The Spillover Effects of Top Income Inequality*

Joshua D. Gottlieb David Hémous
Jeffrey Hicks Morten Olsen†

May 2025

Abstract

Top income inequality in the United States has increased considerably within many occupations. This phenomenon has led to a search for a common explanation. We instead develop a theory where increases in income inequality originating within a few occupations can "spill over" through consumption into others. We show theoretically that such spillovers occur when an occupation provides non-divisible services of heterogeneous quality to consumers. Examining local income inequality across U.S. regions, we find evidence that such spillovers exist for physicians, dentists, other medical occupations, and real estate agents. Estimated spillovers for other occupations are consistent with the predictions of our theory. Calibrating our model, we show that spillovers amplify a given shock to top income inequality by at least 16%. Spillovers dampen the increase in consumers' welfare inequality compared with the change in income inequality.

JEL: D31; J24; J31; O15. Keywords: Income inequality, Assignment model, Occupational inequality, Superstars

---

*We are deeply indebted to Jeffrey Clemens for his major role collaborating with us on earlier versions of this paper. Olsen gratefully acknowledges the financial support of the European Commission under the Marie Curie Research Fellowship program (Grant Agreement PCIG11-GA-2012-321693) and the Spanish Ministry of Economy and Competitiveness (Project ref: ECO2012-38134). Gottlieb gratefully acknowledges support from the Becker-Friedman Institute, the Upjohn Institute for Employment Research, and SSHRC. Hémous gratefully acknowledges support from the University Research Priority Program "URPP Equality of Opportunity" of the University of Zurich. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1756 (releases CBDRB-FY22-P1756-R9679 and CBDRB-FY22-P1756-R9845). We are grateful to Frank Limehouse, Shahin Davoudpour, Daniel Polsky, Michael Levin, Leemore Dafny, CIVHC, and the staff at the Chicago Federal Statistical Research Data Center for their generosity in sharing data and facilitating data access. We thank Scott Blatte, Aidan Buehler, Kenta Baron-Furuyama, Oscar Chan, Innessa Colaiacovo, Erin May, Hugh Shiplett, Emilie Vestergaard and Andrew Vogt for outstanding research assistance. Finally, we thank Paul Beaudry, Roland Bénabou, Marianne Bertrand, Austin Goolsbee, David Green, Neale Mahoney, Stefanie Stantcheva, Joachim Voth and numerous seminar participants for valuable comments.

†University of Chicago & NBER; University of Zürich & CEPR; University of Toronto; University of Copenhagen

1 Introduction

The increase in top earnings since the 1980s has been accompanied by growing inequality within the top of the distribution, both in aggregate and within occupations (Bakija, Cole, and Heim, 2012). At first glance, this pattern suggests that any explanation for rising inequality—whether globalization, technology, deregulation, or changes to the tax structure—would have to apply to occupations as diverse as bankers, doctors, and CEOs (Kaplan and Rauh, 2013). We argue instead that an increase in income inequality originating within some occupations can spill over into others, driving broader changes in income inequality. Our prime example is physicians, who comprise 13% of the top one percent of wage earners.

Our first contribution is theoretical. We characterize conditions under which an increase in one group's top income inequality increases top income inequality for service providers. This occurs when the services provided are heterogeneous in quality and consumers cannot perfectly substitute quality with quantity. These spillovers are geographically local when the services are non-tradable.

We examine our model's predictions in U.S. labor market data. Using a shift-share strategy, we show that a region's top income inequality spills over into top income inequality among physicians, dentists, other medical occupations, and real estate agents. The effect is large, with standardized coefficients of 1.1 to 1.5. We do not observe such spillovers for occupations that do not meet the model's requirements (such as engineers and managers). Using a broader set of occupations and characteristics of occupations, we show that the spillover patterns are consistent with the model's predictions.

Our analysis begins in Section 2 by documenting that the rise in top income inequality is driven primarily by growing income inequality within occupations. We decompose wage income changes from 1980 to 2012 and find that three-quarters of the rise in the 99th-to-90th-percentile income ratio is within-occupation.

In Section 3, we develop a theory under which income inequality can spill over from one occupation to another. In our model, widget makers with heterogeneous incomes buy services from doctors who have heterogeneous abilities and thus provide medical services of heterogeneous quality. Consumption of medical services is non-divisible: Each widget maker needs to consume one unit of one doctor's services. In addition, production is not scalable: Each doctor can only serve a fixed number of widget makers. This gives rise to a positive assortative matching mechanism. When both groups' ability distributions have Pareto tails, doctors' income distribution also has a Pareto tail. An (exogenous) increase in income inequality among the widget makers increases relative demand for the services of the highest-ability doctors and increases top income inequality among doctors. Interestingly, increasing inequality of doctors' ability may actually *reduce* doctors' top income inequality. Our results generalize to the case when doctors have a finite positive supply elasticity, and to the case where quantity and quality of medical service are partly (but not perfectly) substitutable. In contrast, with perfect substitution between quality and quantity, any change in widget makers' income distribution only affects the price per unit of quality-adjusted medical service with no consequence for doctors' inequality.

Our baseline model deliberately focuses on local consumption spillovers by considering a single economy and it abstracts from occupational mobility at the top of the income distribution. Our results are robust to allowing for both occupational and geographical mobility of doctors; in both cases, increasing local income inequality of widget makers increases local income inequality for doctors. In contrast, when we allow for trade in medical services, spillovers occur at the national level so that local top income inequality of doctors is independent of local top income inequality.

Section 4 introduces our empirical analysis of inequality spillovers in U.S. local labor markets. We use restricted-access Census and American Community Survey data from 1980 to 2012 to build a panel of labor market areas (LMAs, which are aggregates of commuting zones) and conduct our analysis at this level. Guided by our model, we measure top income inequality as the inverse Pareto parameter for individuals in the top 10% of the local income distribution. We measure inequality for each occupation, *e.g.* physicians, in each region. We then regress local top income inequality among physicians on local top income inequality in the rest of the population.

An OLS regression could suffer from several endogeneity concerns, so we rely on a shift-share strategy. For each LMA, we compute a weighted average of national occupational inequality. The weights correspond to the importance of each occupation in each LMA at the beginning of our sample. In other words, we only exploit the changes in local income inequality that arise from the occupational distribution in 1980 combined with the nationwide trends in occupation-specific inequality. This weighted average serves as our instrument for general inequality in the LMA. We follow the identification assumption of Goldsmith-Pinkham, Sorkin and Swift (2020), and assume that the occupational shares are uncorrelated with changes in the outcome variable other than through their effect on local top income inequality.

The model predicts local inequality spillovers for occupations providing services that are heterogeneous in quality, non-divisible, and non-tradable. In Section 5, we focus on three high-earning occupations satisfying these criteria: physicians, dentists, and real estate agents. We find positive spillover coefficients, and the effect is economically significant: a 1 standard deviation increase in local top income inequality raises doctors' top income inequality by 1.1 standard deviation, and dentists' and real estate agents' top income inequality by 1.5 standard deviation.

We next estimate spillover coefficients for an additional 27 occupations common in the top 10% of the income distribution. Most of these occupations do not entirely fit the requirements of our theory, and, with a few exceptions, we do not find spillovers for these occupations. In line with our theory, the spillover coefficients are positively correlated with measures of the importance of customer service and of working directly with the public from O*NET and negatively correlated with a measure of tradability. Together, physicians, dentists, and real estate agents already account for 16% of the top 1% of earners. We consider them as poster-child occupations where spillovers are relatively easy to identify, but we hypothesize that spillovers may affect additional occupations even if our empirical strategy fails to detect them: Spillovers may occur at the national level, they may only affect subcategories within an occupation (*e.g.* personal lawyers but not corporate lawyers), or we may simply have too few observations to detect them. Beyond occupations common in the top 10%, we find evidence of spillovers for a range of medical occupations (therapists, veterinarians, etc.).

We conduct numerous robustness checks and extensions on these core results. In one noteworthy exercise, we leverage the financial deregulation of the 70s, 80s, and 90s as one specific policy-induced shock to consumer income inequality, rather than using our shift-share instrument which aggregates many shocks. With this policy-driven instrument, we continue to find inequality spillovers to doctors and dentists.

Finally, we calibrate our model in Section 6 to quantify its implications using New York State data. The calibrated model first shows that shifts in consumer income can explain the observed rise in doctors' income inequality. Second, spillovers to doctors alone amplify a given shock to top inequality by 16%.[1] Third, it quantifies how much spillovers dampen the increase in welfare inequality relative to nominal

[1] For this calculation, top inequality is measured by the ratio of aggregate income earned by those in the top 1% to aggregate income earned by those in the top 10%.

income inequality for consumers.

This paper contributes to a large literature on the rise in top income inequality and its causes (Piketty and Saez, 2003; Atkinson, Piketty and Saez, 2011). We build on the "superstars" idea of Rosen (1981), who explains how small differences in talent may lead to large differences in income. The key element in his model is the indivisibility of consumption, which leads to a "many-to-one" assignment problem as each consumer only consumes from one performer (singer, comedian, etc.), but each performer can serve a large market (see also Sattinger, 1993).[2] Income inequality among performers increases because technological change or globalization allows the superstars to serve a much larger market—that is, to scale up production.[3] Specifically, let $w(z)$ denote the income of an individual of talent z, $p(z)$ the average price for her services, and $q(z)$ the quantity provided, so $w(z) = p(z)q(z)$. The standard interpretation of "superstars" is that they have very large markets (high $q(z)$). This makes such a framework poorly suited for occupations where output is not easily scalable.

In contrast, we study an assignment model that is "constant-to-one" where superstars are characterized by a high price $p(z)$ for their services. This makes us closer to Gabaix and Landier (2008). They argue that, since executives' talent increases the overall productivity of firms, the best CEOs are assigned to the largest firms. They show empirically that the increase in CEO compensation can be fully attributed to the increase in firms' market size. Grossman (2007) and Terviö (2008) present models with similar results. Our baseline model with a Cobb-Douglas utility between medical services and the outside good is similar to their production function which is multiplicative in CEO skill and firm productivity, but we focus on a second moment of the income distribution (the Pareto tail instead of the mean) and consider a consumption problem. Importantly, we extend the analysis beyond the baseline model by considering general utility functions, an intensive margin, different entry margins, geographical reallocation, and limited substitutability between quality and quantity.[4]

Our theory offers an amplification mechanism where any shock to top income inequality can spill over to other occupations. The "original" shock may arise from

[2]Adding network effects, Alder (1985) goes further and writes a model where income can drastically differ among artists of equal talents.

[3]Koenig (2021) provides empirical evidence for entertainers using the roll-out of television.

[4]Relatedly, Määttänen and Terviö (2014) build an assignment model for housing, which they calibrate to six U.S. metropolitan areas. They find that the increase in income inequality has led to an increase in house price dispersion (see also Landvoigt, Piazzesi and Schneider, 2015).

various channels: technological change affecting firm size (Geerolf, 2017), globalization (Bonfiglioli, Crino and Gancia, 2018), an increase in innovation (Jones and Kim, 2018, Aghion et al., 2019), tax system changes (Piketty, Saez and Stantcheva, 2014), or increased occupational specialization (Edmond and Mongey, 2021). Our theory is agnostic about which explanation matters most for the original rise in income inequality, and focuses on the resulting spillovers to other occupations. Nevertheless, we trace out the spillover effects of a specific shock, namely deregulation in the financial sector as studied by Philippon and Reshef (2012).

Finally, our paper relates to a literature on demand spillovers. In the sociology literature, Wilmers (2017) shows in OLS panel regressions a positive association between wage inequality and dependence on high-income consumers at the industry level. Manning (2004) and Mazzolari and Ragusa (2013) relate the polarization of labor markets to an increase in high-skill workers' demand for low-skill services. Leonardi (2015) argues that high-skill workers also demand relatively more services from other high-skill workers, a pattern that can amplify increases in the skill premium.[5] Importantly, our focus is not the skill premium across occupations, but on changes in top income inequality within an occupation.

2 The Rise of Within-Occupation Top Income Inequality

We begin by showing the importance of within-occupation trends in top income inequality. Among workers with positive wage and salary income, the ratio of incomes at the 98^{th} to 90^{th} percentile rose from 1.7 to 2.0 between 1980 and 2012.[6] The ratios also increased for physicians—from 1.5 to 1.8—and for dentists and real estate agents.

To systematically understand the role of occupations, Figure 1 decomposes overall changes in wage income from 1980 to 2012 into within- and between-occupation components. The green series in Panel (a) shows the overall change in log wage income at each percentile of the distribution. This reproduces the well-known fact that incomes have grown the fastest in the top of the distribution during this time period. We then adapt the within- and between-firm decomposition of Song et al. (2019) to use occupations instead of firms.[7] The green series shows that the average log income of

[5]In Buera and Kaboski (2012), structural change leads to a rise in the skill premium as the demand for skill-intensive service increases with income.

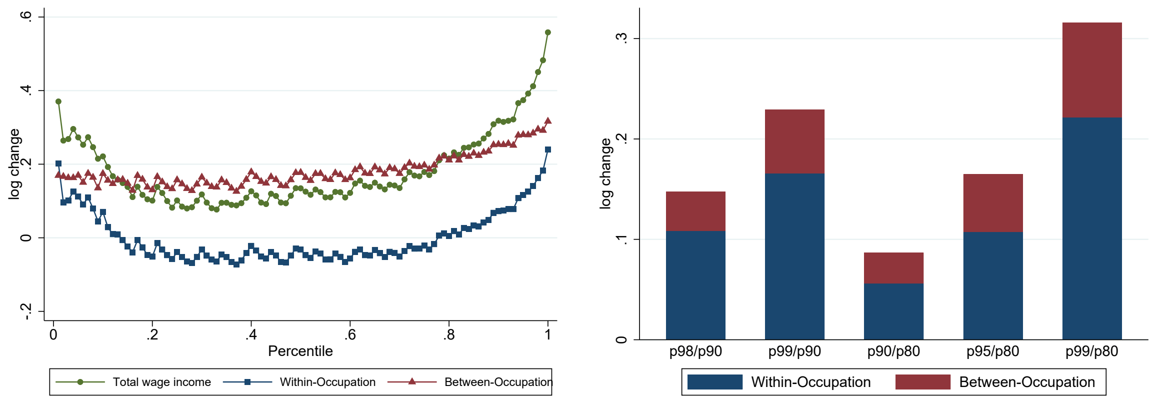
[6]We rely on the Decennial Census (1980, 1990, 2000) and the American Community Survey 2010-2014 waves (henceforth 2012). Appendix Table D.1 shows the corresponding changes for selected occupations. Details on the data are in Section 4 and Appendix B.1.

[7]To find the effect of within-occupation changes, we hold the average log wage income for oc-

Figure 1: Changes in wage income from 1980 to 2012 - between and within occupations

(a) Changes along the income distribution

(b) Decomposition for percentile ratios



Notes: Panel (a) shows the log change in wage income between 1980 and 2012 for all percentiles of the income distribution (green line). Following the method of Song et al. (2019), we decompose this change into changes attributable to between-occupation (red line) and within-occupation (blue line) changes. Panel (b) shows the contributions of between- and within- occupation effects for top income inequality. p99/p90 is the ratio of the top 1% to the top 10%.

the top 1% rose by 0.56 log points during this period, and the blue series shows that increases in within-occupation income inequality drove 0.24 log points of this total.

Regardless of how we measure top income inequality, inequality within occupations plays a central role. Panel (a) shows that income at the 99th and 90th percentiles rose by 0.56 and 0.32 log points, respectively, implying a 0.24 increase in their difference. Within-occupation factors account for 0.17(= 0.24–0.07) of this change (blue bar), *i.e.* 70% of the total. Panel (b) shows similar patterns for other percentile ratios; 65–75% of the rise in top inequality reflects within-occupation changes. Within-occupation inequality rose most in the top, motivating our focus on the top 10%.[8]

3 Theory

Motivated by these patterns, we build an assignment model between doctors and their patients to study inequality spillovers across occupations. Section 3.1 presents a special case and Section 3.2 the more general results. Section 3.3 relaxes several assumptions, including allowing for occupational and geographical mobility. Section

cupations fixed at the level of 1980 and only include the changes in the distribution around the averages. For the between-occupation changes we hold the distribution around the average constant, but change the average log wage for occupations. Due to the binning, these two effects don't identically sum to the total change, though in practice the differences are small. See Appendix B.1 and Song et al. (2019) Online Appendix E for further details.

[8]This is consistent with Edmond and Mongey (2021) who, using CPS data, find that residual income inequality has risen for high-skill workers but fallen for low-skill workers. See also Erosa et al. (2025).

3.4 summarizes our empirical predictions.

3.1 The Cobb-Douglas special case

We consider an economy populated by two types of agents: widget makers of mass 1 and (potential) doctors of mass μ.

Production. Widget makers represent the general population. They produce widgets, a homogeneous numeraire good. A widget maker of ability x can produce x widgets. The ability distribution is Pareto with parameter $\alpha_x > 1$ on $x \geq x_{\min}$, such that a widget maker has ability $X > x$ with probability $P(X > x) = \left(\frac{x_{\min}}{x}\right)^{\alpha_x}$. The Pareto parameter, α_x, is an (inverse) measure of the spread of abilities. We treat α_x as exogenous throughout and a fall in α_x captures a general increase in top income inequality.[9] Such a change could arise from globalization or new technology and directly impacts widget makers but not doctors. We set $x_{\min} = \frac{\alpha_x - 1}{\alpha_x}\widehat{x}$ to fix the mean at $\widehat{x}$ when α_x changes.

Doctors produce health services and can each serve λ customers, where we impose $\lambda > \mu^{-1}$ so that there are enough doctors to serve everyone. Potential doctors differ in their ability z, according to a Pareto distribution with shape α_z. They have ability $Z > z$ with probability $P(Z > z) = \left(\frac{z_{min}}{z}\right)^{\alpha_z}$. All potential doctors can alternatively work as widget makers and produce widgets at some constant ability, which, without loss of generality, we set at x_{min}. (In Section 3.3.2 we instead let an individual's potential ability as a doctor and a widget maker be perfectly correlated). Unlike Rosen (1981), the ability of a doctor does not change how many patients she can treat. Instead, her skill increases the utility patients get from her care.

Consumption. Widget makers are also doctors' patients. Their preferences over the two goods are represented by the Cobb-Douglas utility function:

$$u(z, c) = z^{\beta} c^{1-\beta}, \tag{1}$$

where c is the consumption of widgets and z is the quality of (one unit of) health care. This quality is equal to the ability of the doctor providing the care. The notion that medical services are not divisible is captured by the assumption that each patient needs to purchase care from exactly one doctor; one cannot substitute quantity for quality. As a result, there need not be a common price per unit of quality-adjusted

[9]Guvenen, Karahan, Ozkan, and Song (2021) show that the top of the income distribution is well-described by a Pareto distribution.

medical services. More generally, "doctors" here stand in for any occupation which produces non-divisible goods or services for the general population, including dentists and real estate agents.[10] The Cobb-Douglas utility function eases exposition but can be generalized, as we do in Section 3.2. For simplicity, doctors only consume widgets, so the patients are exclusively widget makers.

3.1.1 Equilibrium

Widget makers. Since a widget maker of ability x produces x homogeneous widgets, widget makers' income must be distributed like their ability. The consumption problem of a widget maker of ability x can then be written as:

$$\max_{z,c} u(z,c) = z^{\beta} c^{1-\beta} \text{ subject to } \omega(z) + c \leq x,$$

where $\omega(z)$ is the price of medical care from a doctor of ability z. Taking first order conditions with respect to z and c yields:

$$\omega'(z)\, z = \frac{\beta}{1-\beta}\left[x - \omega(z)\right]. \tag{2}$$

With Cobb-Douglas preferences, no widget maker spends all their income on health care, so equation (2) implies that $\omega(z)$ must be increasing: Higher-ability doctors earn more per patient. Importantly, the non-divisibility of medical services makes doctors "local monopolists" who compete directly only with doctors of slightly higher or lower ability, and $\omega(z)$ need not be linear in z.

As long as the utility function has positive cross-partial derivatives, the equilibrium involves positive assortative matching between widget makers' income and doctors' ability (see Appendix A.1). We denote the matching function as $m(z)$: a doctor of ability z will be hired by a widget maker whose income is $x = m(z)$.

Doctors. Since there are more doctors than needed, those with lowest z will work as widget makers rather than as physicians. Letting z_c be the ability level of the least able practicing doctor, $m(z)$ is defined over $[z_c, \infty)$ and $m(z_c) = x_{\min}$; the worst

[10] Although we refer to the "quality" of the good, nothing in our model relies on the "high-quality" goods being objectively superior. It is really "quality as perceived by top-earning patients." So a pediatrician who can assuage an anxious parent might have a higher z than one with better diagnostic skills but fewer interpersonal skills. That said, the empirical literature using revealed-preference finds it to be correlated with measures of medical outcomes (*e.g.*, Dingel et al., 2023).

doctor is hired by a patient with income $x_{\min}$. Market clearing at each z implies:

$$P\left(X > m\left(z\right)\right) = \lambda\mu P\left(Z > z\right), \ \forall z \geq z_c. \tag{3}$$

There are $\mu P\left(Z > z\right)$ doctors with an ability higher than z, each of these doctors can serve λ patients, and there are $P\left(X > m\left(z\right)\right)$ patients whose income is higher than $m\left(z\right)$. With Pareto distributions, we can write the matching function explicitly:

$$m\left(z\right) = x_{\min}\left(\lambda\mu\right)^{-\frac{1}{\alpha_x}}\left(\frac{z}{z_{min}}\right)^{\frac{\alpha_z}{\alpha_x}}. \tag{4}$$

Intuitively, if top talent is relatively more abundant among doctors than widget makers ($\alpha_z < \alpha_x$,), then the matching function is concave. Conversely, it is convex if $\alpha_z > \alpha_x$. We then obtain the cutoff value $z_c = \left(\lambda\mu\right)^{\frac{1}{\alpha_z}} z_{min}$.

Let $w\left(z\right)$ denote the income of a doctor of ability z and note that $w(z) = \lambda\omega(z)$ since each doctor provides λ units of health services. As a potential doctor of ability z_c is indifferent between working as a doctor or a widget-maker earning x_{min}, we must have $w\left(z_c\right) = x_{\min}$. Combining (2) and (4), we obtain a differential equation that the wage function $w\left(z\right)$ must satisfy:

$$w'\left(z\right)z + \frac{\beta}{1-\beta}w\left(z\right) = \frac{\beta}{1-\beta}x_{\min}\left(\frac{\lambda^{\alpha_x - 1}}{\mu}\right)^{\frac{1}{\alpha_x}}\left(\frac{z}{z_{min}}\right)^{\frac{\alpha_z}{\alpha_x}}. \tag{5}$$

Using the boundary condition at $z = z_c$, we obtain a single solution for the wage profile of doctors. Appendix A.2.1 demonstrates that this function is:

$$w\left(z\right) = x_{\min}\left[\frac{\lambda\beta\alpha_x}{\alpha_z\left(1-\beta\right)+\beta\alpha_x}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x}} + \frac{\alpha_z\left(1-\beta\right)+\beta\alpha_x\left(1-\lambda\right)}{\alpha_z\left(1-\beta\right)+\beta\alpha_x}\left(\frac{z_c}{z}\right)^{\frac{\beta}{1-\beta}}\right]. \tag{6}$$

As expected, the wage profile $w\left(z\right)$ is increasing in doctors' ability z, and $w\left(z_c\right) = x_{\min}$. The first term on the right hand side of (6) dominates for large $\frac{z}{z_c}$ and implies an asymptotic Pareto distribution, so that for large $\frac{z}{z_c}$, we get:

$$w\left(z\right) \approx x_{\min}\frac{\lambda\beta\alpha_x}{\alpha_z\left(1-\beta\right)+\beta\alpha_x}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x}}. \tag{7}$$

Equation (7) shows that the wage schedule at the top is concave in z if $\alpha_z < \alpha_x$;

that is, if talent is relatively more abundant among physicians than widget makers. To see why, suppose counterfactually that $\omega(z) \propto z$. Widget makers would then spend a rising share of income on medical services. However, this is in conflict with Cobb-Douglas utility for linear pricing schedules which gives constant spending shares. Hence, this cannot be an equilibrium: demand for high-ability doctors would have to drop, reducing their prices, and resulting in a concave payment schedule. Non-divisibility of medical services is crucial: if high-earning widget makers could simply substitute the services of one doctor of ability z with two doctors of ability $z/2$, a linear pricing schedule would emerge. Conversely, the payment schedule would be convex if talent were relatively scarce among doctors. This wage schedule then guarantees that consumers spend asymptotically a constant share of their income on health care (to see this, combine (4) and (7)). We can thus derive:

Proposition 1. *Doctors' income is asymptotically Pareto distributed with the same shape parameter as for the widget makers. In particular, an increase in top income inequality for widget makers increases top income inequality for doctors. Health expenditures grow proportionately with income: $h(x) \approx \eta x$, where η is a constant.*

To see this, we first define the relevant distribution. For active doctors, the probability that income exceeds w_d is given by $P_{doc}(W_d > w_d) = \left(\frac{w^{-1}(w_d)}{z_c}\right)^{-\alpha_z}$. Using equation (7), for w_d large enough, we can approximate this distribution as:

$$P_{doc}(W_d > w_d) \approx \left(\frac{x_{\min}\lambda\beta\alpha_x}{\alpha_z(1-\beta)+\beta\alpha_x}\frac{1}{w_d}\right)^{\alpha_x}. \tag{8}$$

That is, the income of (active) doctors is Pareto distributed at the top. Importantly, in this Cobb-Douglas setting, the shape parameter is inherited from the widget makers, and is independent of α_z, the spread of doctor ability. A decrease in α_x directly translates into a decrease in the Pareto parameter for doctors' income distribution. In other words, the increase in top income inequality spills over from one occupation (the widget makers) to another (doctors). At the top, widget makers' inequality also increases doctors' incomes—a decrease in α_x leads to an increase in $P(W_d > w_d)$ for w_d high enough.[11] Further, a decrease in the mass of potential doctors μ (equivalent

[11]Not all doctors benefit, though, as we combine a decrease in α_x with a decrease in x_{min} to keep the mean constant. As a result the least able active doctor, whose income is $x_{\min}$, sees a decrease in her income. Had we kept $x_{\min}$ constant so that a decrease in α_x also increases the average widget maker income, then all doctors would have weakly gained.

to an increase in the mass of widget makers) increases the share of doctors who are active (z_c decreases) and their wages (as $w(z)$ increases if z_c decreases) but has no effect on doctors' top income inequality.

Our results directly generalize to the case where patient income and doctor ability are only asymptotically Pareto distributed and where potential doctors may (but need not) also consume medical services (see details in Appendix A.3). Our results also hold when doctors' ability distribution has a tail fatter than Pareto. When the tail is thinner than Pareto, a spillover result still holds, although doctors' income is no longer Pareto distributed (see Appendix A.4).

Taking stock. Proposition 1 establishes the central theoretical result: Changes in widget makers' income inequality translate directly into doctors' income inequality.

3.1.2 Welfare inequality

The lack of a uniform quality-adjusted price implies that prices vary along the income distribution. Heterogeneity in consumption patterns implies that people at different points of the income distribution face different price indices (Deaton, 1998). A given increase in income inequality thus translates into a lower increase in welfare inequality. The assignment mechanism implies that as inequality increases, the richest widget makers cannot obtain better health services; in fact, they pay more for health services of the same quality. This mechanism limits the increase in welfare inequality.[12]

To assess this formally, we use a consumption-based measure of welfare. We compute the level of consumption of the homogeneous good $eq(x)$ that, when combined with a fixed level of health quality z_r, gives the same utility to the widget maker as what she actually obtains. That is, we define $eq(x)$ through $u(z_r, eq(x)) = u(z(x), c(x))$. This yields (with a proof in Appendix A.2.2):

Proposition 2. *For x large enough, welfare eq(x) is Pareto-distributed with shape parameter* $\alpha_{eq} \equiv \frac{\alpha_x}{1+\frac{\alpha_x}{\alpha_z}\frac{\beta}{1-\beta}}$. *Thus* $\frac{d\ln\alpha_{eq}}{d\ln\alpha_x} = \frac{1}{1+\frac{\alpha_x}{\alpha_z}\frac{\beta}{1-\beta}} < 1$, *so an increase in widget makers' income inequality translates into a less-than-proportional increase in their welfare inequality. The mitigation is stronger when health services matter more (high* β*) or when doctors' abilities are more unequal (low* α_z*).*

[12]Moretti (2013) can be viewed as discussing a similar assignment mechanism: high earners locate in high-cost cities, so real inequality is lower than nominal inequality across space.

3.2 Generalizing the utility function

Our results obtain in a much more general case than the Cobb-Douglas utility assumed so far. In Appendix A.3.3, we show that they generalize to any homothetic utility function that admits positive and finite limits to the elasticity of substitution between health care quality and the homogeneous good when z/c tends to either infinity or 0. For simplicity, we focus here on the case when preferences have a constant elasticity of substitution (CES). That is, we replace the utility function of equation (1) with:

$$u(z,c) = \left(\beta z^{\frac{\varepsilon-1}{\varepsilon}} + (1-\beta)\, c^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{9}$$

where ε is the elasticity of substitution between health care quality and the homogeneous good. We also only assume that widget makers' and doctors' ability distributions are asymptotically Pareto. With CES preferences, the equilibrium still features positive assortative matching. In Appendix A.3.2, we show:

Proposition 3. *If either (i) $\varepsilon > 1$ and $\alpha_z^{-1} \leq \alpha_x^{-1}$, or (ii) $\varepsilon < 1$ and $(1-\varepsilon)\,\alpha_z^{-1} < \alpha_x^{-1} \leq \alpha_z^{-1}$, doctors' wages are asymptotically Pareto distributed with shape parameter*

$$\alpha_w^{-1} = \frac{1}{\varepsilon}\alpha_x^{-1} + \left(1 - \frac{1}{\varepsilon}\right)\alpha_z^{-1}, \tag{10}$$

so doctors' top income inequality increases with widget makers top income inequality. In addition, log health expenditures grow proportionately with log income: $\ln h(x) \approx \left[\left(1 - \frac{\alpha_x}{\alpha_z}\right)\frac{1}{\varepsilon} + \frac{\alpha_x}{\alpha_z}\right]\ln x + \hat{\eta}$*, where $\hat{\eta}$ is a constant.*

Proposition 3 restricts attention to two cases. We argue below that, beyond Cobb-Douglas ($\varepsilon = 1$), these are the two empirically relevant cases. In both cases, doctors' income is asymptotically Pareto distributed with an inverse Pareto parameter that is a linear combination of those for doctors' ability and widget makers' income. An increase in widget makers' top income inequality (α_x^{-1}) leads to an increase in doctors' top income inequality, with a coefficient that has a direct interpretation as the inverse elasticity of substitution between health care quality and the homogeneous good. Health care expenditures increase less than proportionately with income for $\alpha_z \neq \alpha_x$.

To understand the results of Proposition 3 intuitively, and why we restrict attention to certain parameter sets, consider the two cases in turn. First, let medical services and the outside good be substitutes ($\varepsilon > 1$) and let top doctors' skill be relatively scarce ($\alpha_z^{-1} < \alpha_x^{-1}$). In the Cobb-Douglas case ($\varepsilon = 1$), the pricing schedule would be convex (see (7)) and widget makers would spend a constant share of their income on health care. With $\varepsilon > 1$, widget makers reduce their demand for relatively expensive health-care services and health expenditures grow less than proportionately with income, since $\left(1 - \frac{\alpha_x}{\alpha_z}\right)\frac{1}{\varepsilon} + \frac{\alpha_x}{\alpha_z} < 1$. As a result, the wage schedule is less convex than in the Cobb-Douglas case, and the inverse Pareto coefficient for doctors' income distribution, α_w^{-1}, is smaller than α_x^{-1}—but still increasing in α_x^{-1}. In contrast, if $\varepsilon > 1$ and doctors are abundant at the top ($\alpha_x^{-1} < \alpha_z^{-1}$), widget makers would increase their demand for health care services, asymptotically spending nearly all their income on health care, which is counterfactual (and therefore ignored in Proposition 3).

The other case considered in Proposition 3 is when medical services and other goods are complements ($\varepsilon < 1$) and when doctors are relatively abundant in the top: $\alpha_x^{-1} < \alpha_z^{-1}$. With relatively abundant medical services and complementarity between goods, health care expenditures rise less than proportionately with income. Doctors' income is then again asymptotically Pareto distributed with shape parameter $\alpha_w^{-1} < \alpha_x^{-1}$, provided doctors are not too abundant at the top ($(1-\varepsilon)\,\alpha_z^{-1} < \alpha_x^{-1}$).[13]

Our empirical analysis suggests that the CES case is more relevant than Cobb-Douglas. First, Appendix C estimates empirical Engel curves for medical care. We find slopes less than 1, in line with Proposition 3 and in contrast with the Cobb-Douglas case. Second, we will estimate spillover coefficients above 1 (though not always statistically significantly above 1), corresponding to an elasticity $\varepsilon < 1$. Importantly, when $\varepsilon < 1$, a growing spread in doctors' ability (an increase in α_z^{-1}) *reduces* doctors' income inequality (α_w^{-1} decreases). This is because as α_z^{-1} increases, more top doctors compete for patients who are spending a declining share of their income on health care as we move into the tail. In our model, skill-biased technical change for widget makers can be captured by an increase in α_x^{-1}. Similarly, an increase in α_z^{-1} can capture technical change that increases the relative ability of doctors at the top of the distribution. Our results therefore show that this form of technical change for doctors can paradoxically decrease their top income inequality in what will appear

[13] Equation (10) always holds as long as the right-hand side lies in the interval $[0, \alpha_x^{-1}]$ and covers the Cobb-Douglas case. Appendix A.3.2 presents the results for the alternative parameter spaces. If $\varepsilon < 1$ and $\alpha_x^{-1} < (1-\varepsilon)\,\alpha_z^{-1}$, doctors' income is not Pareto distributed and the slope of the Engel curve is asymptotically zero ($\ln h(x) / \ln x \to 0$) for high incomes. If $(\varepsilon - 1)\left(\alpha_z^{-1} - \alpha_x^{-1}\right) > 0$, widget makers asymptotically spend all their income on health care and doctors' income is Pareto distributed with shape parameter α_x. Appendix C shows that the slope of the (log) Engel curve is asymptotically positive and finite, ruling out these two cases.

to be the most relevant case empirically.

Finally, α_z and α_x are likely to be positively correlated empirically: places with more talent dispersion for widget makers are likely to also have more talent dispersion for doctors. Therefore, without a control for the (unobserved) physician ability distribution, the coefficient of an OLS regression of physician income inequality on widget makers' income inequality would suffer from a downward bias when the true coefficient is above 1—consistent with our findings in Section 5.1.

3.3 Extensions

Our model makes several assumptions about preferences, the structure of labor markets and production. To devise appropriate empirical tests for spillovers, we need to establish which assumptions drive the results and which are innocuous. We now show that our results are robust to introducing scalability for medical services, geographical mobility, and occupational mobility for doctors. In contrast, two assumptions are necessary for local inequality spillovers: sufficiently low substitutability between quality and quantity of medical services and non-tradability.

3.3.1 Scalability

While doctors' supply was inelastic in the baseline model, we now allow them to increase output at some cost. Specifically, doctors pay effort costs $k\lambda^{1/\varepsilon^S+1}/\left(1/\varepsilon^S+1\right)$ where $k>0$, $\varepsilon^S>0$, and λ is the number of patients treated. (Results are identical if doctors pay monetary costs.) Doctors' utility maximization problem immediately implies that their scale depends on the price they can charge as

$$\lambda\left(z\right)=\left(\omega\left(z\right)/k\right)^{\varepsilon^S}, \tag{11}$$

so that ε^S is the (intensive margin) supply elasticity of medical services. Doctors' income is then given by $w\left(z\right)=\lambda\left(z\right)\omega\left(z\right)=\omega\left(z\right)^{1+\varepsilon^S}/k^{\varepsilon^S}$. The widget makers' consumption problem remains identical and there is still positive assortative matching, though market clearing takes into account that doctors serve different number of patients. In Appendix A.5, we solve the model and prove:

Proposition 4. *Doctors' income is asymptotically Pareto distributed when:*
utility is CES with either (i) $\varepsilon=1$, *(ii)* $\varepsilon<1$ *and* $\left(1-\varepsilon\right)\alpha_x<\alpha_z<\varepsilon^S+\alpha_x$, *or (iii)*

$\varepsilon > 1$ *and* $\alpha_z > \varepsilon^S + \alpha_x$. *In all cases the shape parameter is*

$$\alpha_w^{-1} = \frac{1+\varepsilon^S}{1+\frac{\varepsilon^S}{\varepsilon}\alpha_x^{-1}}\left(\left(1-\frac{1}{\varepsilon}\right)\alpha_z^{-1} + \frac{1}{\varepsilon}\alpha_x^{-1}\right). \tag{12}$$

An increase in widget makers' top income inequality increases doctors' income top income inequality ($\frac{d\alpha_w^{-1}}{d\alpha_x^{-1}} > 0$). *A higher supply elasticity* ε^S *increases doctors' top income inequality if and only if* $\varepsilon\alpha_x > 1$.

This Proposition shows that the results of Propositions 1 and 3 generalize as long as the supply elasticity of doctors is finite.[14] An increase in widget makers' top income inequality increases doctors' top income inequality, by increasing inequality in both health care prices and quantities supplied (with "weights" $\frac{1}{1+\varepsilon^S}$ and $\frac{\varepsilon^S}{1+\varepsilon^S}$, respectively).

A higher supply elasticity for widget makers would unambiguously increase their income inequality, but a higher ε^S has an ambiguous effect on doctors' top income inequality. Top doctors can disproportionately expand their scale, but the increased supply of high-quality services depresses prices. This price effect is stronger when health care and the homogeneous good are more complementary (ε low), so higher ε^S reduces doctors' income inequality if $\varepsilon < \alpha_x^{-1}$. Therefore, technological change (*e.g.* greater task delegation) that allows top doctors to serve more patients (as in the classic superstar story) need not increase top income inequality. Empirically, a uniform increase in ε^S across regions would not bias our coefficient; but if areas with rising widget makers' inequality also experience larger increases in ε^S, our OLS estimates would be biased upwards when $\varepsilon\alpha_x > 1$.

3.3.2 Occupational mobility

We have assumed so far that a potential doctor choosing to work as a widget maker earns the minimum widget maker income, x_{min}. In practice, those succeeding as doctors may have succeeded in other occupations as well (Kirkeboen, Leuven and Mogstad, 2016). To capture this, we now switch to the opposite extreme and assume perfect correlation between an individual's ability as a doctor and a widget maker. More specifically, we assume that there is a mass of 1 of agents with a unidimensional (Pareto) distribution of skills who decide whether to be doctors or widget

[14] Gottlieb et al. (2025) estimate a supply elasticity of 0.4 when payments for a doctor's services change. Looking across a broader set of supply responses—beyond those relevant for the Proposition—Clemens and Gottlieb (2014) estimate supply elasticities around 1. There is no contention in the literature that supply is perfectly elastic.

makers. While Section 3.3.1 examined the intensive labor supply margin (doctors increasing their scale), this extension studies the extensive labor supply margin (more talented individuals choosing to become doctors). For simplicity, we focus on the Cobb-Douglas case (though similar results can be obtained in the CES case).

Appendix A.6 shows that Proposition 1 still applies: Doctors' incomes are Pareto distributed with the same coefficient α_x as widget makers, as long as $\lambda^{\alpha_x - 1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta}+1\right)^{-\alpha_x} < 1$. This condition ensures that in equilibrium, above a certain threshold, individuals of a given ability choose to become both doctors and widget makers (otherwise all top individuals choose to be doctors). Intuitively, once individuals have decided their occupation, the ability distribution of doctors is again Pareto, so the results of the baseline model still apply.[15]

3.3.3 Geographical mobility

We now extend the baseline model to allow doctors to move across two regions, A and B, of equal size. Medical services are non-tradable and patients are immobile. The regions differ only in the Pareto shape parameter of widget makers' ability distributions, with A more unequal than B ($\alpha_x^A < \alpha_x^B$). Doctors' abilities are Pareto distributed with a common parameter α_z. For simplicity, we focus on the Cobb-Douglas case, though similar results can be derived in the CES case (as well as settings with multiple regions and heterogeneous masses of potential doctors and widget makers).

In autarky, each region mirrors the baseline model: doctors' incomes are asymptotically Pareto distributed with shape parameters matching the local widget maker distribution. With doctor mobility, however, wages $\omega(z)$ must equalize across regions. Since incomes are higher at the top in region A, high-ability doctors from B migrate to A, while labor market clearing requires lower-ability doctors to move from A to B. As a result, A has a more unequal distribution of doctors' ability than B, but both equilibrium ability distributions are asymptotically Pareto. Appendix A.7 proves:

Proposition 5. *Once doctors have relocated, the income distribution of doctors in region A is asymptotically Pareto with coefficient* α_x^A*, and the income distribution of doctors in region B is asymptotically Pareto with coefficient* α_x^B*.*

[15] With occupational mobility, note that doctors and widget makers interact through both a demand effect and a labor supply effect. Since the wage level is directly determined by doctors' outside option, one may think that the mechanism which leads to spillovers in income inequality is very different compared to the demand-side mechanism of the baseline model. In Appendix A.6.1, we split the role of widget makers into two (patients and an "outside option") and show that, with Cobb-Douglas preferences, doctors' income inequality is only driven by their patients' inequality.

Thus, doctors' observed income inequality continues to reflect local consumer inequality even with doctor mobility. Our empirical analysis therefore does not require us to model geographic mobility of doctors explicitly.[16]

3.3.4 Quality-quantity tradeoff

Our baseline model assumes that healthcare is non-divisible, *i.e.* quantity cannot substitute for quality of health care. How crucial is this assumption? To answer this, we consider the baseline model with Cobb-Douglas preferences between the outside good and health care consumption but now let health care, H, be a CES aggregate of quality z and quantity q. We assume that a patient must consume health care services of the same quality. We replace the utility function in (1) by

$$u(q,z,c) = H^{\beta} c^{1-\beta} \text{ with } H \equiv \left((1-\gamma)^{\frac{1}{\theta}} q^{\frac{\theta-1}{\theta}} + \gamma^{\frac{1}{\theta}} z^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}},$$

where θ measures the elasticity of substitution between quality and quantity. In Appendix A.8, we show that the equilibrium still features positive assortative matching if $\theta < 1$, and as long as $\alpha_z > \alpha_x - 1$, doctors' income distribution is asymptotically Pareto distributed with top income inequality spillovers from consumers to doctors (with $\alpha_z < \alpha_x - 1$ doctors' income is bounded). In contrast, if $\theta = 1$, there is no positive assortative matching and doctors' income is proportional to their ability.

3.3.5 Tradability

Finally, we permit trade in medical services across regions. Since our empirical analysis relies on local spillovers of income inequality, this extension explicitly addresses predictions when services are not sold in a local market. Consider the baseline model of Section 3.1, but with several regions indexed by $s \in \{1, \ldots, S\}$. We allow some patients (a positive share of rich widget makers) to purchase their medical services across regions. The distribution of potential doctors' ability is the same in all regions, and so is the number of patients served per doctor, λ. The other parameters—in particular the Pareto shape parameter of widget makers' income α_x^s—are allowed to differ across regions. It follows that in the top, national income is asymptotically distributed with

[16] Interestingly, the nature of the spillover effect is different with and without mobility: In autarky, local income inequality affects doctors' pay schedule and thereby doctors' local income inequality. With mobility of doctors and many regions, a change in local income inequality does not affect the pay scale for doctors as a function of their ability at the top (except for the most unequal region), but it changes the local *ex post* ability distribution and thereby doctors' local income inequality.

the lowest α_x^s, that of the most unequal region.[17]

The cost of high-quality health care services must be the same everywhere; otherwise, the widget makers who can travel would go to the region with the cheapest health care. Therefore top talented doctors must all earn the same wage for the same ability. Since national top income inequality for widget makers is $\min_s \alpha_x^s$, doctors' income in all regions is asymptotically Pareto distributed with shape parameter $\min_s \alpha_x^s$ (see details in Appendix A.9). That is, there are national but not local spillovers.[18]

It is important to recognize that spillovers still exist. An increase in income inequality for the region with the highest income continues to determine the income inequality of physicians (in all regions). However, an empirical analysis based on local spillovers will fail to find an effect at least in the tail. Empirically, whether the service provided is "local" (non-tradable) or "non-local" (tradable) will depend on the occupations of interest. We will use these results to guide our empirical analysis.

3.4 Empirical predictions and institutional details on health care

To summarize, our model makes the following predictions (where "doctors" represent all occupations that fit the assumptions). (1) High-earning consumers are treated by more expensive doctors. (2) An increase in local inequality will increase local inequality for doctors if they serve the general population directly and their services are non-divisible (or more generally feature limited substitution between quality and quantity). (3) This is true regardless of whether (i) doctors can move across regions and (ii) whether doctors' ability is positively correlated with the income they would receive in alternative occupations. (4) If medical care is tradable, doctors' income in each region does not depend on local income inequality, but on national inequality.

While "doctors" are a primary example of an occupation with inequality spillovers, the medical industry in the U.S. is not perfectly described by the flexible price-setting of our model. The government plays a substantial role through Medicare and Medicaid, the insurance sector acts as an intermediary, and there is information asymmetry between patients and doctors. But these institutional intricacies need not inhibit

[17] If doctors are mobile and medical services are also tradable, the geographic location of agents is undetermined in general, and we would need a full spatial equilibrium model to generate empirical predictions.

[18] Formally, we show in a model with a continuum of agents, that the income distribution of doctors is approximately Pareto with shape parameter $min_s \alpha_x^s$ above a certain cutoff for any positive share of mobile patients. That cutoff depends positively on the share of mobile patients. In practice, national spillovers would be negligible if only a small fraction of patients travel.

market forces—including our spillovers—from operating. In fact, they may offer a mechanism that implements the forces our model discusses. Providers' negotiations with private insurers generally lead to higher prices in the private market than under public insurance (Clemens and Gottlieb, 2017). Despite asymmetric information, patients often have clear beliefs about who the "best" local doctor in a specific field is, whether or not these beliefs relate to medical skill or health outcomes (Kolstad, 2013; Epstein, 2006; Steinbrook, 2006). Finally, though Dingel et al. (2023) document patients sometimes traveling for care, the distance sensitivity is high.[19] Therefore, despite these complications, the structure of the health industry may embody enough flexibility to incorporate the local economic pressures implied by our model.

Appendix C provides evidence in favor of our model's first prediction, namely that there is positive assortative matching between patients' income and medical providers' prices, using medical claims data. We also document that health spending increases with income, using a nationally representative survey. In the rest of the paper, we test our core prediction: within a local geographic market, inequality spills over into occupations that provide non-divisible services with heterogeneous quality.

4 Empirical Strategy to Identify Spillovers

This section introduces the main empirical test of our model, using geographical variation in income inequality across LMAs in the U.S. Our goal is to estimate the causal effect of general income inequality on income inequality within a specific occupation.

4.1 Income data

We use the Decennial Censuses from 1980, 1990, and 2000 and the 2010-2014 (referred to as "2012") waves of the American Community Survey (ACS). We refer to this combined data as Census data. We use 2010-2014 as opposed to 2008-2012 to avoid the immediate aftermath of the Great Recession, which had a large impact on top incomes. We access the restricted-use versions of each which contain a larger sample of respondents and less income censoring than public-use versions.[20] Data from before

[19]Dingel et al. (2023) show that in 2017 around four-fifths of medical care is consumed in the same Hospital Referral Region (roughly the size of the Labor Market Areas we use) where it is produced. Our strategy focuses on an earlier period, 1980–2012, where trade in medical services was lower. It also more heavily weights large metropolitan areas, which tend to have more types of care and expert physicians available, so patients from these regions travel for care even less frequently. Regardless, our theoretical analysis suggests that national travel would reduce our estimated regional spillovers.

[20]The detailed Decennial Census long-form surveys 1/6th of the population. Each of the five ACS samples is 3%, so combining them gives a sample of 15%. We inflation-adjust incomes to 2000

1980 has substantially smaller samples, and we exclude them from the main analysis. Appendix B.2 discusses the definitions of occupations and geographic locations (Labor Market Areas, LMAs) that we use.

Motivated by our theoretical model, we measure a distribution's top income inequality by its estimated Pareto parameter. Consider a set of observations $\mathcal{N} = \{x_i\}_{i=1}^N$ drawn from a Pareto distribution with two parameters: the minimum value and the Pareto parameter. The maximum likelihood estimate for the minimum value is $\hat{x}_{min} = \min\{x_i\}_{i=1}^N$ and for the Pareto parameter:

$$\hat{\alpha}^{-1} = \frac{1}{N}\sum_{i\in\mathcal{N}} \ln\left(\frac{x_i}{\hat{x}_{min}}\right), \tag{13}$$

That is, the estimate of the inverse Pareto parameter is the average log distance from the chosen minimum value. So $\hat{\alpha}^{-1}$ is a measure of income inequality even if the distribution is not exactly Pareto. We adjust (13) for the small number of censored observations (see Appendix B.1). Guvenen, Karahan, Ozkan, and Song (2021) and Jones and Kim (2014) also use α^{-1} as a measure of inequality.

Our focus on top income inequality requires choosing x_{min}. Since the Pareto distribution is a good fit in the top decile, we set x_{min} as the 90^{th} percentile of the local income distribution of employed adults (age ≥ 25) with positive wage income.

Our analysis is demanding of the data: for a given occupation in a given location, computing top income inequality requires many observations. To get enough observations, we use the same x_{min} to compute all occupation-specific $\hat{\alpha}_o^{-1}$ measures. For example, to compute top income inequality of physicians in New York, we use physicians in the top 10% of New York's overall distribution, which is much more than 10% of New York's physicians because they are high-earners. This approach also implies that we use a consistent slice of the population—those in the upper decile of their area's distribution—when considering different outcome occupations (*e.g.* dentists, real estate agents). This aligns with the motivating observation in Figure 1a that rising within-occupation inequality is concentrated at the top of the distribution.

For our regressions, we will use a balanced panel of the 50 most populous LMAs (as of 1980). These areas tend to have a sufficient number of observations to reliably

dollars. Whereas the publicly available data is censored at around the 99.5th percentile of the overall income distribution, the restricted data has very little censoring. For instance, in New York State only around the top 0.1% is censored. Among physicians the number is well below 0.5%.

Table 1: Wage income: Ratio 98/90: actual values and predicted values

Year	General Population α^{-1}	General Population $\frac{98}{90}$	General Population $\widehat{\frac{98}{90}}$	Physicians α^{-1}	Physicians $\frac{98}{90}$	Physicians $\widehat{\frac{98}{90}}$	Dentists α^{-1}	Dentists $\frac{98}{90}$	Dentists $\widehat{\frac{98}{90}}$	Real Estate Agents α^{-1}	Real Estate Agents $\frac{98}{90}$	Real Estate Agents $\widehat{\frac{98}{90}}$
1980	0.34	1.70	1.72	0.25	1.50	1.50	0.29	1.54	1.61	0.42	1.94	1.97
1990	0.38	1.87	1.85	0.40	1.89	1.90	0.35	1.90	1.76	0.47	2.17	2.12
2000	0.42	2.00	1.96	0.33	1.75	1.71	0.33	1.69	1.70	0.54	2.40	2.37
2012	0.42	1.99	1.96	0.34	1.72	1.72	0.36	1.74	1.78	0.50	2.17	2.22

Notes: The inverse Pareto parameter, α^{-1}, is calculated in the top 10% of the relevant national wage income distribution (general population, physicians, dentists, real estate agents). $\frac{98}{90}$ is the ratio of the 98th to 90th percentile of income. $\widehat{\frac{98}{90}}$ is the $\frac{98}{90}$ ratio predicted based on the estimated Pareto parameter as $5^{\alpha^{-1}}$.

compute income inequality for the occupations of interest.[21] Including less populated LMAs starts to introduce imbalance in the panel for the smaller occupations of interest. We discuss robustness to these choices below.

4.2 Income distribution statistics

We now present summary statistics on income distributions and discuss how well they fit the Pareto distribution. Our main income measure is pre-tax wage and salary income. Appendix Table D.2 shows descriptive statistics in 2000 for the most common occupations in the top decile of the national income distribution. It reports each occupation's mean income and the share of the top 1%, 5%, and 10% that the occupation represents. Physicians are one of the most common occupations at the top, accounting for 13% of the top 1% (more than the three financial occupations together) and 4% of the top 10%.

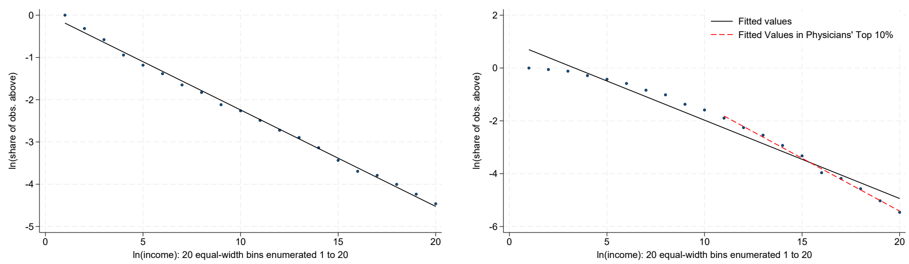
We next analyze whether national income distributions are Pareto at the top. With a large number of observations at the national level, we can estimate $\hat{\alpha}^{-1}$ in the top 10% for doctors, dentists, and real estate agents, respectively. We also compute $\hat{\alpha}^{-1}$ for the general population in their top decile. Table 1 reports the results. $\hat{\alpha}^{-1}$ is mostly larger for the general population than for doctors and dentists which is consistent with the CES model of Section 3.2. To assess the Pareto fit, we compare the ratio of the 98^{th} and 90^{th} income percentiles against the predicted 98/90 ratio assuming income is Pareto distributed (calculated as $5^{1/\alpha}$). The predicted and actual ratios agree closely, consistent with upper tail incomes being Pareto distributed.

Finally, we assess the extent to which the general and physician income distributions are Pareto at the sub-national level. The Pareto distribution implies a linear

[21] For a Pareto distribution, the maximum likelihood estimator $\hat{\alpha}^{-1}$ has a standard error of $\hat{\sigma} = N^{-1/2}\hat{\alpha}^{-1}$. When $N \geq 20$, the ratio $\hat{\sigma}/\hat{\alpha}^{-1} < 0.22$. In Table 6, we indicate how many LMA-year pairs have an outcome $\hat{\alpha}^{-1}$ calculated using fewer than 20 observations. Using the top 50 most populous LMAs ensures that $N \geq 20$ is nearly always the case for key occupations of interest.

Figure 2: Fit of the Pareto Distribution - New York State (2000)

(a) All Occupations **(b)** Physicians



Notes: This figure shows the quality of fit of the empirical income distribution to the Pareto distribution using observations from New York State in 2000. The left panel shows the top 10% among all occupations, and the right panel shows physicians in the top 10% of the total population. The horizontal axis shows the logarithm of incomes split into 20 equal-width bins. The vertical axis shows the logarithm of 1-CDF(income). By construction, the line is downward-sloping, and linear if the underlying distribution is Pareto. The lines are a linear fit to the binned observations. Panel (b) also shows the linear fit among the subset of bins that correspond to the top 10% of physician earners. Source: Authors' calculations using Decennial data.

relationship between log income and the log of 1-CDF(income). The slope of the relationship is the Pareto coefficient α.[22] We test this prediction in New York State (disclosure rules impede testing it at the LMA level). We bin New York's top decile of earners into 20 equally-spaced log income bins. Within each bin, we calculate the share of observations with income in that bin or higher. Figure 2 shows the relationship between the log income bins and log observation shares. Panel 2a shows the relationship for the general population and Panel 2b for physicians specifically. The Pareto fit is excellent for the general population.

For physicians, the distribution is also Pareto in their top 10%, as illustrated by the dashed line. However, the fit worsens lower in the distribution. The α^{-1} estimated on physicians in the top 10% of New York's overall income distribution (as in our regression) is higher than when using the top 10% of physicians. Nevertheless, even when the data are not exactly Pareto, the estimated α^{-1} is a reasonable measure of income inequality.

Appendix Figure B.1 shows that, although the estimate of α^{-1} can be sensitive to the choice of sample, the change between years—the source of identifying variation in our regressions—is much less so. Section 5.3 presents robustness checks to ensure that our results are not driven by deviations from the Pareto distribution.

[22] With N observations of wages drawn from a Pareto distribution, the expected share of observations with a value higher than x, N_x/N, is $(x/x_{min})^{-\alpha}$. This means $\ln(N_x/N) = -\alpha \ln(x) + \alpha \ln(x_{min})$.

4.3 Empirical strategy

Regression framework. We aim to estimate the causal effect of a change in general (population-wide) top income inequality in a region s on the change in top income inequality for a particular occupation o in that region, say, physicians. Let $\alpha_{o,t,s}^{-1}$ be top income inequality for occupation o at time t for geographical area s and $\alpha_{-o,t,s}^{-1}$ be the corresponding value for the general population in s except for o (both computed for individuals in the local top 10%). Let γ_s be a dummy for the geographical area, γ_t a time dummy, and $X_{t,s}$ a vector of controls, including the area's population and average income. The regression for occupation o, at the area-by-year level, is:

$$\alpha_{o,t,s}^{-1} = \gamma_s + \gamma_t + \beta_o \alpha_{-o,t,s}^{-1} + X_{t,s}\delta + \epsilon_{o,t,s}. \tag{14}$$

The coefficient β_o measures the inequality spillover. That is, by how much top income inequality for occupation o increases when top income inequality for the general population increases. We use the Census income data described above to compute $\alpha_{o,t,s}^{-1}$ and $\alpha_{-o,t,s}^{-1}$ for 1980, 1990, 2000 and 2012. We cluster standard errors by LMA and weight LMA-years by the number of persons in occupation o (above x_{min}). With 4 periods and 50 LMAs we have a balanced panel of 200 observations.

Instrument. A natural worry when estimating equation (14) is endogeneity. Even controlling for LMA and year fixed effects, a positive correlation between general income inequality and inequality for a specific occupation might reflect deregulation, tax change, or common local economic trends—rather than a causal effect. Our mechanism itself could generate reverse causality: inequality within the outcome occupation might spill over into other occupations on the right-hand side. And as Section 3.2 explained, unobserved positive correlation between the ability distribution of the occupation of interest and the general population could lead to a downward bias.

To address these concerns, we use a "shift-share" instrument (Bartik, 1991) based on the occupational distribution across geographic areas in 1980. We define:

$$I_{-o,t,s} = \sum_{\kappa \in K_{-o}} \omega_{\kappa,1980,s} \alpha_{\kappa,t,-s}^{-1}, \text{ for } t \in \{1980,\ 1990,\ 2000,\ 2012\}. \tag{15}$$

K_{-o} is the set of most important occupations in the top 10% (excluding o), defined as the union of the 10 most common occupations in the top 10% of each LMA s in 1980. K_{-o} contains around 30 occupations (disclosure rules require rounding to the nearest 10). The weight $\omega_{\kappa,1980,s}$ is the share of occupation κ in LMA s's top 10 in 1980. The shift, $\alpha^{-1}_{\kappa,t,-s}$ is the inverse Pareto coefficient for occupation κ in year t for the entire U.S. excluding LMA s. We estimate equation (14) via 2SLS using $I_{-o,t,s}$ as an instrument for $\alpha^{-1}_{-o,t,s}$. Since K_{-o} does not include all occupations, the weights do not sum to 1. Following Borusyak, Hull, and Jaravel (2022), we control for the sum of the excluded weights interacted with year dummies in all specifications. We explore alternative occupation sets K_{-o} in robustness checks.

The instrument variation is best illustrated by a decomposition of the endogenous variable, income inequality for the general population. Let $O_{-o,s}$ be the set of all occupations excluding the occupation of interest o. The estimator of the inverse Pareto parameter in (13) implies that we can decompose the common estimate of $\alpha^{-1}_{-o,t,s}$ for a set of occupations O_{-o} as $\alpha^{-1}_{-o,t,s} = \sum_{\kappa \in O_{-o}} \omega_{\kappa,t,s} \alpha^{-1}_{\kappa,t,s}$ (where $\alpha_{\kappa,t,s}$ is irrelevant if an occupation does not appear in year t). We exploit this to write our right-hand side measure of inequality as:

$$\alpha^{-1}_{-o,t,s} = \underbrace{\sum_{\kappa \in O_{-o}} \omega_{\kappa,1980,s} \alpha^{-1}_{\kappa,t,-s}}_{\text{National trends}} + \underbrace{\sum_{\kappa \in O_{-o}} \left(\alpha^{-1}_{\kappa,t,s} - \alpha^{-1}_{\kappa,t,-s}\right) \omega_{\kappa,1980,s}}_{\text{Local inequality shocks}} + \underbrace{\sum_{\kappa \in O_{-o}} \alpha^{-1}_{\kappa,t,s} \left(\omega_{\kappa,t,s} - \omega_{\kappa,1980,s}\right)}_{\text{Changes in occupational composition}}, \tag{16}$$

which decomposes the change in local inequality into three terms. The first term captures national trends in occupational income inequality, on which we base our instrument (using the part of the sum associated with occupations in $K_{-o} \subset O_{-o}$).[23] Our IV estimation therefore only exploits the changes in local income inequality arising from the 1980 occupational distribution combined with nationwide trends in occupational inequality. Our instrument relies on national trends that reflect shocks exogenous to a given LMA—such as globalization, technological change or deregulation—but which affect LMAs differently depending on occupational composition, leading to a change in α_x in our model. The other two terms in equation (16) capture local LMA changes: the second term reflects local occupational income inequality relative to U.S. trends, and the third captures changes in local occupational distribution. Neither can plausibly be considered exogenous to the occupation of interest o.

We adopt the Goldsmith-Pinkham, Sorkin and Swift (2020) framework for evalu-

[23]Mazzolari and Ragusa (2013) use a similar approach to instrument for the *level* of income for high-earners in cities based on pre-sample city-specific occupational distribution and national trends in top income growth.

Table 2: Summary Statistics For Regression Variables

	Physicians		Dentists		Real Estate Agents		
	Mean	SD	Mean	SD	Mean	SD	N
α_o^{-1}	0.82	0.12	0.67	0.09	0.50	0.07	200
α_{-o}^{-1}	0.39	0.06	0.40	0.06	0.40	0.06	200
I	0.28	0.02	0.30	0.02	0.30	0.02	200

Notes: Summary statistics for the regression variables. $\alpha_o^{-1}$and α_{-o}^{-1} are the inverse Pareto parameters for the occupation of interest and for the local population excluding the occupation of interest, respectively, in a given LMA×year. Both are based on the top 10% of the wage income distribution in a given LMA×year. I is the instrument: the projected income inequality in a given LMA×year based on the 1980 occupational distribution (see details in text). N is the number of LMA-years.

ating shift-share instruments. They show that a sufficient condition for the validity of a shift-share instrument is that the original weights are conditionally exogenous. Our instrument will be valid if the original occupational composition only affects changes in local top inequality for the occupation of interest through changes in local top income inequality (changes, rather than levels, because we include LMA fixed effects). One concern would be that occupation composition may also affect changes in average incomes, which is why we directly control for this channel. Our identification assumption would be violated if, for instance, in areas which *initially* have more financial managers, physicians are able to get better access to credit *over time*, enabling them to expand their offices, earn higher incomes, and disproportionately so at the top (this corresponds to an increase in ε^S in the model of Section 3.3.1 when $\varepsilon\alpha_x > 1$). We run robustness checks where we exclude financial managers from the instrument.[24] We discuss our shift-share setting further in Section 5.3 and in Appendix D.2.

Summary statistics for the regression variables. Our first regression results focus on three occupations for which we expect to see spillovers: physicians, dentists, and real estate agents. Table 2 presents summary statistics for the main regression variables. The average inverse Pareto coefficients differ from the ones reported in Table 1 particularly for physicians because, as discussed above, we compute them for individuals in the top 10% of the overall local population instead of the top 10% of the occupation.

5 Empirical Spillover Estimates

This section presents our empirical estimates of spillovers across occupations. In Section 5.1, we first focus on a set of high-earning occupations for which we expect to

[24]The alternative view on shift-share instruments articulated by Borusyak, Hull and Jaravel (2022) relies on the exogeneity of the shocks, namely here, trends in occupational inequality. This assumption is likely violated in our case; inequality trends are likely correlated across occupations for other reasons than our spillover through consumption mechanism.

find spillovers: physicians, dentists and real estate agents. Section 5.2 then considers the 30 most common occupations in the top 10% and demonstrates that spillovers are consistent with the predictions of our model. Section 5.3 carries out several robustness checks. Finally, in Section 5.4 we use financial sector deregulation as a specific exogenous shock to income inequality.

5.1 Physicians, dentists and real estate agents

Our central occupations of interest where we expect spillovers are physicians, dentists, and real estate agents. Physicians are a major occupation in the top of the U.S. income distribution (see Appendix Table D.2). They fit our theory well since they provide a service that is heterogeneous in quality and non-divisible. Dentistry is similar but with fewer intermediaries and regulations. Real estate agents are also common in the top of the U.S. income distribution. Real estate services are non-divisible since home sellers/buyers usually only contract with one real estate agent.[25] All three occupations also primarily serve their local market, a necessary condition for our empirical test.

Physicians. Table 3 presents the estimates of equation (14) for physicians. Column (1) shows the OLS regression of physicians' income inequality on general income inequality including year and LMA fixed effects. We find a coefficient of 0.41. This coefficient increases slightly in column (2), where we include controls for LMA population and average wage income among those with positive wage income. Columns (3) and (4) show the first stage regression. The instrument has a reasonable predictive effect on the endogenous variable with an F-statistic around 14. Columns (5) and (6) present IV results: Income inequality from the broader population spills over to physician income inequality with an estimated coefficient of 1.54 in the model without controls and 2.29 in the model with controls. Log population has little conditional relationship with physician inequality, whereas log average income predicts lower inequality. Appendix Figure D.1 visualizes the IV results and demonstrates that they are not driven by outliers.

The magnitude of the spillover relationship is sensible in light of the CES model of Section 3.2. To understand the quantitative implications for the overall increase in physician inequality, we use these estimates and calibrate our model to data from New York State in Section 6. Alternatively, a model-free approach to interpreting

[25] Furthermore, the fee structure in real estate is often proportional to housing prices (Miceli, Pancak and Sirmans, 2007) and the increase in the spread of housing prices is consistent with the increase in income inequality (Määttänen and Terviö, 2014).

Table 3: Spillover estimates for Physicians

	OLS		First Stage		IV	
Dependent Variable:	α_o^{-1}	α_o^{-1}	α_{-o}^{-1}	α_{-o}^{-1}	α_o^{-1}	α_o^{-1}
	(1)	(2)	(3)	(4)	(5)	(6)
α_{-o}^{-1}	0.41	0.58			1.54	2.29
	(0.18)	(0.16)			(0.74)	(0.63)
Ln(Average Income)		-0.33		0.06		-0.47
		(0.10)		(0.04)		(0.13)
Ln(Population)		-0.03		-0.03		0.03
		(0.04)		(0.02)		(0.04)
I			4.93	4.44		
			(1.29)	(1.22)		
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
$(1-\sum_{\kappa\in K_{-o}}\omega_\kappa)\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	200	200	200	200	200	200
F-Statistic					14.6	13.21

Note: This table shows OLS and IV regressions of local top income inequality among physicians on top income inequality in the local population. Top income inequality for physicians is measured by the inverse Pareto parameter α_o^{-1} estimated in each LMA among physicians in the top 10% of the local income distribution. Local income inequality is measured by the the inverse Pareto parameter α_{-o}^{-1} for the local non-physician in the top 10% of the local income distribution. I is our instrument and captures the projected income inequality by interacting local occupational composition with national occupation trends in inverse Pareto parameters (see details in text). Column (1) shows the OLS relationship controlling for LMA and year fixed effects and for the share of the LMA's upper tail in 1980 that is not included in the instrument occupations interacted with year indicators. Column (2) adds controls for average income and population among persons with positive income. Columns (3) and (4) show the first stage regressions. Columns (5) and (6) show the IV regressions. N is the number of observations. Observations are weighted by the number of physicians in the top 10% of the local income distribution.

the magnitude of the spillover estimates is to express them in terms of standard deviations (SD) of each variable (as reported in Table 2). A 1 SD increase in local income inequality increases local physicians' income inequality by 0.8 and 1.1 SDs (from estimates without or with controls, respectively). So the spillover mechanism explains a sizable share of variation in regional physician inequality.

Dentists and real estate agents. We show the corresponding results for dentists and real estate agents in Table 4. In both cases, we estimate spillover coefficients greater than 1 in line with the CES model. The IV coefficients with controls are 2.27 for dentists, which is very similar to the corresponding coefficient for physicians, and 1.71 for real estate agents, which is bit smaller. In both cases, a 1 standard deviation increase in local income inequality leads to a 1.5 standard deviation increase in occupational inequality—so the spillover mechanism can again account for a sizable share of the variation in local occupational inequality.

Relationship between IV and OLS. For physicians and dentists, the IV coefficients are much higher than the OLS correlations. What may explain this? First, the CES model of Section 3.2 predicts that an increase in doctors' ability at the top, α_z^{-1}, which could be interpreted as skill-augmenting technical change for doctors, actually

Table 4: Spillover estimates for Dentists and Real Estate Agents

	Dentists				Real Estate Agents			
	OLS		IV		OLS		IV	
α_{-o}^{-1}	1.08	1.09	2.29	2.27	1.43	1.43	1.69	1.71
	(0.33)	(0.34)	(0.71)	(0.93)	(0.21)	(0.21)	(0.62)	(0.65)
Ln(Avg. Income)		0.08		0.00		0.01		-0.01
		(0.12)		(0.14)		(0.09)		(0.09)
Ln(Pop.)		0.04		0.08		0.02		0.02
		(0.05)		(0.08)		(0.03)		(0.03)
N	200	200	200	200	200	200	200	200
F-Statistic			23.63	13.36			12.72	7.124

Notes: This table mimics Table 3, except rather than physicians as the outcome occupation, the outcome occupations are dentists and real estate agents, respectively.

reduces top income inequality if the elasticity of substitution $\varepsilon < 1$. That is, technical change that may have increased income inequality for most occupations may actually decrease income inequality for physicians and dentists, countering the spillover effect. This, in turn, implies that a positive unobserved correlation between inequality in local doctors' ability and local consumers' ability would bias the OLS coefficient downward.[26] We discuss this further using our calibrated model in Section 6. Second, there are numerous potential omitted variables. For example, more unequal places might have higher taxes and spend more public money on health care. This would support the incomes of worse-off physicians and bias the OLS coefficient downward. Third, due to sampling noise, we estimate inequality in the general population with some error which we expect to bias downward the OLS estimate. For real estate agents, the gap between OLS and IV is much smaller, which in the framework of the CES model, is consistent with a higher elasticity of substitution between the non-divisible service and other goods, ε, and with a smaller spillover coefficient.[27]

5.2 Other occupations

We next analyze spillovers for a broader set of occupations. Since Figure 1a shows that rising within-occupation inequality primarily affects the top of the income distribution, we focus on high-earning occupations. First, as a placebo test, we estimate spillover coefficients for three occupations that do not fit our model: financial man-

[26] For instance, using (10), the ratio Cov $\left(\alpha_z^{-1}, \alpha_x^{-1}\right)/\text{Var}\left(\alpha_x^{-1}\right)$ would have to be 1.3 to account for the entire gap between OLS and IV coefficients in columns (2) and (6) in Table 3.

[27] Proposition 3 also showed that the slope of the Engel curve is less steep when ε is lower. In line with a higher ε for housing than medical services, Figure C.1 implies an elasticity of the Engel curve of around 0.29 for medical expenditures, while for real estate, Zabel (2004) finds Engel curve elasticities of 0.64–0.70 for high-income families.

agers, managers excluding real estate, and engineers. We then examine the 30 most common occupations in the top 10% of the income distribution. We find that (a) spillovers concentrate in occupations fitting our model and (b) spillover coefficients correlate with occupational traits indicating local service provision. Finally, we also find spillovers for a broader class of medical and related occupations that are less common at the top of the income distribution.

Table 5: Spillover estimates for Financial managers, Managers and Engineers

	Financial Managers				Managers Excl. Real Estate				Engineers			
	OLS		IV		OLS		IV		OLS		IV	
α_{-o}^{-1}	1.63	1.31	1.12	0.77	0.22	0.28	0.05	-0.02	0.32	0.45	-0.13	-0.09
	(0.47)	(0.35)	(0.97)	(0.92)	(0.09)	(0.09)	(0.12)	(0.15)	(0.11)	(0.13)	(0.26)	(0.30)
Ln(Avg. Inc.)		0.25		0.29		0.04		0.06		-0.05		-0.02
		(0.14)		(0.16)		(0.03)		(0.03)		(0.04)		(0.05)
Ln(Pop.)		-0.14		-0.16		0.05		0.03		0.07		0.05
		(0.05)		(0.07)		(0.01)		(0.01)		(0.02)		(0.02)
N	200	200	200	200	200	200	200	200	200	200	200	200
F-Statistic			19.92	11.22			22.15	16.49			23.97	27.91

Notes: This table shows OLS and IV regressions analogously to Table 3 but for occupations for which we do no predict spillovers. Columns (1)-(4) look at financial managers, Columns (5)-(8) at managers excluding real estate, and Columns (9)-(12) at engineers. N is the number of observations.

Placebo occupations. Financial managers, managers (excluding those in real estate), and engineers serve as useful placebos. Workers in these occupations generally do not produce a non-divisible good or service for the local population. Managers and engineers work for firms that produce a variety of goods and services for both the local and national markets.[28] Likewise, financial managers also work for firms: they "plan, direct, or coordinate accounting, investing, banking, insurance, securities, and other financial activities of a branch, office, or department of an establishment" according to the Standard Occupational Classification scheme.

Table 5 reports the OLS and IV results for these three occupations. The OLS estimates are significant throughout but the IV coefficients are statistically indistinguishable from zero and smaller than those for physicians, dentists and real estate agents. The positive OLS estimates and non-significant IV estimates demonstrate that spurious correlation between general inequality and occupational inequality at the local level is likely but that our instrument addresses this concern.

Largest occupations in the top 10% of the income distribution. Table 6 reports the results for the most common 30 occupations in the top 10% of the income distri-

[28] An assignment mechanism may exist for managers but then managers' income inequality would reflect firm size inequality (as in Gabaix and Landier, 2008) rather than *local* income inequality.

Table 6: Spillover estimates for a broad set of occupations

	OLS	OLS SE	IV	IV SE	F stat	t stat	N Small
Accountants and auditors	0.90***	0.12	0.25	0.39	13.26	0.63	0
Airplane pilots and navigators	0.66***	0.25	3.54**	1.73	6.88	2.04	40
Chief executives, public admin., and legislators	0.77***	0.20	0.30	0.60	12.21	0.50	86
Computer programmers	0.43*	0.23	0.24	0.43	16.87	0.55	16
Computer systems analysts and computer scientists	0.77***	0.23	0.70**	0.35	21.87	1.97	7
Dentists	1.09***	0.34	2.27**	0.93	13.36	2.45	13
Driver/sales workers and truck Drivers	0.84***	0.25	1.62	1.12	4.17	1.45	17
Electricians	0.28	0.25	-0.17	1.06	4.44	-0.16	56
Engineers	0.45***	0.13	-0.09	0.30	27.91	-0.30	0
Financial managers	1.31***	0.35	0.77	0.92	11.22	0.84	0
Financial service sales occupations	1.86***	0.30	1.11	1.14	6.07	0.97	21
Insurance sales occupations	1.25***	0.17	1.08*	0.62	8.07	1.73	2
Lawyers and judges	0.56**	0.27	-0.33	0.64	14.21	-0.52	0
Managers of properties and real estate	0.70**	0.31	1.32*	0.76	15.80	1.75	36
Managers, Excl. Real Estate	0.28***	0.09	-0.02	0.15	16.49	-0.10	0
Office supervisors	1.00***	0.29	1.41***	0.30	15.91	4.70	0
Other financial specialists	0.82***	0.28	0.32	0.81	15.49	0.40	0
Personnel, HR, training, and labor rel. specialists	0.67***	0.23	0.82	0.60	9.58	1.36	2
Pharmacists	-0.16	0.26	-0.49	0.83	25.01	-0.60	20
Physicians	0.58***	0.16	2.29***	0.63	13.21	3.63	0
Police and detectives, public service	0.47***	0.10	0.30	0.28	27.29	1.08	47
Primary/Secondary School Teachers	0.13	0.28	-0.44	0.64	7.98	-0.68	5
Production supervisors or foremen	0.63***	0.18	-0.09	0.75	9.93	-0.12	2
Real estate sales occupations	1.43***	0.21	1.71***	0.65	7.12	2.64	6
Registered nurses	0.16	0.21	0.71	0.53	27.07	1.36	28
Sales occupations and sales representatives	0.51***	0.11	0.28	0.23	8.03	1.20	0
Sales supervisors and proprietors	0.60***	0.12	0.53	0.43	20.95	1.23	0
Sales workers	0.69***	0.17	0.12	0.43	10.03	0.28	0
Subject instructors, college	0.57***	0.11	-0.04	0.28	19.98	-0.15	0
Supervisors of construction work	0.73***	0.27	1.27	1.05	6.41	1.21	7

Notes: This table shows the OLS and IV coefficients for regressions of local top income inequality for some occupations on top income inequality in the local population excluding that occupation. The occupations shown are the 30 most common occupations in the top 10% of the income distribution. Each row corresponds to the regressions for a given occupation. The variables are defined analogously to Table 3 and estimates are from the specification with population and average income controls. Columns (1) and (2) show the OLS coefficient and standard error. Columns (3) and (4) shows the IV coefficient and standard error. Column (5) is the F statistic for the instrument, column (6) the IV coefficient t statistic. The final column shows the number of LMA-years in which there were fewer than 20 persons used to construct the outcome variable. The total number of LMA-years is 200. *: p<0.1, **: p<0.05, ***: p<0.01.

bution, including the 6 already considered. Whereas most of the occupations have a positive OLS coefficient, the IV is positive only for a small subset. Besides physicians, dentists and real estate agents, we see significant (with $p < 0.1$) positive coefficients for managers of properties and real estate, likely reflecting the same mechanism as for real estate agents, and for insurance sales specialists. The latter often sell their products directly to individuals and, in the occupation characteristics used below, they rank high for customer service, working with the public, and non-tradeability. Therefore, these two occupations plausibly fit our spillover mechanism. We also find significant relationships for pilots, computer system analysts, and office supervisors, which we view as false positives; with 25 placebo occupations, we would expect 2.5 false positives at $p < 0.1$. Note that for pilots, we have fewer than 20 observations to compute income inequality for 20% of the LMA-years.

The remaining occupations can be classified in three groups. First, those occupations for which our theory does not apply—*e.g.* the placebo occupations already mentioned, computer programmers, and production supervisors. Second, occupations for which our theory may apply but at the national level (such as college professors). Third, occupations (for instance "lawyers and judges") that include both subcategories where our theory should apply (personal attorneys) and where it should not (corporate lawyers). Similarly, financial sales occupations cover both individuals working for firms, for whom our theory may not apply, and personal finance managers for which our theory is more likely to apply, albeit perhaps at the national level.

Relationship with occupation characteristics. Our model predicts a higher local spillover coefficient for occupations where production has to be local, and for those that directly serve the public. To test these predictions, we correlate the IV spillover coefficients with these occupational traits. To quantify local production, we treat Blinder's (2009) measure of offshorability as an (inverse) measure of the extent to which an occupation serves the local market. It codes 18 out of the the 30 occupations as completely not offshoreable. To quantify direct public interaction, we rely on measures from the Occupational Information Network (O*NET) of the level of *customer service* and *working with the public*. Appendix B.3 gives further details.

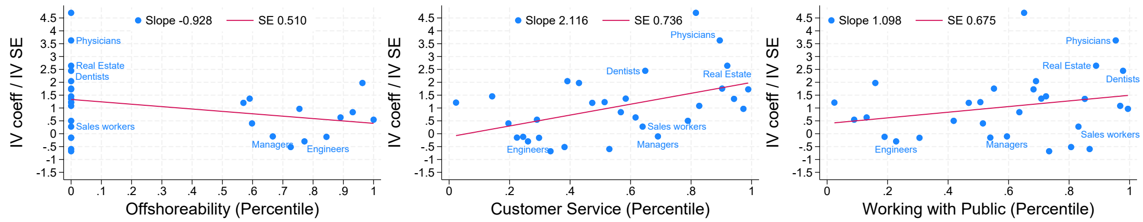
Since spillover estimates have varying precision, we divide them by their standard errors; that is, we correlate the t-statistics with occupational traits. This is equivalent to a regression of the spillover coefficient on the occupational trait weighted by the inverse standard error of the spillover coefficient. Figure 3 shows the results.

Consistent with our model, Panel 3a shows a negative relationship between the spillover and offshorability ($p = 0.07$). Our three focal occupations are non-offshoreable and the non-offshoreable group's average spillover t-statistic is 1.36 compared to only 0.57 in the (at least partly) offshoreable occupations — and this difference is significant at $p < 0.1$. Panels 3b and 3c show positive relationships between the spillover and measures of customer service and working with the public ($p < 0.01$ and $p = 0.11$). Appendix Table D.3 presents the corresponding regressions, along with the "importance" (rather than "level") of the customer service and working with public variables, both of which show statistically significant positive relationships.

These patterns support our model: With a few exceptions, we only observe local spillovers for occupations that fit the model's predictions of delivering non-divisible goods or services of heterogeneous quality in the local market. These relationships also buttress our identification strategy; most potential omitted variables would not drive spillover effects for specifically this type of occupations. In particular, neither "keeping up with the Joneses" (Bertrand and Morse, 2016) nor rich individuals sorting into high-amenity locations would imply this spillover coefficient pattern.

Figure 3: IV estimates and occupational characteristics.

(a) Offshorability **(b)** Customer service **(c)** Work w. Public



Notes: This figure shows the relationship between the t-stat of the spillover coefficients shown in Table 6 and three occupation characteristics: a measure of offshorability from Blinder (2009) and two measures from O*NET: Level of "Customer service and personal service" from Knowledge Requirements and level of "Performing for or working directly with the public" from Work Activities. The measures are rescaled as percentiles.

Other medical occupations less common in the top. While within-occupation inequality primarily affects top earners, our model does not require occupations receiving spillovers to be common in the top. We test this using four health-related occupation groups less common in the top than physicians or dentists: "veterinarians", "physical and occupational therapists", "psychologists", and "optometrists, podiatrists, and other health-diagnosing occupations". A stacked regression across these groups (allowing for group-specific LMA and year fixed effects) reveals positive spillover effects approximately half the magnitude of our focal occupations (Table 7). Due to sample size limitations, we also present results restricted to the 30 most populous LMAs.[29]

Taking stock. To sum up, we find evidence of spillovers for physicians, dentists, and real estate agents, which together account for a significant share of individuals in the top of the income distribution: 15.8% of the top 1% and 5% of the top 10%. We view this as a lower bound on the applicability of our theory. First, spillovers for occupations with tradable output occur nationally, rather than locally, and therefore would not be captured by our empirical strategy. Second, some occupations common in the top 10%, such as personal lawyers or personal wealth managers, fit the requirements of our model (heterogeneous and non-divisible output) but are not easily identified in Census occupation coding. Moreover, spillovers in theory need not be restricted to occupations in the upper tail, such as we show with other medical occupations. With

[29] When looking at all 50 MSAs, around 60% of observations have fewer than 20 observations included in our calculation of top income inequality. In the second panel, this falls to 44%.

that said, the rise in within-occupation inequality is concentrated at the top of the income distribution, suggesting that spillovers may be particularly prominent (and empirically identifiable) among high earning occupations.

Table 7: Spillover estimates for broader medical occupations

	50 Most Populous LMAs						30 Most Populous LMAs					
	OLS		First Stage		IV		OLS		First Stage		IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
α_{-o}^{-1}	0.58	0.54			1.02	1.16	0.05	-0.02			1.18	1.26
	(0.17)	(0.17)			(0.57)	(0.56)	(0.19)	(0.21)			(0.51)	(0.55)
Ln(Avg. Inc)		0.00		0.05		-0.05		0.02		0.04		-0.06
		(0.07)		(0.04)		(0.08)		(0.06)		(0.05)		(0.08)
Ln(Pop.)		-0.05		-0.03		-0.03		-0.05		-0.03		0.00
		(0.03)		(0.02)		(0.03)		(0.04)		(0.02)		(0.04)
I			4.09	3.97					5.17	4.80		
			(0.80)	(0.99)					(1.17)	(1.32)		
N	800	800	800	800	800	800	500	500	500	500	500	500
F Statistic					26.15	15.99					19.35	13.03

Notes: This table shows regression estimates that "stack" four health-related occupations: Veterinarians, Physical and Occupational Therapists, Psychologists, and the combination of Optometrists, Podiatrists and "Other Health Diagnosing Occupations". Stacking means that for each occupations, we calculate outcome inequality, explanatory inequality, and the instrument. We then stack the data sets together, and run our baseline regression with LMA and Year FEs interacted with indicators for the outcome occupation. N is the number of observations rounded to the nearest integer divisible by 50 as required by disclosure rules.

5.3 Robustness checks

We now discuss robustness checks to alternative hypotheses, specification choices, and inference methods, with corresponding tables in Appendix D.2.

Earned income. Gottlieb et al. (2025) show that business income is sizable among top physicians. The Census data report wage income, business income, and capital income; we define "earned income" as the sum of the first two. We calculate top income inequality using earned income for physicians, dentists and real estate agents in the same manner as for wage income and replace the dependent variable. We leave the instrument and the independent variables unchanged. The IV coefficients, reported in Appendix Table D.4, remain similar to that of Table 3, though the coefficient for real estate agents declines by around 40%.

Occupational and geographical mobility. Section 3.3 showed theoretically that our spillover mechanism is robust to allowing for either occupational or geographical mobility. We now investigate these issues empirically. First, we restrict attention to physicians that are less occupationally or geographically mobile. Given the cost and time required to enter the medical profession, those older than 35 will have decided to become a medical doctor more than 10 years ago, and cannot easily respond to economic trends over the preceding 10 years. We therefore estimate a regression that restricts attention to doctors older than 35. Similarly for geographical mobility, we consider doctors who have not moved within the past 5 years. Appendix Table D.5 columns (1) and (2) report the results: in both cases, the IV coefficients are a bit higher at 3.1 and 3.3, respectively.

Second, we directly study entry by computing the employment share of the occupation of interest. This measure can vary because of geographical or occupational mobility. In Appendix Table D.6, we run two sets of regressions, both using our standard IV specification: First, we use employment share as the dependent variable and show that rising inequality does to some extent encourage the entry of new physicians. Nevertheless, when we include it as an additional control in our standard IV regressions with inequality as the outcome we find that our spillover coefficient is virtually unchanged. This is not surprising as entry along the entire ability distribution of physicians would leave physician inequality unchanged while decreasing their mean income. We perform the same analysis with dentists and real estate agents and find that the inclusion of the employment share leaves the spillover coefficient basically unchanged but marginally reduces precision for real estate agents ($p = 0.11$).

Scale. Our theory emphasizes the importance of price inequality in determining doctors' income inequality. Section 3.3.1 demonstrates that our spillover mechanism remains robust even when production scalability is introduced, allowing consumer inequality to raise top performers' relative scale. Yet, technological change that affects the (intensive margin) supply elasticity, ε^S, could affect doctors' income inequality, potentially biasing our IV estimates if correlated with the instrument. Census data lacks direct measures of physician scale. In Appendix C.3, we use data from the National Ambulatory Medical Care Survey to build a proxy for physicians' scale and find that there is no trend in the standard deviation of that measure between 1992 and 2011. This suggests that, at the aggregate level, there has not been a large shock to ε^S.[30] Similarly, for real estate agents, Gilbukh and Goldsmith-Pinkham (2024)

[30]Further, recall that in our model, an increase in ε^S increases doctors' top income inequality if and only if $\varepsilon\alpha_x > 1$. We can perform a back-of-the-envelope computation to check whether this condition holds. Using (12), we get

$$\beta^{IV} = \frac{d\alpha_w^{-1}}{d\alpha_x^{-1}} = \frac{\varepsilon^S\left(1-\alpha_w^{-1}\right)+1}{\varepsilon^S\alpha_x^{-1}+\varepsilon}.$$

Using the average values in Table 1 for α_w^{-1} and α_x^{-1}, column (6) of Table 3 for β^{IV}, and $\varepsilon^s = 0.4$ from Gottlieb et al. (2025), we get $\varepsilon = 0.398$, so $\varepsilon\alpha_x = 1.02$. Taking our estimates and model seriously, parameters are very near the values where top income inequality is independent of ε^S .

find that the share of transactions executed by high commission rate agents has been relatively constant since 2000 (their Appendix Figures F3 and F4).

An increase in dentists' scale is often associated with increased specialization. We proxy for this phenomenon by computing the ratio of dental hygienists to dentists in an LMA. We add this control and our main coefficient is nearly unchanged (Appendix Table D.5, column (4))—suggesting again that a shock to ε^S is not driving our results.

Medical specialties. Since pay varies across medical specialties and our physician category combines all specialties, compositional effects could influence results. We address this by controlling for specialty shares (detailed in Appendix D.2). As Appendix Table D.5 column (3) shows, the IV coefficient remains virtually unchanged.

Deviations from Pareto distribution. To ensure enough observations and a consistent population slice across occupations, we calculate α_o^{-1} using individuals from the occupation within the top 10% of the general local population. LMA $\times$ year $\times$ occupation samples are not necessarily Pareto, but this does not undermine identification. First, α_o^{-1} is an inequality measure regardless of distributional form. Our regression coefficient identifies spillovers for that inequality measure provided that our instrument is exogenous—though the IV coefficient may not exactly identify $1/\varepsilon$. Section 6 shows how to calibrate $1/\varepsilon$ using the spillover estimate. Second, as we discussed previously, although α_o^{-1} depends somewhat on the cutoff x_{min}, changes in α_o^{-1}, which we use for identification, are much less sensitive. Third, in Appendix Table D.7, for our three focal occupations and the three main "placebo" occupations, we control for the share of the occupation in the top 10% of the general population, with little effect on results. Finally, Appendix Table D.8 shows that calculating α_o^{-1} using only the top 10% of physicians yields a spillover coefficient of 1 ($p = 0.12$), or 1.23 ($p < 0.05$) when restricted to the 30 most populous LMAs to reduce measurement noise.[31]

Other sample restrictions. Appendix Table D.9 reports robustness to other sample choices. First, we use the top 5% of the general population instead of the top 10%. Second, we use 30 or 70 LMAs rather than 50. Third, we construct the instrument using the union of the 13 or 7 (rather than 10) biggest occupations in each LMA. Results are robust to these changes, with one exception: using the 5% cutoff for real agents and dentists yields qualitatively similar but less precise coefficients ($p < 0.15$),

[31] In unreported regressions, we ran the same exercise for real estate agents and dentists. We did not find significant spillovers, but note that in both cases we have fewer observations (and much fewer for dentists) than for doctors or for the baseline specification.

consistent with having fewer individuals to calculate inequality.

Shift-share robustness checks. Since our identification relies on the exogeneity of the occupational shares, we follow Goldsmith-Pinkham et al. (2020) and show the 16 largest (in absolute terms) Rotemberg weights in Appendix Table D.10. The three occupations with the largest weights are Financial Service Sales (0.35), Financial Managers (0.22), and Airplane Pilots and Navigators (0.21). We exclude, in turn, the five occupations with the highest Rotemberg weight from the IV in our baseline regression for physicians. Appendix Table D.11 reports the results and shows that the spillover coefficient is very stable.

Alternative inference. Finally, we report alternative inference methods in Appendix Table D.12. First, we compute Adão, Kolesár and Morales (2019) standard errors that account for potential correlation of residual errors across LMAs with similar occupational composition. Second, we calculate confidence intervals following Lee et al. (2023) that are valid when the instrument is weak. Neither approach changes the conclusions. In fact, the latter confidence intervals are sometimes shorter than under usual inference, which Lee et al. (2023) find is common in many applications.

5.4 Using financial deregulation as an instrument

This section investigates spillovers from one specific shock to income inequality: the substantial financial deregulation from the 1970s–1990s.[32] Systematically identifying the source of the original shocks is beyond our scope, but we consider this one in particular since finance occupations are the most influential component of the shift-share instrument according to Rotemberg weights. Financial deregulation was followed by substantial increases in average income for financial occupations (Philippon and Resheff, 2012). We show that it was also associated with rising income inequality. We use this policy variation, combined with varying concentrations of finance occupations across LMAs, as an exogenous source of variation in local inequality. This exercise requires extending our Census data back to 1960 so we have a pre-treatment period.[33]

[32]Financial deregulation broadly had three components, roughly in this chronological order: removal of restrictions on intra-state branching; removal of restrictions on inter-state banking; erosion of regulatory barriers that separated insurance companies, commercial banking, and investment banks. Many states started intra-state deregulation in the 1970s, most states then experienced inter-state deregulation during the 1980s and early 1990s (Hoffmann and Stewen, 2020). Removal of barriers between commercial banking, investment banking, and insurance occurred throughout the 1990s, culminating in 1999 with the repeal of the Glass-Steagal Act.

[33]Census years 1960 and 1970 contain data quality inconsistencies. For instance, in 1960, some data was lost and later restored, and different county ID schemes were used. In 1970, mail-in forms

Let $\omega_{f,1970,s}$ denote the share of an LMA's employed population that works in a finance occupation in 1970, defined as: (i) Financial Managers, (ii) Financial Service and Sales Occupations, and (iii) Other Financial Specialists. The share serves as a measure of exposure to deregulation. We first estimate a continuous difference-in-difference to measure the differential changes in local income inequality (excluding doctors) between areas with high vs. low finance occupation shares relative to 1970:

$$\alpha^{-1}_{-o,t,s} = \theta_s + \theta_t + \sum_{h \neq 1970} \theta_{o,h} \mathbf{1}_{h=t} \omega_{f,1970} + v_{o,t,s}, \tag{17}$$

where $\theta_{o,h}$ capture the effect of interest, and θ_s and θ_t are LMA and year fixed effects.

Two identification assumptions are required. First, absent deregulation, areas with more finance would have experienced similar trends in income inequality as areas with less finance. Second, we require that any *direct* effect of finance deregulation on physician inequality is uncorrelated with $\omega_{f,1970,s}$. To illustrate, consider a scenario where deregulation improved physicians' access to credit, which in turn increased physician income inequality. Our identification requires this effect to be uniform across LMAs. If inter-state banking deregulation expanded physicians' credit access more in Columbus than in New York, our spillover estimates would be biased.[34]

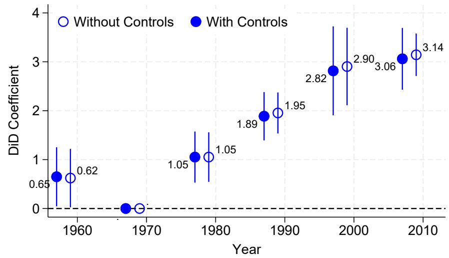
Figure 4a visualizes the first stage by plotting estimates of $\theta_{o,h}$. These indicate that consumers' inequality rose more from 1970 to 2000 in areas that had higher finance shares in 1970. A 10 percentage point higher share of people working in finance in 1970 predicts a 0.105 increase in α^{-1} for the general population by 1980 and a 0.29 increase by 2000. Panels (b)-(d) plot the reduced form regressions where the outcome in (17) is replaced with income inequality among physicians, dentists, or real estate agents. The finance share predicts changes for physicians and dentists and the reduced form coefficients' pattern approximately tracks the first stage coefficients in Panel (a). One can back-out an implicit spillover by dividing the reduced form by the first stage. For doctors in 2000, for example, the implied spillover coefficient is $5.26/2.82 = 1.87$, similar to the main spillover estimates in the shift-share approach. The same exercise for dentists gives a spillover coefficient of 1.39—a bit lower than our main estimate.

were introduced for urban areas, and by 1980, nearly all enumeration took the modern form of mailed questionnaires. This motivates our focus on 1980 and later in the main analysis.

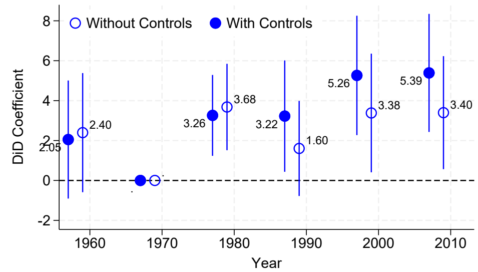
[34]Recall that our spillover estimates are largely unchanged when we remove each of the financial occupations (Appendix Table D.11), so our main results are unlikely to suffer from such a bias. Moreover, such a bias could not give rise to the systematic relationship between spillover coefficients and occupation characteristics shown in Figure 3.

Figure 4: Finance Deregulation Reduced Form

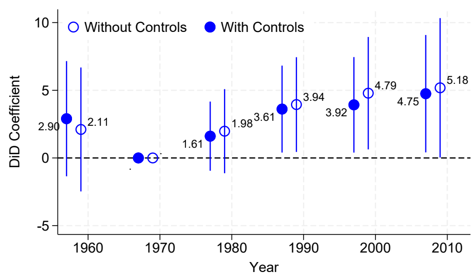
(a) First Stage



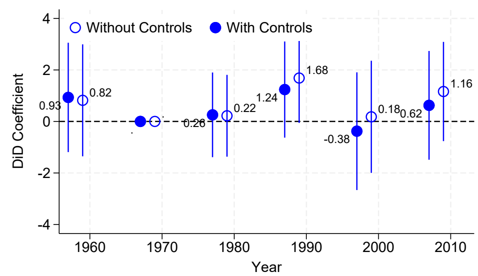
(b) Physicians



(c) Dentists



(d) Real Estate Agents



Notes: For our baseline set of LMAs, we calculate the share of their employed population that works in a finance occupation as of 1970: (i) Financial Managers, (ii) Financial Service and Sales Occupations, and (iii) Other Financial Specialists. Denote that share $\omega_{f,s,t}$. Panel (a) shows estimates of $\theta_{o,h}$ from $\alpha^{-1}_{-o,t,s} = \theta_s + \theta_t + \sum_{h\neq 1970} \theta_{o,h}\omega_{f,1970,s}\mathbf{1}\{h = t\} + v_{o,t,s}$ where o is physicians. The "with controls" version adds ln(population) and ln(average income) controls. Panels (b)-(d) show reduced form estimates from $\alpha^{-1}_{o,t,s} = \pi_s + \pi_t + \sum_{h\neq 1970} \pi_{o,h}\omega_{f,1970,s}\mathbf{1}_{h=t} + e_{o,t,s}$, where the outcome is income inequality among physicians, dentists, real estate agents, respectively,, π_t and π_s are year and LMA fixed effects, and $\pi_{o,h}$ are effects of financial deregulation based on pre-existing finance concentrations ($\omega_{f,1970,s}$). 90% confidence intervals are shown.

These results offer supportive evidence for our main shift-share strategy for doctors and dentists, though the estimates for real estate agents are insignificant.

6 Quantitative Investigation of the Spillover Mechanism

Our theory establishes that changes in consumers' income inequality translate into changes in doctors' income inequality, particularly in the tail of the distribution. In this section, we solve our model numerically to analyze how these spillovers affect the entire distribution of doctors' incomes and quantify the welfare implications. We use income distributions for both doctors and non-doctors (whom we call "consumers") in New York State for 1980 and 2012.[35] This calibration serves three purposes. First, we solve for the endogenous response of doctors' income distribution to the observed change in consumers' income between 1980 and 2012, revealing that our model largely

[35]This exercise is most sensible within one geographical market. Disclosure requirements do not permit us to extract data for a specific LMA, so we use New York State instead of New York City.

reproduces the observed change in doctors' distribution over this period. Second, we quantify the importance of the spillover mechanism in driving income inequality at the top of the distribution. Third, we quantify how much spillovers dampen the increase in welfare inequality relative to income inequality for the consumers. Appendix E provides details on the data preparation and calibration procedure.

Data and calibration approach. We observe four empirical income distributions for New York State: doctors and consumers for both 1980 and 2012 (inflation-adjusted to 2000 dollars). We drop doctors with medical resident-level earnings and the lowest-earning 10% of consumers. We rescale income values by the lowest consumer income in 1980. For each distribution, we fit a flexible kernel to the bottom 90% and a Pareto tail on the top 10%, with the parameter α^{-1} estimated from the same top 10%. Table 8 presents the α^{-1}estimates for all four distributions.

Our calibration procedure focuses on matching the 2012 distributions. We then replace the 2012 consumer income distribution with the 1980 distribution and compute the resulting physician income distribution implied by the model. This allows us to assess how physician income inequality changed between 1980 and 2012 due to spillovers, taking as given changes in doctors' ability distribution. We use the version of our model from Section 3.2 with CES preferences and where both consumer income and doctor ability distributions are only restricted to be Pareto in the tail. This more flexible specification allows us to match the full income distribution rather than just its tail behavior. We allow the minimum income of doctors x_{min}^{dr} to differ from that of consumers to match the significantly higher income of doctors.

While we observe the consumer income distribution, doctors' ability distribution is unobserved and must be calibrated. Our model has four key parameters beyond this ability distribution: (1) the number of patients a doctor can treat, λ; (2) the minimum income of doctors, x_{min}^{dr}; (3) the preference weight on medical services, β; and (4) the elasticity of substitution between medical services and other goods, ε.[36] We set $\lambda = 175$ to match the observed fraction of doctors in New York State in 2012 (excluding residents), and x_{min}^{dr} to match doctors' minimum income. The elasticity parameter ε is particularly important, since lower values strengthen spillovers. We set $\varepsilon = 0.35$ in the calibration which, we demonstrate below, replicates a spillover

[36] μ measures the mass of potential doctors and cannot be identified separately from $F_z(z)$ except to ensure that there is a sufficient number of doctors to treat all $\mu\lambda > 1$.

Table 8: Empirical, fixed and calibrated parameters for 2012

Year	α^{-1} (drs)	α^{-1} (cons)	ε	α_z^{-1}	λ	β	$\log(x^{dr}_{min})$
1980	0.32	0.34	0.35	0.51	175	0.01	1.92
2012	0.42	0.48	0.35	0.51	175	0.01	2.01

Note: α^{-1} for doctors and consumers is estimated on the top 10% of the corresponding empirical distributions. ε is set exogenously (see text). λ is chosen to fit the number of doctors in New York State in 2012. $\log(x^{dr}_{min})$ is set to the minimum income of doctors and is scaled by lowest income of consumers in 1980. β is calibrated along with the ability distribution $F_z(z)$ (including tail α_z^{-1}) to fit the empirical distribution of doctors in 2012. The calibrated values of ε, α_z^{-1}, λ and β are kept fixed at 2012 values for 1980.

of around 1.5 as in the regression without controls.[37] This leaves us $F_z(z)$ and β to calibrate to match doctors' observed 2012 income distribution. To compute doctors' predicted 1980 income distribution, we keep the calibrated parameters fixed while substituting the 1980 consumer income distribution and x^{dr}_{min} value.

Results. Figure 5 shows the main calibration results. Panel (a) presents smoothed consumer income distributions in 1980 and 2012. New York State follows the national pattern in Figure 1; income at the 90th percentile rose by 0.40 log points, with smaller increases at lower percentiles. Panel (b) shows the corresponding distributions for doctors; income growth was similarly higher at upper percentiles of the doctor distribution, in addition around 80 percent of doctors (post medical residency) fall in the top 10 percent of the general income distribution. The 2012 predicted distribution is visually indistinguishable from the empirical one.

The solid black line in Panel (b) presents our model's prediction for the 1980 doctor income distribution, using parameters calibrated to 2012 data while changing only the consumer income distribution and x^{dr}_{min}.[38] Using this model, we quantify the spillover coefficient as the ratio of changes in α^{-1} between consumers and doctors, akin to our regressions. Table 8 shows a change in α^{-1} of 0.14 for consumers and Appendix Figure E.3 shows a change of 0.19 for the doctors using the set of doctors in the top 10% of the overall distribution, implying a spillover coefficient of 1.38. Using only the top 10% of doctors we get a similar coefficient of 1.56, both in line with the spillover coefficient without controls in Table 3. The theory replicates the empirical finding of Appendix Figure B.1 that, although the estimate of α^{-1} can be sensitive to the choice of sample, the difference between periods is much less so.

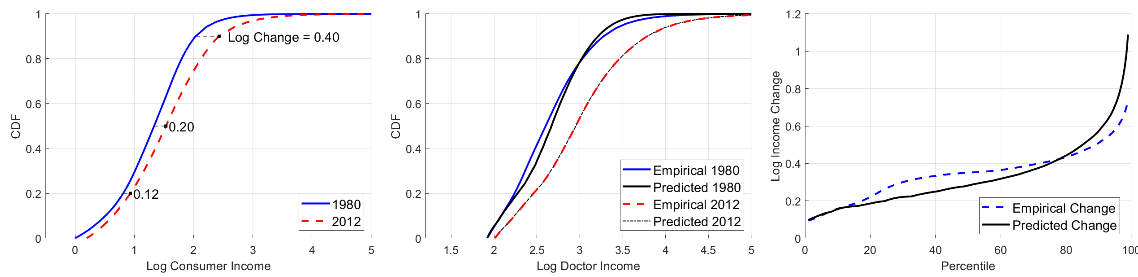
Our model is calibrated to the IV spillover coefficient from our panel regressions,

[37] At first glance, the empirical results in Section 5 along with Proposition 3 would suggest values between $\varepsilon = 1/2.29 = 0.44$ (with controls) and $\varepsilon = 1/1.54 = 0.65$ (without controls). However, since we do not estimate α^{-1} on the extreme tail, the regression-based spillovers somewhat overstate ε.

[38] Recall that x^{dr}_{min} is pinned down by the lowest income of doctors. Whether we fix x^{dr}_{min} or change it in 1980 only impacts the very bottom of the predicted distribution

Figure 5: CDFs of income for consumers and physicians in New York State

(a) Consumer income distribution | **(b)** Doctor income distribution | **(c)** Change in Doctors income



Notes: Panel (a) is the consumers' income distribution. Panel (b) shows the *empirical* distribution of doctors in 1980 and 2012. Parameters are calibrated to match 2012 exactly. "Predicted 1980" uses these parameters to predict the distribution in 1980. Panel (c): Empirical and predicted changes in log income for doctors along the percentile of doctors' income distribution. All values are normalized by the lowest consumer income in 1980.

and does not target actual changes in physicians' income distribution in New York. Nevertheless, Figure 1b shows that the shift in consumers' income alone in our model replicates well the overall empirical shift in doctors' income distribution, though it somewhat over-predicts changes at the very top and under-predicts changes in the middle. Panel (c) makes this comparison explicit by plotting the log income difference between 2012 and 1980 at each percentile of the doctor distribution. At the median, our model predicts an increase of 0.28 log points compared to the empirical increase of 0.35, while for the 95th percentile, it predicts 0.57 versus the observed 0.51.

This pattern—where our model over-predicts increases at the top—is consistent with doctors' ability becoming more unequal between 1980 and 2012, as discussed in Section 3.2 after Proposition 3. With complementarity between medical services and other goods ($\varepsilon < 1$), a rise in doctors' ability inequality (higher α_z^{-1}) dampens their income inequality. Appendix Figure E.2 shows the 1980 and 2012 doctors' ability distributions necessary to perfectly match both years' doctor income distributions. The 2012 distribution has a fatter tail: α_z^{-1} needs to rise from 0.35 to 0.51—a change comparable to the 0.14 increase in consumers' α^{-1}.

Spillover effects and within-top income inequality. We next analyze how spillovers increase top income inequality for the combined income distribution (doctors plus consumers). We calculate the empirical share of income within the top 10% going to various subgroups, as shown in Table 9. In 2012, the top 10% earned 34.4% of total income, with the top 1% capturing 29.9% of income within the top 10%.

We then compare two scenarios for 1980: our baseline model with spillovers and a counterfactual where we keep the shape of doctors' income distribution fixed at 2012

Table 9: Spillover effects and within-top income inequality

Income Group	2012 (%)	Doctors: 4.6% of top 10% Δ (Spill)	Δ (Mean)	Doctors + others: 9.2% of top 10% Δ (Spill)	Δ (Mean)
		percentage points		percentage points	
Top 10%	34.4	8.1	7.9	8.1	7.7
Top 5% / top 10%	69.7	6.1	5.6	5.8	4.9
Top 1% / top 10%	29.9	8.1	7.0	8.0	6.0
Top 0.1% / top 10%	8.9	4.3	3.6	4.4	3.1

Notes: Income inequality measures are for the whole distribution (consumers + doctors). The first column shows empirical measures of income inequality within the top for New York State in 2012. The second is the model-predicted difference between 1980 and 2012 and show changes (positive numbers means value higher in 2012 than 1980). The third shows analogous predictions where we shut off spillovers by keeping the shape of doctors' income distribution constant at 2012 level, but shift it down by the same mean. The previous two columns considers doctors as the only spillover occupation (4.6% of top 10%), whereas the final two consider a larger size of spillover occupations (2× actual size of doctors). Top 5% / 10% refers to the share of top 10% income accruing to the top 5% earners.

levels but shift it down by the same log mean, eliminating the spillover effect. The spillover model predicts a 8.1 percentage point increase in the share of the top 10% that accrues to the top 1% between 1980 and 2012, compared to only 7.0 points in the counterfactual without spillovers. This substantial difference exists despite doctors comprising only 4.6% of workers in the top 10% of the overall distribution.

To illustrate the potential broader impact of spillovers, we scale up the population experiencing spillovers. Real estate occupations (agents and managers) and dentists are around half as numerous as physicians. We thus double the number of physicians, representing a larger set of occupations subject to spillovers (henceforth, spillover occupations).[39] The model without spillovers now has a larger group with no within-group increase in income inequality and predicts a 6.0 percentage point increase in the top 1% share. The model with spillovers predicts a 8.0 percentage point increase—a difference of more than 30%. So even when spillover occupations are a relatively small share of the top income distribution, they can generate substantial additional within-top inequality.[40] To streamline exposition, for the remainder of this section we only focus on the case with this broader category of "spillover occupations".

Nominal and real income inequality. Our theoretical model showed that spillovers reduce welfare inequality for consumers. We build on that here by distinguishing between income and welfare changes. For each percentile of the consumer distribution, we calculate the equivalent variation—the extra income needed in 1980 to reach the

[39]Formally, we double the distribution of potential doctors and halve λ to capture the larger mass of spillover occupations. We recalibrate the model, which leaves $F_z(z)$ unchanged and doubles β.

[40]These calculations are for our baseline model where spillover occupations don't consume their own services. In an (unreported) analysis where they do, quantitative results are almost identical.

same utility level as in 2012. That is, we calculate the equilibrium utility value for each percentile, p, of consumers $v_p^t(x_p^t) = u(x_p^t - \omega^t(z^t(x_p^t)), z^t(x_p^t))$ for $t = 1980, 2012$, where x_p^t is the income of a consumer at percentile p in year t and $\omega^t(z^t(x_p^t))$ is the spending on medical services at the equilibrium choice of doctor quality $z^t(x_p^t)$ in year t. We then calculate the equivalent variation, EV_p, as the extra income required in 1980 to bring an agent at percentile p to the utility of 2012, that is:

$$v_p^{1980}(x_p^{1980} + EV_p) = v_p^{2012}(x_p^{2012}). \tag{18}$$

Figure 6 shows both the observed change in real income and the welfare-based EV measure. For consumers in the bottom the difference is below 0.01. However, for those above the 80th percentile, the log difference between income and welfare changes grows to $0.02 - 0.04$ (Panel (b)). This difference reflects the disproportionate increase in prices for high-quality services faced by high-income consumers. Although the share of spending on services provided by spillover occupations declines with income, this effect is quantitatively minor compared to the price effect.

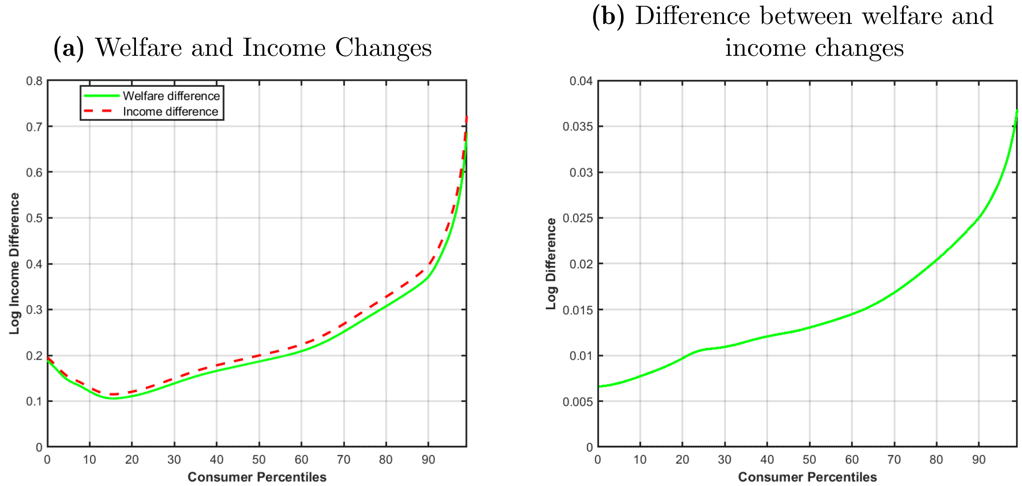
The spillover effect implies that conventional measures of income inequality using a common price deflator overstate the increase in welfare inequality. While the measured 90/50 income ratio rose by 0.20 log points, the corresponding welfare ratio increased by only 0.18 log points.

These two exercises show that—although spillovers increase top income inequality—they also imply that changes in income inequality overstate changes in welfare inequality. Appendix E brings these two elements together and shows that, compared to the scenario where doctors all experience the same proportional income change, spillovers reduce welfare inequality outside of the top 1% of the combined income distribution, but increase it in the very top. In essence, high-earning consumers lose from spillover effects through higher prices on the services they buy. But, in our setting, the service providers are even higher up the income distribution and the gains from spillovers accumulate in the top 1% of the income distribution.

7 Conclusion

This paper documents that the majority of the increase in top income inequality in the U.S. is within occupations. We develop a new theoretical framework where an increase in top income inequality in one occupation can spill over through consumption to other occupations that provide non-divisible services directly to customers, such

Figure 6: Changes in income and welfare for consumers, 1980 to 2012



Notes: In Panel (a), "Income difference" is the log change in real income (with a common CPI) for each percentile of consumers and "Welfare difference" is the extra income required in 1980 to reach the utility for the same percentile in 2012. Panel (b) shows the gap between Income difference and Welfare difference shown in Panel (a).

as physicians, dentists and real estate agents. We show empirically that changes in local income inequality do indeed spill over to these occupations, with standardized coefficients ranging from 1 to 1.5. Calibrating our model to New York State, we find that spillovers can account for the entire increase in income inequality for physicians.

Our analysis suggests that the increase in top income inequality observed across most occupations since 1980 may not require a common explanation. Increases in inequality for bankers or CEOs due to deregulation or globalization may have spilled over to other high-earning occupations, increasing top income inequality broadly. While we have emphasized positive results, the theory has an important normative implication: Increasing inequality in the prices of non-divisible services implies that welfare inequality does not rise as much as nominal income inequality. Similar spillover effects could exist for scarce goods such as luxury wine, or sport teams, a question we leave to future research.

References

R. Adão, M. Kolesár, and E. Morales. Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, 134(4):1949–2010, 2019.

P. Aghion, U. Akcigit, A. Bergeaud, R. Blundell, and D. Hémous. Innovation and top income inequality. *Review of Economic Studies*, 86(1):1–45, 2018.

M. Alder. Stardom and talent. *The American Economic Review*, 75(1):208–212, 1985.

P. Armour, R. Burkhauser, and J. Larrimore. Using the pareto distribution to improve estimates of topcoded earnings. *Economic Inquiry*, 54:1263–1273, 2016.

J. Bakija, A. Cole, and B. Heim. Jobs and income growth of top earners and the causes of changing income inequality: Evidence from u.s. tax return data. 2012.

T. J. Bartik. Boon or boondoggle? the debate over state and local economic development policies. *Upjohn Institute for Employment Research*, pages 1–16, 1991.

G. Becker and G. Lewis. On the interaction between the quantity and quality of children. *Journal of Political Economy*, 81(2):279–288.

M. Bertrand and A. Morse. Trickle-down consumption. *The Review of Economics and Statistics*, 98(5):863–879, 2016.

A. S. Blinder. How many us jobs might be offshorable? *World Economics*, 10(2): 41–78, 2009.

A. Bonfiglioli, R. Crino, and G. Gancia. Betting on exports: Trade and endogenous heterogeneity. *The Economic Journal*, 128(609):471–916, 2018.

K. Borusyak, P. Hull, and X. Jaravel. Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213, 2022.

F. Buera and J. Kaboski. The Rise of the Service Economy. *American Economic Review*, 102(6):2540–2569, 2012.

J. Clemens and J. D. Gottlieb. Do physicians' financial incentives affect treatment patterns and patient health? *American Economic Review*, 104(4):1320–1349, 2014.

J. Clemens and J. D. Gottlieb. In the shadow of a giant: Medicare's influence on private payment systems. *Journal of Political Economy*, 125(1):1–39, 2017.

J. Clemens, J. D. Gottlieb, and T. L. Molnár. Do health insurers innovate? evidence from the anatomy of physician payments. *Journal of Health Economics*, 55, 2017.

A. Deaton. Getting prices right: What should be done? *Journal of Economic Perspectives*, 12(1):37–46, 1998.

D. J. Deming. The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640, 2017.

J. I. Dingel, J. D. Gottlieb, M. Lozinski, and P. Mourot. Market size and trade in medical services. Working Paper No. 31030, NBER, March 2023.

D. Dorn. Essays on inequality, spatial interaction, and the demand for skills, ph.d. dissertation, university of st. gallen, 2009.

C. Edmond and S. Mongey. Unbundling labor. mimeo, 2021.

A. J. Epstein. Do cardiac surgery report cards reduce mortality? assessing the evidence. *Medical Care Research and Review*, 63(4):403–426, 2006.

A. Erosa, L. Fuster, G. Kambourov, and R. Rogerson. Labor market polarization and inequality: a roy model perspective. *NBER WP 33687*, 2025.

X. Gabaix and A. Landier. Why has ceo pay increased so much? *The Quarterly Journal of Economics*, 123 (1):49–100, 2008.

F. Geerolf. A theory of Pareto distributions. *Working Paper*, 2017.

S. Gilbukh and P. Goldsmith-Pinkham. Heterogeneous real estate agents and the housing cycle. *The Review of Financial Studies*, 37(11):3431–3489, August 2024.

P. Goldsmith-Pinkham, I. Sorkin, and H. Swift. Bartik instruments: What, when, why and how. *American Economic Review*, 110(8):2586–2624, 2020.

J. Gottlieb, M. Polyakova, K. Rinz, H. Shiplett, and V. Udalova. The earnings and labor supply of u.s. physicians. *The Quarterly Journal of Economics*, May 2025.

V. Grossman. Firm size, productivity and manager wages: a job assignment approach. *The B.E. Journal of Theoretical Economics*, 7(1), 2007.

F. Guvenen, F. Karahan, S. Ozkan, and J. Song. What do data on millions of u.s. workers reveal about life-cycle earnings risk. *Econometrica*, 89(5):2303–2339, 2021.

M. Hoffmann and I. Stewen. Holes in the dike: The global savings glut, us house prices, and the long shadow of banking deregulation. *Journal of the European Economic Association*, 18(4):2013–2055, 2020.

C. Jones and J. Kim. A schumpeterian model of top income inequality. *Journal of Political Economy*, 126(5):1785–1826, 2018.

S. Kaplan and J. Rauh. It's the market: the broad-based rise in the return to top talent. *Journal of Economic Perspectives*, 27(3):35–56, 2013.

L. J. Kirkeboen, E. Leuven, and M. Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

F. Koenig. Technical change and superstar effects: Evidence from the rollout of television. *American Economic Review: Insights*, 5(2), 2021.

J. T. Kolstad. Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7):2875–2910, 2013.

W. Kopczuk, E. Saez, and J. Song. Earnings inequality and mobility in the united states: Evidence from social security data since 1937. *Quarterly Journal of Economics*, 125(1):91–128, 2010.

T. Landvoigt, M. Piazzesi, and M. Schneider. The housing market(s) of san diego. *American Economic Review*, 105(4):1371–1407, 2015.

D. S. Lee, J. McCrary, M. J. Moreira, J. R. Porter, and L. Yap. What to do when you can't use '1.96' confidence intervals for iv. Working Paper 31893, National Bureau of Economic Research, November 2023.

M. Leonardi. The effect of product demand on inequality: Evidence from the united states and the united kingdom. *AEJ: Applied Economics*, 7(3):221–247, 2015.

N. Määttänen and M. Terviö. Income distribution and housing prices: an assignment model approach. *Journal of Economic Theory*, 151:381–410, 2014.

A. Manning. We can work it out: The impact of technological change on the demand for low-skill workers. *Scottish Journal of Political Economy*, 51(5):581–608, 2004.

S. Manson, J. Schroeder, D. V. Riper, and S. Ruggles. *IPUMS National Historical Geographic Information System: Version 12.0*. Minneapolis: University of Minnesota, 2017. doi: http://doi.org/10.18128/D050.V12.0.

F. Mazzolari and G. Ragusa. Spillovers from high-skill consumption to low-skill labor markets. *Review of Economics and Statistics*, 95(1):74–86, 2013.

Medical Group Management Association (MGMA). Physician compensation and production survey: 2009 report based on 2008 data., 2009.

T. Miceli, K. Pancak, and C. Sirmans. Is the compensation model for real estate brokers obsolete? *Journal of Real Estate Finance and Economics*, 35:7/22, 2007.

M. Mogstad and M. Wiswall. Testing the quantity-quality model of fertility: Estimation using unrestricted family size models. 7(1):157–192.

E. Moretti. Real wage inequality. *AEJ: Applied Economics*, 5(1):65–103, 2013.

T. Philippon and A. Reshef. Wages and human capital in the u.s. finance industry: 1909-2006. *Quarterly Journal of Economics*, 127 (4):1551–1609, 2012.

T. Piketty and E. Saez. Income inequality in the united states: 1913-1998. *Quarterly Journal of Economics*, 118(1):1–39, 2003.

T. Piketty, A. B. Atkinson, and E. Saez. Top income in the long run of history. *Journal of Economic Literature*, 49(1):3–71, 2011.

T. Piketty, E. Saez, and S. Stantcheva. Optimal taxation of top labor incomes: A tale of three elasticities. *AEJ: Economic Policy*, 6(1):230–271, 2014.

S. Rosen. Hedonic prices and implicit markets: Product differentiation in pure competition. 82(1):34–55.

S. Rosen. The economics of superstars. *American Economic Review*, 71(5), 1981.

M. Sattinger. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31 (2):831–880, 1993.

J. Song, D. Price, F. Guvenen, N. Bloom, and T. von Wachter. Firming up inequality. *The Quarterly Journal of Economics*, 134(1):1–50, 2019.

R. Steinbrook. Public report cards—cardiac surgery and beyond. *New England Journal of Medicine*, 355(18):1847–1849, 2006.

C. M. Tolbert and M. Sizer. U.s. commuting zones and labor market areas. a 1990 update. *Economic Research Service Staff Paper No. 9614*, 1996.

N. Wilmers. Does consumer demand reproduce inequality? high-income consumers, vertical differentiation, and the wage structure. *Am. Journal of Sociology*, 123(1).

J. Zabel. The demand for housing services. *Journal of Housing Economics*, 13(1): 16–35, 2004.

Online Appendix to supplement: "The Spillover Effects of Top Income Inequality"

Joshua D. Gottlieb, David Hémous, Jeffrey Hicks, and Morten Olsen

Contents

A Theory Appendix

A.1 Positive assortative matching in equilibrium

Since CES and Cobb-Douglas functions have positive cross-partial derivatives, the following lemma applies to the utility functions considered in our paper:

Lemma 1. *The equilibrium features positive assortative matching between the income of the patient and the skill of the doctor if the utility function has a positive cross-partial derivative.*

Proof. We prove the result by contradiction. Consider two individuals, 1 and 2, with income $x_1 < x_2$ whose consumption bundles are so that $z_1 > z_2$ and $c_1 < c_2$. Utility depends on z and the remaining disposable income $x - \omega(z)$. Since widget maker 1 chooses a doctor of quality

z_1, it must be that: $u(z_1, x_1 - \omega(z_1)) \geq u(z_2, x_1 - \omega(z_2))$. Further, we have:

$$
\begin{aligned}
& u(z_1, x_2 - \omega(z_1)) - u(z_2, x_2 - \omega(z_2)) \\
& = u(z_1, x_2 - \omega(z_1)) - u(z_1, x_1 - \omega(z_1)) + u(z_1, x_1 - \omega(z_1)) - u(z_2, x_1 - \omega(z_2)) \\
& \quad + u(z_2, x_1 - \omega(z_2)) - u(z_2, x_2 - \omega(z_2)) \\
& = \int_{x_1 - \omega(z_1)}^{x_2 - \omega(z_1)} \left(\frac{\partial u}{\partial c}(z_1, c) - \frac{\partial u}{\partial c}(z_2, c) \right) + u(z_1, x_1 - \omega(z_1)) - u(z_2, x_1 - \omega(z_2)).
\end{aligned}
$$

If the utility function has a positive cross-partial derivative, then the first term is positive as $z_1 > z_2$. Since the second term is also weakly positive, then $u(z_1, x_2 - \omega(z_1)) > u(z_2, x_2 - \omega(z_2))$. In other words, widget maker 2 would rather pick a doctor of ability z_1. This is a contradiction and it must be that $z_1 < z_2$. □

A.2 Proofs for the baseline model

A.2.1 Solving equation (5)

We look for a specific solution to equation (5). We find that $w(z) = K_1 z^{\frac{\alpha_z}{\alpha_x}}$ is one if

$$K_1 = x_{\min} \frac{\beta \alpha_x \lambda}{\alpha_z (1-\beta) + \beta \alpha_x} \left(\frac{1}{z_c} \right)^{\frac{\alpha_z}{\alpha_x}}.$$

The solutions to the differential equation $w'(z) z + \frac{\beta}{1-\beta} w(z) = 0$ are given by $K z^{-\frac{\beta}{1-\beta}}$ for any constant K. We get that all solutions to (5) take the form:

$$w(z) = \frac{x_{\min} \beta \alpha_x \lambda}{\alpha_z (1-\beta) + \beta \alpha_x} \left(\frac{z}{z_c} \right)^{\frac{\alpha_z}{\alpha_x}} + K z^{-\frac{\beta}{1-\beta}}.$$

We then obtain (6) by using that $w(z_c) = x_{\min}$ which fixes

$$K = x_{\min} z_c^{\frac{\beta}{1-\beta}} \frac{\alpha_z (1-\beta) + \beta \alpha_x (1-\lambda)}{\alpha_z (1-\beta) + \beta \alpha_x}.$$

A.2.2 Proof of Proposition 2

Combining (4) and (6), we can derive spending on health care as

$$h(x) = \frac{\beta \alpha_x}{\alpha_z (1-\beta) + \beta \alpha_x} x + x_{\min} \frac{\alpha_z (1-\beta) + \beta \alpha_x (1-\lambda)}{\lambda (\alpha_z (1-\beta) + \beta \alpha_x)} \left(\frac{x_{\min}}{x} \right)^{\frac{\alpha_x \beta}{\alpha_z (1-\beta)}}. \tag{19}$$

Combining (19) with (1) and (4), we obtain the utility of a widget maker with income x:

$$u(x) = \left(\frac{\alpha_z (1-\beta) x}{\alpha_z (1-\beta) + \beta \alpha_x} - \frac{(\alpha_z (1-\beta) + \beta \alpha_x (1-\lambda)) x_{\min}}{\lambda (\alpha_z (1-\beta) + \beta \alpha_x)} \left(\frac{x_{\min}}{x} \right)^{\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}} \right)^{1-\beta} z_c^{\beta} \left(\frac{x}{x_{\min}} \right)^{\frac{\beta \alpha_x}{\alpha_z}}.$$

Therefore $eq(x)$ obeys

$$eq(x) = \left(\frac{\alpha_z (1-\beta) x}{\alpha_z (1-\beta) + \beta \alpha_x} - \frac{(\alpha_z (1-\beta) + \beta \alpha_x (1-\lambda)) x_{\min}}{\lambda (\alpha_z (1-\beta) + \beta \alpha_x)} \left(\frac{x_{\min}}{x} \right)^{\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}} \right) \left(\frac{z_c}{z_r} \frac{x}{x_{\min}} \right)^{\frac{\alpha_x \beta}{\alpha_z (1-\beta)}},$$

which implies that for x large enough:

$$eq(x) \approx \left(\frac{z_c}{z_r} \right)^{\frac{\alpha_x \beta}{\alpha_z (1-\beta)}} \frac{\alpha_z (1-\beta) x_{\min}^{-\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}}}{\alpha_z (1-\beta) + \beta \alpha_x} x^{1+\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}}.$$

The distribution of real income, EQ, obeys $\Pr(EQ > e) = \Pr(X > eq^{-1}(e))$, so that for e large enough, we obtain:

$$\Pr(EQ > e) \approx \left(\frac{z_c}{z_r} \right)^{\frac{\alpha_x}{1+\frac{\alpha_z}{\alpha_x} \frac{1-\beta}{\beta}}} \left(\frac{x_{\min} \alpha_z (1-\beta)}{\alpha_z (1-\beta) + \beta \alpha_x} \frac{1}{e} \right)^{\frac{\alpha_x}{1+\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}}}.$$

Therefore asymptotically, real income is Pareto distributed with a shape parameter $\alpha_{eq} \equiv \frac{\alpha_x}{1+\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}}$. Moreover, we obtain: $\frac{d \ln \alpha_{eq}}{d \ln \alpha_x} = \frac{1}{1+\frac{\alpha_x}{\alpha_z} \frac{\beta}{1-\beta}}$.

A.3 Generalization to other utility functions

We now consider a generalized version of the model. There is a mass 1 of patients and a mass μ of potential doctors. Potential doctors may consume medical services with the same utility function as other agents (in which case the mass of widget makers is $1-\mu$) or not (the mass of widget makers is 1). The technology for health services is the same as before and we keep $\lambda > \mu^{-1}$. Agents not working as doctors produce a composite good which is the numeraire, and potential doctors can work as widget makers with the lowest productivity $x_{\min}$ as an alternative.

Patients' income is asymptotically Pareto distributed:[41] $P_x(X > x) = \overline{G}_x(\overline{x}) \overline{G}_{x,\overline{x}}(x)$, where $\overline{G}_{x,\overline{x}}(x)$ is the conditional counter-cumulative distribution above $\overline{x}$, $\overline{G}_x(\overline{x})$ is the unconditional counter-cumulative distribution, and for $\overline{x}$ large enough, $\overline{G}_x(x, \overline{x}) \approx (\overline{x}/x)^{\alpha_x}$ with $\alpha_x > 1$. The ability distribution of potential doctors is also asymptotically Pareto distributed.

We assume that patients' utility features positive cross-partial derivative (and put more structure in the following subsections), so that the equilibrium still features assortative matching and we still denote the matching function $m(z)$. Market clearing at every z can still be written as (3). The least able potential doctor who actually works as a doctor will have ability $z_c = \overline{G}_z^{-1}(1/(\lambda\mu))$, which is independent of α_x. Therefore, equation (3) implies that $m(z)$ is defined by $m(z) = \overline{G}_x^{-1}\left(\overline{G}_{z,z_c}(z)\right)$. For z above some threshold, $\overline{z}$, both doctors' talents and incomes

[41]If potential doctors do not consume health care services, this is an assumption on an exogenous object, the income distribution of widget makers. If potential doctors do consume health care services, this is an assumption on the equilibrium, which will be verified if the (exogenous) income distribution of widget makers is asymptotically Pareto.

are approximately Pareto distributed, which allows us to rewrite the previous equation as:

$$\overline{G}_x\left(m\left(\overline{z}\right)\right)\left(\left(\frac{m\left(\overline{z}\right)}{m\left(z\right)}\right)^{\alpha_x}+o\left(\left(\frac{m\left(\overline{z}\right)}{m\left(z\right)}\right)^{\alpha_x}\right)\right)=\overline{G}_{\overline{z},z_c}\left(\overline{z}\right)\left(\left(\frac{\overline{z}}{z}\right)^{\alpha_z}+o\left(\left(\frac{\overline{z}}{z}\right)^{\alpha_z}\right)\right),$$

$$\Rightarrow m\left(z\right)=Bz^{\frac{\alpha_z}{\alpha_x}}+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right)\text{ with }B=m\left(\overline{z}\right)\left(\frac{\overline{G}_x\left(m\left(\overline{z}\right)\right)}{\overline{G}_{\overline{z},z_c}\left(\overline{z}\right)\overline{z}^{\alpha_z}}\right)^{\frac{1}{\alpha_x}}. \tag{20}$$

A.3.1 Cobb-Douglas case

We now assume a Cobb-Douglas utility as in the baseline model. Solving for the patient problem still leads to the differential equation (2). Plugging (20) in (2) gives:

$$w'\left(z\right)z+\frac{\beta}{1-\beta}w\left(z\right)\approx\frac{\beta}{1-\beta}\lambda Bz^{\frac{\alpha_z}{\alpha_x}}.$$

Up to a constant, the problem is identical to the baseline for high z, so that Proposition 1 applies. Doctors' income is asymptotically Pareto distributed with shape parameter α_x.

Proof. We can rewrite (2) as

$$w'\left(z\right)z=\frac{\beta}{1-\beta}\left(\lambda Bz^{\frac{\alpha_z}{\alpha_x}}-w\left(z\right)\right)+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right). \tag{21}$$

We define $\overline{w}\left(z\right)\equiv\frac{\beta\alpha_x}{\alpha_z(1-\beta)+\beta\alpha_x}\lambda Bz^{\frac{\alpha_z}{\alpha_x}}$ which is a solution to the differential equation without the negligible term, and $\widetilde{w}\left(z\right)\equiv w\left(z\right)-\overline{w}\left(z\right)$, which must satisfy

$$\widetilde{w}'\left(z\right)z=-\frac{\beta}{1-\beta}\widetilde{w}\left(z\right)+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right).$$

This gives

$$\widetilde{w}'\left(z\right)z^{\frac{\beta_z}{1-\beta_z}}+\frac{\beta}{1-\beta}\widetilde{w}\left(z\right)z^{\frac{\beta_z}{1-\beta_z}-1}=o\left(z^{\frac{\alpha_z}{\alpha_x}}z^{\frac{\beta_z}{1-\beta_z}-1}\right)$$

Integrating we obtain: $\widetilde{w}\left(z\right)=Kz^{-\frac{\beta_z}{1-\beta_z}}+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right)$ for some constant K, so that $\widetilde{w}\left(z\right)$ is negligible in front of $\overline{w}\left(z\right)$. This ensures that

$$w\left(z\right)=\frac{\beta\alpha_x}{\alpha_z\left(1-\beta\right)+\beta\alpha_x}\lambda Bz^{\frac{\alpha_z}{\alpha_x}}+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right). \tag{22}$$

Therefore, for $\overline{w}_d$ large enough, doctors' income is distributed according to

$$P\left(W_d>w_d|w_d>\overline{w}_d\right)\approx\left(\overline{w}_d/w_d\right)^{\alpha_x}: \tag{23}$$

doctors' income follows a Pareto distribution with shape parameter α_x. When potential doctors consume medical services, this result is consistent with the initial assumption that patients' income is asymptotically Pareto distributed with shape parameter α_x. □

A.3.2 CES case and Proof of Proposition 3

We now assume that patients' utility is CES (9) with $\varepsilon \neq 1$. The first order condition for the patient's problem can be written as:

$$\frac{\partial u}{\partial z} = \omega'(z) \frac{\partial u}{\partial c}. \tag{24}$$

Using (9) and (4), and with $w(z) = \lambda\omega(z)$, we find that for high levels of z the wage function obeys a differential equation given by

$$w'(z) = \frac{\lambda^{\frac{\varepsilon-1}{\varepsilon}}\beta}{1-\beta} z^{-\frac{1}{\varepsilon}} \left(\lambda B z^{\frac{\alpha_z}{\alpha_x}} - w(z)\right)^{\frac{1}{\varepsilon}} (1 + o(1)). \tag{25}$$

The asymptotic distribution of doctors' wages is given either by Proposition 3 or by the following Proposition, which studies the remaining cases.

Proposition 6. *1) If either (i) $\varepsilon > 1$ and $\alpha_x^{-1} < \alpha_z^{-1}$, or (ii) $\varepsilon < 1$ and $\alpha_z^{-1} < \alpha_x^{-1}$, then doctors' wages are asymptotically Pareto distributed with the same shape parameter as widget makers: $\alpha_w = \alpha_x$. Further, asymptotically, widget makers spend all their income on health.*

2) Assume that $\varepsilon < 1$. Then for $\alpha_x^{-1} < (1-\varepsilon)\alpha_z^{-1}$, doctors' wages are bounded. For $\alpha_x^{-1} = (1-\varepsilon)\alpha_z^{-1}$, doctors' wages are asymptotically exponentially distributed. In both cases, the elasticity of health expenditures with respect to income tend to 0, $\frac{\ln h(x)}{\ln x} \to 0$.

Proof. We now establish Propositions 3 and 6. Since consumption of the homogeneous good must remain positive then $\lim \lambda B z^{\frac{\alpha_z}{\alpha_x}} - w(z) \geq 0$, which means that $w(z)$ cannot grow faster than $z^{\frac{\alpha_z}{\alpha_x}}$. We can then distinguish 2 cases: $w(z) = o\left(z^{\frac{\alpha_z}{\alpha_x}}\right)$ and $w(z) \propto z^{\frac{\alpha_z}{\alpha_x}}$.

Case with $w(z) = o\left(z^{\frac{\alpha_z}{\alpha_x}}\right)$. Then for z high enough, one obtains that

$$w'(z) = \lambda \frac{\beta}{1-\beta} B^{\frac{1}{\varepsilon}} z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}} + o\left(z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}\right). \tag{26}$$

Integrating, we obtain that for $\left(\frac{\alpha_z}{\alpha_x} - 1\right)\frac{1}{\varepsilon} \neq -1$

$$w(z) = K + \lambda \frac{\beta}{1-\beta} \frac{B^{\frac{1}{\varepsilon}}}{\left(\frac{\alpha_z}{\alpha_x} - 1\right)\frac{1}{\varepsilon} + 1} z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1} + o\left(z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}\right),$$

where K is a constant. Note that to be consistent, we must have $\left(\frac{\alpha_z}{\alpha_x} - 1\right)\frac{1}{\varepsilon} + 1 < \frac{\alpha_z}{\alpha_x}$, that is $(\alpha_z - \alpha_x)(\varepsilon - 1) > 0$: this case is ruled out if $\alpha_z \geq \alpha_x$ and $\varepsilon < 1$ or if $\alpha_z \leq \alpha_x$ and $\varepsilon > 1$.

If $\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1<0$ then $w(z)$ is bounded by K.

If $\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1>0$, then we get that

$$w(z)=f^w(z)=\lambda\frac{\beta}{1-\beta}\frac{B^{\frac{1}{\varepsilon}}}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}+o\left(z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}\right),$$

where the notation f^w is introduced for clarity. Therefore one gets, for $\overline{w}$ large:

$$\Pr(W>w)=\Pr\left(Z>(f^w)^{-1}(w)\right)=\overline{G}_w(\overline{w})\left(\frac{\overline{w}}{w}\right)^{\frac{\alpha_z}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}}+o\left(w^{-\frac{\alpha_z}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}}\right),$$

so that w is Pareto distributed asymptotically with a coefficient $\alpha_w^{-1}=\frac{1}{\varepsilon}\alpha_x^{-1}+\left(1-\frac{1}{\varepsilon}\right)\alpha_z^{-1}$, which is increasing in α_x^{-1} (and we have $\alpha_w^{-1}<\alpha_x^{-1}$).

If $\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1=0$, then $\alpha_z=\alpha_x(1-\varepsilon)$, and integrating (26), one obtains

$$w(z)=f^w(z)=\lambda\frac{\beta}{1-\beta}B^{\frac{1}{\varepsilon}}\ln z+o(\ln z).$$

Therefore

$$\begin{aligned}\Pr(W>w)&=\Pr\left(Z>\left(\exp\left(\frac{1-\beta}{\lambda\beta B^{\frac{1}{\varepsilon}}}w\right)+o(\exp(w))\right)\right)\\&=\overline{G}_{z,z_c}(\overline{z})\,\overline{z}^{\alpha_z}\exp\left(-\frac{\alpha_z(1-\beta)}{\lambda\beta B^{\frac{1}{\varepsilon}}}w\right)+o(\exp(-\alpha_z w))\end{aligned}$$

In that case, w is distributed exponentially.

Case where $w(z)\propto z^{\frac{\alpha_z}{\alpha_x}}$. That is we assume that

$$w(z)=Az^{\frac{\alpha_z}{\alpha_x}}+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right) \tag{27}$$

for some constant $A>0$. Then, we have that

$$\Pr(W>w)=\Pr\left(Z>\left(\left(\frac{w}{A}\right)^{\frac{\alpha_x}{\alpha_z}}+o(w)^{\frac{\alpha_x}{\alpha_z}}\right)\right)=\overline{G}_w(\overline{w})\left(\frac{\overline{w}}{w}\right)^{\alpha_x}+o(w)^{\frac{\alpha_x}{\alpha_z}}$$

That is w is Pareto distributed with coefficient α_x. Plugging (27) in (25), we get:

$$A\frac{\alpha_z}{\alpha_x}z^{\frac{\alpha_z}{\alpha_x}-1}+o\left(z^{\frac{\alpha_z}{\alpha_x}-1}\right)=\lambda^{\frac{\varepsilon-1}{\varepsilon}}\frac{\beta}{1-\beta}(\lambda B-A)^{\frac{1}{\varepsilon}}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}+o\left((\lambda B-A)^{\frac{1}{\varepsilon}}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}\right). \tag{28}$$

First, if $\alpha_z=\alpha_x$, then we obtain $A=\lambda^{\frac{\varepsilon-1}{\varepsilon}}\frac{\beta_z}{\beta_c}(\lambda B-A)^{\frac{1}{\varepsilon}}$.

Second, assume that $\alpha_z\neq\alpha_x$. If $\lambda B\neq A$ then (28) is impossible when $\varepsilon\neq 1$, and we must have that $\lambda B = A$. This equation then requires that

$$\frac{\alpha_z}{\alpha_x} - 1 < \left(\frac{\alpha_z}{\alpha_x} - 1\right)\frac{1}{\varepsilon} \Leftrightarrow (\alpha_z - \alpha_x)(\varepsilon - 1) < 0.$$

In fact, for $(\alpha_z - \alpha_x)(\varepsilon - 1) < 0$, one gets that

$$w(z) = \lambda B z^{\frac{\alpha_z}{\alpha_x}} - \lambda \left(B \frac{\alpha_z}{\alpha_x} \frac{\beta_c}{\beta_z}\right)^{\varepsilon} z^{\varepsilon\left(\frac{\alpha_z}{\alpha_x} - 1\right) + 1} + o\left(z^{\varepsilon\left(\frac{\alpha_z}{\alpha_x} - 1\right) + 1}\right)$$

satisfies (25) as long as the function $o\left(z^{\varepsilon\left(\frac{\alpha_z}{\alpha_x} - 1\right) + 1}\right)$ solves the appropriate differential equation.

Collecting the different cases together gives Propositions 3 and 6. In addition, since the income distribution of doctors never has a fatter tail than a Pareto with shape parameter α_x, the results are always consistent with patients' income being Pareto distributed with shape parameter α_x for the case where potential doctors consume medical services. □

A.3.3 Homothetic utility function

We now consider a general homothetic utility function u. In that case, the ratio of marginal utilities $\frac{\partial u}{\partial z}/\frac{\partial u}{\partial c}$ only depends on the ratio c/z. Using patient's budget constraint and the matching function (20), we can then write (24) as

$$w'(z) = \lambda \frac{\partial u}{\partial z} / \frac{\partial u}{\partial c} \equiv \lambda f\left(B z^{\frac{\alpha_z}{\alpha_x} - 1} - \frac{w(z)}{z\lambda}\right). \tag{29}$$

We assume that the utility function admits positive and finite limits to its elasticity of substitution when z/c goes to either 0 or infinity. That is:

$$\lim_{z/c \to \infty} - \frac{d \ln\left(\frac{\partial u}{\partial z}/\frac{\partial u}{\partial c}\right)}{d \ln(z/c)} = \frac{1}{\varepsilon_\infty} \text{ and } \lim_{z/c \to 0} - \frac{d \ln\left(\frac{\partial u}{\partial z}/\frac{\partial u}{\partial c}\right)}{d \ln(z/c)} = \frac{1}{\varepsilon_0},$$

where $\varepsilon_k \in (0, \infty)$ for $k \in \{0, \infty\}$. Then, we can write that for z/c arbitrarily large ($k = \infty$) or small ($k = 0$):

$$\ln\left(\frac{\partial u}{\partial z}/\frac{\partial u}{\partial c}\right) = \left(\frac{1}{\varepsilon_k} \ln\left(\lambda B z^{\frac{\alpha_z}{\alpha_x} - 1} - \frac{w(z)}{z}\right) + \ln \beta\right)(1 + o(1)),$$

where β is a constant. In these two cases, we can then rewrite (29) as:

$$w'(z) = \lambda^{\frac{\varepsilon_k - 1}{\varepsilon_k}} \beta \left(\lambda B z^{\frac{\alpha_z}{\alpha_x} - 1} - \frac{w(z)}{z}\right)^{\frac{1}{\varepsilon_k}} (1 + o(1)), \tag{30}$$

which is the same expression as (25) in the CES case (except that there are two potential values for ε_k). We obtain:

Proposition 7. *Propositions 3 and 6 apply to any homothetic utility function which admit positive and finite local elasticities of substitutions as the ratio z/c tend to 0 or infinity. The relevant elasticity is ε_0 when $\alpha_z > \alpha_x$ and ε_∞ when $\alpha_z < \alpha_x$ (Proposition 1 applies when $\varepsilon_k = 1$).*

Proof. If $\alpha_z < \alpha_x$, then $z/c \to \infty$, regardless of $w(z)$, and the logic of Propositions 3 and 6 immediately applies with ε_∞ (and Proposition 1 applies if $\varepsilon_\infty = 1$).

Consider now the case $\alpha_z > \alpha_x$. To establish the result, we need to check that $z/c \to 0$ (that is $c/z = \lambda B z^{\frac{\alpha_z}{\alpha_x}-1} - w(z)/z \to \infty$) in all cases. If $\varepsilon_0 = 1$, wages are Pareto distributed and health expenditures are an interior share of total income, which ensures that $\lambda B z^{\frac{\alpha_z}{\alpha_x}-1} - w(z)/z \to \infty$ (so Proposition 1 applies). If $\varepsilon_0 > 1$, then, following Proposition 3, health expenditures become a negligible share of total income, which ensures that $\lambda B z^{\frac{\alpha_z}{\alpha_x}-1} - w(z)/z \to \infty$. If $\varepsilon_0 < 1$, then, following Proposition 6, health expenditures are asymptotically equal to total income. Therefore, we must have $w(z) = \lambda B z^{\frac{\alpha_z}{\alpha_x}} - g(z)$, where $g(z)$ is negligible compared with $z^{\frac{\alpha_z}{\alpha_x}}$. Plugging this expression in (30) gives:

$$\frac{\alpha_z}{\alpha_x}\lambda B z^{\frac{\alpha_z}{\alpha_x}-1} - g'(z) = \lambda^{\frac{\varepsilon-1}{\varepsilon}}\beta\left(\frac{g(z)}{z}\right)^{\frac{1}{\varepsilon}}(1+o(1)).$$

Assume that $g(z)/z$ is bounded, then we would get that $g'(z) \to \alpha_z \lambda B z^{\frac{\alpha_z}{\alpha_x}-1}/\alpha_x$, but this contradicts the assumption that $g(z)$ is negligible in front of $z^{\frac{\alpha_z}{\alpha_x}}$. Therefore, $g(z)/z$ is unbounded, so that $\lambda B z^{\frac{\alpha_z}{\alpha_x}-1} - w(z)/z \to \infty$ in that case as well. □

A.4 Generalized ability distribution

We consider the set-up of the baseline model with a Cobb-Douglas utility function but generalize the ability distribution to any unbounded distribution with a counter-CDF denoted $\overline{G}_z$. (We keep the widget makers' income distribution Pareto but this could be generalized as well.) We assume that $\lim_{z\to\infty} \frac{z g_z(z)}{\overline{G}_z}$ exists. If this limit is positive and finite then the ability distribution is asymptotically Pareto (with a shape parameter equal to that limit) and this case is treated in Appendix A.3.1. We focus here on the case where the limit is either 0 or infinite. As before, there is a cut-off value z_c above which all potential doctors choose to be doctors. We then define $\widetilde{G}(z) = \overline{G}_z(z)/\overline{G}_z(z_c) = \lambda\mu\overline{G}_z(z)$, which is the counter-cumulative ability distributions of individuals who actually choose to be doctors (and $\widetilde{g}$ is the corresponding conditional PDF). We have $\lim \frac{z\widetilde{g}(z)}{\widetilde{G}} = \lim \frac{z g_z(z)}{\overline{G}_z}$. Equation (4) is then replaced by $m(z) = x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$, which allows to derive the differential equation for the wage function as:

$$w'(z)\,z = \frac{\beta}{1-\beta}\left(\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}} - w(z)\right), \tag{31}$$

instead of (5). We then establish:

Proposition 8. *If the ability distribution has a tail at least as fat as Pareto ($\lim_{z\to\infty} \frac{zg_z(z)}{\widetilde{G}_z}$ exists and is finite), doctors' income is asymptotically Pareto distributed with shape parameter α_x. If the ability distribution has a tail thinner than Pareto ($\lim_{z\to\infty} \frac{zg_z(z)}{\widetilde{G}_z} = \infty$), doctors' income is not asymptotically Pareto distributed but $\frac{\ln(P(W>w))}{\ln w}$ decreases with α_x for w large enough.*

For a Pareto distribution, $\ln(P(W > w)) / \ln w = -\alpha_w$, so the statement that $\ln(P(W > w)) / \ln w$ decreases with α_x for high w directly generalizes Proposition 1: top income inequality spills over from the consumers' distribution to the doctors' income distribution even when the ability distribution has a tail thinner than Pareto.

Proof. We consider in turn two cases: either $w(z) \to A\lambda x_{\min} \left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$ for some constant $A \in (0, 1]$ or $w(z)$ is dominated by $\lambda x_{\min} \left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$.

Case 1: $w(z) \to A\lambda x_{\min} \left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$. Then

$$P(W > w) \to \widetilde{G}\left(\widetilde{G}^{-1}\left(\frac{w}{A\lambda x_{\min}}\right)^{-\alpha_x}\right) = \left(\frac{w}{A\lambda x_{\min}}\right)^{-\alpha_x},$$

so that the doctors' income distribution is Pareto distributed with shape parameter α_x. We can write $w(z) = A(z)\lambda x_{\min} \left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$ where $A(z)$ tends toward a positive constant so that in the limit $A'(z) = 0$. Plugging this in (31), one gets:

$$A'(z)z + \frac{1}{\alpha_x}\frac{z\widetilde{g}(z)}{\widetilde{G}(z)}A(z) = \frac{\beta_z(1 - A(z))}{1 - \beta_z}. \tag{32}$$

Since $A(z)$ tends toward a constant, we must have $\lim A'(z)z = 0$ (if $A'(z)z$ were bounded below above 0, then $A(z)$ would grow faster than the log function). When $\lim \frac{z\widetilde{g}(z)}{\widetilde{G}(z)}$ is positive and finite, we recover the asymptotic Pareto case that we have already studied. If $\lim \frac{z\widetilde{g}(z)}{\widetilde{G}(z)} = 0$, then we must have that $A(z) \to 1$. This is consistent with equation (32) since both the right-hand and left-hand sides tend toward 0. In contrast, if $\lim \frac{z\widetilde{g}(z)}{\widetilde{G}(z)} = \infty$, the left-hand side is unbounded and the right-hand side is bounded which yields a contradiction: so that $w(z)$ must be dominated by $\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$ in that case.

Case 2: $w(z) = o\left(\lambda x_{\min} \left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}\right)$. Then, (31) leads to

$$w'(z) = \frac{\beta}{1 - \beta}\lambda x_{\min}\frac{1}{z}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}(1 + o(1)). \tag{33}$$

If $\lim \frac{zg_z(z)}{\widetilde{G}_z(z)} = 0$, then for any $K > 0$, we get that for z high enough, $\widetilde{G}(z) > Kz\widetilde{g}(z)$, so that

$$w'(z) > K\frac{\beta}{1-\beta}\lambda x_{\min} z\widetilde{g}(z)\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}-1}.$$

This directly implies that $w(z) > K\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$ for any K, which is a contradiction. Since we have also ruled out $\lim \frac{zg_z(z)}{\widetilde{G}_z(z)}$ positive and finite, then we must have that $\lim \frac{zg_z(z)}{\widetilde{G}_z(z)} = \infty$.

In return when $\lim \frac{zg_z(z)}{\widetilde{G}_z(z)} = \infty$, (33) implies that $\frac{d}{dz}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}/w'(z) \to \infty$ as $\frac{d}{dz}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}} = z\widetilde{g}(z)\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}-1}$. This justifies the assumption that $w(z) = o\left(\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}\right)$. Integrating (33), we can write

$$w(z) = w(z_M) + \frac{\beta(1+o(1))}{1-\beta}\lambda x_{\min}\left(F_{\alpha_x}(z) - F_{\alpha_x}(z_M)\right).$$

for some z_M, where $F_{\alpha_x}(z)$ is a primitive of $\frac{1}{z}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$. As $\widetilde{G}(z)$ has a thinner tail than Pareto, we get in particular that $\widetilde{G}(z) \leq z^{-2\alpha_x}$, so that $\frac{1}{z}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}} > z$. As a result $F_{\alpha_x}(z)$ and $w(z)$ go to infinity. We can then rewrite:

$$w(z) = \frac{\beta(1+o(1))}{1-\beta}\lambda x_{\min} F_{\alpha_x}(z),$$

so that for large w, $z(w) \approx F_{\alpha_x}^{-1}\left(w\frac{1-\beta_z}{\beta_z\lambda x_{\min}}\right)$. We then get

$$P(W > w) = \widetilde{G}(z(w)) \approx \widetilde{G}\left(F_{\alpha_x}^{-1}\left(w\frac{1-\beta}{\beta\lambda x_{\min}}\right)\right).$$

By definition, we can rewrite

$$w\frac{1-\beta_z}{\beta\lambda x_{\min}} = \int_{z_M}^{F_{\alpha_x}^{-1}\left(w\frac{1-\beta}{\beta\lambda x_{\min}}\right)} \frac{1}{\zeta}\left(\widetilde{G}(\zeta)\right)^{-\frac{1}{\alpha_x}} d\zeta.$$

Differentiating with respect to α_x, one gets:

$$\frac{\partial F_{\alpha_x}^{-1}(w)}{\partial \alpha_x}\frac{1}{F_{\alpha_x}^{-1}(w)}\left(\widetilde{G}\left(F_{\alpha_x}^{-1}(w)\right)\right)^{-\frac{1}{\alpha_x}} = \int_{z_M}^{F_{\alpha_x}^{-1}(w)} \frac{1}{\alpha_x}\frac{1}{\zeta}\left(\widetilde{G}(\zeta)\right)^{-\frac{1}{\alpha_x}-1} d\zeta.$$

Therefore F_{α_x} is increasing in α_x. As $\widetilde{G}$ is decreasing, $P(W > w)$ is decreasing in α_x. □

A.5 Scalability in health care: proof of Proposition 4

This section presents the proof of Proposition 4.[42] Health care market clearing now takes into account that doctors serve different number of patients. Still denoting z_c the ability of the least able doctor and using the Pareto assumptions, we get:

$$(m_z/x_{\min})^{-\alpha_x} = \int_z^{\infty} \lambda(\zeta)\, \alpha_z \frac{1}{\zeta} (\zeta/z_c)^{-\alpha_z}\, d\zeta, \tag{34}$$

Cobb-Douglas case. In the Cobb-Douglas case, we combine (2), (11) and (34), and we obtain the differential equation:

$$\omega'(z)\, z + \frac{\beta}{1-\beta}\omega(z) = \frac{\beta}{1-\beta} x_{\min} \left(\int_z^{\infty} \left(\frac{\omega(\zeta)}{k}\right)^{\varepsilon^S} \alpha_z \frac{1}{\zeta} \left(\frac{\zeta}{z_c}\right)^{-\alpha_z} d\zeta \right)^{-\frac{1}{\alpha_x}}. \tag{35}$$

We verify that a solution to the problem of the form $\omega(z) = C_1 z^{\psi}$ exists. Plugging this expression in (35), we obtain a solution with $\psi = \frac{\alpha_z}{\alpha_x + \varepsilon^S}$ and some constant C_1. In fact, the solution must asymptotically behave like $C_1 z^{\psi}$, otherwise the left-hand and right-hand sides of (35) cannot be of the same order. Doctors' income can then be written as

$$w(z) = \lambda(z)\, \omega(z) \to C_2 z^{\frac{\alpha_z\left(1+\varepsilon^S\right)}{\alpha_x+\varepsilon^S}},$$

where C_2 is another constant. Therefore for w, large enough, we obtain:

$$\Pr(W > w) \approx \Pr\left(Z > (w/C_2)^{\frac{\alpha_x+\varepsilon^S}{\alpha_z(1+\varepsilon^S)}} \right) \approx z_c^{\alpha_z} (w/C_2)^{-\frac{\alpha_x+\varepsilon^S}{1+\varepsilon^S}}.$$

Doctors' incomes are asymptotically Pareto distributed with inverse Pareto parameter α_w^{-1}:

$$\alpha_w^{-1} = \frac{1+\varepsilon^S}{1+\varepsilon^S \alpha_x^{-1}} \alpha_x^{-1} > \alpha_x^{-1}.$$

We note that α_w^{-1} is increasing in α_x^{-1} and in ε^S (since $\alpha_x > 1$). Besides, we get $\lambda(z) = C_1^{\varepsilon^S} z^{\psi\varepsilon^S}/k^{\varepsilon^S}$, so that

$$\Pr(\Lambda > \lambda) \approx \Pr\left(Z > \left(\lambda k^{\varepsilon^S}/C_1^{\varepsilon^S}\right)^{\frac{\alpha_x+\varepsilon^S}{\alpha_z \varepsilon^S}} \right) \propto \lambda^{-\frac{\alpha_x+\varepsilon^S}{\varepsilon^S}}.$$

[42] For the cases not covered by Proposition 4, we get (proofs omitted) that doctors' income distribution is also asymptotically Pareto distributed with $\alpha_w^{-1} = \frac{1+\varepsilon^S}{1+\varepsilon^S \alpha_x^{-1}} \alpha_x^{-1}$ if $\varepsilon < 1$ and $\alpha_z > \varepsilon^S + \alpha_x$, or ii) $\varepsilon > 1$ and $\alpha_z < \varepsilon^S + \alpha_x$. In addition, it is bounded if $\varepsilon < 1$ and $(1-\varepsilon)\, \alpha_x > \alpha_z$.

λ is Pareto distributed with inverse Pareto parameter $\alpha_\lambda^{-1} = \frac{\varepsilon^S \alpha_x^{-1}}{1+\alpha_x^{-1}\varepsilon^S}$ increasing in α_x^{-1}.

CES case. We now consider the CES case. We can rewrite the first-order condition (24) as

$$\omega'(z) = \frac{\beta}{1-\beta} z^{-\frac{1}{\varepsilon}} \left(m(z) - \omega(z)\right)^{\frac{1}{\varepsilon}}.$$

We assume (and verify) that when either i) $\varepsilon < 1$ and $(1-\varepsilon)\alpha_x < \alpha_z < \varepsilon^S + \alpha_x$, or ii) $\varepsilon > 1$ and $\alpha_z > \varepsilon^S + \alpha_x$, we are on the empirically relevant case where heath care expenditures rise less fast than income so that $\omega(z)/m(z) \to 0$. In that case, we get that for z sufficiently high:

$$\omega'(z) \approx \frac{\beta}{1-\beta} z^{-\frac{1}{\varepsilon}} m(z)^{\frac{1}{\varepsilon}}. \tag{36}$$

We guess and verify that the solution features $m(z) = m_0 z^{m_1} + o(z^{m_1})$ with $m_1 > 0$. Integrating (36) gives

$$\omega(z) \approx \frac{\beta}{1-\beta} \frac{1}{\frac{m_1-1}{\varepsilon}+1} m_0^{\frac{1}{\varepsilon}} z^{\frac{m_1-1}{\varepsilon}+1} + K + o\left(z^{\frac{m_1-1}{\varepsilon}+1}\right),$$

for some constant K. Plugging that expression in (34) and using (11) gives

$$(m(z))^{-\alpha_x} x_{\min}^{\alpha_x} = \frac{\alpha_z z_c^{\alpha_z}}{k^{\varepsilon^S}} \int_z^\infty \left(\frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}} \zeta^{\frac{m_1-1}{\varepsilon}+1}}{\frac{m_1-1}{\varepsilon}+1} + K + o\left(z^{\frac{m_1-1}{\varepsilon}+1}\right)\right)^{\varepsilon^S} \zeta^{-\alpha_z-1} d\zeta. \tag{37}$$

We need to consider three cases. i) First, assume that $\frac{m_1-1}{\varepsilon}+1 > 0$, then K is negligible in front of $\frac{m_1-1}{\varepsilon}+1$ and we can rewrite (37) as

$$(m_0 z^{m_1})^{-\alpha_x} x_{\min}^{\alpha_x} \approx \frac{\alpha_z z_c^{\alpha_z}}{k^{\varepsilon^S}} \left(\frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}}}{\frac{m_1-1}{\varepsilon}+1}\right)^{\varepsilon^S} \frac{1}{\left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S - \alpha_z} z^{\left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S - \alpha_z}. \tag{38}$$

This integration is only possible if $\left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S - \alpha_z < 0$ and this (approximate) equality can only hold if

$$\alpha_x m_1 = \alpha_z - \left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S \Leftrightarrow m_1 = \frac{\alpha_z + \left(\frac{1}{\varepsilon}-1\right)\varepsilon^S}{\alpha_x + \frac{\varepsilon^S}{\varepsilon}}.$$

For this value of m_1, we can verify that $m_1 > 0$ since we assumed that $\alpha_z > \varepsilon^S + \alpha_x$ when $\varepsilon > 1$. In addition,

$$\frac{m_1-1}{\varepsilon}+1 = \frac{\alpha_z + (\varepsilon-1)\alpha_x}{\varepsilon\alpha_x + \varepsilon^S} > 0,$$

since we have assumed that when $\varepsilon < 1$ then $(1-\varepsilon)\alpha_x < \alpha_z$. Moreover,

$$\left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S - \alpha_z = \frac{\alpha_x\left[\varepsilon^S(\varepsilon-1) - \alpha_z\varepsilon\right]}{\varepsilon\alpha_x + \varepsilon^S},$$

which is negative if $\varepsilon < 1$ but also if $\varepsilon > 1$ as in that case we have assumed that $\alpha_z > \varepsilon^S + \alpha_x$. Finally, we get that

$$\omega(z) \approx \frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}}}{\frac{m_1-1}{\varepsilon}+1} z^{\frac{m_1-1}{\varepsilon}+1} \approx \frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}}}{\frac{m_1-1}{\varepsilon}+1} z^{\frac{\alpha_z+(\varepsilon-1)\alpha_x}{\varepsilon\alpha_x+\varepsilon^S}}.$$

We have that

$$\frac{\omega(z)}{m(z)} \approx \frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}-1}}{\frac{m_1-1}{\varepsilon}+1} z^{\left(\frac{\alpha_z-\varepsilon^S-\alpha_x}{\alpha_x+\frac{\varepsilon^S}{\varepsilon}}\right)\left(\frac{1}{\varepsilon}-1\right)} \to 0,$$

since we have assumed that $\alpha_z < \varepsilon^S + \alpha_x$ when $\varepsilon < 1$ and $\alpha_z > \varepsilon^S + \alpha_x$ when $\varepsilon > 1$. Therefore health care expenditures do not scale up with income asymptotically as assumed initially and this case is internally consistent.

ii) Assume that $\frac{m_1-1}{\varepsilon}+1 < 0$ and $K \neq 0$ (or $\frac{m_1-1}{\varepsilon}+1 = 0$), then (37) leads to

$$m_0^{-\alpha_x} z^{-m_1\alpha_x} x_{\min}^{\alpha_x} \approx \frac{z_c^{\alpha_z}}{k^{\varepsilon^S}} K^{\varepsilon^S} z^{-\alpha_z}. \tag{39}$$

Therefore, we must have $m_1 = \frac{\alpha_z}{\alpha_x}$. This implies that

$$\frac{m_1-1}{\varepsilon}+1 = \frac{1}{\varepsilon}\left(\frac{\alpha_z}{\alpha_x} - 1 + \varepsilon\right) > 0,$$

since we have assumed that $\alpha_z > (1-\varepsilon)\alpha_x$ when $\varepsilon < 1$. This leads to a contradiction, so that this case is impossible.

iii) Assume now that $\frac{m_1-1}{\varepsilon}+1 < 0$ and $K = 0$, then (37) also leads to (38), which as argued below implies $\frac{m_1-1}{\varepsilon}+1 > 0$ leading to a contradiction.

Therefore only case i) is possible. We then get that doctors' income is

$$w(z) = \frac{1}{k^{\varepsilon^S}} (\omega(z))^{1+\varepsilon^S} \to C_1 z^{\frac{\alpha_z+(\varepsilon-1)\alpha_x}{\varepsilon\alpha_x+\varepsilon^S}\left(1+\varepsilon^S\right)},$$

for some constant C_1, so that doctors' income is asymptotically Pareto distributed with

$$\alpha_w = \alpha_z \frac{\varepsilon\alpha_x + \varepsilon^S}{(\alpha_z + (\varepsilon-1)\alpha_x)(1+\varepsilon^S)},$$

which we can rewrite as (12). We get that

$$\frac{d\alpha_w^{-1}}{d\alpha_x^{-1}} = \frac{\varepsilon^S+1}{\varepsilon\left(\frac{\varepsilon^S}{\varepsilon}\alpha_x^{-1}+1\right)^2}\left[1-\left(1-\frac{1}{\varepsilon}\right)\alpha_z^{-1}\varepsilon^S\right] > 0,$$

which is obvious if $\varepsilon < 1$ and holds if $\varepsilon > 1$ since $\alpha_z > \varepsilon^S + \alpha_x$. Moreover,

$$\frac{d\alpha_w^{-1}}{d\varepsilon^S} = \frac{\left(1 - \frac{\alpha_x^{-1}}{\varepsilon}\right)\left(\left(1 - \frac{1}{\varepsilon}\right)\alpha_z^{-1} + \frac{1}{\varepsilon}\alpha_x^{-1}\right)}{\left(\frac{\varepsilon^S}{\varepsilon}\alpha_x^{-1} + 1\right)^2},$$

which is positive if and only if $\varepsilon\alpha_x > 1$. Finally, we get that $\lambda(z) \propto w(z)^{\varepsilon^S/\left(1+\varepsilon^S\right)} \propto z^{\frac{\alpha_z+(\varepsilon-1)\alpha_x}{\varepsilon\alpha_x+\varepsilon^S}\varepsilon^S}$, so that λ is Pareto distributed with inverse Pareto parameter $\alpha_\lambda^{-1} = \frac{\varepsilon^S\left(\left(1-\frac{1}{\varepsilon}\right)\alpha_z^{-1}+\frac{1}{\varepsilon}\alpha_x^{-1}\right)}{\frac{\varepsilon^S}{\varepsilon}\alpha_x^{-1}+1}$, which is increasing in α_x^{-1}. This establishes the proof.

A.6 Occupational mobility

We now analyze the model briefly described in Section 3.3.2. Individuals' abilities as doctors and widget makers are positively (in fact perfectly) correlated so that there can be occupational mobility along the entire ability distribution. Formally, we keep a similar set-up as in the baseline model but we assume that there is a mass 1 of agents who decide whether to be doctors or widget makers. We rank agents in descending order of ability and use i to denote their rank. For two agents i and i' with $i < i'$, i will be better both as a widget maker and as a doctor than i'. Both ability distributions are Pareto with parameters $(x_{\min}, \alpha_x)$ for widget maker and $(z_{\min}, \alpha_z)$ for doctors. An agent i can choose between becoming a widget maker earning $x(i)$ or being a doctor providing health services of quality $z(i)$ and earning $w(z(i))$. Those working as doctors also need the services of doctors. We assume that $\lambda > 1$ to ensure that everyone can get health services. By definition of the rank we have that the counter-cumulative distribution functions for x and z obey $\overline{G}_x(x(i)) = \overline{G}_z(z(i)) = i$.

Assume that below a certain rank, some individuals choose to be widget makers and some doctors. This holds in equilibrium under a condition specified below. Then, individuals must be indifferent between the two occupations, so that for i low enough, we have $w(z(i)) = x(i)$. Therefore the wage function must satisfy $w(z) = \overline{G}_x^{-1}\left(\overline{G}_z(z)\right)$ for z high enough. As both ability distributions are Pareto, we get:

$$w(z) = x_{\min}(z/z_{\min})^{\alpha_z/\alpha_x}. \tag{40}$$

Doctor wages grow in proportion to what they could earn as a widget maker.

Let $\mu(z) \in [0,1]$ denote the share of individuals with medical ability z who choose to be doctors. Market clearing in medical services implies that:

$$(x_{\min}/m(z))^{\alpha_x} = \int_z^\infty \lambda\mu(\zeta)\, g_z(\zeta)\, d\zeta, \tag{41}$$

where $m(z)$ denotes the income of the patient of a doctor of quality z.

The first order condition on health care consumption (2) still applies. For z sufficiently high, μ is interior, and (40) holds, combining these expressions with (41), we obtain:

$$\int_z^\infty \mu(\zeta)\,\alpha_z \zeta^{-\alpha_z-1} d\zeta = \lambda^{\alpha_x-1} z^{-\alpha_z}\left(\frac{\alpha_z}{\alpha_x}+\frac{\beta}{1-\beta}\right)^{-\alpha_x}.$$

Differentiating with respect to z, we find that μ is a constant: $\mu = \lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}+\frac{\beta}{1-\beta}\right)^{-\alpha_x}$. Intuitively, with a constant μ, doctors' wages grow proportionately with patients' incomes, in line with the Cobb-Douglas assumption. To be consistent with our assumption of an interior equilibrium, we must have $\lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta}+1\right)^{-\alpha_x} < 1$.[43]

With a constant share of individuals choosing to be doctors (above a threshold), we get that $P_{doc}(W_d > w_d) = P\left(Z > w^{-1}(w_d)\right)$ for w_d high enough so that the observed distribution for doctor wages is Pareto with a shape parameter α_x. Therefore, Proposition 1 still applies:[44]

Proposition 9. *Assume that* $\lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta}+1\right)^{-\alpha_x} < 1$, *then doctors' income is Pareto distributed above a threshold with the same shape parameter as for widget makers. Therefore, an increase in top income inequality for widget makers increases top income inequality for doctors.*

Therefore the models with and without occupational mobility are observationally equivalent for top income inequality: doctors' top income inequality perfectly traces that of widget makers. Finally, note that with occupational mobility, doctors and widget makers interact through two channels: a demand side and an outside option side. Appendix A.6.1 below presents an additional model that separates the two. It highlights that the demand effect drives the result. Intuitively, if top income inequality increases for the outside option, higher-ability doctors will move to the outside option. This generates an increase in the relative pay of the remaining high-ability doctors, which, under Cobb-Douglas preferences, exactly compensates for the change in ability distribution of active doctors, leaving the observed income distribution unchanged.[45]

A.6.1 Disentangling supply side and demand side effects

To disentangle demand side and outside option effects, we now assume that doctors have an outside option positively correlated with their ability but patients are a separate group. There are two types of agents: a mass 1 of consumers and a mass M of potential doctors. Consumers have the same utility as before (1), and their income, x, is Pareto distributed with shape parameter α_x. Potential doctors only consume the homogeneous good. They are ranked in descending order of ability and agent i can choose between being a doctor providing health

[43] If $\lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta}+1\right)^{-\alpha_x} > 1$, then all individuals above a certain ability threshold choose to be doctors while all those below it choose to be widget makers. This is counterfactual.

[44] If the distributions of x and z are only asymptotically Pareto, then Proposition 1 applies asymptotically.

[45] This intuition does not generalize to the CES case where part of the effect of a rise in income inequality on doctors' income inequality results from the outside option effect instead of purely the demand side.

services of quality $z(i)$ and earning $w(z(i))$ or working in the homogeneous good sector earning $y(i)$. y and z are distributed according to the counter-cumulative distributions:

$$\overline{G}_y(y(i)) = \overline{G}_z(z(i)) = i \text{ with } \overline{G}_y = (y_{\min}/y)^{\alpha_y} \text{ and } \overline{G}_z = (z_{\min}/z)^{\alpha_z}.$$

Further $\lambda M > 1$ so that all consumers can get health services.

Assume that the equilibrium is such that for individuals of a sufficiently high level of ability, some choose to be doctors and others to work in the homogeneous good sector. Then, for i low enough, agents must be indifferent between becoming a doctor or working in the homogeneous good sector, so that $w(z(i)) = y(i)$. Hence, the wage function obeys:

$$w(z) = y_{\min}(z/z_{\min})^{\alpha_z/\alpha_y}. \tag{42}$$

Market clearing for health care services above z implies:

$$\left(\frac{x_{\min}}{m(z)}\right)^{\alpha_x} = \lambda M \int_z^{\infty} \mu(\zeta)\, g_z(\zeta)\, d\zeta, \tag{43}$$

where $\mu(\zeta)$ denotes the share of potential doctors who choose to be doctors. Plugging this expression in the first order condition (2) together with (42), we obtain:

$$\int_z^{\infty} \mu(\zeta)\, g_z(\zeta)\, d\zeta = \frac{1}{\lambda M}\left(\frac{\frac{\beta}{1-\beta}\lambda x_{\min}}{\left(\frac{\alpha_z}{\alpha_y}+\frac{\beta}{1-\beta}\right) y_{\min}}\right)^{\alpha_x}\left(\frac{z}{z_{\min}}\right)^{-\alpha_x\frac{\alpha_z}{\alpha_y}}. \tag{44}$$

Taking the derivative with respect to z, we get:

$$\mu(z) = \frac{\alpha_x}{\alpha_y}\frac{1}{\lambda M}\left(\frac{\frac{\beta}{1-\beta}\lambda x_{\min}}{\left(\frac{\alpha_z}{\alpha_y}+\frac{\beta}{1-\beta}\right) y_{\min}}\right)^{\alpha_x}\left(\frac{z}{z_{\min}}\right)^{\alpha_z\left(1-\frac{\alpha_x}{\alpha_y}\right)}. \tag{45}$$

Since $\mu(z) \in (0,1)$, this case is only possible if $\alpha_y \leq \alpha_x$, that is consumers' income distribution has a fatter tail than the outside option for potential doctors (and, if $\alpha_y = \alpha_x$, $\frac{\alpha_x}{\alpha_y}\frac{1}{\lambda M}\left(\frac{\alpha_y\beta\lambda x_{\min}}{(\alpha_z(1-\beta)+\beta\alpha_y)y_{\min}}\right)^{\alpha_x} \leq 1$). Then, for w high enough, doctors' income distribution obeys:

$$\Pr(W > w) = \int_{z_{\min}\left(\frac{w}{y_{\min}}\right)^{\frac{\alpha_y}{\alpha_z}}}^{\infty} \mu(\zeta)\left(\frac{z_{\min}}{\zeta}\right)^{\alpha_z}\frac{d\zeta}{\zeta} = \frac{1}{\lambda M \alpha_z}\left(\frac{\alpha_y\beta\lambda x_{\min}}{\alpha_z(1-\beta)+\beta\alpha_y}\right)^{\alpha_x} w^{-\alpha_x}.$$

Therefore, for w high enough, doctors' income is distributed like their patients' income.

With $\alpha_y > \alpha_x$ or $\alpha_y = \alpha_x$ together with $\frac{\alpha_x}{\alpha_y}\frac{1}{\lambda M}\left(\frac{\alpha_y\beta\lambda x_{\min}}{(\alpha_z(1-\beta)+\beta\alpha_y)y_{\min}}\right)^{\alpha_x} > 1$, then above a certain threshold, all potential doctors will choose to be doctors, so that the model behaves like that of Section 3.1, and the outside option is "mute".

Therefore, in all cases, income is Pareto distributed at the top with shape parameter α_x. Changes in α_y have no impact on doctors' top income inequality.

A.7 Doctors moving: Proof of Proposition 5

With no trade in goods between the two regions, we can normalize the price of the homogeneous good to 1 in both. As doctors only consume the homogeneous good, doctors' nominal wages must be equalized in the two regions and the price of health care of quality z must be the same in both regions. From the first order condition on health care consumption, the matching function is also the same: doctors of quality z provide health care to widget makers of income $m(z)$ in both regions. Moreover, the least able potential doctor who decides to become a doctor must have the same ability z_c in both regions.[46]

We define by $\varphi(z)$ the *net* share of doctors initially in region B with ability *at least* z who decide to move to region A. Labor market clearing in A implies that, for $z \geq z_c$,

$$\left(x_{\min}^{A}/m(z)\right)^{\alpha_x^A} = \lambda\mu\left(1+\varphi(z)\right)\left(z_{\min}/z\right)^{\alpha_z}. \tag{46}$$

There are initially $\mu\left(\frac{z_{\min}}{z}\right)^{\alpha_z}$ doctors with ability at least z in each region and a share $\varphi(z)$ of those move from region B to region A. As each doctor serves λ patients, after doctors have relocated the total supply over a quality z in region A is given by the right-hand side of (46). Total demand corresponds to region A patients with an income higher than $m(z)$, of which there are $P(X > m(z))$. The same equation, replacing $\varphi(z)$ by $-\varphi(z)$, holds in region B:

$$\left(x_{\min}^{B}/m(z)\right)^{\alpha_x^B} = \lambda\mu\left(1-\varphi(z)\right)\left(z_{\min}/z\right)^{\alpha_z}. \tag{47}$$

Since the two regions are of equal size, total demand for health services must be the same and on net, no doctors move: $\varphi(z_c) = 0$. Summing up the market clearing equations (46) and (47) for $z = z_c$, we obtain $z_c = (\lambda\mu)^{\frac{1}{\alpha_z}} z_{min}$, as in the baseline model.

Similarly, combining (46) and (47) for any z, we obtain

$$x_{\min}^{A}\left(1+\varphi(z)\right)^{-\frac{1}{\alpha_x^A}} = x_{\min}^{B}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^B}-\frac{\alpha_z}{\alpha_x^A}}\left(1-\varphi(z)\right)^{-\frac{1}{\alpha_x^B}}. \tag{48}$$

Since $\alpha_x^B > \alpha_x^A$, we find that $\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^B}-\frac{\alpha_z}{\alpha_x^A}}$ tends towards 0. As a net share $\varphi(z) \in (-1,1)$. If $\varphi(z) \to -1$, the left-hand side tends toward infinity and the right-hand side toward 0, which is

[46] Here, potential doctors who decide to work in the homogeneous good sector would go to region B since $\alpha_x^A > \alpha_x^B$ implies that $x_{\min}^A < x_{\min}^B$. This is without consequences: alternatively, we could have assumed that the outside option of doctors is to produce $\widehat{x}$, which is identical between the two regions. In that case potential doctors who work in the homogeneous sector would not move.

a contradiction. Therefore $1+\varphi(z)$ must be bounded below, which ensures that the left-hand side is bounded above 0. If $\varphi(z) \not\to 1$, then the right-hand side tends toward 0, which is also a contradiction. Therefore asymptotically, we must have that $\varphi(z) \to 1$: nearly all the best doctors move to the most unequal region.

Plugging (46) in (2), we get that in region A:

$$w'(z)\,z + \frac{\beta}{1-\beta} w(z) = \frac{\beta\lambda}{1-\beta}\left(1+\varphi(z)\right)^{-\frac{1}{\alpha_x^A}}\left(\frac{z_c}{z}\right)^{-\frac{\alpha_z}{\alpha_x^A}}.$$

Therefore, asymptotically:

$$w(z) \to \frac{\lambda\beta\alpha_x^A 2^{-\frac{1}{\alpha_x^A}}}{\alpha_z(1-\beta)+\beta\alpha_x^A}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^A}} \tag{49}$$

As $\varphi(z) \to 1$, doctors' talent is asymptotically distributed with Pareto coefficient α_z in region A after the location decision. For z high enough, there are $2\mu\left(\frac{z_{\min}}{z}\right)^{\alpha_z}$ doctors eventually located in region A. Then, as in the baseline model, doctors' income is asymptotically Pareto distributed with coefficient α_x^A in A. Further, using (48), we get:

$$1-\varphi(z) \to 2^{\alpha_x^B/\alpha_x^A}\left(x_{\min}^B/x_{\min}^A\right)^{\alpha_x^B}(z/z_c)^{\alpha_z\left(1-\alpha_x^B/\alpha_x^A\right)}. \tag{50}$$

Therefore, the *ex post* talent distribution of doctors in region B is still Pareto but now with a coefficient $\alpha_z' = \alpha_z\frac{\alpha_x^B}{\alpha_x^A}$. In region B, the probability that a doctor earns at least $\widetilde{w}$ obeys:

$$P_{doc}^B(W>\widetilde{w}) = \frac{\mu P\left(Z>w^{-1}(\widetilde{w})\right)\left(1-\varphi\left(w^{-1}(\widetilde{w})\right)\right)}{\mu P(Z>z_c)},$$

where w above denotes the wage function. Indeed, there are initially $\mu P\left(Z>w^{-1}(\widetilde{w})\right)$ doctors in region B with a talent sufficient to earn $\widetilde{w}$. A share of $1-\varphi\left(w^{-1}(\widetilde{w})\right)$ of these doctors stay in B. Moreover, the total mass of active doctors in region B is given by $\mu P(Z>z_c)$, since overall there is no net movement of actual doctors. Using (49), we get:

$$w^{-1}(\widetilde{w}) \to z_c\left(\widetilde{w}\frac{\alpha_z(1-\beta)+\beta\alpha_x^A}{\lambda\beta\alpha_x^A}2^{\frac{1}{\alpha_x^A}}\right)^{\frac{\alpha_x^A}{\alpha_z}}.$$

Using this expression and (50) we get that:

$$P_{doc}^B(W>\widetilde{w}) = \left(\frac{z_c}{w^{-1}(\widetilde{w})}\right)^{\alpha_z}\left(1-\varphi\left(w^{-1}(\widetilde{w})\right)\right) \to \left(\frac{x_{\min}^B}{x_{\min}^A}\frac{\lambda\beta\alpha_x^A}{\alpha_z(1-\beta)+\beta\alpha_x^A}\frac{1}{\widetilde{w}}\right)^{\alpha_x^B}.$$

Therefore, doctors' income in region B is Pareto distributed with shape parameter α_B as in the baseline model. This establishes Proposition 5.

A.8 Quality-quantity tradeoff

We consider the model of Section 3.3.4.[47] With $\omega(z)$ the price per unit of health care of quality z, the budget constraint faced by a consumer with income x is:

$$\omega(z)\, q + c \leq x,$$

We restrict attention to the case $\theta \leq 1$: for $\theta > 1$, the problem is ill-defined because consumers can achieve infinite utility by purchasing an infinitesimal amount of health-care of infinite quality. We establish:

Proposition 10. *1) In the Cobb-Douglas case, $\theta = 1$, there is not positive assortative matching. Doctors' income is Pareto distributed with inverse Pareto parameter $\frac{\gamma}{1-\gamma}\alpha_z^{-1}$, independent of widget makers' top income inequality.*

2) When quality and quantity are complements ($\theta < 1$), and $\alpha_z > \alpha_x - 1$, there is positive assortative matching; doctors' income is asymptotically Pareto distributed with inverse Pareto parameter

$$\alpha_w^{-1} = \alpha_x^{-1}\left(1 + \alpha_z^{-1}\right) - \alpha_z^{-1} < \alpha_x^{-1}.$$

Doctors' top income inequality is increasing in widget makers' top income inequality.

3) When quality and quantity are complements ($\theta < 1$), and $\alpha_z < \alpha_x - 1$, there is positive assortative matching but doctors' income is bounded.

Proof. We solve in turn the complement and then Cobb-Douglas cases.

Complement case: $\theta < 1$. We solve for the consumer maximization problem in two steps. First, we solve for the optimal allocation between quantity and quality for given healthcare expenditures. Second, we solve for the consumer's problem taking into account the relationship between quality and quantity. In a first step, a consumer maximizes:

$$H = \left((1-\gamma)^{\frac{1}{\theta}} q^{\frac{\theta-1}{\theta}} + \gamma^{\frac{1}{\theta}} z^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}} \text{ such that } \omega(z)\, q \leq h.$$

Plugging the budget constraint in the objective function, we get that the consumer solves:

$$\max_z H(z,h) = \left((1-\gamma)^{\frac{1}{\theta}} \left(\frac{h}{\omega(z)}\right)^{\frac{\theta-1}{\theta}} + \gamma^{\frac{1}{\theta}} z^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}. \tag{51}$$

[47] Utility functions where the quality and quantity of goods do not aggregate in a Cobb-Douglas way have been considered by Rosen (1974) or Becker and Lewis (1973) in the context of fertility and education decision. In the latter case, Mogstad and Wiswall (2016) specifically look at the CES case. None of these papers, however, consider a matching mechanism like ours.

We obtain the first order condition:

$$\frac{\partial H}{\partial z} = \left(\gamma^{\frac{1}{\theta}} z^{-\frac{1}{\theta}} - (1-\gamma)^{\frac{1}{\theta}} h^{\frac{\theta-1}{\theta}} \omega(z)^{\frac{1-2\theta}{\theta}} \omega'(z)\right) H^{\frac{1}{\theta}} = 0 \tag{52}$$

which leads to the optimal z satisfying:

$$z^* = \frac{\gamma}{1-\gamma} h^{1-\theta} \omega(z^*)^{2\theta-1} \omega'(z^*)^{-\theta}. \tag{53}$$

The solution is interior and a maximum provided that the second order condition holds which stipulates that at the optimum we must have $\frac{\partial^2 H}{(\partial z)^2} < 0$. We have that

$$\begin{aligned} \frac{\partial^2 H}{(\partial z)^2}(z^*, h) &= -\gamma^{\frac{1}{\theta}} g(z^*) \frac{\omega'(z)}{\omega(z^*)} (z^*)^{-\frac{1}{\theta}} H^{\frac{1}{\theta}} \\ \text{with } g(z) &\equiv \frac{1}{\theta} \frac{1}{\varepsilon_\omega(z^*)} + \frac{1-2\theta}{\theta} + \frac{\omega(z^*)\omega''(z^*)}{(\omega'(z))^2} \end{aligned} \tag{54}$$

where we introduced $\varepsilon_\omega(z) \equiv \frac{z\omega'(z)}{\omega(z)}$ the elasticity of the price function. Therefore at an interior optimum, we must have $g(z^*) > 0$. Differentiating (52), we obtain:

$$\frac{dz^*}{dh} = \frac{\frac{\partial^2 H}{\partial h \partial z}}{\left(-\frac{\partial^2 H}{(\partial z)^2}\right)} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{\theta}} \frac{1-\theta}{\theta g(z^*)} (z^*)^{\frac{1}{\theta}} h^{\frac{-1}{\theta}} \omega(z^*)^{\frac{1-\theta}{\theta}}.$$

Therefore, we have that for an interior solution health care quality increases with health care expenditures, $\frac{dz^*}{dh} > 0$ since $\theta < 1$. Taking that relationship into account, we can rewrite

$$H(z) = z\gamma^{\frac{1}{\theta}} \left(\frac{1}{\varepsilon_\omega(z)} + 1\right)^{\frac{\theta}{\theta-1}}, \tag{55}$$

$$\text{with} \frac{dH}{dz} = \gamma^{-\frac{1-\theta}{\theta^2}} g(z) \frac{\theta}{1-\theta} \left[\frac{H(z)}{z\gamma^{\frac{1}{\theta}}}\right]^{\frac{1}{\theta}},$$

$$\text{and } \frac{dH}{dz} = \frac{\partial H}{\partial z} + \frac{\partial H}{\partial h}\frac{dh}{dz} = \frac{\partial H}{\partial h}\frac{1}{\frac{dz}{dh}} = \gamma^{-\frac{1-\theta}{\theta^2}} g(z^*) (z^*)^{-\frac{1}{\theta}} \frac{H^{\frac{1}{\theta}}\theta}{1-\theta} > 0, \tag{56}$$

from the second order condition. In addition, the quantity of health care demanded by a consumer consuming quality z is then given by

$$q(z) = z\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}} (\varepsilon_\omega(z))^{\frac{\theta}{1-\theta}}. \tag{57}$$

Using (53), we can then rewrite the original problem as

$$\max u(z,c) = \gamma^{\frac{\beta}{\theta}} \left(\frac{\omega(z^*)}{\omega'(z^*)} z^{-\frac{1}{\theta}} + \gamma^{\frac{1}{\theta}} z^{\frac{\theta-1}{\theta}} \right)^{\frac{\theta}{\theta-1}\beta} c^{1-\beta},$$

$$\text{s.t.} \quad \left(\frac{1-\gamma}{\gamma} z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)^{\frac{1}{1-\theta}} + c = x.$$

Taking the ratio of the two first order conditions, we obtain:

$$\frac{\beta \frac{dH}{dz} c}{(1-\beta) H} = \left(\frac{1-\gamma}{\gamma} \right)^{\frac{1}{1-\theta}} \frac{d \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)^{\frac{1}{1-\theta}}}{dz}.$$

As before, we denote by $m(z)$ the income of patients of a doctor of quality z. Then plugging in the budget constraint in the previous expression, we get:

$$\begin{aligned} & \beta \frac{dH}{dz} \left(m(z) - \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)^{\frac{1}{1-\theta}} \right) \\ = & \left(\frac{1-\gamma}{\gamma} \right)^{\frac{1}{1-\theta}} \frac{d \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)^{\frac{1}{1-\theta}}}{dz} (1-\beta) H. \end{aligned} \quad (58)$$

We note that

$$\frac{d \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)}{dz} = \theta g(z) \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right) \frac{\omega'(z)}{\omega(z)}.$$

Using this expression, together with (56) into (58), and noting that at the optimum we always have $g(z) \neq 0$, we get:

$$\begin{aligned} & \beta z^{-\frac{1}{\theta}} \left(m(z) - \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)^{\frac{1}{1-\theta}} \right) \\ = & \gamma^{\frac{1-\theta}{\theta^2}} \left(\frac{1-\gamma}{\gamma} \right)^{\frac{1}{1-\theta}} \left(z\omega(z)^{1-2\theta} \omega'(z)^{\theta} \right)^{\frac{1}{1-\theta}} \frac{\omega'(z)}{\omega(z)} (1-\beta) H^{\frac{\theta-1}{\theta}}. \end{aligned}$$

Then using (55), we obtain:

$$\beta m(z) = z\omega(z) \left(\varepsilon_{\omega}(z) \right)^{\frac{\theta}{1-\theta}} \left[\beta + \left(\frac{1-\gamma}{\gamma} \right)^{\frac{1}{1-\theta}} (1-\beta) + \left(\frac{1-\gamma}{\gamma} \right)^{\frac{1}{1-\theta}} (1-\beta) \varepsilon_{\omega}(z) \right]. \quad (59)$$

This differential equation characterizes the optimal behavior of consumers.

Next, market clearing in health care imposes that the quantity of health care provided by doctors of quality at least z must match consumption by consumers of income at least $m(z)$:

$$\lambda\mu\left(\frac{z}{z_{\min}}\right)^{-\alpha_z}=\int_{m(z)}^{\infty}q\left(m^{-1}(x)\right)\alpha_x x^{-(\alpha_x+1)}x_{\min}^{\alpha_x}dx.$$

Differentiating with respect to z, we obtain:

$$-\alpha_z\lambda z^{-\alpha_z-1}\mu z_{\min}^{\alpha_z}=-m'(z)\,q(z)\,\alpha_x m(z)^{-(\alpha_x+1)}\,x_{\min}^{\alpha_x}.$$

Plugging in (57), we get a second differential equation that characterizes market clearing:

$$z^{-\alpha_z-2}\alpha_z\lambda\mu z_{\min}^{\alpha_z}=m'(z)\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(\varepsilon_\omega(z))^{\frac{\theta}{1-\theta}}m(z)^{-(\alpha_x+1)}\alpha_x x_{\min}^{\alpha_x}. \tag{60}$$

Together with initial conditions determining the cut-off z_c and ensuring that $w(z_c)=x_{\min}$, equations (59) and (60) fully characterize the equilibrium.

We are interested in the asymptotic behavior of the solution. Multiplying equations (59) and (60) with each other, we get

$$\begin{aligned}&\left[\left(\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}}+1-\beta\right)\omega(z)\,z^{-\alpha_z-1}+(1-\beta)\,z^{-\alpha_z}\omega'(z)\right]\alpha_z\lambda\mu z_{\min}^{\alpha_z} \\ =\;& m'(z)\,m(z)^{-\alpha_x}\,\beta\alpha_x x_{\min}^{\alpha_x}.\end{aligned} \tag{61}$$

We assume that the solution behaves regularly in the sense that $\varepsilon_\omega(z)$ admits a limit in $[0,\infty]$ (we know that $\varepsilon_\omega(z)\geq 0$). We consider in turn the three possible cases of an infinite limit, a finite but positive limit and a null limit.

Case 1: $\varepsilon_\omega(z)\to\infty$. Then, $\omega(z)$ must grow faster than any power function and $\omega(z)\,z^{-\alpha_z-1}$ becomes negligible in front of $z^{-\alpha_z}\omega'(z)$. Using (61), we can write:

$$\left[-\alpha_z z^{-\alpha_z-1}\omega(z)+z^{-\alpha_z}\omega'(z)\right](1-\beta)\,\alpha_z\lambda\mu z_{\min}^{\alpha_z}\sim\alpha_x m'(z)\,m(z)^{-\alpha_x}\,x_{\min}^{\alpha_x}\beta.$$

Integrating, we get

$$\alpha_z\lambda\mu z_{\min}^{\alpha_z}(1-\beta)\,\omega(z)\,z^{-\alpha_z}+K\sim\frac{\alpha_x}{1-\alpha_x}m(z)^{1-\alpha_x}\,x_{\min}^{\alpha_x}\beta,$$

for some constant K. However, while the RHS tends toward 0, the LHS becomes unbounded, leading to a contradiction. This case cannot be an equilibrium.

Case 2: $\varepsilon_\omega(z)\to\omega_1$ with $\omega_1>0$ and finite. Then we can write $\omega(z)=\omega_0 z^{\omega_1}+o(z^{\omega_1})$ and

$\omega'(z) = \omega_0\omega_1 z^{\omega_1-1} + o(z^{\omega_1-1})$. Using (61), we can write

$$\omega_0\left[\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta + \omega_1(1-\beta)\right] z^{\omega_1-\alpha_z-1}\alpha_z\lambda\mu z_{\min}^{\alpha_z} \sim m'(z)\, m(z)^{-\alpha_x}\, \beta\alpha_x x_{\min}^{\alpha_x}.$$

Integrating, we obtain

$$\frac{\omega_0\left[\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta + \omega_1(1-\beta)\right]}{\omega_1-\alpha_z} z^{\omega_1-\alpha_z}\alpha_z\lambda\mu z_{\min}^{\alpha_z} + K \sim \frac{1}{1-\alpha_x} m(z)^{1-\alpha_x}\, \beta\alpha_x x_{\min}^{\alpha_x},$$

for some constant K. As the RHS tends toward 0, the LHS must also tend toward 0, which requires both that $K = 0$ and that $\omega_1 < \alpha_z$. Then, we get that $m(z) \sim m_0 z^{\frac{\alpha_z-\omega_1}{\alpha_x-1}}$ with

$$m_0 = \left[\frac{(\alpha_z-\omega_1)\,\beta\alpha_x x_{\min}^{\alpha_x}}{(\alpha_x-1)\,\omega_0\left[\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta + \omega_1(1-\beta)\right]\alpha_z\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1}{\alpha_x-1}} > 0.$$

Plugging that expression in (59), we get

$$\beta m_0 z^{\frac{\alpha_z-\omega_1}{\alpha_x-1}} \sim \omega_0 z^{\omega_1+1}\omega_1^{\frac{\theta}{1-\theta}}\left[\beta + (1-\beta)\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(1+\omega_1)\right].$$

This implies that we must have

$$\frac{\alpha_z-\omega_1}{\alpha_x-1} = \omega_1 + 1 \Rightarrow \omega_1 = \frac{\alpha_z+1-\alpha_x}{\alpha_x}.$$

The solution assumes that $\omega(z)$ is increasing, so that this is only an equilibrium if $\omega_1 > 0$, that is for $\alpha_z > \alpha_x - 1$. As $\alpha_x > 1$, we note that the condition $\omega_1 < \alpha_z$ is automatically satisfied. ω_0 is then determined by ensuring that $\beta m_0 = \omega_0\omega_1^{\frac{\theta}{1-\theta}}\left[\beta + (1-\beta)\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(1+\omega_1)\right]$.

We still need to check that $g(z) > 0$ to satisfy the second order condition. To do so, we first compute $\omega''(z)$. Log-differentiating (59), we get

$$\frac{zm'(z)}{m(z)} = \frac{1}{1-\theta} + \frac{1-2\theta}{1-\theta}\varepsilon_\omega(z) + \frac{\theta}{1-\theta}\frac{z\omega''(z)}{\omega'(z)} + \frac{(1-\beta)\,\varepsilon_\omega(z)\left(1+\frac{z\omega''(z)}{\omega'(z)} - \varepsilon_\omega(z)\right).}{\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta + (1-\beta)\,\varepsilon_\omega(z)} \quad (62)$$

From (60), we get that $m'(z)$ is asymptotically a power function and that $\frac{zm'(z)}{m(z)}$ must tend toward a constant. Then (62) implies that $\frac{z\omega''(z)}{\omega'(z)}$ also tends toward a constant, which must be

$\omega_1 - 1$, so that we can write $\omega''(z) = \omega_0\omega_1(\omega_1 - 1) z^{\omega_1 - 2} + o(z^{\omega_1 - 2})$. Using (54), we get

$$g(z) \sim \frac{1}{\theta}\frac{1}{\omega_1} + \frac{1-2\theta}{\theta} + \frac{\omega_1 - 1}{\omega_1} \sim \frac{1-\theta}{\theta}\frac{1+\omega_1}{\omega_1} > 0,$$

so that the SOC is satisfied.

Doctors' income $w(z) = \lambda\omega(z)$ is then asymptotically Pareto distributed with

$$\Pr(W > w) = \Pr\left(Z > \left(\frac{w}{\lambda\omega_0}\right)^{\frac{1}{\omega_1}}\right) \sim z_c^{\alpha_z}\left(\frac{w}{\lambda\omega_0}\right)^{-\alpha_w},$$

where $\alpha_w \equiv \alpha_z/\omega_1$. We can rewrite the latter as:

$$\alpha_w^{-1} = \frac{\alpha_z + 1 - \alpha_x}{\alpha_z\alpha_x} = \alpha_x^{-1}\left(1 + \alpha_z^{-1}\right) - \alpha_z^{-1},$$

which is increasing in both α_x^{-1} and α_z^{-1}. In addition, we note that $\frac{d\alpha_w^{-1}}{d\alpha_x^{-1}} = 1 + \alpha_z^{-1} > 1$.

Case 3: $\varepsilon_\omega(z) \to 0$, then $\omega(z)$ grows slower than any power function and $\omega'(z)$ is negligible relative to $\frac{\omega(z)}{z}$. Using (61), we can write:

$$\lambda\mu z_{\min}^{\alpha_z}\left(\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta\right)\left(\alpha_z z^{-\alpha_z - 1}\omega(z) - z^{-\alpha_z}\omega'(z)\right) \sim \alpha_x m'(z)\, m(z)^{-\alpha_x} x_{\min}^{\alpha_x}\beta.$$

Integrating, we get

$$\lambda\mu\left(\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta\right) z_{\min}^{\alpha_z} z^{-\alpha_z}\omega(z) + K \sim \frac{1}{\alpha_x - 1}\alpha_x m(z)^{1-\alpha_x} x_{\min}^{\alpha_x}\beta,$$

for some constant K. $z^{-\alpha_z}\omega(z)$ and $m(z)^{1-\alpha_x}$ tend toward 0, so that $K = 0$. We then obtain:

$$m(z) \sim \left[\frac{z^{\alpha_z}}{\omega(z)}\frac{\beta\alpha_x}{\alpha_x - 1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu\left(\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}} + 1 - \beta\right) z_{\min}^{\alpha_z}}\right]^{\frac{1}{\alpha_x - 1}}. \tag{63}$$

Plugging this in (59), we obtain

$$\beta\left[\frac{z^{\alpha_z}}{\omega(z)}\frac{\beta\alpha_x}{\alpha_x - 1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1}{\alpha_x - 1}} \sim z\omega(z)\left(\varepsilon_\omega(z)\right)^{\frac{\theta}{1-\theta}}\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}\frac{1}{\alpha_x - 1}}\left[\beta + \left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(1-\beta)\right]^{\frac{\alpha_x}{\alpha_x - 1}}.$$

Rearranging terms, we get

$$\left[\frac{\beta^{\alpha_x}\alpha_x}{\alpha_x-1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1-\theta}{(\alpha_x-1)\theta}} z^{\frac{\alpha_z-\alpha_x+1}{\alpha_x-1}\frac{1-\theta}{\theta}-1} \tag{64}$$

$$\sim \omega'(z)\,\omega(z)^{\frac{\alpha_x}{\alpha_x-1}\frac{1-\theta}{\theta}-1}\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{\theta}\frac{1}{\alpha_x-1}}\left[\beta+\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(1-\beta)\right]^{\frac{(1-\theta)\alpha_x}{\theta(\alpha_x-1)}}.$$

Integrating, we get

$$K-\frac{\alpha_x}{\alpha_x-1-\alpha_z}\left[\frac{\beta^{\alpha_x}\alpha_x}{\alpha_x-1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1-\theta}{\theta}\frac{1}{\alpha_x-1}} z^{-\frac{\alpha_x-1-\alpha_z}{\alpha_x-1}}$$

$$\sim \left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{\theta}\frac{1}{\alpha_x-1}}\left[\beta+\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(1-\beta)\right]^{\frac{(1-\theta)\alpha_x}{\theta(\alpha_x-1)}}\omega(z)^{\frac{1-\theta}{\theta}\frac{\alpha_x}{\alpha_x-1}},$$

for some constant K. The RHS is increasing in $\omega(z)$ and the LHS is only increasing in $\omega(z)$ if $\alpha_x-1>\alpha_z$, so that this condition needs to be met for the equilibrium to be in case 3. Assuming that this is the case, we get

$$\omega(z)\sim\Omega-\omega_0 z^{-\omega_1} \text{ with } \Omega,\omega_0>0 \text{ and } \omega_1=\frac{\alpha_x-1-\alpha_z}{\alpha_x-1}>0.$$

Therefore the price of health care quality is bounded above and so is doctors' income.

We then need to check that the SOC holds. From (64), we get

$$\omega'(z) \tag{65}$$

$$\sim \Omega^{1-\frac{\alpha}{\alpha_x-1}\frac{1-\theta}{\theta}}\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{\theta}\frac{1}{\alpha_x-1}}\left[\beta+\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}(1-\beta)\right]^{\frac{(\theta-1)\alpha_x}{\theta(\alpha_x-1)}}\left[\frac{\beta^{\alpha_x}\alpha_x}{\alpha_x-1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1-\theta}{(\alpha_x-1)\theta}} z^{\frac{\alpha_z}{\alpha_x-1}\frac{1-\theta}{\theta}-\frac{1}{\theta}}.$$

Equation (62) still holds and plugging (60) in it, we can write:

$$\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}}\frac{m(z)^{\alpha_x}z^{-\alpha_z-1}\alpha_z\lambda\mu z_{\min}^{\alpha_z}}{\left(\frac{z\omega'(z)}{\omega(z)}\right)^{\frac{\theta}{1-\theta}}\alpha_x x_{\min}^{\alpha_x}}$$

$$=\frac{1}{1-\theta}+\frac{1-2\theta}{1-\theta}\varepsilon_\omega(z)+\frac{\theta}{1-\theta}\frac{z\omega''(z)}{\omega'(z)}+\frac{(1-\beta)\,\varepsilon_\omega(z)}{\beta\left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{1-\theta}}+1-\beta+(1-\beta)\,\varepsilon_\omega(z)}\left(1+\frac{z\omega''(z)}{\omega'(z)}-\varepsilon_\omega(z)\right).$$

Using (63) and (65), we can write

$$\omega''(z)\sim\left(\frac{1-\theta}{\theta}\frac{\alpha_z}{\alpha_x-1}-\frac{1}{\theta}\right)\frac{\omega'(z)}{z}.$$

Plugging this expression and (65) in (54), we obtain

$$g(z) \sim \frac{1-\theta}{\theta} \frac{\alpha_z}{\alpha_x - 1} \frac{1}{\varepsilon_\omega (z^*)} > 0$$

which ensures that the second order condition is asymptotically satisfied.

Bringing together cases 1), 2) and 3) established Proposition 10 2) and 3).

Cobb-Douglas case ($\theta = 1$). Then, the consumer maximizes

$$\max_z H(z,h) = h^{1-\gamma} \frac{z^\gamma}{(\omega(z))^{1-\gamma}}. \tag{66}$$

The solution of this problem is independent of z, so, if not indifferent, all consumers would consume exactly the same health care quality. This cannot be an equilibrium, so consumers must be indifferent across levels of health care quality and we must have $\omega(z) \propto z^{\frac{\gamma}{1-\gamma}}$. Therefore, there exists a quality adjusted price of health care taken as given by doctors. Doctors' income is Pareto distributed with shape parameter $\alpha_z \frac{1-\gamma}{\gamma}$. □

This Proposition shows that our results can be generalized as long as the possibility to substitute quantity for quality remains limited ($\theta < 1$). Intuitively, in the Cobb-Douglas case ($\theta = 1$), it is possible to rewrite the health care aggregate as a function of quality-adjusted quantity of health care ($\widetilde{q} \equiv z^{\frac{\gamma}{1-\gamma}} q$ so that $H = \widetilde{q}^{1-\gamma}$). Then, health care has a common quality-adjusted price, and the payment schedule for doctors is log-linear in their ability. As such, doctors' income inequality is entirely determined by their ability distribution and not widget makers' income distribution.

In contrast, in the complement case, there is positive assortative matching as in the baseline model. High-income widget makers consume not only higher quality but now also more health care. With a fixed supply of health care, doctors of a given ability are matched with patients of a lower income than in the baseline model ($m(z)$ is proportional to $z^{\frac{\alpha_z+1}{\alpha_x}}$ instead of $z^{\frac{\alpha_z}{\alpha_x}}$ in the baseline). As long as top doctors are not too abundant ($\alpha_z > \alpha_x - 1$), doctors' incomes are still asymptotically Pareto distributed—though top income inequality is lower among doctors than widget makers ($\alpha_w^{-1} < \alpha_x^{-1}$). An increase in widget makers' top income inequality spills over into doctors' top income inequality more than one-for-one because the demand for quantity of health care services also increases disproportionately at the top: $\frac{d\alpha_w^{-1}}{d\alpha_x^{-1}} > 1$. Moreover, doctors' top income inequality decreases as their ability distribution becomes more unequal: $\frac{d\alpha_w^{-1}}{d\alpha_z^{-1}} < 0$, as high-quality health care becomes cheaper when it becomes more abundant. Interestingly, the elasticity of substitution θ has no effect on the size of the spillovers.

A.9 Partly tradable health care

Consider the set-up of Section 3.3.5. Without loss of generality, assume that region 1 is the most unequal; that is $\alpha_x^1 = \min_s \{\alpha_x^s\}$. Again without loss of generality, assume that $\alpha_x^s > \alpha_x^1$ for $s \neq 1$. Denote by $\kappa_s(x)$ the share of widget makers of ability x who are mobile in region s. We assume that $\lim_{x\to\infty} \kappa_1(x) = \underline{\kappa} > 0$; that is, a positive mass of patients travel in the richest region. In all regions $s \neq 1$, for x large enough, the set of potential patients with income above x will be dominated by traveling patients from region 1. Since the equilibrium still features positive assortative matching, for all doctors with z high enough in all regions, most doctors will be matched with a patient from region 1. The analysis of the baseline model (specifically Section A.3.1) applies and doctors' income in each location is asymptotically Pareto distributed with shape parameter α_x^1.

B Data Appendix

B.1 Details of data construction and Figure 1

Sample Selection. We restrict to Census/ACS respondents who are age 25 or older, and (1) have positive income and are categorized as "employed, at work" according to the variable ESR (employment status recoded) or (2) have positive income, are not in the labor force, and are age 65 or older. The latter group approximates retirees. For the positive income restriction, income refers to whichever definition is being used; typically wage income, but in robustness checks we also use all earned income. All constructed variables use Census weights (perwt).

Construction of Figure 1. The green series (with circles) shows the actual change in log income within each income percentile from 1980 to 2012. To decompose this into between- and within-occupation effects, we calculate two sets of statistics for 1980 and 2012. First, the percentiles of average occupational log wage income where occupations are weighted by occupation size. Then within each of these occupational percentile groups, we rank individuals based on their income and group them in 500 bins, and compute the deviation between log income in that bin and that occupational percentile group's average log earnings. The "Between-Occupation Effects Only" series (with red triangles) shows the counterfactual shift in the income distribution if occupational percentiles' incomes had changed to 2012 levels, but the corresponding bins of differences around each percentile (500 for each of the 100 percentiles) had remained unchanged. The "Within-Occupation Effects Only" series (with blue squares) shows the counterfactual shift in the income distribution if the occupational percentiles' wage incomes had remained constant at the 1980 level, but the bins around each percentile had changed to 2012 levels.

Independent Variable Construction in the Regression Analysis. For a given occupation of interest (*e.g.* doctors) and a given percentile cutoff that defines the upper tail of the income distribution, such as the 90th (local) percentile, we do the following:

1. Among *uncensored* observations, calculate the income at that percentile in that LMA-year among all persons regardless of occupation. Drop observations below that income level.

2. Then, drop the occupation of interest, that is, the occupation in the dependent variable.

3. Then, calculate the inverse Pareto parameter as described in the main text. We adjust equation 13 to account for a few censored observations. Consider a sample of draws of a random variable $\tilde{X}$ which follows a Pareto distribution $P(\tilde{X} > \tilde{x}) = (\tilde{x}/x_{min})^{-\alpha}$. With censoring, the observed value is $x = \min\{\tilde{x}, \bar{x}\}$ for some known censoring point $\bar{x} > x_{min}$. Denote N_{cen} and N_{unc}, the number of censored and uncensored observations, respectively, and $\mathcal{N}_{unc}$ the set of uncensored observations. The maximum likelihood estimator of α^{-1} is $\frac{1}{N_{unc}}\left[\sum_{i\in\mathcal{N}_{unc}} \ln\left(\frac{x_i}{x_{min}}\right) + N_{cen}\ln\left(\frac{\bar{x}}{x_{min}}\right)\right]$. Armour, Burkhauser and Larrimore (2016) use this method with Current Population Survey data (March supplement) to show that income inequality trends match those found by Kopczuk, Saez and Song (2010) using Social Security data.

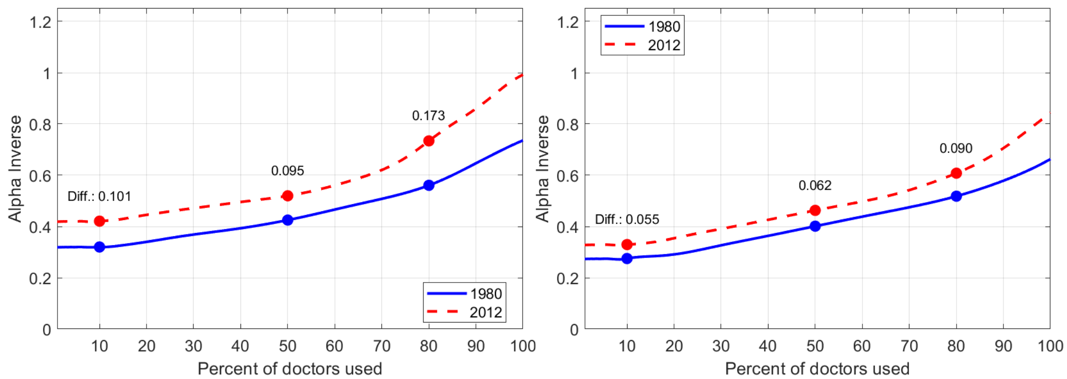
Dependent Variable Construction in the Regression Analysis. For a given occupation of interest and a given cutoff that defines the upper tail of the income distribution, such as the 90th percentile of the local general population, we do the following:

1. Repeat step 1 of the construction of the independent variable.
2. Then, keep only observations from the outcome occupation of interest (*e.g.* doctors).
3. Then, calculate the inverse Pareto parameter correcting for censoring.

Instrument Construction in the Regression Analysis. For each occupation of interest o, we construct our shift-share instrument as follows. We first identify the 10 most common occupations (excluding o) in the upper tail of each LMA in 1980, where the upper tail corresponds to the 90th percentile of the local uncensored observations. We define the set of shift-share occupations, K_{-o}, as the union of all these occupations, which corresponds to around 30 occupations. Then, for each occupation κ in K_{-o}, we calculate the income at the percentile cutoff defining the upper tail (*e.g.*, 90th) among uncensored observations *nationwide*. Observations with income below that level are dropped. Then, for each region s, we drop respondents residing in that region and calculate the nationwide inverse Pareto parameter for each occupation ($\alpha^{-1}_{\kappa,t,-s}$). We then calculate the weights (or shares) for occupation $\kappa \in K_{-o}$ in the instrument as the fraction of the upper tail population that is employed in κ in 1980. The weights are not normalized, and thus sum to less than one. Retirees are excluded from the instrument.

Measures of α^{-1} and sample selection. In our main specifications we estimate α^{-1} on the sample of the occupation that is in the top 10% of income of the general population. For the samples of New York State and nationwide we illustrate how the estimated α^{-1} for doctors depends on the size of the sample (we use New York State and not New York LMA for disclosure reasons). Our data here is the same as that used for the calibration exercise in Section 6. We defer a

Figure B.1: Estimated α^{-1} for doctors for various cutoffs.
(a) New York State **(b)** National



Notes: For New York State (Panel (a)) and the entire United States (Panel (b)) we calculate α^{-1} with an increasing fraction of doctors used (excluding those with residents level incomes). Around 80 per cent of doctors are in the top 10% of the general population in 2012. The numbers give the difference between 1980 and 2012 for the particular per cent of doctors used.

detailed discussion of data and data cleaning until then.

Figure B.1.a shows the α^{-1} estimated for 1980 and 2012 on increasing fractions of doctors in New York State once we restrict attention to doctors with incomes higher than those in medical residency. When using the highest earning 10% of doctors we get a difference between 1980 and 2012 of 0.101. When we use half the observations we get a difference of 0.09. The fraction of (non-resident) doctors who are in the top 10% of the general population is around 80% in 2012. When we use this fraction we get 0.173.[48] Panel (b) performs the same calculation for the nation as a whole with the same points highlighted.

Regression Details. Regressions are weighted by the number of observations in the outcome occupation above the cutoff in that LMA-year. Only the 50 most populous LMAs as of 1980 are included. We estimate the regressions with the Stata commands reghdfe and ivreghdfe. Standard errors are clustered at the LMA level.

B.2 Occupation and LMA definitions

Occupations. We use the occupation classification constructed by Deming (2017) which ensures consistent occupational groups throughout our sample. We create some additional groupings: We combine (1) all engineering occupations into one, (2) all managers (excluding those working in real estate) into one, (3) combine primary and secondary school teachers together, (4) respiratory, occupational, physical, speech and not-elsewhere-classified therapists into one, (5) hairdressers and barbers into one, and (6) waiters and bartenders.

Geography. We use Labor Market Areas (LMAs) defined by Tolbert and Sizer (1996). These are aggregates of the 741 Commuting Zones (CZs) popularized by Dorn (2009). Both commuting zones and labor market areas are defined based on the commuting patterns between counties.

[48]For New York State the fraction of (non-resident) doctors in the top 10% is around 90% in 1980. When we use the year-specific cutoff we get a difference of 0.13.

CZs are unrestricted in size, whereas LMAs aggregate CZs to ensure a population of at least 100,000. LMAs are constructed such that they can be driven through in a matter of a few hours, *e.g.* Los Angeles or New York. Given that our estimation strategy relies on a relatively high number of observations of a particular occupation, LMAs are a more natural choice.

B.3 Occupational characteristics

We use the Occupational Information Network (O*NET) 10.0 June 2006 release of occupation characteristics, drawing on two traits in particular: "Customer and Personal Service" from the list of "Knowledge" traits, and "Performing for or Working Directly with the Public" from the list of "Work Activities". Each trait is scored on a scale from 0 to 7 based on the "level" of skill in that trait required for the occupation. We crosswalk from O*NET-SOC codes to 2000 Census occupation groups (occ2000) using the SOC-to-occ2000 weights used by Acemoglu and Autor (2011) (available at `https://economics.mit.edu/people/faculty/david-h-autor/data-archive`). We then collapse from occ2000 to the occupation definition used in this paper (developed by David Deming (2017), which we call occ1990dd) as described in the main text. For this collapse, we take the weighted average of each O*NET trait at the `occ1990dd` level, where the weights correspond to the fraction of each occ1990dd population derived from each occ2000 category. We calculate these weights using the 2000 public-use Census sample. Finally, once we have O*NET traits at the occ1990dd level, we normalize the traits to be in percentile rankings (from 0 to 1) rather than on the 0 to 7 scale. Each occupation is assigned their percentile ranking in the occupation distribution of the trait (weighted by occupation size). These two normalized traits are the bases of Figure 3.

We use Blinder (2009)'s measure of offshorability as an (inverse) measure of the extent to which an occupation serves the local market. Based on O*NET characteristics for each occupation (using the 2006 version), he manually assigns an ordinal score between 0 and 100 for 817 occupations, with 100 being completely offshorable, and anything less than 25 being completely non-offshorable. In most instances, the O*NET occupation codes correspond one-for-one with the Standard Occupation Classification (SOC) from the US Department of Labor. In his appendix, he lists the SOC occupations scored as greater than 25. When two or more O*NET occupation codes correspond to a single SOC code, and those O*NET occupations are deemed to be substantially different in their offshorability, he keeps the O*NET occupation codes separate, as opposed to aggregating them to the SOC code level. In the few cases in which that occurs, he only reports the O*NET occupation codes scored as greater than 25. For example, Financial Managers are a single category in SOC, but are split into three categories in the O*NET classification: 11-3031.00 Financial Managers, 11-3031.01 Treasurers and Controllers, and 11-3031.02, Financial Managers, Branch or Department. Blinder only reports the one sub-type of financial manager that he considers partially offshorable, and does not list which of the three O*NET sub-types it represents, just the higher-level SOC code.

C Assortative Matching in Healthcare and Physicians' Scale

This Appendix examines the model's first prediction: positive assortative matching: We use health spending and physician price data to construct income gradients of medical spending and medical prices. In addition, we discuss the evolution of physicians' scale.

C.1 Spending and price data

We use medical spending data from the Medical Expenditure Panel Survey (MEPS), a nationally representative survey of families' health insurance coverage and medical spending conducted annually by the Agency for Healthcare Research and Quality. We use the survey waves 2010-2014 to match the five-year ACS data used in our main results. For each family in each year, we calculate total medical spending and total dental spending. For families present in the survey for multiple years, we average their annual spending and annual family income. We then take logs of spending and income to estimate the income elasticity of spending.

We measure service provider *prices* using the Colorado All-Payer Claims Data (APCD-CO), described in Clemens, Gottlieb and Molnár (2017). This data provides details on patient visits for medical care, including the service provided (a 5-digit code established by the Healthcare Common Procedure Coding System (HCPCS)) and the identity of the treating provider. It covers "the majority of covered lives in the state across commercial health insurance plans, Medicare (Fee-for-Service and Advantage), and Health First Colorado (Colorado's Medicaid program)". To match the MEPS years, we select all payments made from 2010 to 2014. Crucially, it indicates the patient's residential zip code and the amount the provider was paid for each service (whether by an insurer, directly by the patient, or both).[49]

We summarize provider prices across procedures and patients by computing markups. We estimate markups as the provider fixed effects in a regression on the insurance claims data that controls for procedure codes. Specifically, denoting $r_{g,j,i}$ as the amount paid to provider g for performing procedure j on patient i, we estimate the following regression:

$$\ln r_{g,j,i} = \varphi_g + \varphi_j + \varepsilon_{g,j,i} \tag{67}$$

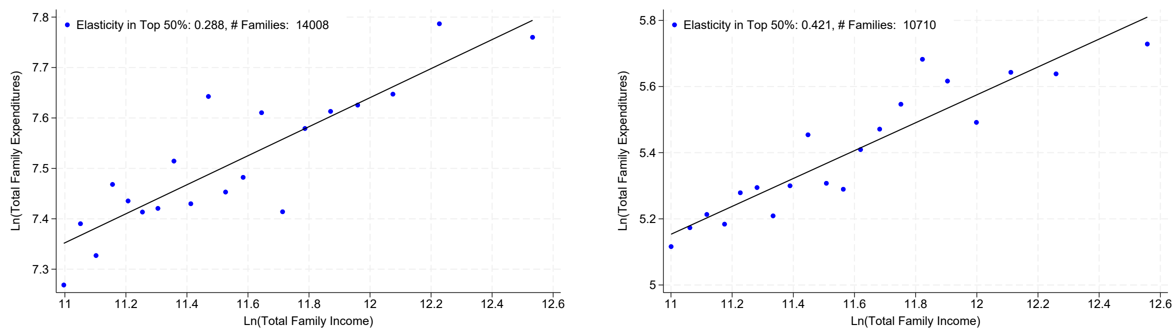
where $\hat{\varphi}_j$ are fixed effects for HCPCS procedure codes and $\hat{\varphi}_g$ are fixed effects for provider g that reflect each provider's average mark-up.

We approximate patients' income using the median family income in their residential zip

[49] Depending on the patient's insurance contract and whether the patient has reached an annual deductible or out-of-pocket maximum, the patient or the insurer may have to pay the physician's fee for a particular treatment. But regardless of who is liable, the amount that the physician expects to receive is governed by the rate negotiated between the physician and the insurer, known as the "allowed charge." We refer the reader to Clemens and Gottlieb (2017) for institutional details about this price setting.

Figure C.1: Engel Curve for Medical Spending

(a) Total Medical Spending **(b)** Dental Spending



Notes: Panel (a) shows the relationship between annual medical spending and annual income, both in logs. The dots show mean values in each vigintile of family income. The line shows a linear regression estimate on the micro-data, and reflects an elasticity of 0.288. Panel (b) shows the same figure but for dental spending only. The elasticity is 0.421. Source: Authors' calculations using data from the Medical Expenditure Panel Survey.

code.[50] We compute the mean of provider markups, $\hat{\varphi}_g$, among all provider visits from patients in a given zip code z: $\overline{\varphi_z} = \frac{1}{N_{visit}} \sum_{visit \in z} \hat{\varphi}_g$. We then estimate $\overline{\varphi_z} = \mu_0 + \mu_1 \ln(\text{income})_z + \varepsilon_z$ while weighting observations by the number of underlying patient claims in z.

C.2 Results: medical spending, physician prices, and income

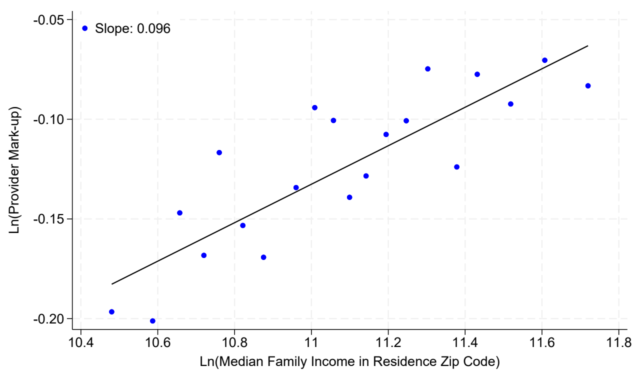
Figure C.1 shows Engel curves of total spending using MEPS data among the top 50% of the family income distribution. Panel (a) shows all spending and Panel (b) shows dental spending, as these are the two spending variables reported in the MEPS survey. The graph shows a binned scatter plot, using 20 vigintiles of family income and the regression line computed on the micro data. The positive relationship reflects an elasticity of 0.288 for total spending. That is, a 10% increase in family income is associated with 2.88% more medical spending. An elasticity less than 1 is consistent with the CES case described in Proposition 3.

Figure C.2 shows Engel curves for medical provider "prices" using the APCD-CO data described above. The first panel shows results using all medical providers, while subsequent panels show curves for different types of providers: physicians, dentists, physical and occupational therapists, and psychologists. The elasticity among all providers combined is 0.096. The gradients are statistically significant and positive for each occupation individually, with the highest gradient among therapists at 0.189. The estimate of the price elasticity is likely to be biased down because we use median family income in a zip code whereas an unbiased estimate would require the average of log income—or, better yet, a link to exact family incomes. This introduces a downward bias when zip codes vary in their local log income inequality. The positive relationship is evident throughout the income distribution and not solely at the top. These results, based on a cross-section, support our assortative matching prediction.

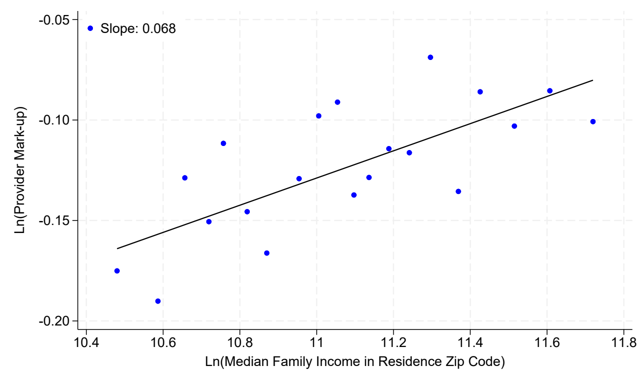
[50] We obtain data on median family income in each Zip Code Tabulation Area (areas that closely approximate zip codes) from IPUMS NHGIS (Manson et al., 2017).

Figure C.2: Engel Curves for Provider Prices

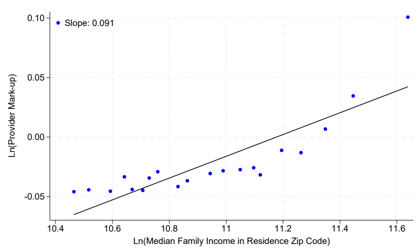
(a) All Providers



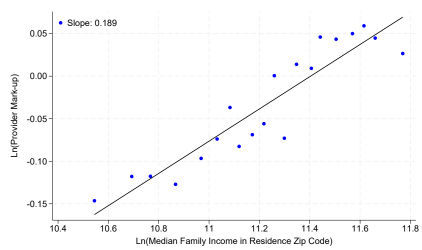
(b) Physicians



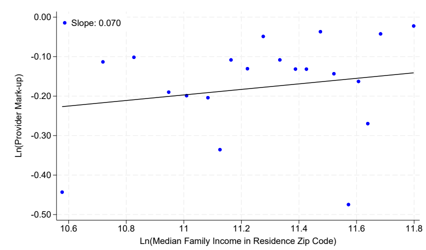
(c) Dentists



(d) Therapists



(e) Psychologists



Notes: Each panel plots the relationship between log mark ups (normalized to mean zero across physicians) charged by the medical provider who treats a patient, and family income (as proxied by the median income in the patient's zip code of residence). Calculation of the mark-up is described in the text above. The figures show a binned scatter plot with 20 vigintiles and the slope of the regression line is the elasticity (shown in each figure). Panel (a) shows the relationship for all medical providers; Panel (b) for physicians; Panel (c) for dentists; Panel (d) for therapists; and Panel (e) for psychologists.

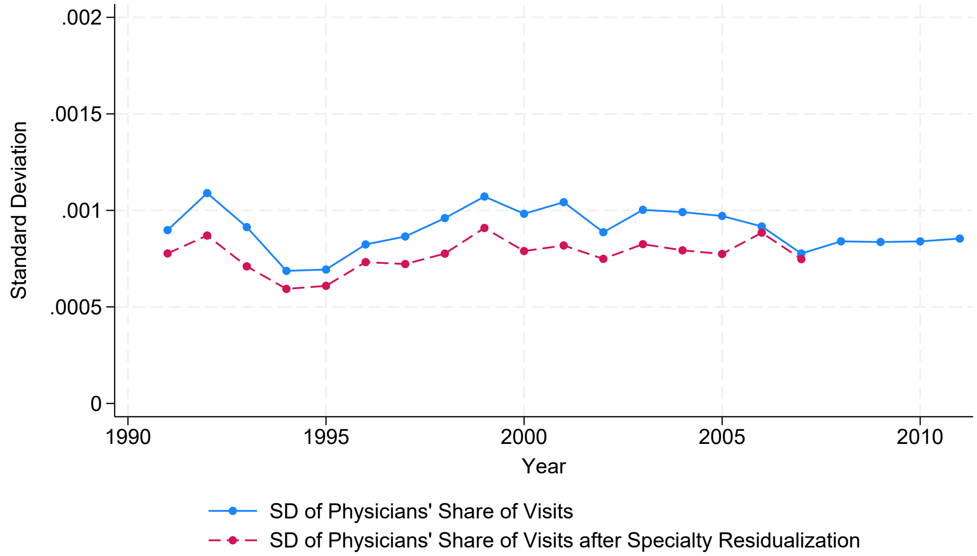
C.3 Some facts about physician scale

When taking our theory to the data, we primarily attribute changes in physicians' income inequality to changing inequality in the prices they charge—with a potential change in inequality of physician scale (*i.e.* the quantity of patients per physician) resulting from the change in prices. We cannot empirically test this assumption because the Census/ACS data do not contain information on the quantity of care provided by physicians.

To provide some context on whether the physician scale (λ in our model) may have changed differently across doctors, we access data from the National Ambulatory Medical Care Survey (NAMCS). NAMCS is a sample of office-based physicians. Physicians are surveyed for a week and report detailed information on their patient visits. For example, in 2000 the NAMCS had 1,200 reporting physicians and recorded 27,369 visits to those physicians during the sampling week. We use this information to examine time trends in the variance of per-physician quantity of care.

In the first step of the survey, physician j estimates their visit volume V_j for the upcoming survey week. We do not observe V_j. Based on V_j, the surveyors then tell the physicians what

Figure C.3: Standard Deviation in Physicians' Share of Total Patient Visits.



Notes: This figure plots the standard deviation in the share of NAMCS-surveyed visits each practitioner accounts for. If all practitioners saw the same number of visits, this would be 0. The dashed series first controls for practitioner specialty and then calculates the standard deviation of the residual. Specialty coding changed in 2008, so the series stops in 2007.

fraction of their visits to record in the survey, F_j, which is a decreasing function of V_j. After the survey is conducted, the surveyor constructs patient visit sampling weights $W_{i(j)}$ (where a visit is indexed by i) that sum to the total estimated nationwide visits. Based on this, in a given year, for physician j we calculate $\hat{V}_j = \sum_i W_{i(j)}$ and treat this as the quantity of medical care provided. The share of all sampled visits performed by physician j is $s_j = \hat{V}_j / \left(\sum_j \sum_i W_{i(j)}\right)$.

Figure C.3 shows the standard deviation of s_j in each year from 1992 to 2011. (The survey sampling and enumeration method changed in 2012). This illustrates whether, on a nationwide basis, some physicians have captured larger shares of overall patient visits. The blue line shows the raw standard deviation and the red line the standard deviation after controlling for practitioner specialty. In both cases, the variance was fairly stable from 1991 to 2011, suggesting that there were no large secular trends in the ability of some doctors to capture disproportionate shares of the market.

D Empirical Appendix

D.1 Additional tables of descriptive statistics

Table D.1 shows the ratio of the 98^{th} to 90^{th} percentiles for selected occupations in 1980 and 2012, as well as for the overall population. Table D.2 gives descriptive statistics for the largest occupations in the top of the income distribution for the year 2000.

D.2 Additional empirical results

Figure D.1 represents the IV results of Table 3 in two binned scatter plots. In Panel D.1a, we show the relationship between the shift-share instrument and non-physician inequality, *i.e.* the first stage regression. Both of these are residuals based on regressions with year and LMA

Table D.1: Ratio of 98^{th} to 90^{th} percentile of wage income for selected occupations

	98th to 90th percentile ratio		
Occupation	1980	2012	Change
Aerospace engineers	1.37	1.46	0.09
Chief executives and public administrators	1.63	2.42	0.80
Dentists	1.54	1.74	0.20
Financial managers	1.62	2.38	0.75
Financial service sales occupations	1.79	2.81	1.02
Lawyers	1.89	2.31	0.42
Managers and administrators, n.e.c.	1.90	1.80	-0.11
Physicians	1.50	1.72	0.23
Primary school teachers	1.26	1.33	0.07
Real estate sales occupations	1.94	2.17	0.23
Registered nurses	1.29	1.48	0.20
All occupations combined	1.70	1.99	0.29

Notes: The ratio of wage income at 98th percentile of the income distribution to wage income at the 90th percentile, for selected occupations. The sample consists of employed workers with positive wage income. Source: Authors' calculations using Decennial Census and American Community Survey data

Table D.2: Descriptive Statistics for Top Occupations in 2000. Wage Income

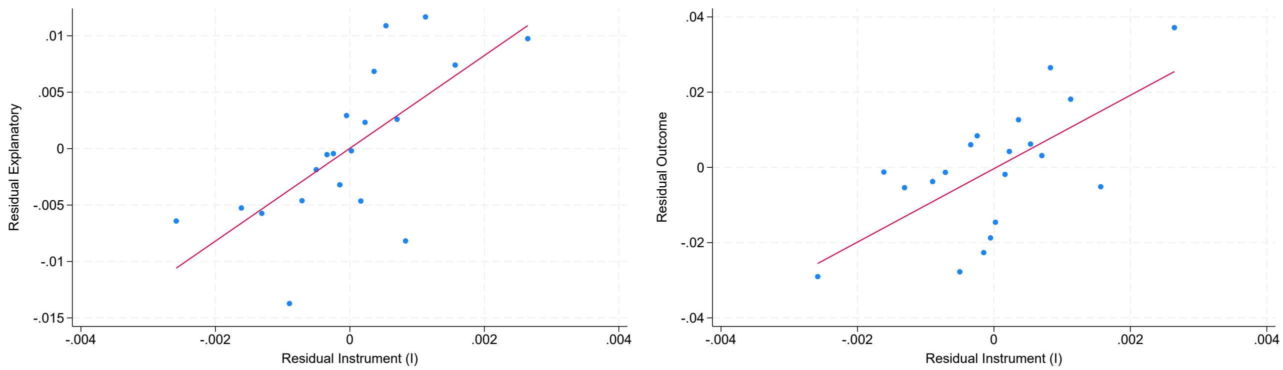
	Mean income	Occupation's share in:		
Occupation	$1000	top 10%	top 5%	top 1%
Managers excl. real estate	61	0.23	0.24	0.18
Chief executives and general administrators, public administration	120	0.06	0.09	0.15
Engineers	63	0.05	0.04	0.01
Computer systems analysts and scientists	57	0.05	0.04	0.02
Lawyers and judges	98	0.04	0.05	0.07
Physicians	136	0.04	0.07	0.13
Sales workers, other commodities	53	0.04	0.04	0.03
Supervisors and proprietors, sales occupations	45	0.04	0.04	0.04
Financial managers	68	0.03	0.03	0.03
Other financial officers	61	0.02	0.02	0.03
Accountants and auditors	47	0.02	0.02	0.02
Postsecondary teachers	43	0.02	0.01	0.00
Securities and financial services sales occupations	102	0.01	0.02	0.04
Computer programmers	57	0.01	0.01	0.00
Real estate sales occupations	53	0.01	0.01	0.02
Supervisors, production occupations	43	0.01	0.01	0.01
Registered nurses	40	0.01	0.01	0.00
Supervisors, general office	37	0.01	0.01	0.01
Teachers, elm., prim., second.	35	0.01	0.00	0.00
Sales workers	29	0.01	0.02	0.01

Notes: For the top twenty occupations in the top ten percent of the national income distribution in 2000, column (1) reports mean income from wage (for the whole population), where the (very few) censored values have been replaced with the state-level mean income among those above the censoring point, and the final three columns show the occupation's share of all earners in the top ten, five, and one percent of the income distribution. Source: Authors' calculations using Decennial Census.

Figure D.1: Graphical Representation of the IV regression: residual top income inequality.

(a) (First Stage) Top income inequality vs. instrument

(b) (Reduced Form) Physician top income inequality vs. instrument



Notes: In Panel (a), we show the relationship between the shift-share instrument and non-physician inequality after residualizing the controls, *i.e.* the first stage regression. Panel (b) shows the relationship between the instrument and physicians' inequality, *i.e.* the reduced form regression. Both are residualized based on regressions with the controls of column (6) in Table 3. For disclosure reasons, we bin our LMA × year observations into 20 equal-frequency bins and plot the average value of the residuals in each.

dummies and the controls of column (6) in Table 3. For disclosure reasons, we bin our LMA × year observations into 20 equal-frequency bins. We plot the average value of the residuals in each bin. Panel D.1b shows the relationship between the instrument and physicians' inequality, *i.e.* the reduced form regression. In both cases, we see strong upward-sloping relationships. The results are not driven by outliers.

Table D.3 shows the regressions results associated with the scatter plots of Figure 3 and also considers the "importance" of customer service and working with the public.

Table D.3: Spillover t-stats and occupational characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Entirely Not Offshoreable	-0.785					
	(0.400)					
Offshore (Blinder)		-0.928				
		(0.510)				
Customer service - level			2.116			
			(0.736)			
Customer service - importance				1.324		
				(0.734)		
Working with public - level					1.098	
					(0.675)	
Working with public - importance						1.520
						(0.656)
Constant	1.359	1.332	-0.121	0.300	0.397	0.129
	(0.338)	(0.331)	(0.408)	(0.401)	(0.370)	(0.370)
Observations	30	30	30	30	30	30

Notes: This table shows the relationship between the t-stat of the spillover coefficients from the IV regressions (from Table 6) and five characteristics of the 30 most common occupations in the top 10% These characteristics are: a measure of offshorability from Blinder (2009) as well as four measures from O*NET: Level and importance of "Customer service and personal service" from Knowledge Requirements and level and importance of "Performing for or working directly with the public" from Work Activities. O*NET and Offshoreability measures are rescaled as percentiles. "Entirely Not Offshoreable" indicates that Blinder (2009) categorized the occupation as effectively impossible to offshore—18 of the 30 occupations our list. *: p<0.1, **: p<0.05, ***: p<0.01.

Table D.4: Earned Income

	Doctors		Dentists		Real Estate Agents	
	OLS	IV	OLS	IV	OLS	IV
α_{-o}^{-1}	0.23	1.73	0.41	2.16	0.71	0.94
	(0.22)	(0.57)	(0.36)	(0.86)	(0.15)	(0.50)
Ln(Average Income)	-0.30	-0.43	0.06	-0.06	-0.03	-0.05
	(0.11)	(0.11)	(0.14)	(0.13)	(0.06)	(0.06)
Ln(Population)	-0.02	0.03	0.04	0.09	0.02	0.02
	(0.03)	(0.03)	(0.04)	(0.07)	(0.03)	(0.03)
N	200	200	200	200	200	200
F-Statistic		13.38		13.35		7.064

Notes: The table contains OLS and IV estimates exactly as in the baseline specification, except the outcome inequality measure is constructed using earned income (instead of wage and salary income)

We next show the results discussed in Section 5.3. Tables D.4, D.7, D.8, D.9, D.10, and D.11 are sufficiently discussed in the text and do not need additional description here.

Table D.5 show robustness checks for Physicians and Dentists. Columns (1)-(3) consider variations of the baseline regressions on physicians: column (1) restricts to physicians who are at least 35, column (2) to those who have not moved within the past 5 years (other mobility questions are not available throughout our sample), and column (3) controls for specialty composition. We use data from the Area Resource File on the composition of specialties across LMAs (we use numbers from 1985 for year 1980) and data from the Medical Group Management Association (2009) on average and standard deviation of income by specialty in 2008. We build four control variables: (1) the share of neurosurgeons, who are the specialty with the highest mean income and the largest standard deviation; (2) the share of physicians in the 8 highest earning specialties (excluding neurosurgeons); (3) the share of physicians in the 7 specialties with the lowest income; and (4) the share of physicians in the 4 specialties with the largest standard deviation in income (excluding neurosurgeons). The rationale behind (2) and (3) is that these specialties share similar average incomes while the next specialty (down or up) in the ranking has a substantially different average income. Column (4) looks at the baseline regression for dentists and adds a control for the ratio of dental hygienists to dentists in the LMA-year to proxy for the increased specialization of dentists and the potential resulting increase in their operating scale.

Table D.6 looks at potential entry effects for physicians, dentists and real estate agents. We denote by E_0/E the employment share of outcome occupation o (*e.g.* physicians) in a given year and LMA. Columns (1), (3), and (5) use E_o/E as the dependent variable in our otherwise identical baseline IV regression. The coefficient on physicians, 0.028, implies that a 1 point increase in α_{-o}^{-1} increases the share of doctors by 2.8 percentage points. In columns (2), (4) and (6) we include this variable as a control to our baseline inequality spillover IV regression. The coefficient of interest for all three occupations only change marginally but we lose a bit of precision for real estate agents (the p-value is still lower than 0.11).

Table D.5: Robustness Checks for Physicians and Dentists

	Physicians			Dentists
	Age 35+	No Recent Move	Specialty Controls	Specialization Control
α_{-o}^{-1}	3.07	3.32	2.01	2.13
	(1.37)	(1.38)	(0.72)	(0.87)
Ln(Avg. Income)	-0.11	-0.05	-0.49	0.02
	(0.22)	(0.22)	(0.09)	(0.14)
Ln(Population)	0.03	0.02	0.00	0.08
	(0.10)	(0.12)	(0.04)	(0.07)
Sh. neurosurgeons			-1.43	
			(5.14)	
Sh. high earning specialties			3.19	
			(1.20)	
Sh. low earning specialties			0.11	
			(0.47)	
Sh. unequal earning specialties			2.12	
			(3.82)	
# hygienists / # dentists				0.04
				(0.03)
N	200	200	200	200
F-Statistic	12.26	12.62	11.03	13.75

Notes: The table contains robustness checks on the baseline IV estimates. The first three columns are for physicians, and the last column is for dentists. Column (1) restricts the sample of physicians to those aged 35 or older. Column (2) restricts the sample of physicians to those who have not recently moved (within the past 5 years). Column (3) uses the baseline sample of physicians but adds four controls for specialties of physicians (see text for details). Column (4) (dentists) controls for the ratio of the number of hygienists to the number of dentists in the area, as a measure of task specialization

Table D.6: Controlling and Testing for Entry Effects

	Physicians		Dentists		Real Estate Agents	
Dependent Variable:	$\frac{E_o}{E}$	α_o^{-1}	$\frac{E_o}{E}$	α_o^{-1}	$\frac{E_o}{E}$	α_o^{-1}
α_{-o}^{-1}	0.028	2.11	0.003	2.44	0.116	1.61
	(0.013)	(0.66)	(0.002)	(0.94)	(0.043)	(0.99)
Ln(Avg. Income)	-0.001	-0.47	-0.001	-0.03	-0.002	-0.01
	(0.002)	(0.12)	(0.000)	(0.15)	(0.008)	(0.09)
Ln(Pop.)	-0.001	0.04	0.000	0.09	-0.002	0.03
	(0.001)	(0.04)	(0.000)	(0.08)	(0.003)	(0.03)
$\frac{E_o}{E}$		6.52		-62.78		0.83
		(11.41)		(49.24)		(4.75)
N	200	200	200	200	200	200
F-Statistic	13.21	14.22	13.36	12.96	7.124	6.285

Notes: E_o is the number of people working in outcome occupation o (*e.g.* physicians) in a given year and LMA. E is the total employed population in the same year and LMA. When E_o/E is included as an independent variable, its purpose is to check that inequality spillovers are not driven by entry. When E_o/E is the outcome variable, we are testing whether the instrument-induced variation in local inequality causes entry. All results are IV estimates with the same specification described for our baseline results

Table D.7: Controlling for the occupational share in the top 10%

Main Occupations			
	Physicians	Dentists	Real Estate Agents
α_{-o}^{-1}	2.27	2.27	1.70
	(0.63)	(0.93)	(0.67)
Outcome Occupation Region-Specific Percentile	0.43	0.67	-0.08
	(0.61)	(0.55)	(1.02)
Ln(Avg. Income)	-0.47	-0.03	-0.01
	(0.12)	(0.14)	(0.09)
Ln(Population)	0.03	0.07	0.02
	(0.04)	(0.08)	(0.03)
N	200	200	200
F-Statistic	12.78	13.33	6.15

Placebo Occupations			
	Financial Managers	Managers Excl. Real Estate	Engineers
α_{-o}^{-1}	0.62	-0.03	-0.10
	(0.78)	(0.15)	(0.30)
Outcome Occupation Region-Specific Percentile	4.50	-0.11	-0.23
	(1.23)	(0.67)	(0.41)
Ln(Avg. Income)	0.17	0.06	-0.02
	(0.11)	(0.03)	(0.05)
Ln(Population)	-0.10	0.03	0.05
	(0.05)	(0.02)	(0.02)
N	200	200	200
F-Statistic	12.64	16.15	28.91

Notes: IV regressions for selected occupations with and addition control for "Outcome Occupation Region-Specific Percentile" which is the fraction of workers in the outcome occupation whose income is in the top 10% of the overall LMA-year income distribution. This fraction will be greater than 10% when the outcome occupation is disproportionately high income (such as physicians)

Table D.8: Using 10% of doctors for dependent variable

	50 LMAs (Baseline)		30 LMAs	
	(1)	(2)	(3)	(4)
α_{-o}^{-1}	1.03	1.00	1.24	1.23
	(0.55)	(0.64)	(0.58)	(0.64)
Ln(Avg. Income)		0.01		-0.01
		(0.12)		(0.12)
Ln(Population)		-0.01		-0.01
		(0.04)		(0.05)
N	200	200	100	100
F Statistic	13.36	13.11	14.61	12.63

Notes: This table shows OLS and IV regression results for physicians using a different approach to calculating physician income inequality. Rather than taking all physicians with income above the LMAs general 90th income percentile, we take all physicians above the physician-specific LMA 90th percentile income. Due to physicians high average earnings, this method results in calculating income inequality using far fewer doctors. N is the number of observations rounded to the nearest integer divisible by 50 as required by disclosure rules.

Table D.9: Robustness to Cutoffs

	Panel (a): Physicians				
	Top 5 pct.	70 LMAs	30 LMAs	Using 13 Occs	Using 7 Occs
α_{-o}^{-1}	2.31	2.16	1.74	2.28	2.28
	(0.95)	(0.87)	(0.51)	(0.64)	(1.37)
Ln(Average Income)	-0.60	-0.49	-0.40	-0.47	-0.48
	(0.19)	(0.10)	(0.13)	(0.12)	(0.14)
Ln(Population)	0.07	0.04	0.02	0.03	0.04
	(0.07)	(0.04)	(0.03)	(0.04)	(0.06)
N	200	300	100	200	200
F Statistic	5.32	9.37	12.34	11.17	1.89
	Panel (b): Dentists				
	Top 5 pct.	70 LMAs	30 LMAs	Using 13 Occs	Using 7 Occs
α_{-o}^{-1}	1.33	2.15	2.32	2.27	2.95
	(0.83)	(0.98)	(1.14)	(0.86)	(1.42)
Ln(Average Income)	-0.18	-0.08	0.03	0.01	-0.07
	(0.21)	(0.13)	(0.15)	(0.13)	(0.15)
Ln(Population)	0.07	0.03	0.09	0.08	0.10
	(0.08)	(0.07)	(0.10)	(0.07)	(0.10)
N	200	300	100	200	200
F Statistic	8.72	10.46	9.16	13.14	4.46
	Panel (c): Real Estate Agents				
	Top 5 pct.	70 LMAs	30 LMAs	Using 13 Occs	Using 7 Occs
α_{-o}^{-1}	0.64	1.49	1.22	1.32	2.43
	(0.44)	(0.64)	(0.55)	(0.59)	(1.15)
Ln(Average Income)	0.09	0.03	0.02	0.04	-0.08
	(0.12)	(0.08)	(0.08)	(0.08)	(0.14)
Ln(Population)	0.00	0.02	0.02	0.02	0.03
	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)
N	200	300	100	200	200
F Statistic	18.37	7.96	5.63	8.22	4.58

Notes: This table shows the IV regressions for physicians (Panel (a)), dentists (Panel (b)), and real estate agents (Panel (c)) for 5 different specifications. Column (1) uses the top 5% (instead of 10%) of the income distribution (for the dependent variable, the independent variable, and the IV). Columns (2) and (3) use 70 and 30 LMAs respectively, instead of 50. Columns (4) and (5) construct the instrument differently: rather than finding the top 10 occupations in each LMA in 1980 then taking the union, column (4) uses the top 13 before taking the union while column (5) uses the top 7 before taking the union. N is the number of observations rounded to the nearest integer divisible by 50 as required by disclosure rules.

Table D.10: Instrument Occupations with Largest Rotemberg Weights

Occupation	Weights	Occupation	Weights
Financial service sales occupations	0.35	Geologists	0.06
Financial managers	0.22	Accountants and auditors	0.06
Airplane pilots and navigators	0.21	Primary/Secondary School Teachers	0.05
Other financial specialists	0.16	Economists, market and survey researchers	-0.05
Sales occupations and sales representatives	0.10	Office supervisors	-0.05
Production supervisors or foremen	0.09	Computer systems analysts and computer scientists	-0.05
Lawyers and judges	0.07	Managers, Excl. Real Estate	-0.07
Driver/sales workers and truck Drivers	0.06	Engineers	-0.13

Notes: Occupations with the sixteen largest Rotemberg weights in absolute value in the instrument for Physician regressions.

Table D.11: Physician regressions: excluding occupations with the highest Rotemberg weights

	Outcome Occupation: Physicians				
Excl. from Instrument:	Financial Managers	Other financial Specialists	Engineers	Airplane pilots and navigators	Financial service sales occupations
α_{-o}^{-1}	2.57	2.39	2.36	2.64	2.54
	(0.80)	(0.73)	(0.60)	(0.84)	(0.92)
Ln(Average Income)	-0.50	-0.48	-0.47	-0.50	-0.49
	(0.14)	(0.13)	(0.13)	(0.14)	(0.13)
Ln(Population)	0.04	0.03	0.03	0.04	0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
N	200	200	200	200	200
F Statistic	9.75	12.66	15.45	12.47	8.57

Notes: This table shows the baseline IV regressions for physicians when we in turn exclude the 5 occupations with the highest Rotemberg weights from the instrument (Rotemberg weights are shown in Table D.10). The column heading denotes which occupation is excluded from the instrument set.

While we follow Goldsmith-Pinkham et al. (2020) and assume that identification in our exercise comes from the exogeneity of the shares (the occupation weights), an alternative identification strategy in shift-share instruments relies on the exogeneity of the shifts (here nationwide occupational inequality). In that case, Adão et al. (2019) emphasize that correlated errors for (in our case) LMAs with similar occupational compositions can lead researchers to reject the null too often, and suggest a formula for standard errors that take this issue into account. In Table D.12, we report their standard errors for our three focal occupations and our three main placebo occupations. For our focal occupations, standards errors only increase by a bit, and the spillover coefficients remain significant (the p-value for real estate agents with controls is 5.1%). Table D.12 also reports the confidence intervals from Lee et al. (2023)which are valid under weak instruments, shown in curly brackets.

E Calibration Appendix

This appendix provides a detailed account of the calibration exercise in Section 6.

E.1 Data

Disclosure restrictions allow us to export 50 bins from each of four distributions in New York State: physicians and consumers (non-physicians) for 1980 and 2012. Each bin represents the average log income within 2 percentiles (in common \$2000 using the CPI). The underlying sample is the same as used for the regression. The consumer data has a rather long left tail and we drop the bottom 10 per cent (those below \$7,080 and \$8,791 in \$2000, for 1980 and 2012, respectively). We fit a kernel to the bottom 90% of each distribution, and impose a Pareto distribution on the top 10, where α^{-1} is estimated using the top of the income distribution. We then choose the scale parameter of the Pareto distribution to ensure ensure a continuous CDF.[51] Panels (b) and (d) in Figure E.1 show histograms drawn from the fitted kernel, with the

[51]Our procedure does not require the CDF to be differentiable in the first point of the Pareto distribution, though due to the good fit of a Pareto distribution in the top it is close to. We confirm that applying the censoring procedure on data drawn from our CDF recreates binned observations that are indistinguishable from

Table D.12: Alternative Inference Approaches for IV Estimator

	Main Occupations					
	Physicians		Dentists		Real Estate Agents	
α_{-o}^{-1}	1.54	2.29	2.29	2.27	1.69	1.71
	(0.74)	(0.63)	(0.71)	(0.93)	(0.62)	(0.65)
	[0.75]	[0.98]	[0.74]	[0.84]	[0.67]	[0.87]
	{0.28, 3.30}	{1.23, 3.72}	{1.03, 3.70}	{0.71, 4.47}	{0.65, 2.92}	{0.64, 3.23}
Ln(Avg. Income)		-0.47		0.00		-0.01
		(0.13)		(0.14)		(0.09)
Ln(Population)		0.03		0.08		0.02
		(0.04)		(0.08)		(0.03)
N	200	200	200	200	200	200
F-Statistic	14.60	13.21	23.63	13.36	12.72	7.12
	Placebo Occupations					
	Financial Managers		Managers		Engineers	
α_{-o}^{-1}	1.12	0.77	0.05	-0.02	-0.13	-0.09
	(0.97)	(0.92)	(0.12)	(0.15)	(0.26)	(0.30)
	[2.34]	[2.03]	[0.17]	[0.15]	[0.51]	[0.43]
	{-0.77, 2.83}	{-1.47, 2.29}	{-0.17, 0.26}	{-0.31, 0.24}	{-0.60, 0.34}	{-0.63, 0.45}
Ln(Avg. Income)		0.29		0.06		-0.02
		(0.16)		(0.03)		(0.05)
Ln(Population)		-0.16		0.03		0.05
		(0.07)		(0.01)		(0.02)
N	200	200	200	200	200	200
F-Statistic	19.92	11.22	22.15	16.49	23.97	27.91

Notes: IV regressions for selected occupations, with additional inference approaches. The first column for each outcome occupation is the IV regression without population and average income controls. The second column adds these controls. The standard errors in parentheses are from standard cluster-robust inference. The standard errors in squared brackets are Adão et al. (2019) standard errors. The curly brackets contain the 95% confidence bounds from Lee et al (2023).

lightly shaded part representing the 10% of observations not included. The figure also shows histograms drawn from the kernels of the doctors. These underlying data is bimodal reflecting a sizable mass of doctors who are still in their medical residencies. Given that wages for medical residents are not driven by consumers' demand for physician skill (Chandra, Khullar, Wilensky, 2014) we exclude them from the analysis, by dropping observations with values lower than the $90th$ percentile of doctors aged 35 and younger ($\$52,052$ and $\$56,387$ in \$2000, respectively). As can be seen from the histograms this is practically equivalent to only including observations after the first "peak" of the distributions. For the analysis we scale all income by the lowest value of consumers' income in 1980, though this figure is prior to scaling.

E.2 Calibration

We calibrate our model for exogenously given λ and ε as described in the main text. Though the underlying ability distribution of potential doctors is given by $F_z(z)$, we can only calibrate the conditional distribution for active doctors, $\hat{F}_z(z)$. We normalize the lowest active doctor to have ability 1. We parameterize this distribution using 12 points along the distribution and interpolating between them.[52] With the tail value α_z and β this gives us 14 parameters

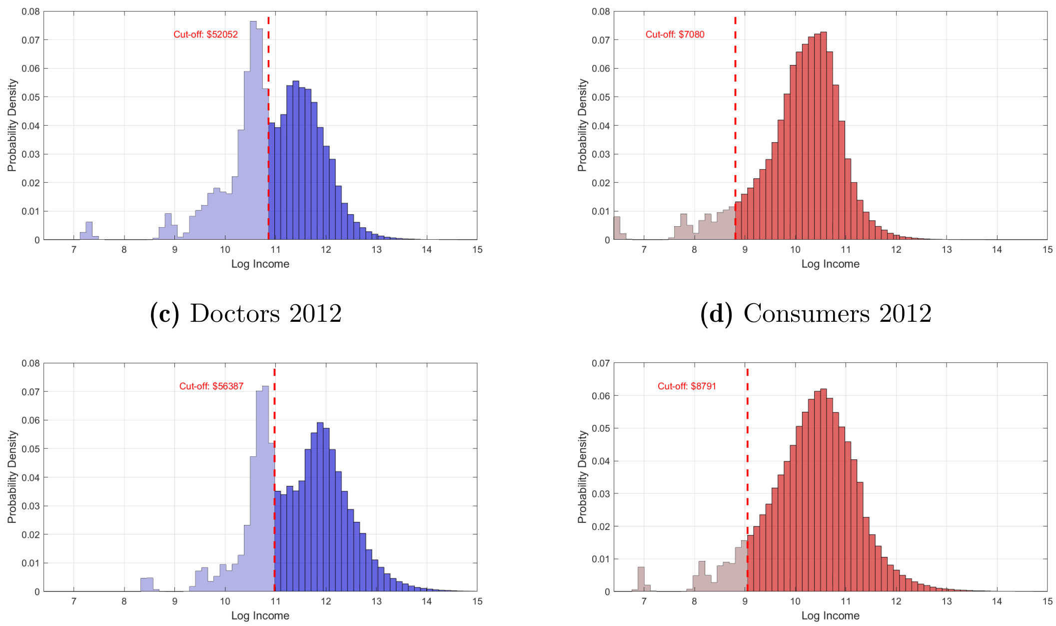
the actual data (not shown).

[52]Formally, we consider 12 points equally spaced along the CDF of active doctors and calibrate the underlying values of z at each of these points. All other points are linearly interpolated over $\{\ln(z), \ln(1 - CDF)\}$ which is linear for a standard Pareto. The top 10 is parameterized using a Pareto distribution with tail α_z.

Figure E.1: Histogram drawn from the CDFs of consumers and doctors

(a) Doctors 1980 **(b)** Consumers 1980

(c) Doctors 2012 **(d)** Consumers 2012



Notes: Histograms are in $2000 dollars and are drawn from the best fitting kernel (with a Pareto tail) on all 50 bins for each population. Lightly shaded areas are observations excluded from the analysis. The cutoff represent the lowest value included. For consumers it is based on bottom 10% percent. For doctors it is based on the 90th percentile of doctors aged 35 and younger.

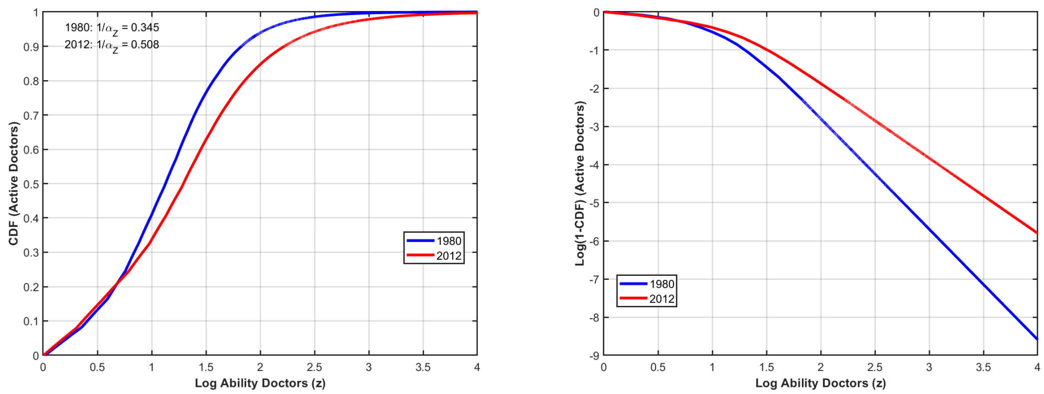
to calibrate for 2012. We minimize the squared deviation of log wages between the empirical kernel and the predicted wage distribution from the model. This gives us the parameters in Table 8. The calibrated $\hat{F}_z(z)$ is shown in Appendix Figure E.2.a. For this figure we also ask the following question: By how much should the ability distribution have changed between 1980 and 2012 to exactly fit the wage distribution in both 1980 and 2012, holding the other parameters, λ, β, and ε constant. The result is shown in the same panel. For the top 70% the distribution shows a clear shift to the right, as well as a higher α_z^{-1} in 2012. This increase in ability inequality is commensurable with the change in α^{-1} for the consumers. Panel (b) shows the same figure in a Pareto plot where the Pareto distribution would imply a straight line.

The measure of α^{-1}. Throughout most of the paper we calculate α^{-1} for an occupation of interest o based on the observations of this occupation in the top of the 10% of the general population. In the following we show how α^{-1} and in particular the change in α^{-1} between 1980 and 2012 depends on the size of the sample chosen. This is analogous to the analysis on the empirical distribution of Appendix Figure B.1, where here we utilize the model-predicted distribution. For reference, slightly more than 80% of (non-resident) doctors are in the top 10% of the overall distribution in 2012, and the difference at this cutoff is 0.193. For 10% of the doctors it is 0.215. These numbers deliver the spillover coefficients of the main text.

Just like for the empirical distribution, the estimate of α^{-1} is somewhat sensitive to the cutoff, but the difference between the two—the identification needed for the analysis—is re-

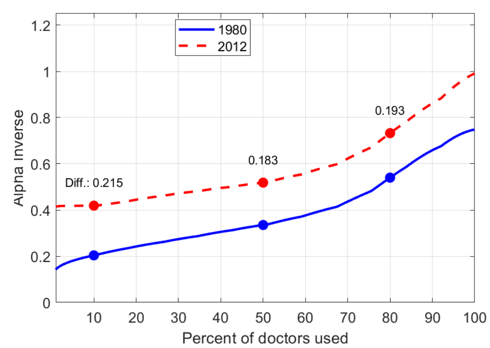
Figure E.2: Calibrated doctor ability distribution

(a) CDF **(b)** Pareto plot



Notes: Panel (a) shows the calibrated ability distribution of the doctors. The 2012 values are used in the analysis in Section 6. The 1980 values are calibrated analogously except the β value is kept constant at the 2012 value. Panel (b) shows the same two figures in a Pareto plot ($\ln(z)$ against $\ln(1 - CDF)$)

Figure E.3: Estimated α^{-1} for doctors for various cutoffs - model predicted distribution



Notes: Calculated α^{-1} on different percentiles of the doctors' distribution (with increasing fraction towards the right) for both 1980 and 2012. The numbers give the difference between 1980 and 2012 at selected samples of doctors used.

markably constant.

Welfare and spillovers. For the analyses behind Table 9, we consider two different scenarios, both based on the same distribution of spillover occupations in 2012, calibrated to fit the doctors' distribution in that year. For the spillover exercise, we consider the model-predicted change to the spillover wages underlying Figure 5. For the counter-factual mean shift we calculate the mean of log income for both 1980 and 2012, and we shift the 2012 distribution down by this mean, preserving the shape. For the exercise in Table 9 we calculate the various inequality measures in the table for 2012 and the two 1980 distributions and consider differences.

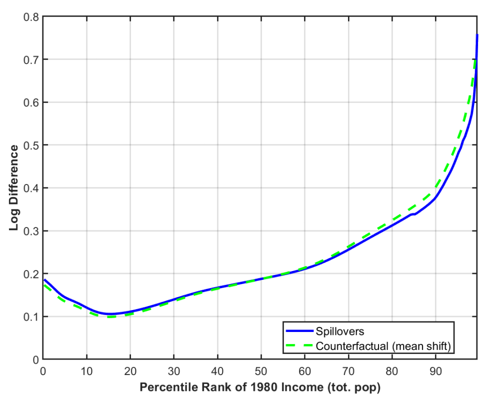
For the exercise comparing EV and income change in Figure 6 we consider only the spillover scenario and compare the income change deflated by the same CPI to the EV calculated based on equation 18. We then plot the difference between EV and income change, both between 1980 and 2012.

Appendix Figure E.4 combines the two analyses, by again considering the two scenarios (the spillovers and the mean (log) shift) and compares the EV of the full distribution. For all consumers we rank them in 1980 and calculate the EV required to get them to the utility of the equivalent percentile of consumers for each of the two scenarios in 2012 (based on equation 18). For all spillover occupations, we calculate the EV required to get them to the same percentile of the spillover occupation in 2012 (equivalent to the income increase due to the utility function). We then take the combined distribution in 1980 and for each percentile we calculate the average EV within that percentile. For the bottom 80% that is only consumers, but for the top 20% that combines both consumers and the spillover occupations.

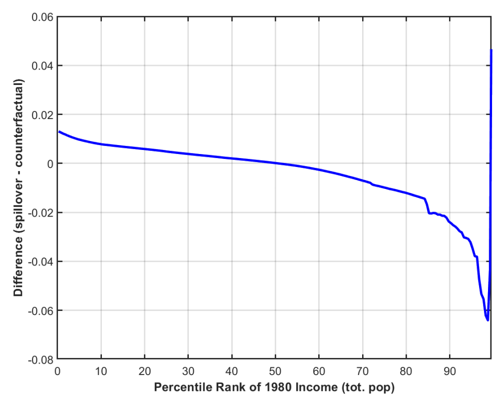
For lower-income consumers, the mean-shift scenario implies higher increases in medical service costs, implying that they benefit from the spillover scenario. The reverse is true for higher-income consumers, who pay higher prices with spillovers. However, because spillover occupations are concentrated at the top of the income distribution, most consumers use services provided by those who earn more than themselves. Consequently, for the highest earners (top 1%), the spillover scenario features higher welfare gains than the mean-shift scenario, as shown more clearly in Panel (b) which shows the difference between the two.[53]

Figure E.4: Differences in welfare for full income distribution (with and without spillovers)

(a) Welfare differences - spillovers and counterfactual



(b) Difference between spillovers and counterfactual



Notes: The Figure considers the full income distribution (consumers plus spillover occupations) and shows the EV welfare measure. Two scenarios are considered: the 'Spillovers' scenario from the baseline model and a counterfactual scenario of the same (log) mean shift of the spillover occupations. Panel (b) shows the difference between the two.

Additional References

Acemoglu, Daron and David Autor, "Skills, tasks and technologies: Implications for employment and earnings," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4 Part B, Elsevier, 2011, pp. 1043–1171.

Chandra, Amitabh; Khullar, Dhruv and Wilensky, Gail, "The Economics of Graduate Medical Education," *The New England Journal of Medicine*, vol 370 issue 25

[53]Given the fatter Pareto tail for the consumers than for the spillover occupations, asymptotically consumers buy the services from those who make less than themselves and the two curves cross again. For these parameters that happens well within the top 0.1%.