relative price variation in general equilibrium, then it *must* be the case that the health sector uses a different production technology than the rest of the economy.

Health-services expenditure is unique amongst consumption for two primary reasons: 1) it is associated with age-specific elasticities and productivities with respect to *delivering* utility (Hall and Jones 2007); 2) health-services consumption today is non-separable with tomorrow's consumption choices, because health-services consumption today impacts the rate at which consumers of age j survive to become consumers of age $j+1$. The model set-up allows us to explore additional, aggregate implications for the rise in the health-services sector. As population aging has been associated with slowing aggregate output growth (Krueger and Ludwig 2007; Backus, Cooley, and Henriksen 2014; Cooley and Henriksen 2018; Kydland and Pretnar 2019; Maestas, Mullen, and Powell 2023), we can also explore the degree to which growth decelerations may result from structural change in the context of the rising health sector.

Among our findings, the main result is that increasing health-sector market concentration relative to economy-wide market concentration appears most responsible for rising relative prices. Baumol's cost disease also plays an important role, given we find that health-sector TFP has grown more slowly than TFP for the rest of the economy (Baumol 1967; Baumol, Blackman, and Wolff 1985), which is consistent with estimates from Blumenthal, Stremikis, and Cutler (2013) and Shatto and Clemens (2022). Population aging does not appear to play a significant role in driving up prices, perhaps because capital-intensity differentials across sectors are small.

We also find that increasing life expectancy over the latter 20th and early 21st centuries is mostly driven by exogenous changes to consumer preferences which provide an impetus toward healthier life styles. Had the structure of aggregate production remained fixed at that of 1960 (i.e., no increasing health-sector market concentration and no TFP growth), life expectancy for newborns in 2015 would be 3.42 life years lower than the model prediction. However, had preferences for healthiness (i.e., increasing returns to health consumption over time) remained fixed, life expectancy of newborns would be over 25 years lower than the model prediction.

Finally, when it comes to aggregate GDP growth, we find mild evidence that undercuts the idea that GDP growth decreases as the population ages. However, since aging can be affected by multiple channels, which channel dominates will determine whether GDP growth slows as the population gets older on average. If the population gets older because new cohorts entering working age remain the same size (exogenous aging), then GDP growth falls relative to the predictive baseline. However, we find little evidence that endogenous aging (i.e., aging due to increased health investment) has weighed negatively on GDP growth.

This paper proceeds as follows. In the next section we present aggregate data pertaining to health expenditure, prices, and population aging, while also discussing the literature. In Section 3 we present a model of household demand for health and non-health consumption and multi-sector production with (possibly) monopolistically competitive health-services production. In Section 4 we engage in several simulation exercises in order to assess which channels are primarily responsible for relative-price growth in a model with parameters calibrated to match several time series of observables from 1960-2015. We also examine model implications for aggregate GDP growth and assess model fit against GDP growth rates as a robustness check. In Section 5 we conclude.

2 Background

2.1 The U.S. Health Services Sector

While much attention has been given to reasons why the aggregate share of expenditure devoted to health services has risen in the U.S., only recently have researchers considered reasons for the rise in the relative price of health services. Further, the degree to which changing consumption patterns due to population aging and/or income effects versus sector-level factors, like differential rates of technical change and/or market concentration, are responsible for the structural transformation of the health sector is still unsettled in the literature.

Generally speaking, the literature can be divided into two broad camps — that which focuses on forces affecting changing *total outlay* devoted to health services and another related strand of literature focusing on the forces affecting health-services *prices* more directly. The measurement of health-services prices and quantities faces several difficulties that do not plague measurements from other sectors. Notably, the Bureau of Labor Statistics' (BLS) Consumer Price Index (CPI) fails to account for the unique structure of how health-services payments are made, often indirectly by employers, insurers, or government agencies. The Bureau of Economic Analysis' (BEA) Personal Consumption Expenditure index (PCE) from the National Income and Products Accounts (NIPA) provides a more holistic accounting of health-service prices and quantities by including in its measurements outlays provisioned to consumers indirectly via third-party payers, be they government or private-sector entities (Dunn, Grosse, and Zuvekas 2016). This index excludes non-health administrative costs associated with government programs or insurance provisions and the value of noncommercial health-services research (Dunn, Grosse, and Zuvekas 2016). Further, both the CPI and PCE are purely economical objects that do not directly account for the varying effectiveness of specific treatments over time or the severity of patient diseases, making quality adjustment less precise than it otherwise would be with a disease-based cost index (Berndt et al. 2001). Disease-based indices have their own limitations, though, as the "quantities" associated with these indices are defined in terms of the number of *patients* treated for specific ailments as opposed to actual real quantities, in terms of some base-year dollar value, of total services rendered (Dunn, Grosse, and Zuvekas 2016). Weighing the different constraints and shortcomings of the various price indices, the BEA's PCE, specifically the PCE health-by-function index, most consistently captures what economists think of as "prices" and "quantities" of final consumption. We thus proceed heretofore referring to prices and quantities as they are measured in the PCE.

Perhaps because quality adjustment in measuring health-services consumption and thus prices is difficult, in approaching the puzzle as to how and why health services have become an increasingly large part of the U.S. economy, many researchers have focused solely on the value of total nominal outlay. Specifically, how has population aging contributed to a rising health services share? Hall and Jones (2007) show that standard consumption preferences, which account for the fact that health investment increases longevity, can reconcile the increase in the health share of U.S. GDP and the simultaneous decline in non-accidental mortality rates (increase in life expectancy) since 1950. The Hall and Jones (2007) framework, while originally novel in allowing for consumer welfare to be affected by health along multiple dimensions (utility and survival probabilities), does not provide a mechanism to decompose the degree to which preference-driven forces like income effects contribute to the rising health share versus changing relative TFP's and market power in the health-services sector. By contrast Acemoglu, Finkelstein, and Notowidigdo (2013) provide statistical evidence that rising incomes are unlikely to have contributed much to the rise in the health share of GDP. It thus remains unsettled in the literature as to the degree to which consumer preferences interacting with rising incomes have contributed to a rising health share, necessitating a falsifiable model that contains various mechanisms (e.g., population aging, technology growth, and rising incomes).

Another strand of literature has grappled with the role that health-service providers have played in contributing to rising health costs. Horenstein and Santos (2019) provide evidence that the relative rise in U.S. health spending compared to the rest of the world is likely due to increasing market concentration in the health sector affecting relative prices, not simply differences in technological transformation. Fonseca et al. (2021) also find only a small role for technological change in explaining rising health spending, yet their partial-equilibrium exercise still provides high-quality support for a deeper, general-equilibrium analysis investigating what has driven rising health-services prices along the growth path. Fonseca et al. (2020) confirm the puzzle presented by Horenstein and Santos (2019) in finding that price wedges can help explain differences in cross-country health expenditure between European nations and the U.S.

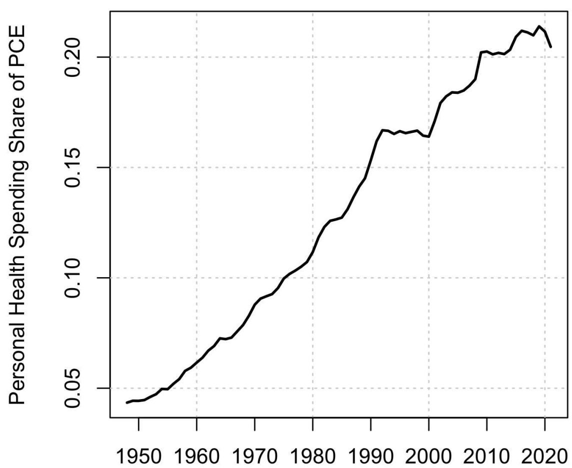
By contrast to work in Fonseca et al. (2020) and Fonseca et al. (2021), Chandra and Skinner (2012) present a model where rising costs (and thus prices) are driven both by technological change advancing the quality of medical procedures but also by consumers with high-quality health insurance demanding evermore advanced treatments. Their model does not contain a role for market concentration, however. Chandra and Skinner (2012) partially build on arguments in Garber and Skinner (2008) that some reasons U.S. health costs have risen are because consumers have incentives to over-utilize high-cost treatments, further incentivizing providers to adopt expensive, cutting-edge medical innovations rapidly even in the face of little clinical evidence regarding such treatments' efficacy.

Horenstein and Santos (2019) estimate that the average U.S. health sector mark-up over the economy-wide mark-up went from -0.5 in the 1970-1977 period, to 3.0 in 1978-1990, and back to 1.8 from 1991-2007.[1] To this point, using micro data, Cooper et al. (2019) show that local hospital market concentration is a strong contributor to relatively higher local health-services prices.[2] Gaynor and Town (2012), Gaynor, Ho, and Town (2015), and Gowrisankaran, Nevo, and Town (2015) also each provide evidence that the hospital industry has become increasingly concentrated

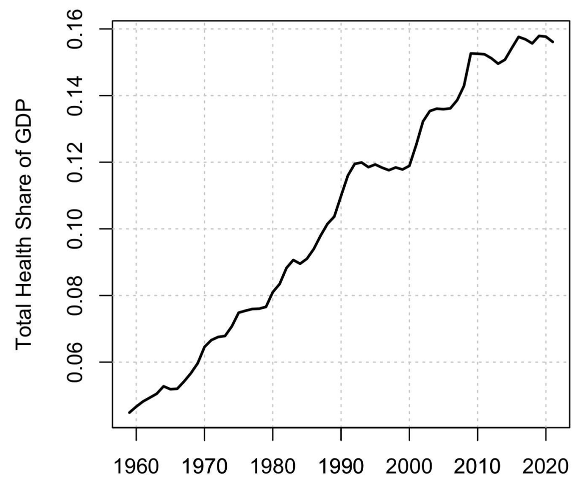
[1] These statistics are measured as follows. Let p_{agg} and mc_{agg} be aggregate price and marginal cost. Let p_h and mc_h be the health-sector price and marginal cost. Then these statistics are $(p_h - mc_h) - (p_{agg} - mc_{agg})$.

[2] There is also evidence that physicians tend to call for a greater number of discretionary procedures and encourage patients to take up more elective surgeries in regions both where health-services prices are effectively higher and physicians earn a relatively greater marginal return on each service provided (Clemens and Gottlieb 2014).

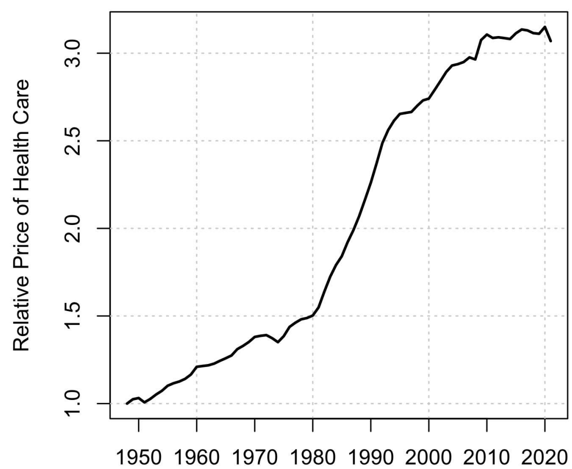
since the early 1980s. This begs the broader quantitative question, to what degree do mark-ups from imperfect competition matter in driving up the relative price of health services when accounting for the general equilibrium implications of changing consumption patterns?



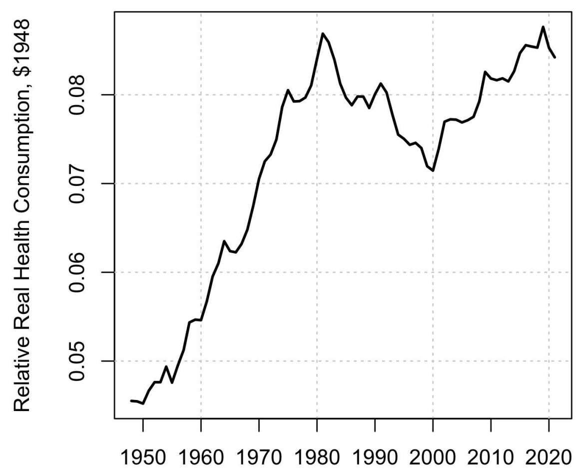
(a) Aggregate Health Share PCE



(b) Aggregate Health Share GDP



(c) Relative Health Price, 1948 = 1



(d) Relative Health Quantities, $1948

Figure 1: Panel (a) shows the share of all personal consumption expenditures devoted to personal health spending with data taken from NIPA Table 1.5.5. Panel (b) shows the share of GDP devoted to all health services outlay, including the value of governmental administrative services not included in the PCE data and which do not directly correspond to health-services rendered directly to the consumer. Panel (c) shows the price of health services PCE relative to all other consumption, computed from NIPA PCE data (see Appendix A.2). Panel (d) shows the real quantities in $1948 of health consumption relative to all other consumption.

Figure 1 demonstrates how health-services outlay, prices, and quantities have changed over time. In panel (a) we observe that health's share of domestic PCE increased from less than 5% in 1947 to over 20% of all domestic consumption expenditure in 2021. In panel (b) we plot the share of GDP devoted to all health-services outlay, which has also risen since 1948 though not by as much as the health-services share of PCE. In panel (c) we normalize all prices to unity in 1948 and plot the price of health services relative to all non-health-services consumption.[3] The relative price of health services has tripled since 1948. Finally, panel (d) shows relative real quantities of health consumption in units of $1948. In 1948 the real, quality-adjusted value (adjusted by the BEA) of health consumption was less than 5% of all non-health consumption, though this value rose to greater than 8% by 2021. Examining panels (a), (c), and (d), it should be apparent that the rise in the PCE share is primarily driven by rising prices.

We plot breakdowns of shares and prices for the sub-components that comprise the health-services PCE aggregate in Appendix A.3. To summarize, we observe that the relative contribution of prescription drugs and medical appliances to price increases has declined over time, while the relative contribution of hospital services to the aggregated health price has risen. This result is consistent with micro-evidence from Cooper et al. (2019) that hospital systems' local pricing power drives up prices faced by consumers. Therefore, with aggregate evidence from Horenstein and Santos (2019) and supporting micro-evidence from Cooper et al. (2019), it is reasonable to conclude that monopolistic pricing has, to some degree, contributed to the rise in the aggregate health-services price level. With our model we will explore this hypothesis.

2.2 Population Aging

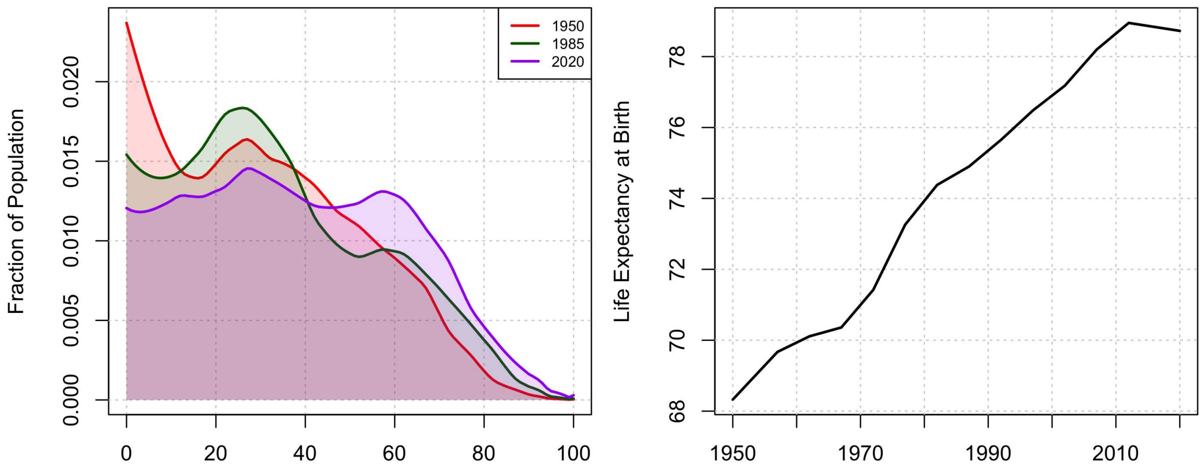
We seek to understand the degree to which population aging has contributed to rising aggregate health-services prices when accounting for possible market concentration and differences in sectoral TFP growth rates. Older individuals require greater amounts of health care and have lower elasticities of health status with respect to health expenditure (Hall and Jones 2007). Thus, as the population ages we expect that demand for health care will increase and the aggregate elasticity of demand with respect to prices will fall, partially contributing to price increases in general equilibrium. To our knowledge nobody has yet to compare the effects of aging to those of market concentration and differential sectoral TFP growth rates in a general equilibrium model where aging is endogenous and can directly impact the relative price of health services.

Jung, Tran, and Chambers (2017) explore the consequences of aging on the U.S. healthcare system, with a particular focus on pressures the Medicare and Medicaid programs are expected to face as a result of aging. Their model endogenizes insurance enrollment decisions but features exogenous survival, though they allow for differential rates of sectoral capital utilization. Heretofore, while Horenstein and Santos (2019) provide the most comprehensive analysis to date regarding the role that market concentration must play to explain the evolution of health-care sectoral aggregate prices relative to the rest of the economy, Jung, Tran, and Chambers (2017) provide the first macroeconomic evidence in the context of a general equilibrium model that aging also affects prices. We complement the analysis in Jung, Tran, and Chambers (2017) in several ways. First, while Jung, Tran, and Chambers (2017) allow for insurance-payer-specific mark-ups, they do

[3] See Appendix A.2 for a list of non-health-services consumption categories.

not consider the implications of sector-wide market concentration. A second difference between our analysis and theirs is that we explicitly allow for both health and non-health sectoral TFP's to grow at different rates. Third, longevity (and thus aging) is endogenized in our model via the Hall and Jones (2007) survival-risk structure that is similar in spirit, though parametrically different, from other approaches in Halliday et al. (2019) and He, Huang, and Ning (2021). Finally, while the analysis in Jung, Tran, and Chambers (2017) is forward-looking, focusing on the implications of aging for the future (up to year 2060), our contribution is backward-looking, exploring what factors have driven up health-services relative prices historically.

Indeed, models like Hall and Jones (2007) which feature endogenous health spending and endogenous population dynamics are common in the health-production sub-literature within macroeconomics. However, to our knowledge heterogeneity in aggregate production (i.e., differential sectoral productivities and/or rates of capital deepening) has heretofore not been included in models that also have endogenous population dynamics. Attanasio, Kitao, and Violante (2010) investigate tax rates required to fully fund U.S. social insurance programs as the population ages. Kuhn et al. (2011) have a social planner choosing health allocations to affect endogenous population levels and thus social welfare, though they abstract from sectoral production heterogeneity. Zhao (2014) explores how the expansion of the U.S. Social Security system has led to increases in health's share of aggregate GDP in a model with age-specific health elasticities but no sectoral production heterogeneity. Halliday et al. (2019) engage in a partial-equilibrium exploration of factors determining health expenditure over the life-cycle. Nygaard (2022) considers how endogenous health affects survival rates, though he focuses on the role that health-consumption decisions have in contributing to frailty risk. Fehr and Feldman (2023) analyze a market economy with endogenous population dynamics and rich health-care financing mechanisms, but exogenously fix the relative price of health-services to all non-health consumption. Our framework, described in detail in Section 3 will be the first to incorporate endogenous population aging into a two-sector model with sectoral production heterogeneity.



(a) Population Distribution by Age Over Time

(b) Life Expectancy at Birth

Figure 2: Panel (a) shows the population distribution by age, where age groups index the horizontal axis. Panel (b) shows the life expectancy of newborns from 1950 to 2020.

Figure 2 shows how the population distribution by age has skewed further to the right over time (a), and that the life expectancy of newborns has also increased (b).[4] The median age of the U.S. population rose from 30.2 in 1950 to 38.3 in 2020. Further, the median age is expected to continue to increase as fertility rates continue to decline. Thus, if population aging is an important contributor to rising health-services prices, then we would expect future aging to continue to drive up prices.

Finally, our work can speak to a broader literature that examines the effects of population aging on long-run aggregate growth rates. Under certain conditions exogenous population aging in general equilibrium overlapping generations models has been shown to contribute to decelerations in long-run GDP growth (Backus, Cooley, and Henriksen 2014; Cooley and Henriksen 2018; Cooley, Henriksen, and Nusbaum 2019; Kydland and Pretnar 2019; Maestas, Mullen, and Powell 2023). The source of aging (i.e., increasing longevity versus declining fertility), however, may affect both the magnitude of second-order changes to GDP as well as the sign of such changes (Prettner 2013). In addition to being able to dissect how different forces are responsible for structural change and rising prices, our framework also allows us to understand how the unique confluence of endogenous aging along with the increasing share of GDP devoted to a low-productivity sector, the outputs of which are demanded at levels which increase in age (i.e., health services), has contributed to declining GDP growth rates.

[4]U.S. population data by age are taken from the United Nations Population Division, Department of Economic and Social Affairs, 2019 World Population Prospects — Total population (both sexes combined) by single age, region, sub-region, and country. Data run from 1950-2019, which is the last year for which the single-age tables are available. The median age comes from the median age by region, sub-region, and country table in the same data series. Life expectancy is taken from the age-specific life tables by year.

2.3 Structural Change and Technological Progress

The structural transformation literature typically studies which forces affect the composition of economic output and/or expenditure, where the focus is on very broad sectors like manufacturing, services, and/or agriculture. Our study is similar in aim except our sectoral classifications are different and the household preference structure built to cater to the unique nature of health-services demand. Nonetheless, we deploy all of the modeling tools from this literature in order to study how the composition of health services *and* its relative price have changed over time. We add to this literature by bringing in imperfect competition, while incorporating modeling aspects from the macro-health literature which generate income effects both over the life-cycle and over time with respect to demand for health services.

The macroeconomic models used to study drivers of, for example, the rising services share or the declining manufacturing share of output in advanced economies, contain one or more of the following features, which are each necessary to varying degrees in order to match the composition of the economy along the historic development path — non-homothetic preferences (Kongsamut, Rebelo, and Xie 2001; Herrendorf, Rogerson, and Valentinyi 2013; Boppart 2014; Comin, Lashkari, and Mestieri 2021), differential rates of sectoral TFP growth (Ngai and Pissarides 2007; Herrendorf, Rogerson, and Valentinyi 2014), or differential rates of capital deepening (Acemoglu and Guerrieri 2008; Alonso-Carrera, Caballé, and Raurich 2015). When all sectors are perfectly competitive, constant returns to scale, and have the same capital intensities, relative prices are exactly the inverse of relative productivities (Ngai and Pissarides 2007; Herrendorf, Rogerson, and Valentinyi 2014). Thus, in such models assuming identical capital intensities means one cannot test the degree to which other forces, aside from productivity variation, like changing demand patterns, have affected relative prices in general equilibrium (Alonso-Carrera, Caballé, and Raurich 2015). Indeed, in most of the structural transformation literature it is assumed that sectoral capital intensities are identical. However, this constraint has recently come under scrutiny, as capital-embodied technical change is both important for understanding employment reallocation and wage inequality (Caunedo, Jaume, and Keller 2023) and aggregate structural transformation (Caunedo and Keller 2023). We will *not* assume the health sector is associated with the same capital intensity as all other production, allowing for the possibility that prices are determined by variation in health-services demand patterns over time due to population aging.

3 Model

Our goal is to understand how the competing effects of market concentration, unbalanced sectoral technical change (i.e., Baumol's cost disease as it pertains to the health sector), and aging have contributed to the rising price of health services in the U.S. relative to all other consumption. Our general equilibrium model includes the following key features: 1) households with endogenous, age-specific mortality risk in an overlapping generations structure; 2) endogenous population aging; 3) possibly imperfect competition in the market for health services; 4) differences in technical change and capital deepening between the health-services sector and non-health-services sectors. A main result is that, in order for changes to the composition of demand to affect the relative market price of health services in general equilibrium, differential capital intensities between health and non-health sectors is required.

The model captures several unique features of the U.S. health-services marketplace. First, independent of population aging, the model is flexible enough so that it can be used to potentially falsify the findings in Horenstein and Santos (2019) that the decline in the relative productivity of health services to other production cannot adequately explain the rising relative price of health services, thus providing a role for imperfect competition in the market for health services. Second, by using the model structure of Hall and Jones (2007), we can explain the rise in life expectancy as endogenously resulting from increased health investment. Third, by linking health outcomes to survival risk we can endogenously account for how changes to aggregate TFP and price mark-ups in the health-services sector contribute to population aging. Fourth, related to the last point, we can then quantify the degree to which aging itself, via general-equilibrium effects, is responsible for the rising relative price of health services.

The model outline proceeds as follows. First, we introduce the model environment. Second, we discuss the decision problem faced by households and outline the state variables that households take into consideration when choosing consumption, health spending, and capital investment. Third, we describe the optimization problem faced by non-health-care producing firms who operate in perfectly competitive markets. We then describe the monopolistically competitive production of health-care services. Fourth, we describe how old-age pensions function. Fifth, we establish market-clearing conditions necessary for general equilibrium to hold. Sixth, we define the model's monopolistically competitive equilibrium. We conclude the model section by defining the general-equilibrium channels through which population aging will cause the relative price of health services to vary.

3.1 Environment

Time, t, is discrete. Each age group is characterized by a single, representative household of age $j \in \{1, \ldots, J\}$. Households can be either of working-age, $j < J_R$, or retired, $j \geq J_R$. They die with certainty after age J, though some fraction of them die each period. They may save by investing in productive capital, and they supply labor inelastically and automatically retire at age J_R. Finally, households receive dividend payments from monopolistically competitive health-services firms which make positive economic profits.

The economy has two sectors. One sector produces non-health-services consumption and investment goods using capital and labor. The other sector is characterized by a continuum of monopolistically competitive health-services firms who are engaged in competition following the Dixit and Stiglitz (1977) framework. Each firm in this sector uses capital and labor to produce its distinguishable health-services product and makes a non-negative (possibly positive) economic profit.

The government's sole purpose is to tax labor and rebate it to retirees using a pay-as-you-go (PAYGO) pension system. In our set-up the government is merely the Social Security and Medicare administrator. We assume the government's budget clears every period, so that receipts from labor taxes are fully rebated to retired households via transfer payments.

3.2 Households

3.2.1 Health

We model the period-t health of an age-j agent as the inverse of the non-accidental mortality rate, as in Hall and Jones (2007). Denote the mortality rate (i.e., the probability an age-j agent dies in period t before becoming an age-$j+1$ agent in period $t+1$) by m_{jt}. Mortality is the sum of accidental mortality (exogenous) and non-accidental mortality (endogenous): $m_{jt} = m_{jt}^{acc} + m_{jt}^{non}$. The survival rate (i.e., the probability an age-j agent in period t becomes an age-$j+1$ agent in period $t+1$) is $s_{jt} = 1 - m_{jt}$.

Denote health status by x_{jt} and health expenditure by h_{jt}. Health status is the inverse of the total mortality rate: $x_{jt} = 1/m_{jt}$. Meanwhile, health spending only reduces the non-accidental mortality rate. Let the inverse of the non-accidental mortality rate be $\widetilde{x}_{jt} = 1/m_{jt}^{non}$, which we take to be the following function of age-specific health expenditure:

$$\widetilde{x}_{jt} = z_t\, \phi_j(\zeta_{jt} h_{jt})^{\theta_j} \tag{1}$$

As in Hall and Jones (2007), z_t is an exogenous aggregate productivity term which explains the efficiency by which health investment can actually be converted to health outcomes. This object is assumed to grow at a constant rate, g_z, so that $z_{t+1} = (1+g_z)z_t$. Meanwhile, ζ_{jt} is an exogenous age-specific productivity term, which we will calibrate to match estimates extracted from Hall and Jones (2007). We also allow the time-independent output intensity, ϕ_j, and elasticity, θ_j, to vary by age. Given (1) survival rates can be written

$$s_{jt} = 1 - m_{jt}^{acc} - 1/\widetilde{x}_{jt} \tag{2}$$

Total health is a function of health expenditure, via $\widetilde{x}_{jt}$, as follows:

$$x_{jt}(h_{jt}) = \frac{1}{m_{jt}^{acc} + \frac{1}{\widetilde{x}_{jt}(h_{jt})}} \tag{3}$$

What does the age-specific productivity $z_t\zeta_{jt}^{\theta_j}$ capture? It describes the efficiency by which health investment, as measured in units of outlay or real consumption dollars, actually scales health outcomes. Any technological improvements to the delivery of health services by medical professionals or health-care institutions will be embedded in the real value of h_{jt}, which is increasing in sectoral TFP, which we will talk more about below. As Hall and Jones (2007) point out, $z_t\zeta_{jt}^{\theta_j}$ captures variation orthogonal to variation in TFP. For example, the efficiency of health investment at generating better health outcomes may increase because of policies that reduce pollution, mandate safer food preparation practices, and enforce water-quality standards. Additionally, along the growth path, people have learned how to live healthier lives (e.g., smoking less, reducing intake of certain kinds of processed foods, exercising regularly, taking vitamins, having safe sex, etc.) so that every dollar invested in health services, as measured in the NIPA tables, goes further toward generating health outcomes and reducing non-accidental mortality rates.

It is also not difficult to think of reasons as to why this term is age specific. Think, for example, of changes to knowledge with regards to how we engage in infant care: babies are not lain prostrate anymore due to risk of sudden infant death syndrome. While such knowledge was indeed acquired via medical research, this knowledge does not directly affect the *quantity* of health-services consumption (i.e., health investment) the consumer purchases, but rather affects the *efficiency* by which such investments may be converted into higher health and lower risk of mortality for humans of a specific age. What about an example of technological improvements to health investment that may affect agents at different ages differently? Well, take the smoking example again: a sudden change to regulations on indoor smoking will affect z_t, yes, but the effect may not be uniform across generations. After all, a generation of middle-aged adults who spent 20-30 years inhaling second hand smoke in indoor spaces will still face significant risks of chronic respiratory illnesses, so that for these adults ζ_{jt} may reduce the impact of the increase in z_t on health outcomes, while for younger people ζ_{jt} may actually scale the impact positively.

3.2.2 Demographics

Let N_{jt} denote the total number of agents of age j that are alive in period t. This object evolves endogenously as follows:

$$N_{j+1,t+1} = s_{jt} N_{jt}, \qquad \forall j > 0 \tag{4}$$

We assume that $s_{Jt} = 0$, forcing $N_{J+1,t+1} = 0$, for all t. The population level at which a cohort enters the economy is denoted by N_{1t} and exogenously grows at net rate g_N. We assume that all migration for each cohort happens at the beginning of working-age life.

3.2.3 Preferences

Households have preferences over their non-health-services consumption, c_{jt}, again following a simplified version of Hall and Jones (2004) and Hall and Jones (2007):

$$u(c_{jt}) = \chi + \xi \, \frac{c_{jt}^{1-\gamma}}{1-\gamma} \tag{5}$$

$\chi > 0$ is an intercept that forces utility to be positive so that consumers have an incentive to live.[5] This will become more apparent when the household's dynamic optimization problem is specified below. ξ scales the contribution of non-durable flow consumption to utility and is needed in order to match the health-services expenditure share. We do not allow for households to have direct preferences over health, but rather health only affects survival rates and thus life expectancy and longevity.

3.2.4 Wealth

Households may save by investing ι_{jt} toward the accumulation of capital assets a_{jt}. Further, given some fraction of each cohort's members perish each period, we must re-distribute the assets bequeathed by those who die. Let b_t denote the bequests received by any living agent at the beginning of period t. These bequests were bequeathed accidentally by those who died in period $t-1$. We assume all living agents receive the same bequests, so we do not index these objects by j. Personal assets evolve according to the law of motion:

$$a_{j+1,t+1} = (1-\delta)(a_{jt} + b_t) + \iota_{jt} \tag{6}$$

$(1-\delta)$ is the gross rate of capital depreciation in the economy. Note that bequests are assumed to be received at the beginning of the period prior to depreciation occurring. Finally, we assume that $a_{J+1,t+1} = 0$ for all t, so that everyone wishes to consume all of their assets prior to entering the terminal phase of life.

3.2.5 Income & Budget Constraint

Working-age households are endowed with one unit of labor, which they supply inelastically and earn after-tax labor income $w_t \eta_j (1-\tau_t)$, where w_t is the economy-wide average wage, η_j captures the hump-shaped life-cycle wage profile following Hansen (1993), and τ_t is the net tax on labor earnings, which the government requires to fund retiree pensions and health spending. Households also earn income on capital holdings, $r_t(a_{jt} + b_t)$, as well as dividends, π_{jt}, from the economic profits of health-services firms operating under oligopolistic competition. Dividends are assumed to be automatically rebated back to households, so that households do not actively choose to invest in health-services firms but are simply credited with some fraction of their profits.

Let p_t be the market price of health-services consumption in units of non-health-services consumption.[6] For working-age adults, the budget constraint is

$$c_{jt} + p_t h_{jt} + a_{j+1,t+1} \leq w_t \eta_j (1-\tau_t) + R_t(a_{jt} + b_t) + \pi_{jt}, \qquad j < J_R \tag{7}$$

Upon retirement households no longer work. Their labor income is then supplanted by transfers,

[5] Note that Hall and Jones (2004) allow for these objects to vary across ages and time, while Hall and Jones (2007), in the published draft, fix $\chi_j = \chi$. We choose the latter approach, as well.

[6] We assume the price of non-health-services consumption is the numeraire each period.

T_t, they receive from the combined Social Security and Medicare administrators. We assume all individual elderly agents receive the same transfer in a given period. They may still choose to save and thus earn returns on capital investments. The budget constraint of retirees is

$$c_{jt} + p_t\, h_{jt} + a_{j+1,t+1} \leq T_t + R_t(a_{jt} + b_t) + \pi_{jt}, \qquad j \geq J_R \tag{8}$$

3.2.6 Choices & Optimization

Households face only idiosyncratic uncertainty over survival. The paths of prices are assumed known because sectoral TFP's driving the evolution of these prices will grow deterministically. Further, N_{1t} and z_t grow deterministically at constant rates, while ζ_{jt} is assumed to grow deterministically as well but at rates that may or may not be constant. Finally, because households know how the population distribution will evolve, they also know how b_t will evolve under the assumption they know all age-specific policy functions. The only endogenous state variable is a_{jt}.

Households choose c_{jt}, h_{jt}, and $a_{j+1,t+1}$ to solve the following recursive optimization problem subject to their age-specific budget constraint:

$$\mathcal{V}_{jt}(a_{jt}) = \max_{c_{jt}, h_{jt}, a_{j+1,t+1}} \left\{ u(c_{jt}) + \beta\, s_{jt}(h_{jt})\, \mathcal{V}_{j+1,t+1}(a_{j+1,t+1}) \right\} \tag{9}$$

where the survival rate, s_{jt}, is a function of health expenditure, h_{jt}. We assume that the value of death is zero.

3.3 Production

3.3.1 Consumption & Investment Producers

The markets for investment products and non-health-services consumption are assumed perfectly competitive and thus can be characterized by a representative firm. Non-health-services consumption and investment is produced with Cobb-Douglas technology, using capital K_{ct} and labor L_{ct}:

$$C_t + I_t = A_{ct} K_{ct}^{\alpha_c} L_{ct}^{1-\alpha_c} \tag{10}$$

3.3.2 Health Care Industry

The health care industry is assumed (possibly) imperfectly competitive with a Dixit and Stiglitz (1977) structure. Assume that each period there exists a unit continuum of health-care firms that produce distinguishable health products, h_{it}, where $i \in (0, 1)$ indexes the firms. The sector is characterized by free entry, though we take no stand on the fixed costs of entry (nor do we need to). These firms are effectively wholesalers, as in the Dixit and Stiglitz (1977) set-up. Let H_t denote the aggregate consumption of health-care services, $N_t = \sum_j N_{jt}$ denote aggregate population, and

h_t denote consumption per-capita, so that

$$h_t = \frac{H_t}{N_t} = \frac{\sum_{j=1}^{J} N_{jt} h_{jt}}{\sum_{j=1}^{J} N_{jt}} \tag{11}$$

We take no stand as to whether the distinguishable varieties of health services are consumed to different degrees by agents of different ages. Instead, as in Dixit and Stiglitz (1977), we assume there exists a single, representative health-services retailer operating in a perfectly competitive market. This retailer buys distinguishable inputs, h_{it}, from wholesalers and sells its output of final health services to consumers according to the CES aggregator:

$$h_t = \left(\int_0^1 h_{it}^{\frac{1}{\mu_t}} \, \mathrm{d}i \right)^{\mu_t} \tag{12}$$

In this set-up μ_t is the period-t aggregate mark-up, which is exogenous. The retailer chooses inputs h_{it} to maximize per-capita profits:[7]

$$\pi_t = \mathcal{P}_t h_t - \int_0^1 p_{it} \, h_{it} \, \mathrm{d}i \tag{13}$$

where $\mathcal{P}_t$ is some price index, encoding the value of per-capita health services expenditure. This is not necessarily the same as p_t. Maximizing (13) yields the standard expression for h_{it} as a fraction of per-capita consumption:

$$h_{it} = \left(\frac{p_{it}}{\mathcal{P}_t} \right)^{\frac{\mu_t}{1-\mu_t}} h_t \tag{14}$$

We can plug (14) into (12), invoking homogeneity, to derive an expression for $\mathcal{P}_t$:

$$\mathcal{P}_t = \left(\int_0^1 p_{it}^{\frac{1}{1-\mu_t}} \, \mathrm{d}i \right)^{1-\mu_t} \tag{15}$$

Now, consider the problem faced by a monopolistically competitive wholesaler. Taking the retailer's demand for input h_{it} as given from (14), producer i chooses the price p_{it} that solves

$$\operatorname{argmax} \left\{ p_{it} \left(\frac{p_{it}}{\mathcal{P}_t} \right)^{\frac{\mu_t}{1-\mu_t}} h_t - mc_{it} \left(\frac{p_{it}}{\mathcal{P}_t} \right)^{\frac{\mu_t}{1-\mu_t}} h_t \right\} \tag{16}$$

Taking demand per-capita, h_t, as given and solving (16), we get the mark-up for variety i as a function of the marginal cost: $p_{it} = \mu_t mc_{it}$.

We assume that all health-care wholesalers use capital and labor in the production of their

[7]Note the distinction between variables associated with the age-j representative agent and per-capita variables: variables associated with the age-j agent have a j subscript, while per-capita variables are still lower case but lack a j subscript.

menu of health-care items. Assume further that prices across wholesalers' varieties are similar enough, so that we can invoke symmetry across the health-care wholesaler marketplace. Then, invoking symmetry, replace mc_{it} with mc_t and p_{it} with p_t, the sector-wide price level of health services. It is well-known that under Cobb-Douglas production a firm's marginal cost can be written

$$mc_t(w_t, r_t\,;\, A_{ht}) = \left(\frac{r_t}{\alpha_h}\right)^{\alpha_h} \left(\frac{w_t}{1-\alpha_h}\right)^{1-\alpha_h} \left(\frac{1}{A_{ht}}\right) \tag{17}$$

where A_{ht} is the total factor productivity of the health services sector. Symmetry and (17) imply that production in the health sector can be written:

$$H_t = A_{ht} K_{ht}^{\alpha_h} L_{ht}^{1-\alpha_h} \tag{18}$$

From (17) it also follows mc_t will vary over time as a result of changing TFP's. The sector-wide mark-up can then be written as a function of the aggregate marginal cost:

$$p_t = \mu_t\, mc_t(w_t, r_t\,;\, A_{ht}) \tag{19}$$

From the perspective of firms, all mark-ups over marginal cost will arise from increases to μ_t, where it must be that $\mu_t > 1$ in order for monopolistic competition to exist. Thus, we must have that the varieties of health-care products on the market are all gross substitutes, but imperfectly so. As $\mu_t \to_+ 1$ mark-ups fall and the varieties become perfect substitutes.

The expression in (19) says that prices can rise either due to mark-ups changing (μ_t rising due to products becoming less substitutable) or the marginal cost changing, where mc_t depends on exogenously varying TFP (A_{ht}) and endogenously varying factor prices (r_t and w_t). Therefore, there are three primary mechanisms that can cause the price of health services to vary:

i. Exogenously varying substitutability of health-care varieties.

ii. Exogenously varying sectoral TFP.

iii. Changes to demand for health services (due possibly to aging) acting via general equilibrium effects on the capital rental rate and aggregate average wages.

Note that exogenous variation in μ_t, as described in (i), would explain several different kinds of changes to the health-services marketplace. If local markets for health services become increasingly concentrated, then μ_t would increase on aggregate causing the aggregate substitutability of health-care varieties, $\mu_t/(\mu_t - 1)$, to fall. As each geographic locale becomes associated with a single monopolistic health-services provider, the substitutability of services falls, since receiving health services for many routine and non-routine treatments still requires consumers to meet in person with medical professionals (e.g., for tests, check-ups, surgeries, therapies, x-rays, scans, and most other examinations). Thus, a rising μ_t captures aggregate implications from increases in the kind of local market concentration described in Cooper et al. (2019).

Exogenously varying sectoral TFP, as described in (ii), will impact changes to the marginal cost, holding input prices fixed. We are also interested in the degree to which the rise in health-services prices can be explained by (iii) absent variation in (i) and (ii). In Section 4 we will dissect the factors contributing to relative-price variation by simulating a calibrated version of our model along transition paths that account for endogenous population aging and possible exogenous changes to mark-ups and relative sectoral TFP's.

3.3.3 Health Care Industry Profits

The substitution elasticities will determine the relative profits of health-service providers. Assume that profits are proportionally distributed as dividends to capital holders. Invoking the symmetry of the unit continuum of health-services producers, aggregate profits generated by the health-services industry are

$$\Pi_t = \mu_t \, mc_t(w_t, r_t \,;\, A_{ht}) \, H_t - mc_t(w_t, r_t \,;\, A_{ht}) \, H_t \tag{20}$$

where $H_t = N_t h_t = \sum_{j=1}^{J} N_{jt} h_{jt}$ is aggregate health-services consumption, and $\Pi_t = N_t \pi_t = \sum_{j=1}^{J} N_{jt} \pi_{jt}$. The consumption and investment producers earn zero economic profits since they operate in perfectly competitive markets.

Agents receive dividends from aggregate profits in proportion to their period-t asset holdings, including accidental-bequests received:

$$\pi_{jt} = \left(\frac{a_{jt} + b_t}{K_t} \right) \Pi_t \tag{21}$$

where K_t is the aggregate capital level. This ensures that dividends enhance chosen investment and is designed to ensure the wealth distribution over the life cycle is realistic.

3.4 Governmental Transfers

The government taxes labor income and rebates it to retirees in the form of lump-sum transfers designed to cover both medical and non-medical expenditure. In each period we re-distribute labor-income tax receipts evenly to all individual retirees:

$$T_t = w_t \, \tau_t \, \frac{\sum_{j=1}^{J_R - 1} N_{jt} \eta_j}{\sum_{j=J_R}^{J} N_{jt}} \tag{22}$$

The government budget constraint then clears by construction.

3.5 Aggregation & Market Clearing

Markets for consumption, health services, and investment satisfy:

$$\sum_{j=1}^{J} N_{jt} c_{jt} = C_t \tag{23}$$

$$\sum_{j=1}^{J} N_{jt} h_{jt} = H_t \tag{24}$$

$$\sum_{j=1}^{J} N_{jt} \iota_{jt} = I_t \tag{25}$$

Capital markets must satisfy:

$$\sum_{j=1}^{J} N_{jt}(a_{jt} + b_t) = K_{ct} + K_{ht} = K_t \tag{26}$$

Total bequests given must satisfy total bequests received:

$$\sum_{j=1}^{J} N_{j,t-1}(1 - s_{j,t-1}) a_{jt} = \sum_{j=1}^{J} N_{jt} b_t \tag{27}$$

Labor markets must satisfy:

$$\sum_{j=1}^{J} N_{jt} \eta_j = L_{ct} + L_{ht} = L_t \tag{28}$$

Finally, health-services profits must satisfy:

$$\sum_{j=1}^{J} N_{jt} \pi_{jt} = \Pi_t \tag{29}$$

3.6 Multi-sector, Monopolistically Competitive Overlapping Generations Equilibrium with Transfers

Given deterministic sequences of Social Security and Medicare tax rates $\{\tau_t\}_t$, deterministic sequences of total-factor productivities $\{A_{ct}, A_{ht}\}_t$, deterministic sequences of newborns $\{N_{1t}\}_t$, deterministic sequences of accidental mortality rates $\{m_{jt}^{acc}\}_{j,t}$, known exogenous sequences of health productivity rates $\left\{z_t, \{\zeta_{jt}\}_j\right\}_t$, and sequences of mark-ups $\{\mu_t\}_t$, a monopolistically competitive equilibrium with transfers consists of:

i. Sequences of household policies $\{c_{jt}, h_{jt}, a_{j+1,t+1}\}_{j,t}$.

ii. Sequences of producers' policies $\{K_{ct}, L_{ct}, K_{ht}, L_{ht}\}_t$.

iii. Sequences of prices $\{p_t, r_t, w_t\}_t$.

iv. Sequences of population distributions $\{N_{jt}\}_{j,t}$.

v. Sequences of bequests $\{b_t\}_t$.

vi. Sequences of dividends $\{\pi_{jt}\}_{j,t}$.

vii. Sequences of transfers $\{T_t\}_t$.

such that

a. Household policies solve the optimization problems for both workers and retirees.

b. The consumption and investment producers maximize profits.

c. Health-services prices satisfy the mark-up condition.

d. Population is exactly the number of survivors from the previous period plus the number of newborns and migrants.

e. Bequests distributed are exactly equal to the leftover assets of the recently deceased.

f. Dividends are distributed in proportion to asset holdings.

g. Transfers equate with total governmental revenues.

h. Markets for consumption, investment, health services, capital, and labor clear.

3.7 The Rising Price of Health Services

How do aging, changing TFP's, and mark-ups contribute to changes in the price of health services relative to all other consumption (i.e., p_t)? Using (17) in (19) plus the marginal cost from the perfectly competitive consumption/investment sector, we can re-write the relative price of health-services consumption as a function of only mark-ups, relative TFP's, and the relative price of capital to labor, r_t/w_t, which sufficiently captures the effects of the changing composition of household expenditure via general equilibrium outcomes:

$$p_t = \mu_t \left(\frac{r_t}{w_t}\right)^{\alpha_h - \alpha_c} \left(\frac{A_{ct}\alpha_c^{\alpha_c}(1-\alpha_c)^{1-\alpha_c}}{A_{ht}\alpha_h^{\alpha_h}(1-\alpha_h)^{1-\alpha_h}}\right) \tag{30}$$

Note that (30) clearly shows that if the relative price of health services, p_t, varies in response to changes in the composition of household expenditure, then it must be the case that the capital intensity of health services production does not equal the capital intensity of production for all other consumption ($\alpha_h \neq \alpha_c$). This is because in (30) all general equilibrium effects act through the ratio, r_t/w_t, since all other components are either exogenously time-varying (μ_t and A_{ct}/A_{ht}) or parameters (α_h and α_c). Taking logs of (30) and differentiating in $\ln(r_t/w_t)$, the elasticity of the relative price of health services with respect to the input-price ratio is $\alpha_h - \alpha_c$. As the economy ages we should expect the ratio of returns and wages (r_t/w_t) to fall (Prettner 2013; Backus, Cooley, and Henriksen 2014; Cooley and Henriksen 2018; Cooley, Henriksen, and Nusbaum 2019). Thus, is $\alpha_h < \alpha_c$, p_t should rise in response to aging.

Alonso-Carrera, Caballé, and Raurich (2015) note that differential capital-intensities in production are required for general-equilibrium outcomes from the changing composition of demand to affect relative prices. Alonso-Carrera, Caballé, and Raurich (2015) show that identical sectoral capital intensities are necessary to guarantee constant relative consumption prices along a growth path. Further, the authors also show that in the absence of constant sectoral capital intensities between the two consumption sectors, constancy of relative consumption prices can still be guaranteed as long as the relative price of labor and capital is itself constant. Thus, $\alpha_h \neq \alpha_c$ is necessary but not sufficient to ensure variation in q_t along the growth path. This is because, with Cobb-Douglas sectoral production functions the re-allocation of total expenditure affects prices via the non-proportional re-allocation of capital and labor inputs (Alonso-Carrera, Caballé, and Raurich 2015; Herrendorf, Herrington, and Valentinyi 2015).

It should be fairly clear to the reader that p_t will always rise as mark-ups rise, but it will also rise as the relative productivity of the production of non-health-care consumption outpaces the growth in health-services total factor productivity, A_{ht}. This is true regardless of the values of α_h and α_c, so that our model set-up can easily quantify the impact on p_t of the two competing forces discussed in Horenstein and Santos (2019) — relatively increasing mark-ups and a relative slowdown in health-services TFP's. In the pages that follow we calibrate the full general equilibrium model and simulate its transition path under different counterfactual scenarios in order to pin down the degree to which population aging, relative TFP variation, and sectoral mark-ups have driven up the relative price of U.S. health services over the last 50 years.

4 Simulation Exercises

We calibrate the model's transition path to an observable time series of data following methods deployed in Krueger and Ludwig (2007). Our primary goal is to match data series over the period 1960-2015 for p_t, life expectancy, the health share of total expenditure, and the working-age population ratio.[8] We take several estimation cues from Hall and Jones (2007), as well as numerical cues from Carroll (2006) and Krueger and Ludwig (2007), which we allude to in the pages that follow. Following calibration we engage in several counterfactual exercises, fixing various variables of interest in order to understand how sensitive model outputs are to various forces driving the structural rise in the health services share.

[8]Note that we abstract from comparing our simulations against data for 2020, intentionally. Our model (and this paper) is about what forces are driving long-run changes to the composition of the economy as it pertains to the role of health-care services. Short-term fluctuations are not the key focus. To this point the 2020 recession, however brief it was, significantly affected households' access to certain essential services, health care being one of them, thus impacting its aggregate share of expenditure. To the extent that such impacts are long-lasting or permanent, we do not know and can only speculate at this point.

4.1 Baseline Calibration

The economy begins with an artificial steady state in the year $t = 1$ which is set to 1950. There are $J = 16$ age groups with agents working until age 65 ($J_R = 10$) when they automatically retire. Agents enter the economy at age 20 ($j = 1$) and die automatically at the end of age 99. Agents are binned into five-year age groups (i.e., 20-24, 25-29, 30-34, ..., 95-99). In all of our quantitative exercises, we simulate the economy's transition path forward from an artificial, initial steady state for $\mathcal{T} = 21$ (1950-2050) periods, assuming market concentration, technologies, and population grow over this period at rates we calibrate. We simulate the model's transition path under the assumption that agents take prices in period t as given and do not anticipate how such prices will evolve in the future as a function of growing state variables, following computational cues from Krueger and Ludwig (2007). We calibrate the model's parameters so that the targeted data moments match moments along the transition path, where our data observations begin in period $t = 3$ (1960) and extend out in five-year intervals to period $t = 14$ (2015). Our solution algorithm and the calibrated age-specific and time-specific parameter sets are presented in detail in Appendix B.

4.1.1 Objects Driving Growth

Since the model is calibrated to a transition path, we must take firm stands on how all technologies and the initial population level grow. We assume that the following objects grow at constant rates along the transition path: N_{1t}, z_t, A_{ct}, and A_{ht}, where we let g denote net growth rates.[9] μ_t is also time-varying but we calibrate this series directly to data using estimates from Horenstein and Santos (2019). From Horenstein and Santos (2019) estimates are only available up to 2005. We assume that after 2005 the mark-up does not change. Based on their findings we set $\mu_t = 1$ until 1990. For 1990, 1995, 2000, and 2005, we then let mark-ups follow the following path, based on the results presented in Figure 7b of their paper: $(1.5, 1.6, 1.75, 1.85)$.[10] After 2005, $\mu_t = 1.85$ always, since data is not available after this date. We calibrate ζ_{jt} directly from Hall and Jones (2007). In benchmarking tests we find that model-implied time series are sensitive to starting levels for z_1 and $A_{h,1}$. We will calibrate these directly, while fixing $N_{1,1} = A_{c,1} = 1$.

4.1.2 Health & Survival

The accidental mortality rates we use are exactly the same as those featured in Figure A.3b in Appendix A.4, but they are adjusted to accommodate our five-year age bins. We have five-year, adjusted accidental mortality rates from 1950-2015. Recall that year 2015 is period $t = 14$ in our model. For $t > 14$ we assume that age-specific accidental mortality rates do not change, so

[9] For example, let $g_\neg$ denote the net growth rate of arbitrary variable $\neg$, so that $\neg_{t+1} = (1 + g_\neg)\neg_t$.

[10] Horenstein and Santos (2019) estimate the relative mark-up of health services compared to economy-wide mark-ups. Prior to 1990 economy-wide mark-ups exceeded health-services mark-ups. We assume $\mu_t = 1$ prior to 1990 because of this fact and then scale μ_t according to the estimated mark-up of health services *over* the economy-wide mark-up for subsequent years.

that $m_{jt}^{acc} = m_{j,14}^{acc}$, $\forall t > 14$. Forcing constancy of this time-varying variable is computationally necessary for achieving long-run stationarity.

We estimate $\{\phi_j, \theta_j\}_j$ and $\{\zeta_{jt}\}_{j,t}$ directly from Hall and Jones (2007) using their data on non-accidental survival rates and health expenditure over the life cycle, which they take from Meara, White, and Cutler (2004). We assume the time-varying, age-specific productivity parameters follow the estimated Hall and Jones (2007) trends using their provided estimation code from the paper's supplementary files.

4.1.3 Calibrated Parameters

Table 1 presents the calibrated parameters and the moments with which they are associated. We primarily target four time series in five-year intervals for periods ranging from $3 \leq t \leq 14$ (1960-2015): 1) relative health-services prices $\{p_t\}_t$; 3) life expectancy LE_t; 3) health share of aggregate consumption expenditure, $p_t H_t/(p_t H_t + C_t)$; 4) working-age to retiree population ratio (WAPR), $\sum_{j=1}^{J_R-1} N_{jt}/\sum_{j=J_R}^{J} N_{jt}$. We give secondary consideration to the following: 1) health spending of those aged 20-24 years relative to the average amongst those aged 75+ years; 2) average health spending of those aged 25-34 years relative to the average amongst those aged 55-64 years; 3) health capital share of total capital, K_{ht}/K_t; 4) health labor share of total labor, L_{ht}/L_t; 5) capital/output ratio of 3; 6) five-year average rate of return on capital of $r = 10.4\%$. In Table 1 externally calibrated parameters are either taken directly from sources or computed directly from the associated data series. Parameters that involve time-series or life-cycle profiles are explicitly presented in Appendix B, except for our construction of accidental mortality rates which is discussed in Appendix A.4. Internally calibrated parameters are designed to primarily match our four main time series along the growth path, while ensuring that $K_t/Y_t \approx 3$ by 2050. As a robustness check we compared model-predicted GDP growth rates to data, in Section 4.3. Note that we do not directly target GDP growth in our calibration procedure, but we can still match average-annual GDP growth over the five-year intervals of our target sample, which will become apparent by reading further.

Our calibrated parameters imply several things. First, the model requires that both the utility weight parameter, ξ, and utility intercept, χ, be large in order to match the health share of spending while ensuring that capital/output ratios converge to reasonable values in the long run. Second, we find that TFP growth in the health services sector is positive but very small (i.e., a net growth rate of ≈ 0.0003). This is consistent with the sector being plagued by substantial Baumol's cost disease (Baumol 1967; Baumol, Blackman, and Wolff 1985; Bates and Santerre 2013; Horenstein and Santos 2019). While such a low, average long-run growth rate may seem extreme, it is still consistent with many other estimates from the literature. Blumenthal, Stremikis, and Cutler (2013) and Shatto and Clemens (2022) find that the health-care sector is associated with waste, driving up costs, and thus suppressing TFP. When looking strictly at the hospital sector, Harper et al. (2010) estimate that from 1987-2006 the sector actually experienced *negative* productivity growth, which is also what Cylus and Dickensheets (2007-2008) find for the period 1981-2005.[11] Our calibrations also suggest that the capital intensity of health-sector production is approximately 0.273 which is very close to what Horenstein and Santos (2019) find (0.25), as well as earlier estimates in Donahoe (2000) (0.26).[12] Finally, we estimate time preferences to be greater than one, which is consistent with the estimate from Hurd (1989) for life-cycle models with survival rates augmenting the discount rate on the value of future consumption.[13]

We present plots of predicted values against data, along with counterfactually simulated time series, in Figure 3, which is featured in the next section. Additionally, Table 2 (also next section) shows how model objects grow and change over time, relative to predicted values. The calibrated model captures the primary qualitative nature of how the four primary data series change over time. The relative price, p_t, is the easiest series to hit. Life expectancy and the working-age to retiree population ratio are each sensitive to starting values and growth rates for health-productivity parameters, z_t and ζ_{jt}.[14] The health share of aggregate expenditure is the most difficult series to target, as calibrating this series requires adjusting ξ for every set of health-productivity parameters. The tradeoff is that as ξ increases the health share falls, but because health-services expenditure affects an agent's survival probability, increasing ξ and subsequently causing total health spending and thus life-expectancy to fall, also causes savings rates and thus K_t/Y_t to fall. In our calibration we manage to achieve a capital/output ratio of approximately 3 in the long run (2050), but the level of the health share is consistently five to eight percentage points below the data target for 1960-2015. However, we can match the growth rate over time of the health share very well. In data we find that the health share of aggregate expenditure rose by 15 percentage points from 1960-2015, and our model predicts it rose by 19 percentage points.

Regarding the secondary moments, the model fits the qualitative nature of their intertemporal variation well. The annualized interest rate is approximately 6% and slightly falling over the 1960-2015 period. In data the average health expenditure of 20-24 year-olds relative to 75+ year olds falls 90% from 1960 to 2015, and 62% in our model. For health spending by 25-34 year-olds relative to 55-64 year-olds, the data suggest a 44% decline, while the model predicts a 9% decline. The model over-predicts increases in the shares of capital and labor devoted to health production from 1960-2015. Data suggest a 2.2 percentage point rise in the capital share and an 8 percentage point rise in the labor share, while the model predicts a 10.6 percentage point rise in the capital

[11] By contrast Romley, Goldman, and Sood (2015) argue that sectoral TFP estimates may be downwardly biased by failing to account for the severity of patients' illnesses when using treatment outcomes to assess productivity. Their approach, however, cannot account for the kinds of near-perfectly discriminatory pricing schema hospital systems engage in, as they index all costs to the administrative prices used by Medicare. This would bias their productivity estimates upward, considering that hospitals can mark-up costs at higher rates for patients of private insurers that have less monopsonistic pricing power than Medicare (Cooper et al. 2019).

[12] Jung, Tran, and Chambers (2017) also use the estimate from Donahoe (2000).

[13] Note that in OLG models featuring some positive probability of accidental death, the effective discount rate is βs_{jt}. In our model, since s_{jt} depends on health investment, this discount rate is endogenous. Since s_{jt} is a 5-year survival rate, β can be greater than one and still yield an effective discount rate which satisfies interiority on the unit interval (i.e., $\beta s_{jt} \in (0, 1)$).

[14] Hall and Jones (2007) had similar difficulties matching the life-expectancy growth rate, as the model under-predicts mid-century values but does better at predicting life expectancy more recently.

Table 1: Calibrated Parameters & Sources/Targeted Moments

		Externally Calibrated Parameters	
Parameter	Value	What	Source
$\{\phi_j\}_j$	Appendix B	Health Intercept	Hall and Jones (2007)
$\{\theta_j\}_j$	Appendix B	Health Elasticity	Hall and Jones (2007)
$\{\eta_j\}_j$	Appendix B	Labor Productivity	Hansen (1993)
$\{\zeta_{jt}\}_{j,t}$	Appendix B	Health Productivity by Age	Hall and Jones (2007)
$\{m_{jt}^{acc}\}_{j,t}$	Appendix A.4	Accidental Mortality	CDC, *Health, United States 2017*
$\{\mu_t\}_t$	Appendix B	Mark-ups	Horenstein and Santos (2019)
γ	2.000	Intertemporal Elasticity	Hall and Jones (2007)
α_c	0.400	Capital Intensity	Horenstein and Santos (2019)
δ	0.185	4% annual depreciation	
g_N	0.033	5-yr. growth newborns 1990-2019	
g_{A_c}	0.035	5-yr. U.S. multi-factor prod. growth 1950-2020	
τ	0.086	1990-2000 S.S. + Medicare Average	

		Internally Calibrated Parameters	
Parameter	Value	Moment	
$A_{h,1}$	0.210	Initial Productivity	Relative health services price + health share
g_{A_h}	0.000	Health TFP Growth	Relative health services price + health share
z_1	4.536	Initial Health Productivity	Life expectancy + health share
g_z	0.123	Health Productivity Growth	Life expectancy + health share
χ	166.549	Utility Intercept	Life expectancy + health share
β	1.097	Time Preference	Capital/output ratio + 10.4% 5-yr r_t
α_h	0.273	Capital Intensity	Capital/labor shares in health prod.
ξ	120.000	Utility Weight	Health share of agg. spending

share and a 16.7 percentage point rise in the labor share.

Since our primary goal is to rank, qualitatively, which forces appear most responsible for rising relative prices (e.g., market concentration, Baumol's cost disease, aging), we are satisfied that our calibration provides a decent benchmark against which to assess how various long-run changes to the economy (i.e., increasing market concentration, cost disease, and aging) have affected health prices and the health share, as well as longevity. The model targets primary moments well, even though models with endogenous aging are notoriously difficult to calibrate. We also believe that the model targets qualitative variation in secondary time series sufficiently in order to justify a qualitative assessment of the causes of relative-price growth. Finally, as a robustness check, we will show in Section 4.3 that, despite not directly targeting GDP growth rates, the model can accurately and precisely predict the long-run decline in GDP growth rates. The calibration is thus sufficiently robust in order to examine our counterfactual results via a qualitative lens.

4.2 Simulated Time Series of Interest

We simulate the baseline model under our calibrated parameters to compare it against data, as well as 11 different counterfactuals. Our counterfactuals are designed to provide insights into how sensitive relative prices are to market concentration, differential TFP growth rates, and aging. Since life expectancy and the working-age to retiree population ratio are both endogenous, while market concentration and TFP are exogenous, we must assess the contribution of aging to price and expenditure-share variation by first figuring out which exogenous drivers of growth (e.g., the birth rate, health productivity rates, TFP growth rates, market concentration, etc.) most affect life expectancy and the working-age to retiree population ratio and then comparing the evolution of p_t for such simulations against those which are *not* associated with significant population aging. The four time series we primarily target (i.e., p_t, LE_T, health share, and WAPR) are featured in Figure 3, which includes data observations (black triangles) and baseline predictions (open circles), while counterfactual simulations are plotted with different colored lines. Table 2, meanwhile, shows how predicted and counterfactual time series' rates of change over the 1960-2015 sample period deviate from data.

Table 2: Counterfactual Simulations Relative to Model Predictions

	Counterfactual	p_t Growth*	LE Lost%	Health Share Change*	WAPR Growth*
(1)	Data	168%	2.53	0.15	-30%
(2)	Predicted	212%	—	0.19	-45%
(3)	$\mu_t = 1, \forall t$	70%	0.02	0.08	-45%
(4)	$g_{A_c} = g_{A_h} = 0$, $\mu_t = 1, \forall t$	9%	3.42	-0.06	-43%
(5)	$g_{A_c} = g_{A_h} = 0$	101%	3.42	-0.02	-43%
(6)	$g_{A_c} = g_{A_h} = g_z = 0$	91%	24.98	0.09	-6%
(7)	$g_{A_c} = 0$	101%	3.42	-0.02	-43%
(8)	$g_{A_h} = 0$	212%	0.00	0.19	-45%
(9)	$g_z = 0$	197%	21.27	0.34	-12%
(10)	$g_N = 0$	220%	0.08	0.17	-63%
(11)	$g_{\zeta_j} = 0$	210%	2.27	0.23	-42%
(12)	$g_{\zeta_j} = g_z = 0$	194%	25.73	0.36	1%
(13)	$g_{\zeta_j} = g_z = g_N = 0$	197%	25.75	0.35	-30%

* Percent change or percentage point difference from 1960-2015.

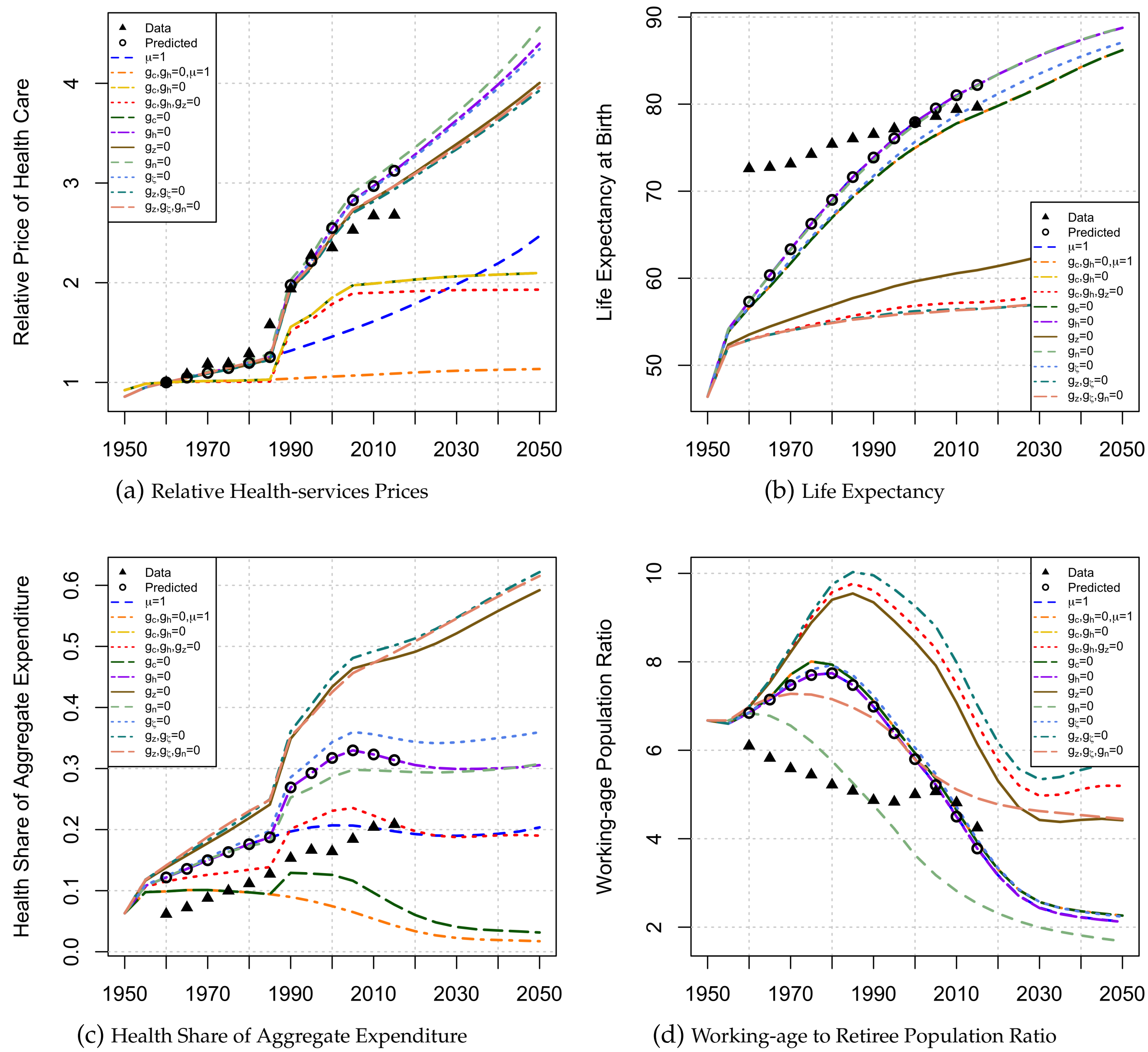
% Life expectancy years lost relative to 2015 baseline prediction. Negative values imply life-years gained.

Our counterfactual simulations roughly fall into three primary constellations, depending on which channel drives the main results. First, we consider situations in which changing market concentration is primarily shut down. These are simulations (3) and (4) of Table 2. Next, we examine different situations associated with different rates of relative sectoral TFP growth, while allowing for market concentration to progress as estimated by Horenstein and Santos (2019). These are simulations (5) through (8) in Table 2. Third, we consider what happens to the four targeted data series when factors that primarily contribute to population aging stop growing. These are simulations (9) through (13) in Table 2.

In simulations where $\mu_t = 1, \forall t$ the health-services market is as competitive as the rest of the economy. We observe that relative price growth is most sensitive to market concentration, since counterfactually growth is 70% when market concentration is completely turned off (line 3) and just 9% when TFP growth rates are also turned off (line 4). The simulation in line (4) essentially freezes the economy, from the production side, at its 1950 productivity and market concentration levels, but still allows the productivity associated with health investments to grow and cause survival rates to increase. Thus, line (4) shows the effects of endogenous aging, independent of market concentration and price growth. Note that life expectancy falls in this simulation where prices barely grow, the health share falls, and the working-age population ratio also falls at about the same rate as the predicted baseline (43% decline versus a 45% decline). Upon examination of line (4) we can see that aging appears to have had little effect on rising p_t and the rising health share.

Lines (5) through (8) collect simulations that aim to tease out the effects of structural transformation due to Baumol's cost disease, while allowing market concentration as well as the primary drivers of population aging to evolve as estimated. In line (5), turning off all aggregate TFP growth, relative prices only double over this period compared to the 212% (tripling) growth in the predicted baseline model, suggesting that TFP growth is responsible for just less than half of the rise in relative prices. The population ages slower in this simulation, but not by much: life expectancy falls by 3.42 years but the working-age to retiree population ratio still falls by 43% compared to 45% in the baseline. Line (6) shows what happens had both health-investment TFP (i.e., what Hall and Jones (2007) attribute to the exogenous affects of pollution mitigation, cleaner standards of living, more sanitary medical procedures, etc.) and aggregate sectoral TFP remained fixed at their mid-century levels. Here, the population does not age in terms of WAPR, but that is because life expectancy does not grow. The effects of just shutting off growth to the low-cost sector (i.e., flipping which sector is associated with Baumol's cost disease) are provided in line (7). Note that these outcomes are almost identical to those in line (5) because $g_{A_h} \approx 0$. Similarly, turning off growth of A_{ht} makes the counterfactual simulation in line (8) look almost exactly like the predicted baseline in line (2).

In lines (9) through (13) we turn off only the growth channels which would most directly affect the aging process. These are g_z, g_N, and g_{ζ_j}, with lines (12) and (13) containing simulations where combinations of these growth channels are shut down. Generally speaking, the simulations collected in lines (9) through (13) suggest that the effects of population aging on relative prices are subtly mixed but somewhat insignificant. First, population aging due to increasing longevity (i.e., growing z_t and ζ_{jt}) appears to have positively impacted relative price growth but only by a few percentage points. This is because relative prices still rise by at least 194% in lines (9), (11), and (12) when either g_z, g_{ζ_j}, or both are shut down and life expectancy still falls. On the other hand, when



(a) Relative Health-services Prices

(b) Life Expectancy

(c) Health Share of Aggregate Expenditure

(d) Working-age to Retiree Population Ratio

Figure 3: In this figure we present data, model-predicted outcomes, and counterfactual model-predicted outcomes for the relative price of health services, normalized to unity in 1960 in panel (a), life expectancy at birth in panel (b), the health share of aggregate expenditure in panel (c), and the working-age to retiree population ratio in panel (d).

we shut down growth of the initial population entering the economy in line (10), we observe that relative prices rise even more and the population distribution ages rapidly with the working-age to retiree population ratio falling by 63% from 1960-2015 as opposed to 45% in the baseline. Thus, the 220% rise in relative prices compared to 212% in the baseline is reflective of the contribution of aggregate demand weighing ever-more on the less elastic, less productive demand for health services of older consumers. The effects of aging via this channel are muted in the baseline model by the fact that younger, healthier consumers are still entering the economy.

Finally, shutting all population growth channels down in line (13) predictably causes aging to slow down, life-expectancy growth to slow, and growth in p_t to slow, despite the fact that the health share rises due to increases in quantities of health-services demanded. Why does the health share of spending rise in this case (as well as in line (12))? The economy is still growing due to rising A_{ct}, so incomes are growing, but the contribution of health investment to longevity is stagnant. Consumers want to live to enjoy their retirement consumption, and since the elasticity of health investment with respect to health is inelastic for all consumers, the income effects drive consumers to devote increasing shares of wallet to ensure they survive.

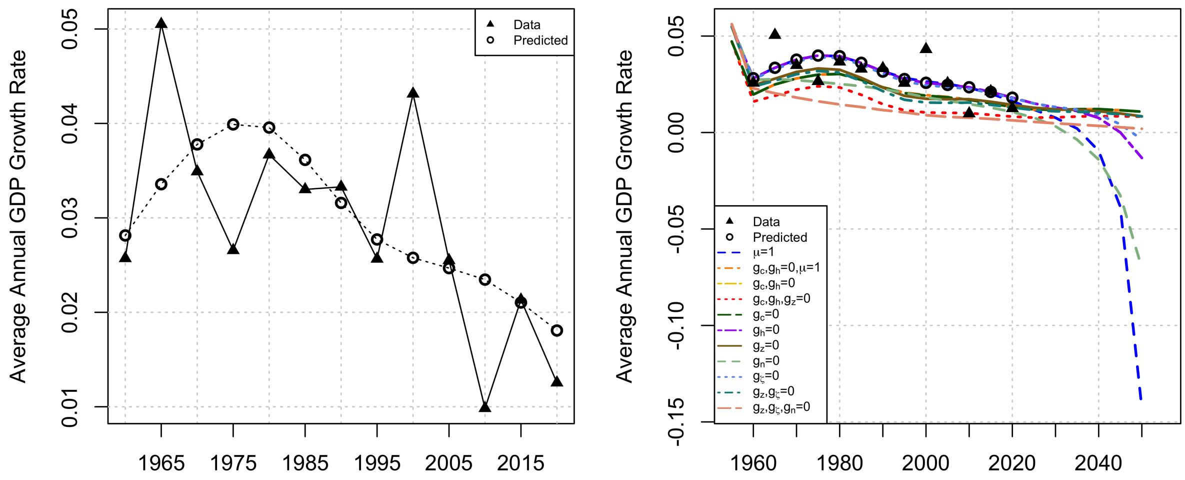
4.3 Implications for Aggregate Growth

Table 3: Simulated GDP Growth Rates

	Counterfactual	1960	1980	2000	2020	2040
(1)	Data	0.026	0.037	0.043	0.013	—
(2)	Predicted	0.028	0.040	0.026	0.018	0.008
(3)	$\mu_t = 1, \forall t$	0.028	0.040	0.026	0.016	-0.010
(4)	$g_{A_c} = g_{A_h} = 0, \mu_t = 1, \forall t$	0.020	0.030	0.019	0.013	0.012
(5)	$g_{A_c} = g_{A_h} = 0$	0.020	0.030	0.019	0.013	0.012
(6)	$g_{A_c} = g_{A_h} = g_z = 0$	0.016	0.023	0.010	0.008	0.008
(7)	$g_{A_c} = 0$	0.020	0.030	0.019	0.013	0.012
(8)	$g_{A_h} = 0$	0.028	0.040	0.026	0.018	0.008
(9)	$g_z = 0$	0.024	0.033	0.018	0.014	0.011
(10)	$g_N = 0$	0.027	0.025	0.018	0.011	-0.014
(11)	$g_{\zeta_j} = 0$	0.028	0.039	0.025	0.018	0.009
(12)	$g_{\zeta_j} = g_z = 0$	0.024	0.031	0.016	0.013	0.010
(13)	$g_{\zeta_j} = g_z = g_N = 0$	0.023	0.015	0.009	0.006	0.003

Our model allows us to understand how aging affects aggregate GDP growth rates. Note that aging is primarily driven by increased longevity, which is itself partially driven by increasing health-services consumption. Further, the health-services sector is a low-productivity, imperfectly competitive sector. As previously noted, aging has been attributed to growth decelerations (Backus, Cooley, and Henriksen 2014; Cooley and Henriksen 2018; Cooley, Henriksen, and Nusbaum 2019; Kydland and Pretnar 2019; Maestas, Mullen, and Powell 2023). How much of this deceleration may result from a low-productivity sector like health services increasing in aggregate share as the population ages? In this section we simulate average-annual growth over five-year intervals under both our baseline calibration and counterfactual simulations to help answer this question. Recall that a model period is five years. We thus compute average annual growth using model-implied GDP, where GDP is $Y_t = C_t + I_t + p_t H_t$. We multiply H_t by p_t to put everything in units of non-health-care consumption. Average annual GDP growth within five-year intervals is then $g_Y = (Y_t/Y_{t-1})^{1/5} - 1$.

Table 3 presents the net GDP growth rates from data, the baseline model, and all counterfac-



(a) g_Y, Baseline and Data

(b) g_Y, Counterfactuals

Figure 4: We plot average-annual GDP growth rates in five-year intervals, starting in the period 1956-1960 and ending in the period 2016-2020. Panel (a) compares baseline predictions to the data. Data points (triangles) and predicted values (circles) are featured in the final year of the five-year interval they correspond to (i.e., the value in 1960 is the average-annual growth rate for the five years from 1956-1960). We drop the initial artificial steady state (1951-1955) from our pictures, so that the simulated values only affect those along the transition path. Panel (b) shows the data, predicted baseline, and counterfactual time series.

tual simulations. Figure 4 shows the model's baseline, predicted average-annual GDP growth rates over five-year intervals against those from data in panel (a), while including the counterfactual time series of growth rates in panel (b). The calibrated model predicts the average decline in growth rates we observe from mid-century to 2020. In fact the model suggests growth rates peaked in the mid-1970s at 4% annually before falling to approximately 1.8% in the 2016-2020 period, which is slightly higher than the data estimate of 1.3%. The quality of fit of the predicted model lends credibility to our counterfactual results, which isolate the different growth channels to understand how their predictions deviate from the predicted baseline. Our counterfactuals allow us to understand how growth rates would have differed had either the population not aged, structural change not occurred, nor market concentration increased.

The results suggest that population aging has had a subtle but mixed impact on GDP growth rates, though evidence that growth is decreasing in aging is undercut by some of our results. In Table 3 when we turn off the economy-wide consumption-side growth parameters (N_{1t} and z_t), we find that average-annual growth rates decline faster than they do in the baseline. Turning off growth in ζ_{jt}, however, has no effect on GDP growth rates. Finally, aggregate growth appears to be most affected by turning off all of the demand-side parameters in line (13). In this case GDP growth slows significantly, though price growth in Table 2 continues to rise and the population continues to age, despite the fact that life expectancy declines precipitously. Further, in this simulation health becomes the dominant sector in the economy by the mid-twenty-first century. This is because, again, in this simulation there are no gains over time to health production, nor are their improvements to health TFP on the supply side. Baumol's cost disease thus manifests itself due to demand-side factors: health production with respect to health investment is inelastic, but investing in health is still required in order to enjoy other consumption. Becaue of this, the health sector comes to dominate, as consumers demand more and more health investment in order to live long enough to enjoy their consumption savings. Inputs of capital and labor which would have been otherwise used for production in the high-productivity sector thus flow to the low-productivity one, driving down aggregate GDP growth rates.

Simulations (3) through (8) reveal other notable aspects of this exercise. Mark-ups appear to have little impact on aggregate growth. Not surprisingly, growth rates fall when the high-productivity sector does not grow (i.e., $g_{A_c} = 0$). But focusing on simulation (6), note that when we turn off all sectoral productivity growth, as well as growth in health-production TFP (z_t), aggregate growth collapses. Comparing simulations (5) and (6), by 2020 when $g_z = 0$ in addition to $g_{A_c} = g_{A_h} = 0$ growth rates are five percentage points lower. This demonstrates that growth in factors orthogonal to production TFP (i.e., improvements to the consumption environment, like eliminating pollution, improving sanitation, etc.) has had almost an equivalent impact on aggregate outcomes as growth in TFP itself. As z_t rises consumers get more for every dollar they spend from services produced by a slow-growing sector, limiting this sector's prominence in the aggregate economy. This exercise thus demonstrates the importance of general equilibrium effects, originating in patterns of household demand, for *determining* aggregate GDP outcomes.

5 Conclusion

Relative health-services prices have risen because the health sector grows slowly and has high market concentration. Population aging driving up demand for health services has played only a small role. Further, increasing health-services prices have not appeared to adversely impact gains to life expectancy, nor is population aging necessarily responsible for declining GDP growth rates.

Our findings have broader implications for policy. Policymakers concerned with rising health costs should actually take solace in our results. If we were to have found that aging is a major contributor to rising prices, policymakers would have less of an impetus to act to make changes because they tend to lack tools to control population aging. However, given increasing market concentration and slow productivity growth are primarily to blame, policymakers can use many tools at their disposal to curb price growth, namely anti-trust regulations and the greater encouragement of technology adoption in the health sector to improve productive efficiency. Further, coupled with micro-level findings in Cooper et al. (2019), our results suggest that greater concentration amongst health services payers, either via monopsonistic insurers or a single-payer system, could also be a solution to help drive down relative prices and counter the market power of health-services providers.

A Data Appendix

This appendix describes the data series used in our both our descriptive and structural analyses. We consider the shares of GDP for both direct health spending and spending on indirect governmental administrative matters. We describe how we construct our relative health-services price index. We also discuss the sources of our accidental and non-accidental mortality data, where we consider the inverse of latter to be a direct function of health spending at the cohort-level.

A.1 Aggregate Health Shares

Nominal consumption expenditure data are taken from the BEA's NIPA Table 2.5.5, which runs through 2021 as of this draft. We focus on the PCE data directly. For those who are unfamiliar with the NIPA PCE data, see Chapter 5 of the BEA's NIPA handbook (*Concepts and Methods of the U.S. National Income and Product Accounts* 2022). PCE data, including health services expenditure, includes new purchases of all goods and services made by households, non-profits acting in service of households, and purchases abroad by U.S. residents while traveling. It also includes expenditure made by third-party payers but the services from which are ultimately utilized by households. This means that "Health" expenditure as it is presented in NIPA Table 2.5.5. is the value of all out-of-pocket expenditures made by households plus the value of all health expenditures financed by either employer-paid or personal health insurance plus the value of health expenditures financed via various government programs, including Medicaid and Medicare.

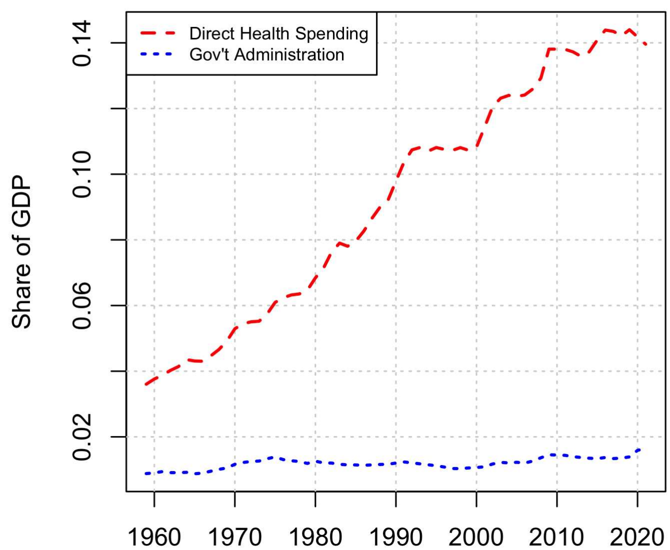


Figure A.1: Direct expenditure on health services as a percentage of GDP grew from less than 5% to approximately 14% from 1959-2021 (red line), while the value of governmental administrative costs associated with Medicaid, Medicare, and other health services (blue line) rose from 0.9% to 1.7% over this same period.

While the price indices in NIPA Table 2.5.4 and the expenditure series in NIPA Table 2.5.5 account for the value of *final* health services ultimately utilized/consumed by households, regardless of their funding source, these tables *do not* account for public administrative costs associated with the Medicaid and Medicare programs. We do not have an explicit governmental administrative sector in our model, but rather the government simply functions as an intermediary facilitating transfers of pension payments from young to old, where such payments implicitly finance both health-care spending and other consumption. Failure to consider the role of Medicaid and Medicare at driving up total costs would be important if we observed that administrative costs associated with Medicaid and Medicare as a fraction of GDP were themselves growing over time, perhaps as a consequence of population aging. Figure A.1 presents a time series breakdown of the shares of GDP for "Direct Health Spending" and "Gov't Administration." The line associated with "Direct Health Spending" contains the PCE value of health outlay from NIPA Table 2.5.5 divided by GDP estimates from NIPA Table 1.5.5. The "Gov't Administration" line contains the value of governmental administrative costs associated with providing health-financing services to households, but it *does not* contain the actual value of governmental outlay paid toward the direct provision of those services, which can be noted by referring to Chapter 9 of *Concepts and Methods of the U.S. National Income and Product Accounts* (2022). Notice that direct spending's share of GDP more than doubled from 1959-2021, while the governmental administrative share remained flat and stagnant, so that administrative costs did not significantly grow over this period. Since governmental administrative costs have not appeared to grow over the 1959-2021 period, we believe our conclusions are robust to abstracting more directly from the government's role in facilitating the provision of health services.

A.2 Relative Health-services Prices

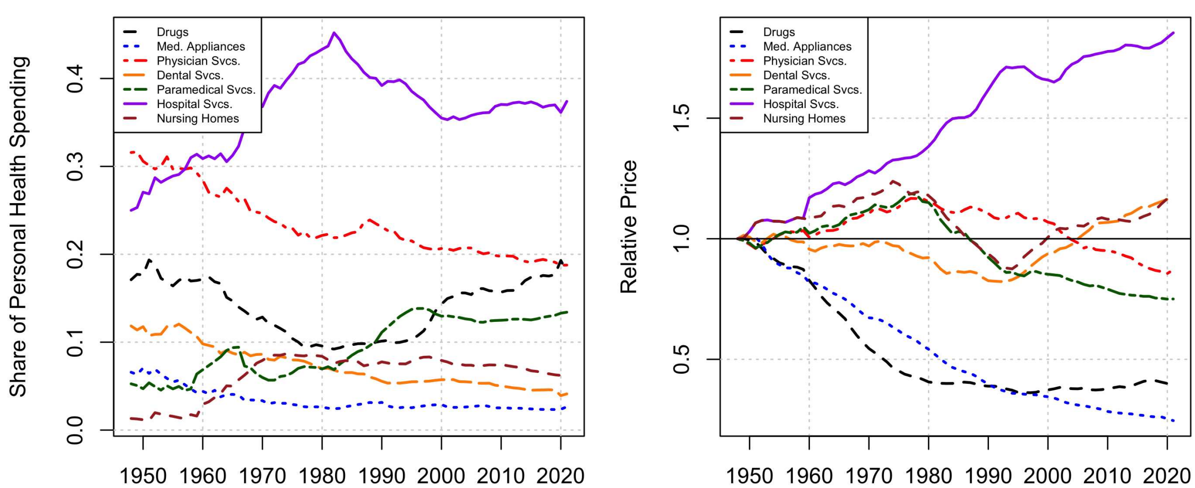
Note that PCE price indices in NIPA Table 2.5.4 are chain-weighted, unlike the Consumer Price Index from the Bureau of Labor Statistics. Chain-weighting is an attempt at accounting for composition effects possibly biasing intertemporal price comparisons. However, to reconstruct price sub-indices for narrower categorizations from chain-weighted ones, we must first unwind the chain-weighted price indices and then rewind them back up according to the alternative consumption categorizations we desire.

For example, NIPA Table 2.5.4 contains one price index for all "Personal consumption expenditures," which itself is comprised of 13 expenditure categories for domestic consumption: 1) "Food and beverages purchased for off-premises consumption," 2) "Clothing, footwear, and related services," 3) "Housing, utilities, and fuels," 4) "Furnishings, household equipment, and routine household maintenance," 5) "Health," 6) "Transportation," 7) "Communication," 8) "Recreation," 9) "Education," 10) "Food services and accommodations," 11) "Financial services and insurance," 12) "Other goods and services," 13) "Final consumption expenditures of nonprofit institutions serving households." Note that we require one price index for "Health" and another index for "Non-health" consumption. While we could, crudely, use the PCE's aggregate index to stand-in for non-health consumption, ideally we need to build a new price index that excludes health's contribution to aggregate PCE in order to arrive at a true estimate of the relative price of health services to non-health-services consumption.

To do this we take the various sub-categories of PCE that are *not* health services and then recombine them all into a new chain-weighted index. We follow procedures described in Whelan (2000) and Whelan (2002) and the appendices of Herrendorf, Rogerson, and Valentinyi (2013) and Bednar and Pretnar (2022). Specifically, non-health-consumption major sub-categories of PCE from NIPA Tables 2.5.4 and 2.5.5 are as follows: 1) "Food and beverages purchased for off-premises consumption," 2) "Clothing, footwear, and related services," 3) "Housing, utilities, and fuels," 4) "Furnishings, household equipment, and routine household maintenance," 5) "Transportation," 6) "Communication," 7) "Recreation," 8) "Education," 9) "Food services and accommodations," 10) "Financial services and insurance," 11) "Other goods and services," 12) "Final consumption expenditures of nonprofit institutions serving households." We let the base year of all prices be 1959, the first year of modern NIPA PCE consumption-category classifications.

A.3 Components of Health-services Prices

For illustration only we include in this Appendix breakdowns of the components of health-services from NIPA Table 2.5.5 to understand which sub-sectors may be most contributing to rising sectoral shares and prices. Figure A.2 features both the shares of nominal outlay by function in panel (a), and the prices of each function of outlay relative to the health-sector aggregate price are in (b).



(a) Contributors to Health PCE

(b) Relative Within-sector Prices

Figure A.2: Panel (a) plots the shares of personal health spending associated with the various sub-categories by function from NIPA Table 2.5.5, while panel (b) plots the relative price of the different components of health services to the PCE aggregate health-services price level from NIPA Table 2.5.4. with all prices normalized to unity in 1948.

Looking at panel (a) notice that hospital services and paramedical services (i.e., ambulances, urgent care, and other emergency services) have been increasing in share since the 1950s. Meanwhile, all other series are either decreasing or flat, with the exception of drugs' share of health-services spending, which declines from 1950-1990 before rising again to achieve its 1950 share by 2020.

Turning to panel (b) it becomes apparent, as well, that increasing prices of hospital services are likely the chief driver of increasing relative health-services prices. Note that because the composition of the expenditure basket within the health-services sector is changing and these indices are chain-weighted, we cannot directly attribute increasing relative health-services prices to increasing hospital-services prices. But, given it is the only sub-component that is increasing relative to the basket's aggregate, *and* given that its share has risen the most as it has come to dominate the basket, the empirical evidence suggests that whatever may be driving up shares and prices in the hospital-services sub-sector is likely important for rising aggregate health-services shares and relative prices — hence, our focus on mark-ups.

A.4 Mortality Data

Sources and implications for the rise in life expectancy and decline in mortality rates since mid-century are thoroughly explored in Murphy and Topel (2006). As in Hall and Jones (2007) our health investment formulation allows the choice of health spending to affect only non-accidental mortality rates, so that accidental mortality is exogenous. In this Appendix we briefly show how both accidental and non-accidental mortality rates have changed over time for different age groups. Up to 2010 accidental mortality had declined for all age groups, coinciding with increased safety regulations both in consumers' personal lives and at the workplace.[15] However, recent data suggest that there has been a reversal in the trend-decline in accidental mortality for some age groups. This has dampened, and for same age groups even reversed, the decline in all-cause mortality since 1950.

For both accidental and non-accidental mortality rates we turn to the CDC's *Health, United States 2017* mortality tables.[16] We classify accidental deaths as those corresponding to drug overdoses, car accidents, homicides, and suicides. All other deaths are indirectly classified as non-accidental and, for our modeling purposes, are considered functions of a living agents' health level which s/he can control via health investment.

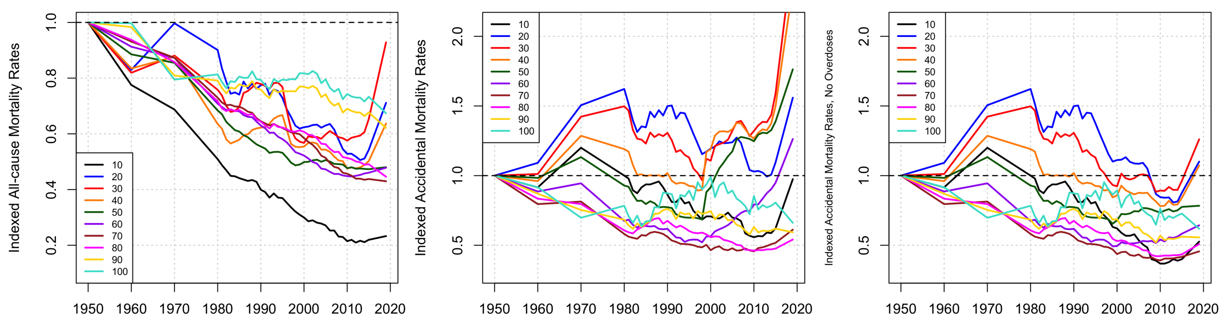
Specifically, we use Table 21 for all cause mortality, Table 27 for drug overdoses (since 1999), Table 28 for car accidents (since 1950), Table 29 for homicides (since 1950), and Table 30 for suicides (since 1950). Since drug-overdose deaths are only available since 1999, we interpolate back to 1950 by age using the trend from 1999 to 2017. Since the CDC reports death rates by age group for broad, 10-year age groups, except those between 1 and 4 years old, we assume that the reported age is the mid-point of the age-range (i.e., if the age range is 35-44, the age associated with

[15] Safety improvements to automobiles come to mind, as well as both a decline in the number of jobs requiring strenuous physical labor with dangerous machines and increased regulation regarding safety procedures associated with using such machines.

[16] The dataset can be accessed here: https://www.cdc.gov/nchs/hus/contents2017.htm.

the datapoint is 40) and we interpolate (linearly) across ages within a period. Then, since data is available only every 10 years until 1980, we interpolate the age-specific mortality rates over years. After interpolating drug overdoses and all non-drug accidental mortality rates across cohorts within a period and within cohorts over time, we arrive at data for accidental mortality rates by age. Finally, for non-accidental mortality rates we use all-cause mortality from Table 21 and interpolate it first within a period over cohorts and then within cohorts over time. We then subtract the accidental mortality rates from the all-cause mortality rate to get the non-accidental mortality rate. For years after 2017, we use linear extrapolation.

Figure A.3 shows how age-specific mortality rates have changed since 1950.[17] All cause mortality rates have declined since 1950 for all age groups, though rates have seen a recent uptick relative to long-run trends for adults aged 30 to 50. Panels (b) and (c) suggest that the uptick may be almost entirely due to recent increases in accidental mortality rates, driven by increased drug overdoses. While we take no direct stand on reasons for the rise of non-accidental mortality rates, concerns over the role of drug overdoses in driving this phenomenon have inspired an emergent new literature in economics, best punctuated by a recent working paper by Greenwood, Guner, and Kopecky (2022).



(a) All Cause Mortality Growth (b) Accidental Mortality Growth (c) Accidental Growth, No Drugs

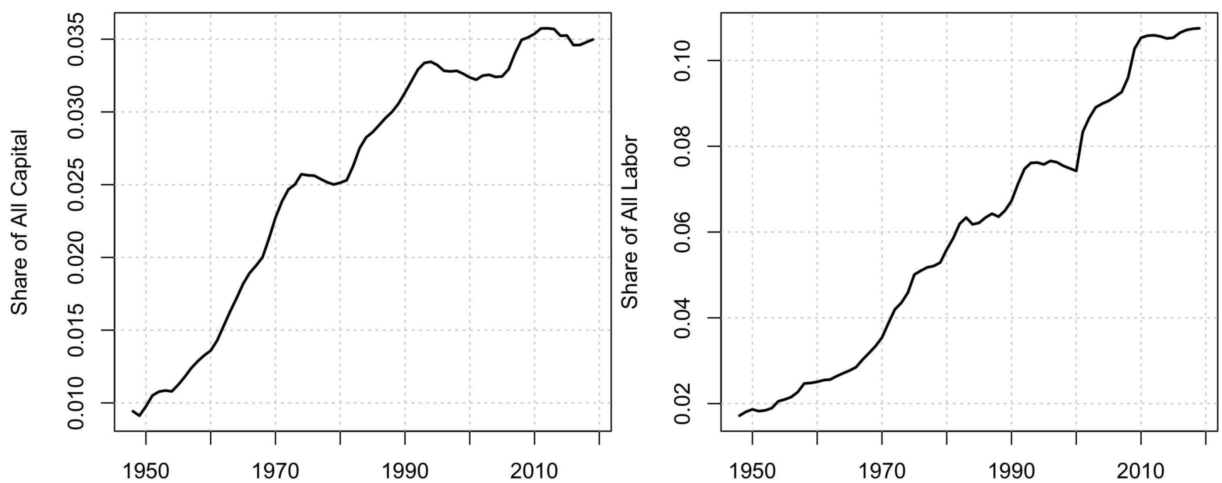
Figure A.3: All time series are normalized to unity in 1950. Panel (a) features all-cause mortality rates. Panel (b) features accidental mortality rates, including drug overdoses. To show that the recent increase in accidental mortality rates is mostly attributable to increases in overdoses, we show accidental mortality rates, excluding overdoses, in panel (c).

A.5 Sectoral Capital and Labor Data

We must compute sectoral capital and labor shares to assess how our model predicts general equilibrium outcomes. Specifically, we want data for the share of all capital in the economy which is used in health-services production, along with the same statistic for labor. For capital by sector we turn to the BEA's Fixed Assets Table 3.1ESI, "Current-cost Net Stock of Private Fixed Assets by Industry." This table accounts for the value of all equipment, structures, and intellectual property used in production. Capital used in the production of health services is taken to be that under the

[17]Note that all rates are normalized to unity in 1950.

category “Health and social assistance.” Total, economy-wide capital is taken to be “Private fixed assets.” For labor by sector we turn to BEA NIPA Tables 6.5B, 6.5C, and 6.5D, which record “Full-time Equivalent Employees by Industry.” Table 6.5B is used to compute the industrial composition of labor from 1948-1987, Table 6.5C is used for 1988-2000, and Table 6.5D is used for 2001 and after. In Table 6.5B and 6.5C the relevant category for health services labor is simply listed as “Health services,” while for Table 6.5D we must add labor records from “Ambulatory health care services,” “Hospitals,” and “Nursing and residential care facilities.”



(a) Share of All Capital in Health (b) Share of All Labor in Health

Figure A.4: Panel (a) presents the share of total, economy-wide capital that is used in the production of health services. Panel (b) shows the share of total, economy-wide labor, in units of full-time equivalent employees, used in the production of health services.

Figure A.4 shows the shares of all capital and all labor used in the production of health services. These values are $K_{ht}/(K_{ht} + K_{ct})$ and $L_{ht}/(L_{ht} + L_{ct})$, respectively. Notice that the share of capital in health-services rose by approximately 2.55 percentage points from 1948 to 2019. For labor the share rose by approximately 9.03 percentage points over this period.

B Calibration Appendix

B.1 Solving the Household’s Problem

All exogenous processes are assumed to evolve deterministically, since this model features a representative agent and is primarily about long-run growth, not short-run fluctuations. In this section we drop time (t) subscripts. For a given set of prices $\{r, w, q, b, \{T_j, \pi_j\}\}$, where b and $\{T_j, \pi_j\}$ are endogenous to the general equilibrium but exogenous from the perspective of the household, we solve the household’s problem using the Carroll (2006) modification of Krueger and Ludwig (2007) as follows:

i. Set a fixed grid for assets, $\mathscr{A}$, and endogenous, age-dependent grid for cash-on-hand, $\mathscr{C}_j$, where $\mathscr{C}_J$ (the grid for the agent in its last period of life) is fixed.

ii. For age J agents, set the consumption policy $c_J = \mathscr{C}_J$ at every point along the grid and compute $\mathcal{V}(c_J) = \chi + \xi\, c_J^{1-\gamma}/(1-\gamma)$. Set $h_J = a'_J = s_J = 0$, where a'_J are assets at age $J+1$, which of course is zero because the agent would be dead.

iii. Iterate back from $j = J-1$ to $j = 1$ doing the following:

   a. Let $\underline{\mathscr{C}}$ denote the smallest cash-on-hand grid point which is fixed for all ages. Set the minimum consumption level (first grid point in c_j) to $c_j = \underline{\mathscr{C}}$ and make the corresponding grid points $h_j = a'_j = s_j = 0$.

   b. Now iterate forward on the elements of $\mathscr{A}$, which should be a strictly increasing grid, setting $cash_{j+1} = w\eta_{j+1}(1-\tau) + (1+r-\delta)(\mathscr{A}+b) + T_{j+1} + \pi_j$ for each element of $\mathscr{A}$. Note that $T_j = 0$ if $j < J_R$, $T_j = T$ (constant) if $j \geq J_R$, and $\eta_j = 0$ for $j \geq J_R$. This is the value of cash an age-j agent today will have at age $j+1$ if they survive and also choose the particular element of $\mathscr{A}$. Thus, there should be as many $cash_{j+1}$ as there are elements on $\mathscr{A}$.

   c. Compute c'_j (future c) and $\mathcal{V}'_j$ (future $\mathcal{V}$) by interpolating each point $cash_{j+1}$ on the endogenous grid $\mathscr{C}_{j+1}$.

   d. With c'_j and $\mathcal{V}'_j$ in hand, calculate $h_j(\mathscr{A})$ and $c_j(\mathscr{A})$ by solving first-order condition:

$$\frac{\partial s_j}{\partial h_j}\mathcal{V}'_j = q\, s_j(h_j)\,(1+r-\delta)\,(c'_j)^{-\gamma}$$

   This is the health choice as a function of the asset grid points. Again, there should be as many elements of h_j as there are elements of $\mathscr{A}$ (one for each point).

   e. Given $h_j(\mathscr{A})$ compute $s_j(\mathscr{A})$ and $c_j(\mathscr{A})$, where s_j is given directly by the survival function, and c_j solves the first-order condition:

$$q\, c_j^{-\gamma} = \beta\, \frac{\partial s_j}{\partial h_j}\mathcal{V}'_j$$

   f. Update the endogenous cash grid and a'_j:

$$\begin{aligned}\mathscr{C}_j &= \mathscr{A} + c_j + q\, h_j \\ a'_j &= \mathscr{A}\end{aligned}$$

   g. Send j to $j-1$ and repeat until all policy functions are in hand.

B.2 Solving the General Equilibrium

We follow Krueger and Ludwig (2007) and iterate on the price vector $\{r, w, p, b, \{T_j, \pi_j\}\}$. Some innovations are required to pin down the allocation of capital and labor to the two sectors. Further, we must ensure the monopolistically competitive mark-up over marginal cost is satisfied. Note our solution of the G.E. requires the assumption that consumers fail to forecast how prices will evolve and are surprised to find out they have changed due to exogenous growth in productivities, mark-ups, and newborns. We thus pre-suppose the G.E. follows an MIT shock process, similar to Krueger and Ludwig (2007), though we do not assume capital fully depreciates each period.

We start by solving an artificial steady state in period $t = 1$ which corresponds to the year 1950. We then iterate forward by setting the prices for period $t - 1$ to be the starting values for period t and moving the growth objects forward one period. Recall, the objects that grow/change over time are $\{N_{1t}, z_t, A_{ct}, A_{ht}, \mu_t, \{\zeta_{jt}\}_j\}_t$, where the population of newborns, economy-wide health-investment productivity, sectoral TFP's, and age-specific health-investment productivities grow at constant rates, while the time series for $\{\mu_t\}_t$ is directly calibrated. Next, we iterate on the household problem under these prices to solve for the policies. We then aggregate the household problem and use a root-solver (Broyden's method) to find the price vector which solves a system of equations that correspond to the equilibrium allocations of capital and labor across sectors in the given period t.

The following algorithm solves the G.E. along the transition path:

i. For a given set of prices $\{r, w, p, b, \{T_j, \pi_j\}\}$, solve for household policy functions as described in Appendix B.1.

ii. To aggregate household decisions we start by assuming that in every period the first generation enters the economy with $a_1 = 0$, though they may have bequests, b. Given this assumption we can use the endogenous grids, $\mathscr{C}_j$, to iterate forward from $j = 2$ to J to solve for the optimal decision paths for a', c, h, and s. This step involves interpolation over equilibrium available cash-on-hand, which is given by $cash_j = w\eta_j(1-\tau) + (1 + r - \delta)(a_j + b) + T_j + \pi_j$. In this first period's steady state a_j is the optimal decision made by the age $j - 1$ agent in the same period. Along the transition path $a_j = a'_{j-1,t-1}$, so that the optimal asset allocation decision made by an age $j - 1$ agent in period $t - 1$ is taken to be the asset level of a living age j agent in period t. This ensures that the aggregate capital used in t is equal to that which the surviving agents from period $t - 1$ carried forward into the future plus $b \sum_j N_j$, which is the total capital bequeathed by those who died at the end of $t - 1$.

iii. To complete aggregation we also need the population levels by age. Note that $N_{1,1} = 0$ in the initial period's steady state. Thereafter, we compute the population by cohort as $N_{jt} = s_{j-1,t-1}N_{j-1,t-1}$, where we re-introduce time subscripts here to show that, along the transition path, the population of cohort j in period t is simply those age $j - 1$ agents from period $t - 1$ who survived to live another year. In steady state only the population evolves according to $N_{jt} = s_{j-1,t}N_{j-1,t}$.

iv. Aggregate the household decisions: $L = \sum_j N_j \eta_j$, $K = \sum_j N_j(a_j + b)$, $K' = \sum_j N_j a'_j$, $H = \sum_j N_j h_j$, $C = \sum_j N_j c_j$. Also, $I = K' - (1-\delta)K$.

v. Now, use the fact that under Cobb-Douglas production marginal costs can be written as functions of input prices:

$$mc_m(w, r\,;\, A_m) = \frac{1}{A_m}\left(\frac{r}{\alpha_m}\right)^{\alpha_m}\left(\frac{w}{(1-\alpha_m)}\right)^{1-\alpha_m}, \qquad \forall m \in \{c, h\}$$

Use the marginal cost condition to compute the fractions of capital and labor going to each sector as:

$$\begin{aligned}
K_c &= \left(\frac{mc_c(w,r\,;\,A_c)\,\alpha_c\,(C+I)}{mc_c(w,r\,;\,A_c)\,\alpha_c\,(C+I) + mc_h(w,r\,;\,A_h)\,\alpha_h\,H}\right)K \\
K_h &= \left(\frac{mc_h(w,r\,;\,A_h)\,\alpha_h\,H}{mc_c(w,r\,;\,A_c)\,\alpha_c\,(C+I) + mc_h(w,r\,;\,A_h)\,\alpha_h\,H}\right)K \\
L_c &= \left(\frac{mc_c(w,r\,;\,A_c)\,(1-\alpha_c)\,(C+I)}{mc_c(w,r\,;\,A_c)\,(1-\alpha_c)\,(C+I) + mc_h(w,r\,;\,A_h)\,(1-\alpha_h)\,H}\right)L \\
L_h &= \left(\frac{mc_h(w,r\,;\,A_h)\,(1-\alpha_h)\,H}{mc_c(w,r\,;\,A_c)\,(1-\alpha_c)\,(C+I) + mc_h(w,r\,;\,A_h)\,(1-\alpha_h)\,H}\right)L
\end{aligned}$$

vi. Compute the pricing relationships between aggregates, including health sector profits:

$$\begin{aligned}
r &= A_c\,\alpha_c\left(\frac{K_c}{L_c}\right)^{\alpha_c - 1} \\
w &= A_c(1-\alpha_c)\left(\frac{K_c}{L_c}\right)^{\alpha_c} \\
p &= \mu\, mc_h(w,r\,;\,A_h) \\
b &= \frac{\sum_j (1-s_j) N_j a'_j}{\sum_j N_j} \\
T_j &= \begin{cases} 0, & j < J_R \\ \frac{\sum_j N_j\, w\, \eta_j\, \tau}{\sum_j N_j}, & j \geq J_R \end{cases} \\
\pi_j &= \left(\frac{a_j + b}{K}\right)\Pi
\end{aligned} \tag{B.1}$$

where Π is defined in (20) in the main text.

vii. Note that (B.1) forms a self-map on the price vector, as in Krueger and Ludwig (2007). Let $\mathscr{P}$ be the collapsed price vector $\{r, w, p, b, \{T_j, \pi_j\}\}$. Further, let $f(\mathscr{P})$ describe the self-map in (B.1) which is a $J+6$-dimensional, vector-valued function.[18] The root-finder seeks to solve for the price set such that $\mathscr{P} = f(\mathscr{P})$.

[18]There are J π_j values, two T_j though one is trivial, and four other prices.

viii. When the root has been found, set $\mathscr{P}_{t+1} = \mathscr{P}_t$ as an initial value, iterate the growth parameters forward, and solve for the next period's prices, returning to part (i).

B.3 Exogenously Calibrated Sets of Parameters

In this appendix we show the following age-dependent, health-production parameters which are taken from Hall and Jones (2007) and Hansen (1993) — $\{\phi_j\}_j$, $\{\theta_j\}_j$, $\{g_{\zeta_j}\}_j$, and $\{\eta_j\}_j$. Note that we assume that growth rates for ζ_{jt} are age-dependent but constant over time. For the labor productivity parameters we take the average of the male/female productivity estimates from Hansen (1993). All of the age-dependent parameter values are in Figure B.2.

The time series of mark-ups, meanwhile, is in Figure B.1. These values are computed by assuming that there are no mark-ups in the health sector whenever the economy-wide mark-up in Horenstein and Santos (2019) exceeds the health-sector mark-up. From there we scale mark-ups after 1990 so that they grow at the same rate as the relative health-sector/economy-wide mark-up in Horenstein and Santos (2019). We stop re-scaling after 2005, since that is the last year for which estimates in Horenstein and Santos (2019) are available.

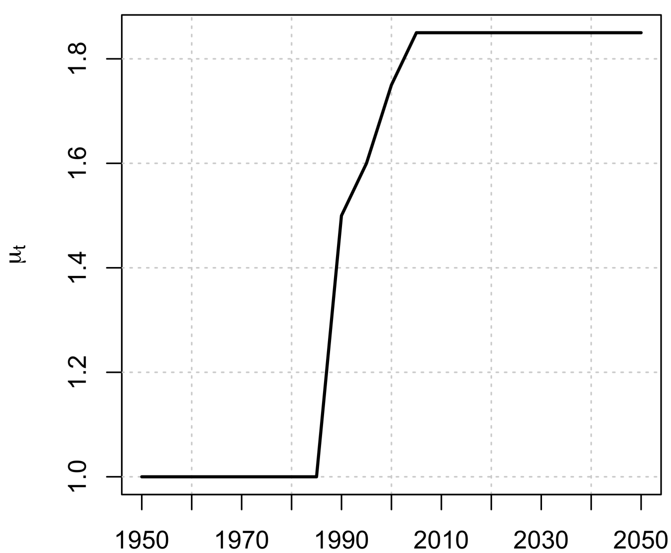
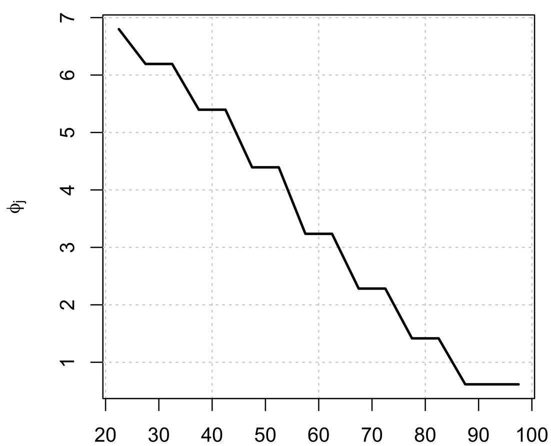
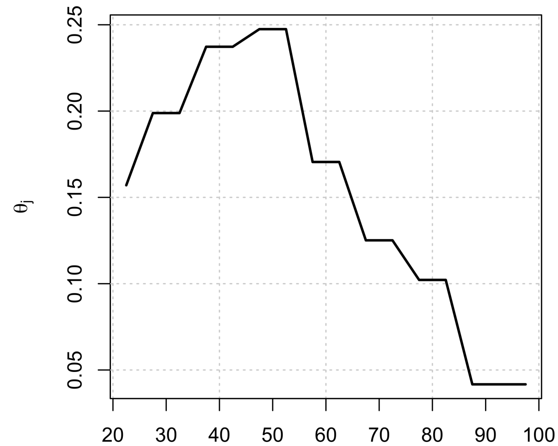


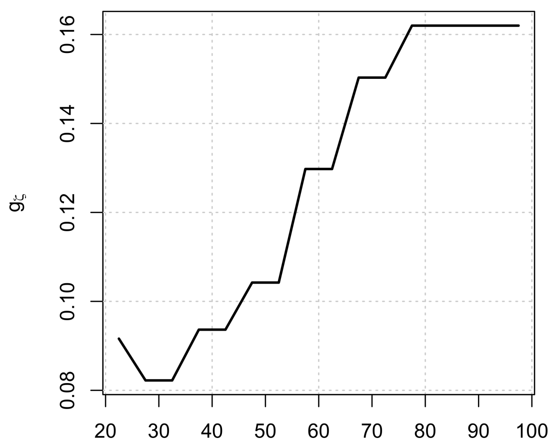
Figure B.1: This figure contains our estimates of health-sector mark-ups, computed from data in Horenstein and Santos (2019). Time is on the horizontal axis.



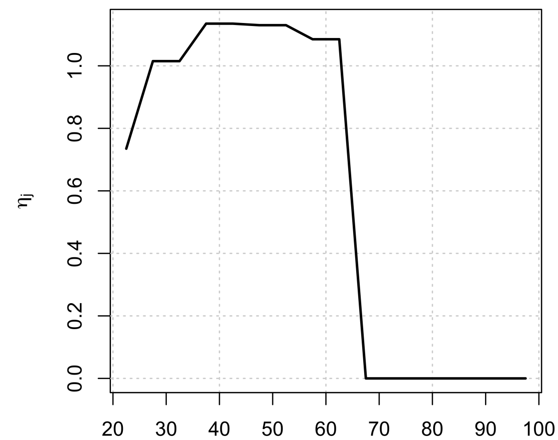
(a) Health Production Intercept, ϕ_j



(b) Health Production Elasticity, θ_j



(c) Health Productivity Growth, g_{ζ_j}



(d) Labor Productivity, η_j

Figure B.2: In this figure we present the age-dependent parameters that are externally calibrated. Panels (a) through (c) contain the health production intercepts, health production elasticities, and health productivity growth rates from Horenstein and Santos (2019). Panel (d) contains labor productivity estimates from Hansen (1993). Age is on the horizontal axis in all panels.

References

Acemoglu, Daron, Amy Finkelstein, and Matthew J. Notowidigdo (2013). "Income and Health Spending: Evidence from Oil Price Shocks". *The Review of Economics and Statistics* 95.4, pp. 1079–1095 (cit. on p. 4).

Acemoglu, Daron and Veronica Guerrieri (2008). "Capital Deepening and Nonbalanced Economic Growth". *Journal of Political Economy* 116.3 (cit. on p. 9).

Alonso-Carrera, Jaime, Jordi Caballé, and Xavier Raurich (2015). "Consumption composition and macroeconomic dynamics". *The B.E. Journal of Macroeconomics* 15 (cit. on pp. 1, 9, 20).

Attanasio, Orazio, Sagiri Kitao, and Giovanni L. Violante (2010). "Financing Medicare: A General Equilibrium Analysis". *Demography and the Economy*. University of Chicago Press, pp. 333–366 (cit. on p. 7).

Backus, David, Thomas Cooley, and Espen Henriksen (2014). "Demography and low-frequency capital flows". *Journal of International Economics* 92 (cit. on pp. 2, 8, 20, 28).

Bates, Laurie and Rexford Santerre (2013). "Does the U.S. health care sector suffer from Baumol's cost disease? Evidence from the 50 states". *Journal of Health Economics* 32, pp. 386–391 (cit. on p. 22).

Baumol, William (1967). "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis". *The American Economic Review* 57.3 (cit. on pp. 2, 22).

Baumol, William, Sue Anne Batey Blackman, and Edward Wolff (1985). "Unbalance growth revisited: asymptotic stagnancy and new evidence". *The American Economic Review* 75.4, pp. 806–817 (cit. on pp. 2, 22).

Bednar, W.L. and Nick Pretnar (2022). "The Evolution of the Consumption Experience: Why the Services Share Has Risen". *Working Paper* (cit. on p. 33).

Berndt, Ernst R. et al. (2001). "Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues". *Medical Care Output and Productivity*. Ed. by David M. Cutler and Ernst R. Berndt. University of Chicago Press. Chap. 4, pp. 141–200 (cit. on p. 3).

Blumenthal, David, Kristof Stremikis, and David Cutler (2013). "Health Care Spending — A Giant Slain or Sleeping?" *The New England Journal of Medicine* 369.26 (cit. on pp. 1, 2, 22).

Boppart, Timo (2014). "Structural Change and the Kaldor Facts in a Growth Model with Relative Price Effects and Non-Gorman Preferences". *Econometrica* 82.6 (cit. on p. 9).

Carroll, Christopher D. (2006). "The method of endogenous gridpoints for solving dynamic stochastic optimization problems". *Economics Letters* 91.3, pp. 312–320 (cit. on pp. 20, 36).

Caunedo, Julieta, David Jaume, and Elisa Keller (2023). "Occupational Exposure to Capital-Embodied Technical Change". *American Economic Review* 113.6, pp. 1642–1685 (cit. on p. 9).

Caunedo, Julieta and Elisa Keller (2023). "Capital Embodied Structural Change". *Working Paper* (cit. on p. 9).

Chandra, Amitabh and Jonathan Skinner (2012). "Technology Growth and Expenditure Growth in Health Care". *Journal of Economic Literature* 50.3, pp. 645–680 (cit. on p. 4).

Clemens, Jeffrey and Joshua D. Gottlieb (2014). "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review* 104.4, pp. 1320–1349 (cit. on p. 4).

Comin, Diego, Danial Lashkari, and Martí Mestieri (2021). "Structural Change with Long-run Income and Price Effects". *Econometrica* 89.1, pp. 311–374 (cit. on p. 9).

Concepts and Methods of the U.S. National Income and Product Accounts (2022). Bureau of Economic Analysis, U.S. Department of Commerce (cit. on pp. 31, 32).

Cooley, Thomas and Espen Henriksen (2018). "The demographic deficit". *Journal of Monetary Economics* 93, pp. 45–62 (cit. on pp. 2, 8, 20, 28).

Cooley, Thomas, Espen Henriksen, and Charlie Nusbaum (2019). “Demographic Obstacles to European Growth”. *NBER Working Paper #26503* (cit. on pp. 8, 20, 28).

Cooper, Zack et al. (2019). “The Price Ain’t Right? Hospital Prices and Health Spending on the Privately Insured”. *The Quarterly Journal of Economics*, pp. 51–107 (cit. on pp. 4, 6, 16, 23, 30).

Cylus, Jonathan D. and Bridget A. Dickensheets (2007-2008). “Hospital Multifactor Productivity: A Presentation and Analysis of Two Methodologies”. *Health Care Financing Review* 29.2 (cit. on p. 23).

Dixit, Avinash K. and Joseph E. Stiglitz (1977). “Monopolistic competition and optimum product diversity”. *The American Economic Review* 67.3, pp. 297–308 (cit. on pp. 1, 10, 14, 15).

Donahoe, Gerald F. (2000). “Capital in the national health accounts”. *Health Care Financing Administration* (cit. on p. 23).

Dunn, Abe, Scott D. Grosse, and Samuel H. Zuvekas (2016). “Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States”. *Health Services Research* 53.1 (cit. on p. 3).

Fehr, Hans and Maria Feldman (2023). “Financing Universal Health Care: Premiums or Payroll Taxes?” *Working Paper: University of Wuerzberg* (cit. on p. 7).

Fonseca, Raquel et al. (2020). “Understanding Cross-country Differences in Health Status and Expenditures”. *NBER Working Paper #26876* (cit. on pp. 1, 4).

Fonseca, Raquel et al. (2021). “Accounting for the Rise of Health Spending and Longevity”. *Journal of the European Economic Association* 19.1, 536–579 (cit. on pp. 1, 4).

Garber, Alan M. and Jonathan Skinner (2008). “Is American Health Care Uniquely Inefficient?” *Journal of Economic Perspectives* 22.4, pp. 27–50 (cit. on p. 4).

Gaynor, Martin, Kate Ho, and Robert J. Town (2015). “The Industrial Organization of Health-Care Markets”. *Journal of Economic Literature* 53.2, pp. 235–284 (cit. on p. 4).

Gaynor, Martin and Robert J. Town (2012). “Competition in Health Care Markets”. *Working Paper No. 12/282, Centre for Market and Public Organisation, Bristol Institute of Public Affairs, University of Bristol* (cit. on p. 4).

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town (2015). “Mergers When Prices Are Negotiated: Evidence from the Hospital Industry”. *The American Economic Review* 105.1, pp. 172–203 (cit. on p. 4).

Greenwood, Jeremy, Nezih Guner, and Karen A. Kopecky (2022). “The Downward Spiral”. *NBER Working Paper #29764* (cit. on p. 35).

Grossman, Michael (1972). “On the Concept of Health Capital and the Demand for Health”. *Journal of Political Economy* 80.2 (cit. on p. 1).

Hall, Robert and Charles Jones (2004). “The Value of Life and the Rise in Health Spending”. *NBER Working Paper No. 10737* (cit. on pp. 12, 13).

— (2007). “The Value of Life and the Rise in Health Spending”. *The Quarterly Journal of Economics* (cit. on pp. 1, 3, 4, 6, 7, 10–13, 20–24, 26, 34, 40).

Halliday, Timothy J. et al. (2019). "Health Investment Over the Life-cycle". *Macroeconomic Dynamics* 23, pp. 178–215 (cit. on p. 7).

Hansen, Gary (1993). "The Cyclical and Secular Behaviour of the Labour Input: Comparing Efficiency Units and Hours Worked". *Journal of Applied Econometrics* 8.1, pp. 71–80 (cit. on pp. 13, 24, 40, 41).

Harper, Michael J. et al. (2010). "Nonmanufacturing industry contributions to multifactor productivity, 1987–2006". *Monthly Labor Review, U.S. Bureau of Labor Statistics* (cit. on p. 22).

He, Hui, Kevin X.D. Huang, and Lei Ning (2021). "Why Do Americans Spend So Much More on Health Care Than Europeans?" *International Economic Review* 62.4, pp. 1363–1399 (cit. on p. 7).

Herrendorf, Berthold, Christopher Herrington, and Ákos Valentinyi (2015). "Sectoral Technology and Structural Transformation". *American Economic Journal: Macroeconomics* 7.4, pp. 104–133 (cit. on p. 20).

Herrendorf, Berthold, Richard Rogerson, and Ákos Valentinyi (2013). "Two Perspectives on Preferences and Structural Transformation". *American Economic Review* 103.7 (cit. on pp. 9, 33).

— (2014). "Growth and structural transformation". *Handbook of economic growth*. Vol. 2. Elsevier, pp. 855–941 (cit. on p. 9).

Horenstein, Alex R. and Manuel S. Santos (2019). "Understanding Growth Patterns in US Health Care Expenditures". *Journal of the European Economic Association* 17.1, pp. 284–326 (cit. on pp. 1, 4, 6, 10, 20–25, 40, 41).

Hurd, Michael (1989). "Mortality Risk and Bequests". *Econometrica* 57.4, pp. 779–813 (cit. on p. 23).

Jung, Jurgen, Chung Tran, and Matthew Chambers (2017). "Aging and health financing in the U.S.: A general equilibrium analysis". *European Economic Review* 100, pp. 428–462 (cit. on pp. 6, 7, 23).

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie (2001). "Beyond Balanced Growth". *Review of Economic Studies* 68, pp. 869–882 (cit. on p. 9).

Krueger, Dirk and Alexander Ludwig (2007). "On the consequences of demographic change for rates of returns to capital, and the distribution of wealth and welfare". *Journal of Monetary Economics* 54, pp. 49–87 (cit. on pp. 2, 20, 21, 36, 38, 39).

Kuhn, Michael et al. (2011). "Externalities in a life cycle model with endogenous survival". *Journal of Mathematical Economics* 47 (cit. on p. 7).

Kydland, Finn and Nick Pretnar (2019). "The Costs and Benefits of Caring: Aggregate Burdens of an Aging Population". *NBER Working Paper #25498* (cit. on pp. 2, 8, 28).

Maestas, Nicole, Kathleen J. Mullen, and David Powell (2023). "The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity". *American Economic Journal: Macroeconomics* 15.2, pp. 306–322 (cit. on pp. 2, 8, 28).

Meara, Ellen, Chapin White, and David Cutler (2004). "Trends in Medical Spending by Age, 1963-2000". *Health Affairs* 23 (cit. on p. 22).

Murphy, Kevin and Robert Topel (2006). "The Value of Health and Longevity". *The Journal of Political Economy* 114.5 (cit. on p. 34).

Ngai, Rachel and Christopher Pissarides (2007). "Structural Change in a Multisector Model of Growth". *American Economic Review* 97.1 (cit. on p. 9).

Nygaard, Vegard M. (2022). "Causes and consequences of life expectancy inequality". *Working Paper* (cit. on p. 7).

Prettner, Klaus (2013). "Population Aging and Endogenous Economic Growth". *Journal of Population Economics* 26.2 (cit. on pp. 8, 20).

Romley, John A., Dana P. Goldman, and Neeraj Sood (2015). "US hospitals experienced substantial productivity growth during 2002–11". *Health Affairs* 34.3, pp. 511–518 (cit. on p. 23).

Shatto, John D. and M. Kent Clemens (2022). *Projected Medicare Expenditures under an Illustrative Scenario with Alternative Payment Updates to Medicare Providers*. Tech. rep. Centers for Medicare & Medicaid Services (cit. on pp. 1, 2, 22).

Triplett, Jack and Barry Bosworth (2004). *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Brookings Institution Press (cit. on p. 1).

Whelan, Karl (2000). "A Guide to the Use of Chain Aggregated NIPA Data". *Federal Reserve Board Working Paper* (cit. on p. 33).

— (2002). "A Guide to U.S. Chain Aggregated NIPA Data". *Review of Income and Wealth* 48.2, pp. 217–233 (cit. on p. 33).

Zhao, Kai (2014). "Social security and the rise in health spending". *Journal of Monetary Economics* 64, pp. 21–37 (cit. on pp. 1, 7).