The Spillover Effects of Top Income Inequality*

Joshua D. Gottlieb   David Hémous
Jeffrey Hicks   Morten Olsen†

June 2023

Abstract

Top income inequality in the United States has increased considerably within 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 to consumers, with physicians our prime example. Examining local income inequality across U.S. regions, the data suggest that such spillovers exist for physicians, dentists, and real estate agents. Estimated spillovers for other occupations are consistent with the predictions of our theory.

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. Morten 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. David 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, 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 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 (Jones and Kim, 2014) and within occupations (Bakija, Cole, and Heim, 2012). At first glance, this pattern suggests that any explanation for rising inequality—whether globalization, deregulation, changes to the tax structure or technology—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 a few 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 certain service providers. This occurs when the services provided are *heterogeneous* in quality and *non-divisible*—*i.e.*, consumers cannot 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 increases in a region's top income inequality spill over into top income inequality among physicians, dentists, and real estate agents, with a spillover elasticity of 1.2 to 1.5. In contrast, we do not observe such spillovers for occupations that do not meet the model's requirements (such as engineers and financial managers). Using a broader set of occupations and characteristics of occupations, we show that the sizes of spillovers are consistent with the model's predictions.

Our analysis begins in section 2 by documenting that the increase in top income inequality is driven primarily by an increase in within-occupation top income inequality. 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 the services of *doctors* of heterogeneous abilities, who 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' abilities are

Pareto distributed, the incomes of both widget makers and doctors are also Pareto distributed. 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.

Non-divisibility in output is necessary for this assignment mechanism to emerge: if medical services were divisible and doctors were to differ in their quality-adjusted quantity of medical service provided, then any change in the income distribution of widget makers would only translate into a change in the price of a unit of quality-adjusted medical service—with no consequence for doctors' inequality. Non-scalability is not necessary; our results generalize to the case when doctors have a positive (though not infinite) supply elasticity.

Our baseline model deliberately focuses on local consumption spillovers by considering a single economy and abstracting 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 the local top income inequality.[1]

In support of our assortative matching mechanism, section 4 presents empirical evidence on how health care spending and physician prices relate to household income. Using a nationally representative survey and detailed medical claims data from one state, we find that patients earning 10% more spend 4.4% more on medical expenditures, with a substantial part of that elasticity reflecting higher prices.

Section 5 introduces our empirical analysis of inequality spillovers in local U.S. labor markets. We use Census and American Community Survey data from 1980 to 2014 to build a panel of labor market areas (LMAs) (aggregates of commuting zones) and conduct our analysis at this level. Guided by our model, we measure top income inequality as the inverse Pareto coefficient 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.

[1] Dingel et al. (2023) show that in 2017 over three-quarters of medical care is consumed in the same Hospital Referral Region (roughly the same size as the Labor Market Areas (LMAs) used in the present paper) where it is produced and that, while there is some trade across regions, the home market drives the pattern of that trade. This local share was even higher before 2017.

An OLS regression would 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 (measured with the inverse Pareto coefficient). The weights correspond to the relative importance of each occupation in each LMA at the beginning of our sample (1980). 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). That is, we assume the occupational shares are not correlated 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 6, we focus on three high-earning occupations satisfying these criteria: physicians, dentists, and real estate agents. We find positive spillover coefficients, with elasticities in the range of 1.2 to 1.5. The parameter estimates suggest that most of the increase in inequality for these occupations can be explained by increases in others' income inequality.

Finally, we estimate spillover coefficients for an additional 25 occupations common in the top 10% of the income distribution. These occupations do not fit the requirements of our theory, and we find significant local inequality spillovers for only one of the 25. We then relate the estimated coefficients to occupational characteristics. In line with our theory, the spillover coefficient is 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 offshorability, our proxy for the tradability of a service.

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). This literature has established that at the top, the income distribution is well-described by a Pareto distribution (see Guvenen, Karahan, Ozkan, and Song, 2021, for recent evidence and Pareto, 1896, for the earliest). Jones and Kim (2014) show that the increase in top income inequality specifically reflects a fattening of the right tail of the income distribution, which corresponds to a decrease in the shape parameter of the Pareto distribution.

More specifically, our paper builds on the "superstars" literature originating with 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 arises from a fixed cost in consumption per unit of quantity. This 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] In that framework, 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, if $w(z)$ denotes the income of an individual of talent z, $p(z)$ denotes the average price for her services, and $q(z)$ is the quantity provided, so $w(z) = p(z)q(z)$, the standard interpretation of "superstars" is that they have very large markets (a high $q(z)$). This makes such a framework poorly suited for occupations where output is not easily scalable, such as doctors, dentists, and real estate agents.

In contrast, we focus on such occupations and study an assignment model that is "constant-to-one" where superstars are characterized by a high price $p(z)$ for their services. This makes our paper 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. These papers use multiplicative production functions in CEO skill and firm productivity, *i.e.* Cobb-Douglas, whereas we consider a consumption problem and extend beyond the Cobb-Douglas setting. Along the same lines, Määttänen and Terviö (2014) build an assignment model for housing. They calibrate their model to six U.S. metropolitan areas and find that the increase in inequality has led to an increase in house price dispersion (see also Landvoigt, Piazzesi and Schneider, 2015).[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]König (2021) provides empirical evidence for entertainers using the roll-out of television.

[4]Määttänen and Terviö (2014) consider an assignment model with CES preferences over housing and other goods. They show that a mean-preserving spread in income will lower the price of lower-priced houses and might increase the prices of more expensive ones.

technological change affecting firm size and thereby executive pay.[5] Globalization is a natural candidate and Bonfiglioli, Crino and Gancia (2018) show how trade costs can affect firm size distribution and top income inequality.[6] Many other explanations exist: Jones and Kim (2018) and Aghion, Akcigit, Bergeaud, Blundell and Hémous (2019) look at the role played by innovation; Piketty, Saez and Stantcheva (2014) argue that low marginal income tax rates divert managers' compensation from perks to wages and increase their incentives to bargain for higher wages; Philippon and Reshef (2012) emphasize the role played by the financial sector; while Edmond and Mongey (2021) argue that as high-skill occupations become more specialized, workers with a comparative advantage in the skills intensively used by one of these occupations earn larger rents, leading to an increase in within-occupation income inequality. 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.

Finally, our paper relates to a literature on demand spillovers. Among others, Manning (2004) and Mazzolari and Ragusa (2013) relate the polarization of labor markets to an increase in the demand of low-skill services by high-skill workers. More closely related to our work, Leonardi (2015) argues that in addition, high-skill workers also demand relatively more services by high-skill workers, a pattern that can amplify increases in the skill premium.[7] Importantly, our focus is not on the skill premium but on changes in inequality within the top of the income distribution.

2 The rise of within-occupation top income inequality

We motivate our analysis by showing the importance of within-occupation changes for the trends in top wage income inequality. Among workers with positive labor income, the ratio of incomes at the 98^{th} to 90^{th} percentile rose from 1.7 to 2.0 from

[5]For instance, Geerolf (2017) builds a span of control model to micro-found the fact that firms' size distribution is Pareto. An increase in the number of layers in the firm is associated with a fatter tail. His model naturally leads to "superstar" effects for managers and a bounded distribution of talents can lead to an unbounded income distribution.

[6]In their model, a larger market incentivizes firms to invest in riskier but bigger projects at the entry stage, resulting in a change in the productivity distribution. Other papers (Manasse and Turini, 2001, Kukharskyy, 2012; Gesbach and Schmutzel, 2014 and Ma and Ruzic, 2020) relate globalization with top income inequality without featuring changes in the Pareto shape parameter of the income distribution.

[7]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.

1980 to 2012.[8] Physicians' ratio increased from 1.5 to 1.8 during the same period, with similar increases for two other occupations on which we focus—dentists and real estate agents.

To understand the role of occupations more systematically, Figure 1 decomposes overall changes in wage income from 1980 to 2012 into within- and between-occupation components. The green series in Panel 1a shows the overall change in log wage income at each percentile of the income distribution. This reproduces the well-known fact that incomes have grown the fastest in the top of the income distribution during this time period. We then adapt the within- and between-firm decomposition of Song, Price, Guvenen, Bloom and von Wachter (2019) but use occupations instead of firms.[9] The green series shows that income at the 99^{th} percentile (the average of log income of 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.[10]

Since we focus on changes in income inequality, we next examine how the within-occupation changes drive relative gains at different points in the distribution. Consider first the log of the ratio of income at the 99^{th} to 90^{th} percentile. Panel 1a shows that income at these percentiles rose by 0.56 and 0.32 log points, respectively (green series), so the difference increased by 0.24. Within-occupation factors contributed 0.17 ($= 0.24 - 0.07$) log points of that (blue series), so around three quarters of the total change in the 99^{th} to 90^{th} ratio is attributable to the increase in within-occupation income inequality. Panel 1b presents an analogous decomposition for four additional ratios of income inequality. We see that between two-thirds and three-quarters of the increase in top income inequality is attributable to changes in within-occupation income inequality.

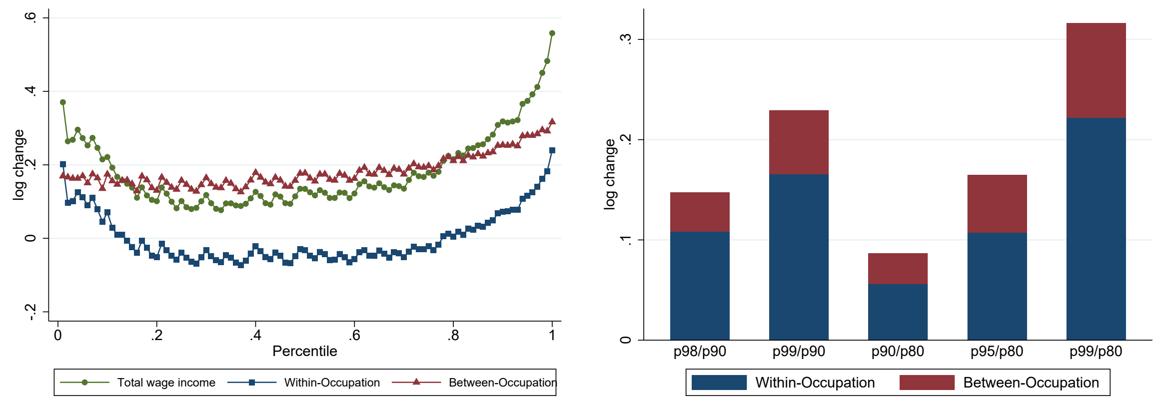
[8] Throughout the paper, we rely on the decennial census (1980, 1990, 2000) and the American Community Survey (average of 2010-2014, which we refer to as 2012). Appendix Table B.1 shows the corresponding changes for a selected set of occupations. Details on the data are in section 5 and Appendix B.1. Panel 1b shows that the 99^{th} and 98^{th} percentiles show similar trends; disclosure rules make it easier to present the latter when looking within individual occupations.

[9] To find the effect of within-occupation changes, we hold the average log wage income for occupations 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.

[10] 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.

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 (blue line). Following the method of Song et al. (2019), it further decomposes this wage change into changes attributable to between-occupation changes (green line) and within-occupation changes (red line). Panel (b) then shows the contributions of between- and within- occupation effects for selected measures of top income inequality. p99/p90 refers to the ratio of the top 1 percent to the top 10%.

3 Theory

To understand these patterns, this section builds a model of inequality spillovers across occupations. Section 3.1 builds an assignment model between doctors and their patients. Section 3.2 relaxes several assumptions, including allowing for occupational and geographical mobility. Section 3.3 summarizes our empirical predictions.

3.1 The Baseline Model

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

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 such that a widget maker has ability $X > x$ with probability $P(X > x) = \left(\frac{x_{\min}}{x}\right)^{\alpha_x}$, with lower bound $x_{\min}$ and Pareto parameter $\alpha_x > 1$. The Pareto parameter, α_x, is an (inverse) measure of the spread of abilities. We treat as α_x as exogenous throughout and a fall in α_x captures a general increase in top income inequality. 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 $\lambda > \mu_d^{-1}$ 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 3.2.1 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 is represented by the Cobb-Douglas utility function:

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

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 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.[11]

We start with the Cobb-Douglas utility function for ease of exposition but investigate more general utility functions in Section 3.2.3. For simplicity, doctors only consume widgets, so the doctors' patients are exclusively widget makers. This assumption is not essential and is relaxed in Appendix A.5.

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

[11] Although we refer to the "quality" of the good, nothing in our model relies on the "high-quality" goods being objectively superior. It is merely "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 other measures (*e.g.*, Dingel et al., 2023).

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 the quality of medical care and the quantity of the homogeneous good consumed gives:

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

With Cobb-Douglas preferences, no widget maker spends all her income on health care, so equation (2) immediately implies that $\omega(z)$ must be increasing: doctors of higher ability earn more per patient. Importantly, the non-divisibility of medical services implies that doctors are "local monopolists" in that they are in direct competition only with doctors of slightly higher or lower ability. Therefore, doctors do not take prices as given and $\omega(z)$ need not be a linear function of 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 $m(z)$ the matching function: 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, the least able doctors will choose to work as widget makers rather than practicing physicians. Let z_c be the ability level of the least able practicing doctor. Thus $m(z)$ is defined over $[z_c, \infty)$ and $m(z_c) = x_{\min}$, where the worst doctor is hired by a patient with income $x_{\min}$. Market clearing at all quality levels then implies:

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

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

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

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

We let $w(z)$ 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 and in the homogeneous good sector earning x_{min}, we must have $w(z_c) = x_{\min}$. Combining this with the matching function, (2), we obtain a differential equation the wage function $w(z)$ must satisfy:

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

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:

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

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

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

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, consider a counter-factual equilibrium with a linear price schedule, $\omega(z) \propto z$. Widget makers would then be spending a rising share of their income on medical services since medical services are abundant in the top. 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, bringing down prices, and resulting in a concave payment schedule for doctors. The non-divisibility of medical services is essential for this result. If a high-earning widget maker could simply substitute the services of one doctor of ability z with two doctors of ability $z/2$, a linear pricing schedule would be the equilibrium. Conversely, the payment schedule would be convex if talent were relatively scarce among doctors.[12] We can thus derive:

Proposition 1 (*Spillovers*)**.** *Doctor 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.*

To see this result, we first define the relevant distribution. Among the set of practicing doctors (potential doctors who work as doctors), let $P_{doc}(W_d > w_d)$ be the probability that income exceeds w. By definition, this is $P_{doc}(W_d > w_d) = \left(\frac{w^{-1}(w_d)}{z_c}\right)^{-\alpha_z}$. 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, the shape parameter is inherited from the widget makers, and is independent of the spread of doctor ability, α_z. In particular, a decrease in α_x directly translates into a decrease in the Pareto parameter for doctors' income distribution: an increase in inequality among widget makers leads to an increase in inequality among doctors. In other words, the increase in top income inequality spills over from one occupation (the widget makers) to another (doctors). At the top it also increases the income of doctors—as a decrease in α_x leads to an increase in $P(W_d > w_d)$ for w_d high enough.[13]

The inequality in skills for the doctors is irrelevant for the Pareto parameter: a decline in α_z—an increase in ability inequality—would imply relatively more high-ability doctors, but their relative pay would correspondingly decline, leaving the overall level of pay inequality unaffected. Further, a decrease in the mass of potential doctors μ_d (equivalent to an increase in the mass of widget makers) does not affect inequality among doctors at the top. But it does increase the share of doctors who are active (z_c decreases) and their wages (as $w(z)$ increases if z_c decreases).

Our results directly generalize to the case where patients' income and doctors' ability distribution are only asymptotically Pareto distributed and where potential doctors may (but need not) also consume medical services (see details in Appendix

[12]Appendix D.1 in the Supplementary Material (available at `http://www.gottlieb.ca`) presents a graphical representation of the model.

[13]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.

A.5). 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.6). Finally, in Appendix A.7, we allow doctors to increase scale at some cost. Intuitively, more elastic supply from each doctor increases the supply of healthcare quality, especially at the top. This reduces the pass-through from widget makers' inequality into price inequality and thus into doctors' income inequality—despite the increase in high-quality healthcare production. As long as the supply elasticity is finite,[14] doctors' income is still asymptotically Pareto distributed with a shape parameter increasing in α_x, though now different from α_x. The size of the spillover coefficient decreases with each doctor's supply elasticity of health care services.

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

3.1.2 Consumer implications

This result has important implications for key outcomes on the consumer side: health expenditures and welfare inequality.

Health expenditures. Using (4) and (7), we compute that a widget-maker with income x has log health care expenditures asymptotically given by:

$$\ln h\left(x\right) \approx \ln x + \ln\left(\frac{\beta\alpha_x}{\alpha_z\left(1-\beta\right)+\beta\alpha_x}\right). \tag{9}$$

That is, the spending share on health care is asymptotically constant. As we show below, this depends crucially on the Cobb-Douglas assumption.

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

[14] Estimates of doctors' supply elasticity are all finite, generally in a range around 1 (Clemens and Gottlieb, 2014) or less than 1 (e.g., Gottlieb et al., 2020) depending on the level at which they are measured. There is no contention that supply is perfectly elastic.

pay more for health services of the same quality. This mechanism limits the welfare increase in inequality.[15]

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 (Welfare inequality)**.** *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).*

3.2 Extensions

The baseline model makes several important assumptions about 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 first change assumptions that are not essential for our results: we allow for geographical and occupational mobility for doctors and we consider a more general CES utility function. We then demonstrate that two assumptions necessary for local inequality spillovers are non-tradability and non-divisibility of medical services.

3.2.1 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. We

[15] Moretti's (2013) work on real wage inequality across cities can be viewed as proposing a similar assignment mechanism causing high earners to locate in high-cost cities. Diamond (2016) argues that the amenities of expensive cities are more valuable to the high earners who choose to live there, so we should not fully adjust incomes for these high costs when calculating welfare. In our context, this critique would apply if high-income widget makers had stronger preferences for high-quality doctors than low-income widget makers.

keep the model as before, except there is a mass of 1 of agents with a uni-dimensional (Pareto) distribution of skills who decide whether to be doctors or widget makers.

Appendix A.3 shows that Proposition 1 still applies; that is, the income distribution of doctors is 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). Therefore the models with and without occupational mobility are observationally equivalent.

Supply- versus demand-side effects. In this model with occupational mobility, doctors and widget makers interact through both a demand effect—widget makers are the clients of doctors—and a labor supply effect—doctors can choose to become widget makers. 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 Supplementary Material available on our webpage,[16] we split the role of widget makers into two: patients, who only serve as consumers of doctors' services, and an "outside option" which serves only to provide doctors with an alternative occupation to medicine. We show that, with Cobb-Douglas preferences, doctors' income inequality is entirely driven by their patients' inequality and is independent of the outside option's inequality.[17] Consequently, the driving force is still the demand side.

3.2.2 Geographical mobility

Returning to the baseline model of section 3.1, we now assume that there are 2 regions of equal size, A and B, and that doctors can move across regions. Medical services are non-tradable and patients cannot move. The two regions are identical except for the ability distribution of widget makers, which is Pareto in both but with different shape parameters. Region A is more unequal than region B; $\alpha_x^A < \alpha_x^B$. Doctors' ability is Pareto distributed with parameter α_z in both regions.[18]

[16] See Appendix D.2 in the Supplementary Material available at `http://www.gottlieb.ca`

[17] 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.

[18] Our results would generalize to a case with multiple regions and heterogeneous masses of potential doctors and widget makers. In contrast, 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

We compare autarky to an equilibrium where the only interaction between regions is the movement of doctors. In autarky, the baseline model of section 3.1 captures each region's equilibrium: doctors' income is asymptotically Pareto distributed with shape parameters given by the local income distribution for widget makers. In contrast, when doctors can move, the equilibrium pay schedule $\omega(z)$ must be the same in both regions; observed inequality can differ only because of changes to the ability distribution. Most rich patients are in region A, where the income distribution has a fatter tail. As doctors' income increases with the incomes of their patients, some high-ability doctors from region B will move to A—and a larger share of higher-ability doctors. Since the original share of doctors and ability distributions are the same in both regions, net migration of doctors is balanced. Therefore, lower-ability doctors move from A to B, and inequality in ability will be higher in A than in B. In equilibrium, the resulting ability distributions are both asymptotically Pareto. We then obtain (proof in Appendix A.4):

Proposition 3. *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.*

In the baseline model, doctors' income in region A is asymptotically Pareto distributed with a shape parameter of α_x^A, and similarly with parameter α_x^B in region B. Consequently, the *observed* income distribution among doctors is equivalent in autarky and with doctors' mobility. However, while differences in doctors' income distribution between A and B originate from differences in the pay scale under autarky, these differences result from different ex-post ability distributions with doctors' mobility. Consequently, our empirical analysis does not require us to take a stand on whether doctors are mobile.

3.2.3 Generalizing the utility function

Our results obtain in a far more general case than the Cobb-Douglas utility assumed above. In Appendix A.5.3, we show that they generalize to any homothetic utility function that admits positive and finite limits to the elasticity of substitution both when $\frac{z}{c}$ tends to infinity and when $\frac{z}{c}$ tends to 0. For simplicity, we focus here on the

spatial equilibrium model to generate empirical predictions. See Dingel et al. (2023) for more on this topic, also emphasizing the role of demand in driving spatial patterns of healthcare delivery.

case when preferences have a constant elasticity of substitution (CES). That is, we replace the utility function of equation (1) with:

$$u(z,c) = \left(\beta_z z^{\frac{\varepsilon-1}{\varepsilon}} + \beta_c c^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{10}$$

where ε is the elasticity of substitution between physician quality and the homogeneous good. With CES preferences, the equilibrium still features positive assortative matching. In Appendix A.5.2, we show:

Proposition 4. *If either (i) $\varepsilon > 1$ and $\alpha_x \leq \alpha_z$, or (ii) $\varepsilon < 1$ and $\alpha_z \leq \alpha_x < \frac{\alpha_z}{1-\varepsilon}$, then wages of doctors are asymptotically Pareto distributed with shape parameter $\alpha_w = \frac{\alpha_z}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}$, which is increasing in the widget maker shape parameter α_x. In addition, log health expenditures grow proportionately with log income: $\ln h(x) \approx \left[\left(1-\frac{\alpha_x}{\alpha_z}\right)\frac{1}{\varepsilon}+\frac{\alpha_x}{\alpha_z}\right]\ln x + \eta$, where η is a constant.*

Proposition 4 restricts attention to two cases. We argue below that, outside of the Cobb-Douglas case ($\varepsilon = 1$), these are the two empirically relevant cases. In both cases, doctors' income is asymptotically Pareto distributed. Although the distribution's shape parameter differs from that of widget makers when $\alpha_z \neq \alpha_x$, doctors' top income inequality continues to increase with widget makers' top income inequality. Health care expenditures increase less than proportionately with income for $\alpha_z \neq \alpha_x$.[19,20]

To understand the results of Proposition 4 intuitively, and why we restrict attention to a specific set of parameters, consider the two cases in turn. First, let medical services and the outside good be substitutes ($\varepsilon > 1$) and let doctors skill be relatively scarce in the top ($\alpha_x < \alpha_z$). 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

[19] In alternative parameter spaces, the equilibria are: If $\varepsilon < 1$ and $\alpha_x \geq \frac{\alpha_z}{1-\varepsilon}$, doctors' income is not Pareto distributed and the slope of the Engel curve is horizontal for high incomes. That is $\frac{\ln h(x)}{\ln x} \to 0$, which contradicts the results in section 4. When $(\varepsilon-1)(\alpha_x-\alpha_z) > 0$, widget-makers asymptotically spend all their income on health care and doctors' income is Pareto distributed with shape parameter α_x. See Appendix A.5.2, which solves for the equilibrium in all cases.

[20] The proposition focuses again on the baseline case with neither geographical nor occupational mobility. It is however possible to derive similar results in the CES case when allowing for either type of mobility. With occupational mobility, one difference from the Cobb-Douglas case is that with CES, part of the effect of a rise in income inequality on doctors' income inequality arises from the outside option effect instead of purely the demand side.

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 linear case, and the Pareto coefficient for doctors' income distribution, α_w, is larger than α_x—but still increasing in α_x.

When $\varepsilon < 1$ and $\alpha_x > \alpha_z$, widget-makers would increase their demand for health care services, asymptotically spending nearly all their income on health care, which is counterfactual (and therefore not considered in Proposition 4).

The other case considered in Proposition 4 is when medical services and other goods are complements $\varepsilon < 1$ and when doctors are relatively abundant in the top: $\alpha_x > \alpha_z$. 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 $\alpha_w > \alpha_x$ provided that doctors are not too abundant at the top ($\alpha_x < \frac{\alpha_z}{1-\varepsilon}$).

Our empirical analysis suggests that the CES case is more relevant than Cobb-Douglas. First, in section 4, we find that the slope of the Engel curve is less than 1, in line with Proposition 4 and in contrast with equation (9) under Cobb-Douglas. Second, we will estimate spillover coefficients above 1 (though generally not statistically significantly above 1). Log-differentiating the expression for α_w with respect to both α_z and α_x, yields:

$$\frac{d\alpha_w}{\alpha_w}=\underbrace{\frac{\frac{\alpha_z}{\alpha_x}\frac{1}{\varepsilon}}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}}_{\text{Term 1}}\frac{d\alpha_x}{\alpha_x}+\underbrace{\frac{1-\frac{1}{\varepsilon}}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}}_{\text{Term 2}}\frac{d\alpha_z}{\alpha_z}. \tag{11}$$

This expression reflects how income inequality among doctors responds to changes in inequality among widget makers $\left(\frac{d\alpha_x}{\alpha_x}\right)$ and in doctors' ability z $\left(\frac{d\alpha_z}{\alpha_z}\right)$. For $\varepsilon = 1$, the expression becomes $\frac{d\alpha_w}{\alpha_w}=\frac{d\alpha_x}{\alpha_x}$ and we recover the expression from Proposition 1.

In the complement case, $\varepsilon < 1$ (and $\alpha_z \leq \alpha_x < \frac{\alpha_z}{1-\varepsilon}$), the spillover effect from income inequality of widget makers on doctors' income inequality, Term 1, is greater than one. In addition, Term 2 is negative: a growing spread in doctors' ability (a decrease in α_z) *reduces* doctors' income inequality (α_w increases). This is because as α_z decreases, more top doctors compete for patients who are spending a declining share of their income on health care as we move into the tail. The reverse holds in the substitute case ($\varepsilon > 1$ and $\alpha_x \leq \alpha_z$), where Term 1 is below 1 and Term 2 is positive.

Empirically, α_z and α_x are likely to be positively correlated: 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 should suffer from a downward bias when the true coefficient is above 1—consistent with our findings in section 6.1.

3.2.4 The role of non-divisibility and non-tradability

We next highlight two assumptions that are necessary for our result of local spillovers: the non-divisibility and the non-tradability of the service.

First, to highlight the role of non-divisibility, consider the baseline model and contrast health care services with a divisible consumption good, "beer". That is, the quality-adjusted quantity of beer enters the utility function. Beer is produced by potential brewers, who, like potential doctors, have Pareto distributed ability with shape parameter α_y. The quality-adjusted quantity of beer produced by a brewer is proportional to their ability. Beer will therefore have a common quality-adjusted price p, and the payment schedule for brewers will be linear in their ability. As such, brewer income inequality is entirely determined by their ability distribution and not the income distribution of widget makers.

Second, we permit trade in medical services across regions. Since our empirical analysis will rely 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 the lowest α_x^s, that of the most unequal region.

The cost of 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 in all regions must asymptotically have income inequality shape parameter of $\min_s \alpha_x^s$ (see details in Appendix A.8). That is, there are national spillovers, but no local spillovers.[21]

[21] Formally, we show in a model with a continuum of agents, that the income distribution of

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. 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.

3.3 Empirical predictions

To summarize, our model makes the following predictions (where "doctors" represent all occupations that fit the assumptions): 1) High-earning patients 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; 3) This is true regardless of whether doctors can move across regions, and regardless of whether doctors' ability is positively correlated with the income they would receive in alternative occupations; and 4) If patients can travel easily, doctors' income in each region does not depend on local income inequality, but on national inequality. The remainder of this paper presents empirical tests of these predictions.

4 Assortative Matching with Health Spending

We begin by examining the model's first prediction: assortative matching. We use a nationally representative survey and medical claims data to measure the income gradient of medical spending. Our mechanism requires not only that high-income patients spend more on medical care but also that they visit higher-priced physicians. We therefore measure physician prices using the method described below, and present the income gradient of physician prices.

4.1 Institutional background and measurement of physician prices

The medical industry in the U.S. is not perfectly described by the flexible price-setting model of section 3.1. The government plays a substantial role through Medicare

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, naturally, one would not expect national spillovers when only a small fraction of patients can travel.

and Medicaid, the insurance sector has an important role as an intermediary, there is substantial information asymmetry between patients and doctors, and patients sometimes travel for medical care. So, at first glance, our model might not seem applicable to this industry. But these institutional intricacies need not inhibit market forces—including our spillovers—from operating. In fact, they may offer a mechanism that implements the forces our model discusses.

Although the government sets prices for those whose care it pays for directly, providers' negotiations with private insurers generally lead to higher prices in the private market (Clemens and Gottlieb, 2017). Even in the presence of 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). And, although patients occasionally travel for care, a patient in Dallas is vastly more likely to seek medical care in Dallas than in Boston (Dingel, Gottlieb, Lozinski and Mourot, 2023).[22] Therefore, despite these complications, the structure of the health insurance industry may embody enough flexibility to incorporate the economic pressures implied by our model.

We summarize physician prices by computing their markups over Medicare rates. That is, if Medicare sets a reimbursement rate of r_j^M for treatment j, Clemens, Gottlieb and Molnár (2017) show that private insurer i's reimbursement to physician group g for that treatment is often determined by $r_{i,g,j} = \varphi_{i,g} r_j^M$, where $\varphi_{i,g}$ is an insurer-physician group constant, and these markups reflect economic pressures such as physician market power. The markups can thus be used as a summary measure of medical prices. Following Clemens et al. (2017), we estimate these markups as the physician fixed effects in a regression on insurance claims data—data that record insurers' payments to provider groups for specific treatments. Specifically, we run the following regression at the level of each treatment j and physician g:

$$\ln r_{g,j} = \varphi_g + \ln r_j^M + \varepsilon_{g,j} \tag{12}$$

and interpret the estimates, $\hat{\varphi}_g$, as physician g's (log) markup over Medicare.[23]

[22] Furthermore, our empirical strategy more heavily weights large metropolitan areas, which are more likely to have a full portfolio of medical specialties available, implying less need to travel. Regardless, our theoretical analysis suggests that national travel would reduce our estimated spillovers.

[23] In the interest of brevity, we refer the reader to Clemens, Gottlieb and Molnár (2017) or Clemens and Gottlieb (2017) for more institutional details about this price setting in the physician context, and Ho (2009) or Gaynor and Town (2011) for the hospital context. Cooper et al. (2019) apply this

4.2 Data on medical spending and physician prices

Medical spending data. We first measure overall health care spending using data from the Medical Expenditure Panel Survey (MEPS). MEPS is a detailed, nationally representative survey of families' health insurance coverage and medical spending. The survey is conducted annually by the Agency for Healthcare Research and Quality and collects information about specific medical expenditures and their costs. We aggregate medical spending in 2015 to the family, the level at which income data are collected, and consider those with incomes above \$50,000 ($N = 5,436$). We take logs of medical spending and family income to estimate the income elasticity of spending.

Health insurance claims data. We measure Engel curves of physician *prices* using the Colorado All-Payer Claims Data (APCD-CO), described in Clemens, Gottlieb and Molnár (2017). This dataset provides details on patient visits for physician care, including the service provided (a 5-digit code established by the Healthcare Common Procedure Coding System (HCPCS)) and the identity of the physician providing treatment. Crucially, it indicates the amount the physician was paid for each service, the insurer providing coverage, whether the physician is in-network, and the patient's residential zip code. We focus on the "allowed charge" for in-network care.[24]

This provides the information necessary to estimate equation (12). We estimate (12) on micro data, and then match the physician fixed effects $\hat{\varphi}_g$ to the patients who see that physician. We approximate the patients' income using the median family income in their residential zip code.[25] We compute the mean of physician markups among all physician visits from patients in a given zip code z: $\overline{\varphi_z} = \frac{1}{N_{visit}} \sum\limits_{visit \in z} \hat{\varphi}_g$. We then regress this on log median family income in the zip code:

$$\overline{\varphi_z} = \mu_0 + \mu_1 \ln \left(\text{median family income}\right)_z + \varepsilon_z.$$

We run this regression at the zip code level, weighting observations by the number of underlying physician visits in that zip code.

method to hospital pricing as well.

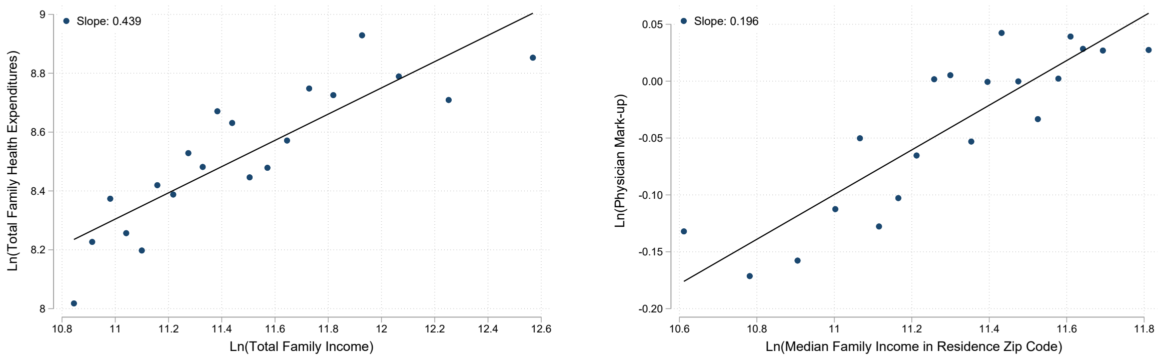
[24]Depending on the details of the patient's insurance contract and whether the patient has reached an annual deductible or out-of-pocket maximum, the patient or the insurer may have to pay the physician's fee for a particular treatment. But regardless of who is liable, the amount that the physician expects to receive is governed by the rate negotiated between the physician and the insurer, known in the industry as the "allowed charge."

[25]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 2: Engel Curve for Medical Spending and Physician Prices

(a) Health Expenditures and Family Income

(b) Physician Prices and Family Income



Notes: Panel A shows the relationship between annual medical spending and annual income, both in logs. The dots show mean values for each bin, computed as twenty vigintiles of family income. The line shows a linear regression estimate on the micro-data, and reflects an elasticity of 0.44. Panel B shows an analogous relationship between log markups (normalized to mean zero across physician groups) charged by the physician who treats a patient, and family income (as proxied by the median income in the patient's zip code of residence). This panel also shows a binned scatter plot with 20 vigintiles, and the regression line reflects an elasticity of 0.20. Source: Authors' calculations using data from the Medical Expenditure Panel Survey and Colorado All-Payer Claims Data.

4.3 Results: medical spending, physician prices, and income

Figure 2 shows the results. Panel 2a plots the Engel curve for family medical spending. 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 is immediate and reflects an elasticity of 0.44. That is, a 10% increase in family income is associated with 4.4% more medical spending. A positive elasticity smaller than 1 is in consistent with the CES case described in Proposition 4.

To examine how much of this elasticity reflects differences in prices, as opposed to quantity or composition of care, we move to the medical claims data from APCD-CO. Panel 2b shows the results, with an elasticity of 0.20 between physician log markups and log median family income.[26]

These facts indicate that the matching relationship implied by the model appears in the data. We now investigate our core prediction: within a local geographic

[26]The estimate of the price elasticity is likely to be biased down: The estimate behind Panel 2b uses the log of 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. Note also that physician-patient matching is more complex in the real world than in our stylized model, with insurers linking physicians to patients through different networks and insurance plans. This provides a mechanism for connecting higher-income patients with higher-priced physicians.

market, inequality spills over into occupations providing non-divisible services with heterogeneous quality.

5 Empirical Strategy to Identify Spillovers

This section introduces the main empirical test of our model. Our goal is to estimate the causal effect of general income inequality on income inequality within a specific occupation. Our empirical strategy uses geographical variation in income inequality across LMAs in the United States.

5.1 Income data

Our data come from the Decennial Census for 1980, 1990, and 2000 and the 2010-2014 waves of the American Community Survey (ACS) (which, combined together, we refer to as 2012). We access the restricted-use versions of each which contain a larger sample of respondents and less income censoring than public-use versions.[27] 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 refer to this combined data set as Census data. Data from before 1980 has substantially smaller samples, and we exclude them from the 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 $\{x_i\}_{i=1}^N$ 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:

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

where $\mathcal{N}$ is the set of observations. That is, the estimate of the inverse Pareto

[27] The detailed Decennial Census micro data long-form surveys 1/6th of the population. Each of the five ACS samples is 3%, so combining them gives a sample of 15%. The income numbers are inflated using the consumer price index such that all numbers are in 2014 prices. 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% of the population is censored. Among physicians the number is well below 0.5%.

parameter is the average log distance of observations from the minimum (the chosen cutoff value). So the estimated Pareto parameter is a measure of income inequality, even if the distribution is not exactly Pareto. We adjust (13) for the small number of censored observations (details in Appendix B).

One benefit of this approach is that the Pareto parameter can easily be translated into relative incomes at different ranks of the income distribution. For a Pareto distribution with parameter α, the relative income of somebody at the 99th percentile compared to somebody at the 95th percentile is $5^{1/\alpha}$. The Gini coefficient is $(2\alpha-1)^{-1}$. Guvenen, Karahan, Ozkan, and Song (2021) and Jones and Kim (2014) also employ α^{-1} as a measure of income inequality.

Our focus on top income inequality requires choosing a threshold x_{min}. We set this at the 90^{th} percentile of the local income distribution for those with positive wage income, since the Pareto distribution is generally a good fit in the top decile. We focus on employed individuals older than 25. We need a reasonable number of observations to compute local occupational income inequality, so we restrict the sample to LMAs with at least 25 observations above x_{min} for each of the four years.[28] Appendix B further discusses this restriction and we provide various robustness checks below.

5.2 Income distribution statistics

We present several summary statistics on the wage income distribution. Our income measure is pre-tax wage and salary income, and we focus on observations with positive income. Appendix Table C.1 shows basic 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 in the top income distribution, accounting for 13% of the top 1%.

Table 1 shows the ratio of income at the 98^{th} percentile to income at the 90^{th} percentile for physicians and the general population as well as the corresponding ratio predicted by the inverse Pareto parameter, $\hat{\alpha}^{-1}$. Here, we estimate $\hat{\alpha}^{-1}$ based on the top 10% of the national income distribution for the general population and for

[28] The regressions are therefore balanced. The number of observations shown in the regression tables are not always divisible by 4 due to disclosure requirements which request that we only release the number of observations rounded to the nearest multiple of 50. The underlying observation count is always a multiple of 4.

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

	General Population			Physicians		
Year	α^{-1}	Actual	Predicted	α^{-1}	Actual	Predicted
1980	0.34	1.70	1.72	0.25	1.50	1.50
1990	0.38	1.87	1.85	0.40	1.89	1.90
2000	0.42	2.00	1.96	0.33	1.75	1.71
2012	0.42	1.99	1.96	0.34	1.72	1.72

Notes: The inverse Pareto parameter, α^{-1}, is calculated based on the top 10% of the relevant national wage income distribution (general or physician). The Predicted 98/90 ratio is calculated based on the estimated Pareto coefficient as: $5^{1/\alpha}$.

doctors, respectively. The predicted and actual ratios agree closely, consistent with a good fit to the Pareto distribution at the top of the income distribution.

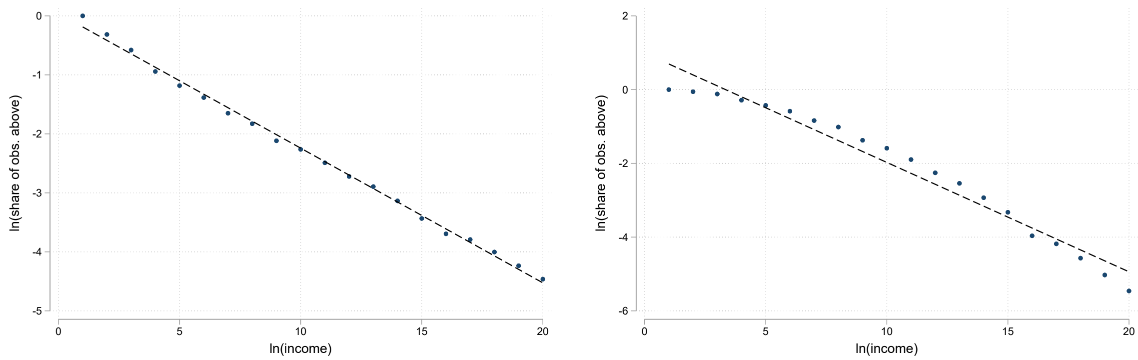
Finally, we assess the extent to which the income distribution is Pareto at the local level. We use the fact that a Pareto distribution implies a linear relationship between log income and the log of the number of observations with a higher value.[29] Due to disclosure regulations we cannot state exact income numbers nor export LMA-specific information. Instead, we focus on the top 10% of the income distribution in New York State. We bin the data into 20 equally spaced bins and plot the average (log) value of the observations within a bin (the estimates of the Pareto parameter in the rest of the paper are based on the underlying observations and not binned data) in Figure 3. Panel 3a shows the relationship between log income and the log number of observations overall, and Panel 3b for physicians specifically. The Pareto fit is excellent for the general population. For physicians, the Pareto coefficient is estimated on those in the top 10% of the entire local income distribution. (This is in line with our regression analysis but differs from Table 1, which focuses on the top 10% of physicians.) Given that a large share of physicians are in the top 10%, the Pareto fit is less good. Regardless of whether the data are exactly Pareto, recall that the inverse Pareto coefficient α^{-1} is a reasonable measure of income inequality. In addition, we run robustness checks where we only use the top 5% of the local income distribution, and Panel 3b shows that the Pareto fit is better at higher incomes.[30]

[29] Formally, for a dataset with N observations of wages drawn from a Pareto distribution, the expected share of observations with a value higher than x, $\frac{N_x}{N}$, is given by $\frac{N_x}{N} = \left(\frac{x}{x_{min}}\right)^{-\alpha}$. Hence, we have: $\ln\left(\frac{N_x}{N}\right) = -\alpha \ln(x) + \alpha \ln(x_{min})$, a linear relationship between $\ln\left(\frac{N_x}{N}\right)$ and $\ln(x)$.

[30] We compute α^{-1} on the top 10% in our regression analysis in order to have a sufficient number of observations.

Figure 3: 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 physician 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 log number of observations in each bin. By construction, the line must be downward-sloping, and linear if the underlying distribution is Pareto. The lines are a linear fit to the binned observations, not the underlying Pareto parameter. Source: Authors' calculations using Decennial data.

5.3 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. 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 of interest, at the area-occupation-year level, is:

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

The coefficient β_o measures the elasticity of top income inequality for our occupation of interest with respect to general income inequality. We estimate regression (14) by using the Census income data described above to compute both $\alpha_{o,t,s}^{-1}$ and $\alpha_{-o,t,s}^{-1}$ for 1980, 1990, 2000 and 2012. Throughout our estimation, we cluster standard errors by LMA and weight LMA-years by the number of underlying members of the occupation of interest (above x_{min}).

Instrument. One would naturally worry about reverse causality and omitted variables when estimating equation (14). Even controlling for LMA and year fixed effects, a positive correlation between general income inequality and income inequality for a specific occupation might reflect deregulation, changes in the tax system, or common local economic trends—rather than a causal effect from general income inequality to inequality for the occupation of interest. Some of these might lead to an upward bias in our estimate. Our mechanism itself could generate reverse causality: inequality within the outcome occupation might spill over into other occupations on the right-hand side of the regression. As section 3.2.3 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:

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

where $K_{-o,s}$ is the set of the 20 most important occupations in the top 10% of the income distribution of LMA s in 1980 (excluding occupation o). The corresponding share in 1980 of these occupations $\kappa \in K_{-o,s}$ is denoted $\omega_{\kappa,1980,s}$. (In robustness exercises we exploit other choices of occupation sets $K_{-o,s}$.) $\alpha^{-1}_{\kappa,t,-s}$ is the inverse Pareto coefficient for occupation κ in year t in the entire U.S., excluding the LMA of interest s. We then estimate equation (14) via two-stage least squares, using $\ln(I_{-o,t,s})$ as an instrument for $\ln\left(\alpha^{-1}_{-o,t,s}\right)$.

The source of variation in our instrument 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 in LMA s for which there are observations with incomes above x_{min} during the time periods of study, excluding the occupation of interest o. Let $\tilde{\omega}_{\kappa,t,s}$ be the corresponding occupation shares where $\sum_{\kappa \in O_{-o,s}} \tilde{\omega}_{\kappa,t,s} = 1$. 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 as $\alpha^{-1}_{-o,t,s} = \sum_{\kappa \in O_{-o,s}} \tilde{\omega}_{\kappa,t,s} \alpha^{-1}_{\kappa,t,s}$ (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:

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

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.[31,32] Our IV estimation therefore only exploits the changes in labor market income inequality that arise from the occupational distribution in 1980 combined with nationwide trends in occupational inequality. By using these national trends, our instrument relies on variation associated with national shocks exogenous to the LMA, such as the effects of globalization, technological change or deregulation. These shocks would affect LMAs differently depending on their occupational composition, in line with a change in α_x in our theoretical model. The two other terms in equation (16) capture local changes in the LMA during our time period. The second term captures changes in local occupational income inequality relative to the average trend in the rest of the U.S. The third term captures changes in the local occupational distribution. Neither of these two terms can plausibly be considered exogenous to a particular occupation of interest o.

We adopt the Goldsmith-Pinkham, Sorkin and Swift (2020) framework for evaluating 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. For instance, it might be that in an area with many financial managers, physicians have been able to get better access to credit over time, which has enabled the best of them to expand their offices and earn higher incomes. As long as this effect doesn't work through

[31] Technically for our instrument, we only include the top 20 occupations in 1980 in $K_{-o,s}$ and normalize the weights to sum to one (which is why $\omega_{\kappa,1980,s}$ and $\tilde{\omega}_{\kappa,1980,s}$ differ).

[32] 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 sales occ.**		
	Mean	Std Dev	N	Mean	Std Dev	N	Mean	Std Dev	N
$\ln(\alpha_o^{-1})$	−0.15	0.17	750	−0.42	0.13	150	−0.69	0.14	200
$\ln(\alpha_{-o}^{-1})$	−0.99	0.16	750	−0.92	0.14	150	−0.93	0.14	200
$\ln(I)$	−1.00	0.07	750	−0.91	0.05	150	−0.92	0.05	200
α_o^{-1}	0.87	0.15	750	0.66	0.08	150	0.51	0.07	200
α_{-o}^{-1}	0.37	0.06	750	0.40	0.06	150	0.40	0.06	200
I	0.26	0.03	750	0.30	0.02	150	0.29	0.02	200

Notes: This table shows basic summary statistics for the variables in our different occupation regressions. $ln(\alpha_o^{-1})$ is the logarithm of the inverse Pareto parameter for the occupation of interest in a given LMA × year. $ln(\alpha_{-o}^{-1})$ is the logarithm of the inverse Pareto parameter for the local population excluding the occupation of interest. Both are based on the top 10% of the wage income distribution in a given LMA × year. $ln(I)$ is the instrument: the logarithm of the projected income inequality in a given LMA × year based on the occupational distribution in 1980 (see details in text). N is the number of observations rounded to the nearest multiple of 50 as required by disclosure rules.

the income inequality of financial managers, this will not bias our results. In addition, differences in credit access would be captured by the LMA fixed effect. We run robustness checks where we exclude financial managers from the instrument.[33] We discuss our shift-share setting further in section 6.2 and in Appendix C.3.

Summary statistics for the regression variables. Our first regression results focus on three occupations for which we expect to see spillovers, namely physicians, dentists, and real estate agents. Table 2 presents summary statistics for the main variables in these regressions. Note that the sample size varies because of our requirement that each LMA has sufficient observations to compute our dependent variable.[34]

6 Empirical Spillover Estimates

This section presents our empirical estimates of spillovers across occupations. We first focus on a set of occupations for which we expect to find spillovers. We present

[33]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.

[34]The average inverse Pareto coefficients in Table 2 differ from the ones reported in Table 1 particularly for physicians. There are two reasons for this: First, in Table 2, Pareto coefficients are calculated for physicians in the top 10% of the overall population instead of the top 10% of physicians. Second, Table 2 shows the unweighted mean of Pareto coefficients in each labour market area (using local cutoffs) instead of a national Pareto coefficient.

the spillover estimates in section 6.1 and conduct several robustness checks in section 6.2. We then compute spillover coefficients for most of the 30 biggest occupations in the top 10% of the income distribution and correlate these spillover coefficients with occupation characteristics in section 6.3.

6.1 Testing the model when spillovers are predicted: 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 C.1). They fit our theory well since they provide a service that is heterogeneous in quality, non-divisible, and primarily serve the local market. Dentistry is similar to other medical services but it involves fewer intermediaries and is less regulated. Real estate agents are another occupation common in the top of the U.S. income distribution. Real estate services are non-divisible since home sellers usually only contract with one real estate agent.[35]

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 an elasticity of 0.16. This elasticity 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 using our instrument. The instrument has a reasonable predictive effect on the endogenous variable with F around 8.

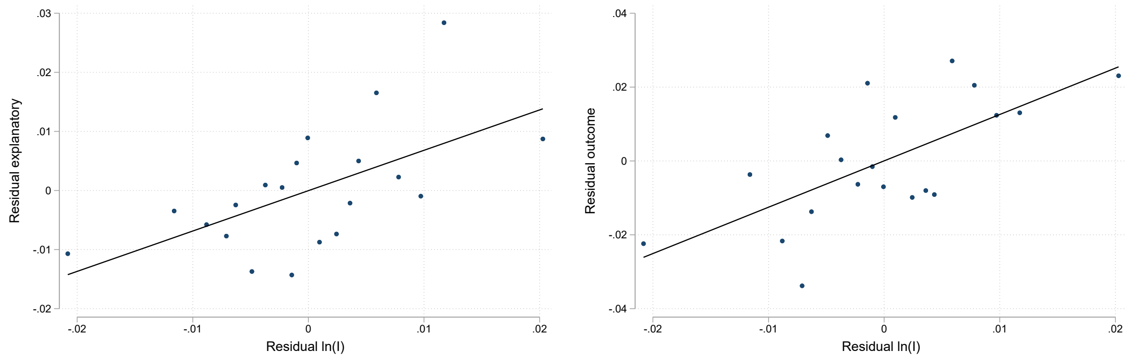
Columns (5) and (6) present IV results: Income inequality from the broader population spills over to physician income inequality with an estimated elasticity of 1.5 in the model with controls. From 1980 to 2012 wage income inequality as measured by the inverse Pareto parameter rose by 24% with a corresponding increase for physicians of 34%. So a spillover elasticity of 1.5 can plausibly explain the entire rise of income inequality for physicians of 36%, although the contribution is measured with uncertainty. Log population size has little conditional relationship with physician income inequality, whereas log average income predicts lower income inequality. Including

[35] Furthermore, the fee structure in real estate often ends up being 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).

Figure 4: 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



the controls does not alter the coefficient of interest in the IV regressions.

Table 3: Spillover estimates for Physicians

	OLS		**1st Stage**		**IV**	
Dependent variable	$\ln(\alpha_o^{-1})$	$\ln(\alpha_o^{-1})$	$\ln(\alpha_{-o}^{-1})$	$\ln(\alpha_{-o}^{-1})$	$\ln(\alpha_o^{-1})$	$\ln(\alpha_o^{-1})$
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\alpha_{-o}^{-1})$	0.16**	0.22***			1.74**	1.50**
	(0.08)	(0.06)			(0.75)	(0.70)
ln(Average Income)		−0.40***		0.17***		−0.60***
		(0.09)		(0.05)		(0.14)
ln(Population)		−0.02		−0.06		0.07
		(0.03)		(0.04)		(0.07)
$\ln(I)$			0.70***	0.70***		
			(0.24)	(0.26)		
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	750	750	750	750	750	750
F-Statistic					8.65	7.43

Notes: 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 log of 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 log of the inverse Pareto parameter α_{-o}^{-1} for the local population excluding physicians in the top 10% of the local income distribution. $ln(I)$ is our instrument and captures the projected occupational income inequality from national trends by interacting local occupational composition with national trends in inverse Pareto parameters for each occupation (see details in text). Column (1) shows the OLS relationship only including LMA and year fixed effects. Column (2) adds controls for average income among individuals with positive income and population. Columns (3) and (4) show the first stage regressions. Finally, Columns (5) and (6) show the IV regressions. N is the number of observations rounded to the nearest multiple of 50. Observations are weighted by the number of uncensored physicians in the top 10% of the local income distribution. *: p<0.1, **: p<0.05, ***: p<0.01.

Figure 4 shows the IV results graphically in two binned scatter plots. In Panel 4a, 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 dummies and the controls of column (6) in Table 3. For disclosure reasons we bin our LMA $\times$ year observations into 20 bins with equal number of underlying individuals. We plot the average value of the residuals in each bin. Panel 4b 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.

Dentists and real estate agents. We next show the corresponding results for dentists and real estate agents. We still require that LMAs contain at least 25 observations in the occupation of interest in the top 10% in order to compute local top income inequality for that occupation. Therefore, the number of LMAs included in the regressions is substantially lower for both dentists and real estate agents, though the F statistics remain similar. The results are given in Table 4. The estimates from the IV regression with controls are 1.33 and 1.22, close to that for physicians. The most notable difference is that the OLS coefficients are substantially higher. The OLS and IV estimates for real estate agents are nearly identical. Based on the estimated IV spillovers, the predicted increases in income inequality are 32% and 29% for dentists and real estate agents, respectively. This is larger than the actual increases of 22% and 18%, respectively, though the actual increase is well within the confidence interval.

Relationship between OLS and IV results. For physicians and dentists, the IV results are substantially higher than the OLS correlations. There are three main reasons. First, the augmented model with a CES utility function predicts a downward bias in the OLS relationship in the empirically relevant case when $\varepsilon < 1$. We show this mathematically in section 3.2.3: the bias arises from unobserved correlations between inequality in local doctors' ability and local consumers' ability.[36] Second, there are numerous potential omitted variables. For example, more unequal places might have

[36]A likely explanation for the similarity of the IV and OLS estimates for real estate agents is the higher elasticity of the Engel curve for housing. Subsection 3.2.3 showed that an elasticity of substitution between medical services and other goods substantially less than one has two implications: (i) The Engel curve has an elasticity substantially below one and (ii) the OLS estimate is biased downward if the ability distribution of doctors and the general population are correlated. For physicians, Figure 2 implies an elasticity of the Engel curve of around 0.44. But for real estate, Zabel (2004) finds elasticities of the Engel curve to be 0.64-0.70 for high-income families. This suggests a higher elasticity of substitution for housing than medical services, therefore less of a downward bias in the real estate OLS coefficient.

higher taxes and spend more public money on health care. This would support the incomes of those physicians who are not at the top of the income distribution. Third, we estimate inequality in the general population with error. Since this is an estimate from a small sample, we would expect the estimated $\alpha^{-1}_{-o,t,s}$ to suffer from classical measurement error, again biasing the OLS estimate down.

Table 4: Spillover estimates for Dentists and Real Estate Agents

Panel A: Dentists

	OLS		1st stage		IV	
Dependent variable	$\ln(\alpha_o^{-1})$	$\ln(\alpha_o^{-1})$	$\ln(\alpha_{-o}^{-1})$	$\ln(\alpha_{-o}^{-1})$	$\ln(\alpha_o^{-1})$	$\ln(\alpha_o^{-1})$
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\alpha_{-o}^{-1})$	0.60***	0.58***			1.29**	1.33*
	(0.20)	(0.19)			(0.63)	(0.69)
ln(Average Income)		0.12		0.10		0.05
		(0.18)		(0.07)		(0.20)
ln(Population)		−0.01		-0.02		0.02
		(0.08)		(0.05)		(0.11)
$\ln(I)$			2.24***	2.20***		
			(0.67)	(0.74)		
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	150	150	150	150	150	150
F-Statistic					11.31	8.89

Panel B: Real Estate Agents

	OLS		1st stage		IV	
Dependent variable	$\ln(\alpha_o^{-1})$	$\ln(\alpha_o^{-1})$	$\ln(\alpha_{-o}^{-1})$	$\ln(\alpha_{-o}^{-1})$	$\ln(\alpha_o^{-1})$	$\ln(\alpha_o^{-1})$
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\alpha_{-o}^{-1})$	1.20***	1.17***			1.25**	1.22**
	(0.15)	(0.15)			(0.54)	(0.57)
ln(Average Income)		0.14		0.16**		0.13
		(0.18)		(0.07)		(0.16)
ln(Population)		0.00		−0.03		0.00
		(0.05)		(0.05)		(0.06)
$\ln(I)$			2.00***	1.91***		
			(0.65)	(0.69)		
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	200	200	200	200	200	200
F-Statistic					9.60	7.56

Notes: This table shows OLS and IV regressions of local top income inequality among dentists (Panel A) and real estate agents (Panel B) on top income inequality in the local population. The variables are defined analogously to Table 3. Column (1) shows the OLS relationship only including LMA and year fixed effects. Column (2) adds average income and population. Columns (3) and (4) show the first stage regressions. Finally, Columns (5) and (6) show the IV regressions. Further details are in Table 3. *: p<0.1, **: p<0.05, ***: p<0.01.

6.2 Robustness checks

This section presents several robustness checks on the spillover results. We first investigate various issues specific to physicians. We then show that our results are robust to measuring inequality in levels instead of the log-linear specification. We show that the results are robust to excluding the occupations with the largest Rotemberg weights from the IV. Finally, we show robustness checks for our various cutoffs. This subsection discusses the results and we refer the reader to Appendix C.2 for tables.

Issues specific to physicians. In Appendix Table C.2, we investigate three issues specific to physicians. First, Gottlieb et al. (2020) show that a substantial share of top physician income is business income. The census data provides information on wage income, business income, and capital income; we define "earned income" as the sum of wage income and business income. We calculate top income inequality for physicians' earned income in the same manner as for wage income and replace the dependent variable. We leave the instrument and the RHS variables unchanged. The IV coefficient remains similar to that of Table 3.

Second, pay varies considerably across different specialties of medicine. Given that our physician occupation category includes all physicians we could potentially pick up compositional effects across specialties. To address this, we build controls for the share of physicians in different specialties.[37] The IV coefficient drops slightly to 1.12 and remains significant.

Specification using levels. Our baseline analysis measures inequality using log of the inverse Pareto parameter. In Appendix Table C.3, we re-run our baseline regressions using the level of income inequality for physicians, dentists, and real estate agents.[38] We similarly find evidence of inequality spillovers for the three occupations with this

[37]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: the share of neurosurgeons, who are the specialty with the highest mean income and the largest standard deviation; the share of physicians in the 8 highest earning specialties (excluding neurosurgeons); the share of physicians in the 7 specialties with the lowest income; and the share of physicians in the 4 specialties with the largest standard deviation in income (excluding neurosurgeons). The rationale behind the two income groups of specialties is that these specialties share similar average incomes while the next specialty (down or up) in the ranking has a substantially different average income.

[38]The weights of the top 20 occupations in the top 10% of the local income distribution are not normalized to sum up to 1 in this specification. We control for the share of individuals in other occupations in the top 10% interacted with a year dummy (Goldsmith-Pinkham et al., 2020).

functional form.

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 15 largest (in absolute terms) Rotemberg (1983) weights in Appendix Table C.4.[39] The three occupations with the largest Rotemberg weights are Financial Service Sales (0.34), Financial Managers (0.22), and Lawyers and Judges (0.19). We exclude, in turn, the five occupations with the highest Rotemberg weight from the IV in our baseline regression for physicians. Appendix Table C.5 reports the results and shows that the coefficient of interest is very stable. In Appendix C.3, we also show a regression which includes the Adão, Kolesár and Morales (2019) standard errors.

Our data construction relies on a number of cutoffs. We use alternative cutoffs in Appendix Table C.6 for physicians. First, we base our regressions on the top 5% of the income distribution instead of top 10%. Our results are, if anything, statistically stronger. We change our rule for selecting LMAs by requiring there to be either at least 40, or at least 15, individuals of the occupation of interest in the top 10% of the local income distribution for each year (compared with 25 in the baseline). Finally, we build our IV based on the 30 or 15 most common occupations in the top 10% of the local income distribution (instead of 20). In all cases, we find similar coefficients.

Appendix Table C.6 reproduces the same exercises for dentists and real estate agents. We find broadly consistent results, though we lose statistical power for dentists and real estate agents when considering the top 5% of the income distribution. This reflects the fact dentists and real estate agents are less numerous in the top 5% so our inequality measures are computed on fewer observations.

6.3 Testing the model for other top occupations

Placebo occupations. We now carry out analogous regressions for other common occupations in the top 10% of the wage income distribution. We first focus on selected occupations for which we do not expect to see spillovers, namely financial managers, managers (excluding those in real estate), and engineers. 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

[39] To do that, we move to a linear setting where we use the same set of occupations in the instrument for all LMAs.

and services for both the local and national markets.[40] 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: Spillover estimates for Financial managers, Managers and Engineers

	Financial Managers			**Managers, excl. real estate**			**Engineers**		
	OLS (1)	1st stage (2)	IV (3)	OLS (4)	1st stage (5)	IV (6)	OLS (7)	1st stage (8)	IV (9)
$\ln(\alpha_{-o}^{-1})$	0.71***		−1.73	0.33***		−0.28	0.51***		−0.86
	(0.16)		(1.23)	(0.04)		(0.21)	(0.09)		(0.60)
ln(Average Income)	0.56***	0.13**	0.83***	0.02	0.16**	0.14**	−0.19	0.06	−0.11
	(0.19)	(0.05)	(0.24)	(0.05)	(0.07)	(0.06)	(0.12)	(0.05)	(0.14)
ln(Population)	−0.15*	−0.05	−0.29	0.12***	−0.12***	0.05	0.20***	−0.07***	0.10
	(0.08)	(0.04)	(0.19)	(0.02)	(0.04)	(0.03)	(0.05)	(0.02)	(0.07)
$\ln(I)$		1.11***			0.51***			0.71***	
		(0.31)			(0.11)			(0.18)	
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	450	450	450	1500	1500	1500	1000	1000	1000
F-Statistic			12.50			19.83			16.38

Notes: This table shows OLS and IV regressions of local top income inequality for selected occupations, where we do not predict spillovers, on top income inequality in the local population. The variables are defined analogously to Table 3. Columns (1)-(3) look at financial managers, Columns (4)-(6) at managers (excluding real estate), and Columns (7)-(9) at engineers. Columns (1), (4) and (7) show OLS regressions. Columns (2), (5) and (8) show first-stage regressions. Columns (3), (6) and (9) show the IV results. All regressions include controls for local population and average income, LMA and year fixed effects. N is the number of observations rounded to the nearest integer divisible by 50. *: p<0.1, **: p<0.05, ***: p<0.01.

Table 5 reports regression results (OLS, first stage, and IV including the controls) for these three occupations. The OLS estimates are significant throughout and comparable to those for physicians, dentists and real estate agents. However, the IV coefficients are statistically indistinguishable from zero and are all negative. 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.

[40] 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

Occupation	OLS	OLS SE	IV	IV SE	F-stat	t-stat	N
Financial managers	0.71***	(0.16)	−1.73	(1.23)	12.50	3.54	450
Managers of properties and real estate	0.92***	(0.32)	1.32	(0.91)	11.00	3.32	100
Managers, excl. real estate	0.33***	(0.04)	−0.28	(0.21)	19.83	4.45	1,500
Accountants and auditors	0.73***	(0.12)	0.16	(0.47)	13.37	3.66	500
Other financial specialists	0.46***	(0.17)	−0.06	(0.68)	12.85	3.58	300
Engineers	0.51***	(0.09)	−0.86	(0.60)	16.38	4.05	1,000
Systems analysts and scientists	1.07***	(0.28)	1.48**	(0.70)	12.24	3.50	250
Physicians	0.22***	(0.06)	1.50**	(0.70)	7.43	2.73	750
Dentists	0.58***	(0.19)	1.33*	(0.69)	8.89	2.98	150
Lawyers and judges	0.20	(0.15)	−0.52	(0.53)	15.67	3.96	450
Computer programmers	0.96**	(0.42)	−2.57	(1.63)	12.42	3.52	150
Sales supervisors and proprietors	0.43***	(0.08)	−0.06	(0.47)	15.30	3.91	950
Real estate sales occupations	1.17***	(0.15)	1.22**	(0.57)	7.56	2.75	200
Financial service sales occupations	0.8 ***	(0.17)	0.70	(0.55)	10.93	3.31	150
Sales occupations and sales representatives	0.62***	(0.07)	−0.09	(0.36)	15.41	3.93	800
Supervisors of construction work	0.49***	(0.15)	0.44	(0.62)	12.92	3.59	450
Production supervisors or foremen	0.69***	(0.19)	−1.53	(1.13)	10.63	3.26	500
Driver/sales workers and truck drivers	0.82***	(0.25)	1.40	(0.92)	9.97	3.16	450

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 28 most common occupations in the top 10% of the income distribution, excluding those with an F statistic smaller than 7. Each row corresponds to the regressions for a given occupation. The variables are defined analogously to Table 3 and regressions include controls for local population, average income, LMA and year fixed effects. Columns (1) shows the OLS coefficient, columns (2) the OLS standard error, column (3) the IV coefficient, Column (4) the IV standard error, Column (5) the F statistic for the excluded instrument, Column (6) the IV t statistic and Column (7) the number of observations, rounded to the nearest multiple of 50. The number of observations varies because we only include LMAs with at least 20 members of the occupation of interest in the top 10% of the local income distribution. *: p<0.1, **: p<0.05, ***: p<0.01.

Top occupations in the top 10% of the income distribution. We now take a more systematic look at the most common 30 occupations in the top 10% of the U.S. income distribution. We restrict attention to occupations with at least 20 LMAs that satisfy our requirement of enough observations in the top 10% of the income distribution to compute local occupational income inequality. This leaves us with 28 occupations, including the six already considered. Table 6 reports the results for all occupations with an F statistic greater than 7.[41] The only occupations with a significant coefficient at the 10% level are our occupations of interest (physicians, real estate agents and dentists) plus "computer systems analysts and computer scientists." Computer scientists do not fit our theory and we think of their case as one false positive out of 28 occupations (unsurprising at $p < 0.05$).

The remaining occupations can be thought of in two groups. First, those occupations for which our theory does not apply: the placebo occupations already mentioned, computer programmers, and production supervisors. Second, a group of occupations

[41] We omit occupations with an F statistic below 7 since their IV estimates are likely subject to weak instruments bias. None of them has a significant coefficient in the second stage.

which includes subcategories for which our theory would apply and subcategories for which it would not. For instance, "lawyers and judges" contain both corporate lawyers—who work for firms and for which we would not expect to see spillovers—and personal attorneys—for which our mechanism would likely apply. Similarly, financial sales occupations cover both individuals working for firms and personal finance managers for which our theory is more likely to apply, albeit perhaps at the national level. The fact that we do not find spillovers for a broad range of occupations also suggests that our results do not originate from a "'keeping up with the Joneses effect" (see, *e.g.*, Bertrand and Morse, 2016).

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) offshorability measure as an (inverse) measure of the extent to which an occupation serves the local market; the extent to which an occupation can be performed abroad is an inverse proxy for the extent to which it has to be performed in the local area. To quantify direct public interaction, we rely on measures from the Occupational Information Network (O*NET) of the importance of *customer service* and *working with the public*. Appendix B.3 gives further details.

Since spillover coefficients are estimated with varying precision, we divide them by their standard errors; that is, we correlate the t-statistic 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. The scatter plots in Figure 5 show the results. Consistent with our model, Panel 5a shows a strong negative relationship between the spillover coefficient and offshorability ($p = 0.019$).[42] Panels 5b and 5c show positive relationships between the spillover coefficient and measures of customer service and working with the public ($p = 0.012$ and $p = 0.028$).[43]

These patterns support our model: With one exception, we only observe local spillovers for the occupations that fit the model's predictions of delivering non-

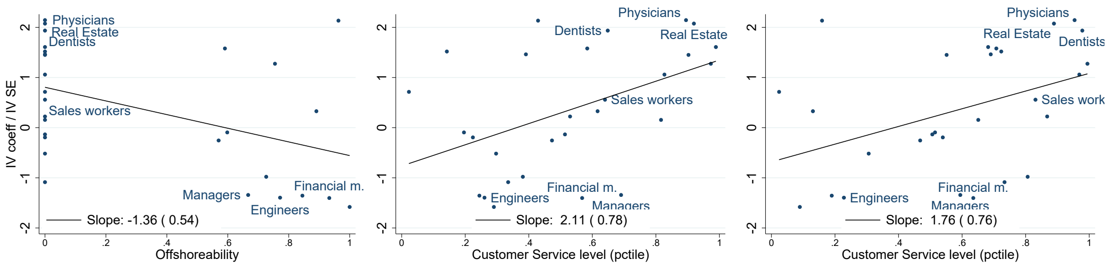
[42] Alternatively, one could split the occupations up into a group of 16 which Blinder (2009) consider non-offshorable and one of 12 with some extent of offshorability. The means of these two groups are statistically significantly different and the coefficient for the offshoreable group is insignificant.

[43] Appendix Table C.8 presents the regressions corresponding to Figure 5, and adds two regressions where we use the importance measures for the customer service and working with the public variables. These additional regressions also show positive relationships. Restricting attention to occupations with $F > 7$ delivers essentially the same results.

divisible goods or services of heterogeneous quality, and serving the local market.

Figure 5: IV estimates and occupational characteristics.

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



Notes: This figure shows the relationship between the t-stat of the spillover coefficients from the IV regressions (from Table 6) and three characteristics of occupations. These are a measure of offshorability from Blinder (2009) as well as 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. O*NET measures are rescaled as percentiles. We use 28 occupations which are those amongst the biggest 30 occupations in the top 10% with at least 20 LMAs for the IV regressions

Quantifying indirect spillovers The workers experiencing spillovers in our model are themselves consumers. Our model implies that increasing doctors' income inequality generates further spillovers on other workers whose output doctors consume. We conclude by quantifying these indirect spillovers, to calculate the effect on total income inequality of an exogenous increase in one occupation's inequality.

Consider the total set of occupations among top earners, Ω, and let $\Omega^{SO} \subset \Omega$, be the subset which experience inequality spillovers. We assume that inequality of occupation $i \in \Omega^{SO}$ follows $\ln(\alpha_i^{-1}) = \beta_i \ln(\alpha^{-1}) + \kappa_i$, as in equation 14, where β_i is the spillover estimated in section 6 for occupation i and κ_i is a constant. The income inequality of the occupations not in Ω^{SO} are exogenous to the model. By definition, aggregate income inequality in the top is $\alpha^{-1} = \sum_{i\in\Omega} \omega_i \alpha_i^{-1}$, where ω_i is occupation i's share of top earners so $\sum_{i\in\Omega} \omega_i = 1$.

Suppose income inequality among workers in occupation $1 \notin \Omega^{SO}$ increases for some exogenous reason; $d\alpha_1^{-1} > 0$. The direct effect of occupation 1's inequality on overall inequality is proportional to the occupation's weight: $d\alpha^{-1} = \omega_1 d\alpha_1^{-1}$. With spillovers, inequality will also increase for occupations in Ω^{SO}, who generate further spillovers on themselves and each other. This is akin to a Keynesian multiplier. Adding up these spillovers yields a total effect of:

$$d\alpha^{-1} = \frac{1}{1 - \sum_{i\in\Omega^{SO}} \beta_i \frac{\omega_i \alpha}{\alpha_i}} \times \omega_1 d\alpha_1^{-1}. \tag{17}$$

For concreteness, consider our focal occupations—physicians, dentists and real estate agents—which together constitute 15.8% of the top one percent of earners. Combining our estimates of $\hat{\beta}_i$ with each occupation's share, we find a multiplier effect of $(1-0.19)^{-1} = 1.24$; that is, the impact on total top inequality is 24% higher than without spillovers. Recall that our estimation procedure only captures local spillovers, and if other occupations have national spillovers, the total multiplier effect would increase.

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 as physicians, dentists and real estate agents. We show empirically that changes in local income inequality do indeed spill over to these occupations. The effect is large enough to explain the increase in income inequality for these occupations. In contrast, we find no such spillover effects for occupations which do not fit our theory.

Our analysis suggests that the increase in top income inequality across most occupations observed in the last 40 years 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.

References

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-Share Designs: Theory and Inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous, "Innovation and Top Income Inequality," *Review of Economic Studies*, 2018, *86* (1), 1–45.

Alder, Moshe, "Stardom and Talent," *The American Economic Review*, 1985, *75(1)*, 208–212.

Armour, Philip, Richard Burkhauser, and Jeff Larrimore, "Using the Pareto Distribution to Improve Estimates of Topcoded Earnings," *Economic Inquiry*, 2016, *54*, 1263–1273.

Bakija, Jon, Adam Cole, and Bradley Heim, "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data," 2012.

Bartik, Timothy J., "Boon or Boondoggle? The Debate Over State and Local Economic Development Policies," *Upjohn Institute for Employment Research*, 1991, pp. 1–16.

Bertrand, Marianne and Adair Morse, "Trickle-Down Consumption," *The Review of Economics and Statistics*, 2016, *98* (5), 863–879.

Blinder, Alan S., "How Many US Jobs Might be Offshorable?," *World Economics*, 2009, *10* (2), 41–78.

Bonfiglioli, Alessandra, Rosario Crino, and Gino Gancia, "Betting on Exports: Trade and Endogenous Heterogeneity," *The Economic Journal*, 2018, *128* (609), 471–916.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental Shift-share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

Buera, Francisco and Joseph Kaboski, "The Rise of the Service Economy," *American Economic Review*, 2012, *102* (6), 2540–2569.

Clemens, Jeffrey and Joshua D. Gottlieb, "Do Physicians' Financial Incentives Affect Treatment Patterns and Patient Health?," *American Economic Review*, April 2014, *104* (4), 1320–1349.

_ and _ , "In the Shadow of a Giant: Medicare's Influence on Private Payment Systems," *Journal of Political Economy*, February 2017, *125* (1), 1–39.

_ , _ , and Tímea Laura Molnár, "Do Health Insurers Innovate? Evidence From the Anatomy of Physician Payments," *Journal of Health Economics*, September 2017, *55*, 153–167.

Cooper, Zack, Stuart V Craig, Martin Gaynor, and John Van Reenen, "The Price ain't Right? Hospital Prices and Health Spending on the Privately Insured," *Quarterly Journal of Economics*, 2019, *134*, 51–107.

Deaton, Angus, "Getting Prices Right: What Should be Done?," *Journal of Economic Perspectives*, 1998, *12* (1), 37–46.

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *American Economic Review*, 2016, *106* (3), 479–524.

Dingel, Jonathan I., Joshua D. Gottlieb, Maya Lozinski, and Pauline Mourot, "Market Size and Trade in Medical Services," Working Paper No. 31030, National Bureau of Economic Research March 2023.

Dorn, David, "Essays on Inequality, Spatial Interaction, and the Demand for Skills, Ph.D. Dissertation, University of St. Gallen," 2009.

Edmond, Chris and Simon Mongey, "Unbundling Labor," 2021. mimeo.

Epstein, Andrew J, "Do Cardiac Surgery Report Cards Reduce Mortality? Assessing the Evidence," *Medical Care Research and Review*, 2006, *63* (4), 403–426.

Gabaix, Xavier and Augustin Landier, "Why has CEO pay increased so much?," *The Quarterly Journal of Economics*, 2008, *123 (1)*, 49–100.

Gaynor, Martin and Robert J. Town, "Provider Competition," in Thomas G. McGuire, Mark Pauly, and Pedro P. Barros, eds., *Handbook of Health Economics, Vol 2*, Amsterdam: Elsevier, 2011.

Geerolf, Francois, "A Theory of Pareto Distributions," *Working Paper*, 2017.

Gesbach, Hans and Armin Schmutzler, "Does Globalization Create Superstars? A Simple Theory of Managerial Wages," *European Economic Review*, 2014, *71*, 34–51.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gottlieb, Joshua, Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova, "Who Values Human Capitalists' Human Capital? Healthcare Spending and Physician Earning," *Working Paper*, 2020.

Grossman, Volker, "Firm Size, Productivity and Manager Wages: a Job Assignment Approach," *The B.E. Journal of Theoretical Economics*, 2007, *7* (1).

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song, "What do Data on Millions of U.S. Workers Reveal about Life-cycle Earnings Risk," *Econometrica*, 2021, *89* (5), 2303–2339.

Ho, Katherine, "Insurer-provider Networks in the Medical Care Market," *The American Economic Review*, 2009, *99* (1), 393–430.

Jones, Charles and Jihee Kim, "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, 2018, *126* (5), 1785–1826.

Kaplan, Steven and Joshua Rauh, "It's the Market: the Broad-based Rise in the Return to Top Talent," *Journal of Economic Perspectives*, 2013, *27* (3), 35–56.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad, "Field of Study, Earnings, and Self-selection," *The Quarterly Journal of Economics*, 2016, *131* (3), 1057–1111.

Koenig, Felix, "Technical Change and Superstar Effects: Evidence from the Rollout of Television," *American Economic Review: Insights*, 2021, *5* (2).

Kolstad, Jonathan T, "Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards," *American Economic Review*, 2013, *103* (7), 2875–2910.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song, "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937," *Quarterly Journal of Economics*, 2010, *125* (1), 91–128.

Kukharskyy, Bohdan, "Trade, Superstars, and Welfare," 2012. BGPE Discussion Paper No 120.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider, "The Housing Market(s) of San Diego," *American Economic Review*, 2015, *105* (4), 1371–1407.

Leonardi, Marco, "The Effect of Product Demand on Inequality: Evidence from the United States and the United Kingdom," *American Economic Journal: Applied Economics*, 2015, *7* (3), 221–247.

Ma, Lin and Dimitrije Ruzic, "Globalization and Top Income Shares," *Journal of International Economics*, 2020, *125*, 1–18.

Määttänen, Niku and Marko Terviö, "Income Distribution and Housing Prices: an Assignment Model Approach," *Journal of Economic Theory*, 2014, *151*, 381–410.

Manasse, Paolo and Alessandro Turrini, "Trade, Wages and 'Superstars'," *Journal of International Economics*, 2001, *54* (1), 97–117.

Manning, Alan, "We Can Work It Out: The Impact of Technological Change on the Demand for Low-Skill Workers," *Scottish Journal of Political Economy*, 2004, *51* (5), 581–608.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles, *IPUMS National Historical Geographic Information System: Version 12.0*, Minneapolis: University of Minnesota, 2017.

Mazzolari, Francesca and Giuseppe Ragusa, "Spillovers from High-skill Consumption to Low-skill Labor Markets," *Review of Economics and Statistics*, 2013, *95* (1), 74–86.

Medical Group Management Association (MGMA), "Physician Compensation and Production Survey: 2009 Report Based on 2008 Data.," 2009.

Miceli, Thomas, Katherine Pancak, and C Sirmans, "Is the Compensation Model for Real Estate Brokers Obsolete?," *Journal of Real Estate Finance and Economics*, 2007, *35*, 7/22.

Moretti, Enrico, "Real Wage Inequality," *American Economic Journal: Applied Economics*, 2013, *5* (1), 65–103.

Pareto, Vilfredo, *Cours d'Economie Politique* 1896.

Philippon, Thomas and Arief Reshef, "Wages and Human Capital in the U.S. Finance Industry: 1909-2006," *Quarterly Journal of Economics*, 2012, *127 (4)*, 1551–1609.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States: 1913-1998," *Quarterly Journal of Economics*, 2003, *118* (1), 1–39.

_ , Anthony B. Atkinson, and Emmanuel Saez, "Top Income in the Long Run of History," *Journal of Economic Literature*, 2011, *49* (1), 3–71.

_ , Emmanuel Saez, and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, 2014, *6* (1), 230–271.

Rosen, Sherwin, "The Economics of Superstars," *American Economic Review*, 1981, *71* (5), 845–858.

Sattinger, Michael, "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 1993, *31 (2)*, 831–880.

Song, Jae, David Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality," *The Quarterly Journal of Economics*, 2019, *134* (1), 1–50.

Steinbrook, Robert, "Public Report Cards—Cardiac Surgery and Beyond," *New England Journal of Medicine*, 2006, *355* (18), 1847–1849.

Tolbert, Charles M. and Molly Sizer, "U.S. Commuting Zones and Labor Market Areas. A 1990 Update," *Economic Research Service Staff Paper No. 9614*, 1996.

Zabel, Jeffrey, "The Demand for Housing Services," *Journal of Housing Economics*, 2004, *13* (1), 16–35.

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

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

A Theory Appendix

A.1 Positive assortative matching in equilibrium

This Appendix establishes the following lemma.

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.*

Since CES and Cobb-Douglas functions have positive cross-partial derivatives, then this lemma applies in particular to the utility functions considered in our paper.

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 zand the remaining disposable income $x - \omega(z)$. Since widget maker 1 chooses a doctor of quality z_1, it must be the case 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 derive equation (6). We look for a specific solution to equation (5) of the type $w(z) = K_1 z^{\frac{\alpha_z}{\alpha_x}}$. We find that K_1 must satisfy

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

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

$$h(x) = \frac{\beta \alpha_x}{\alpha_z (1-\beta) + \beta \alpha_x} x + x_{\min} \frac{\alpha_z (1-\beta) + \beta \alpha_x (1-\lambda)}{\lambda (\alpha_z (1-\beta) + \beta \alpha_x)} \left(\frac{x_{\min}}{x} \right)^{\frac{\alpha_x \beta}{\alpha_z (1-\beta)}}. \tag{18}$$

Combining (18) with (1) and (4), we get that the utility of a widget maker with income x is given by

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

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:

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

Then 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 Occupational Mobility

In this Appendix, we analyze the model briefly described in Section 3.2.1. 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:

$$w(z) = x_{\min}(z/z_{\min})^{\alpha_z/\alpha_x}. \tag{19}$$

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:

$$(x_{\min}/m(z))^{\alpha_x} = \int_z^\infty \lambda\mu(\zeta)\,g_z(\zeta)\,d\zeta, \tag{20}$$

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 (19) holds, combining these expressions with (20), 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 $\lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta} + 1\right)^{-\alpha_x} < 1$.[44]

With a constant share of individuals choosing to be doctors (above a threshold), we get that $P_{doc}(W_d > w_d) = P(Z > w^{-1}(w_d))$ 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:[45]

Proposition 5. *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 D.2 in the Supplementary Material available at `http://www.gottlieb.ca` presents an additional model that separates the two. It highlights that the demand effect drives the result.

A.4 Doctors moving: Proof of Proposition 3

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. As a result 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]

[44] 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.

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

[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.

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 $z \geq z_c$,

$$\left(x_{\min}^A / m(z)\right)^{\alpha_x^A} = \lambda \mu_d \left(1 + \varphi(z)\right) \left(z_{\min}/z\right)^{\alpha_z}. \tag{21}$$

There are initially $\mu_d \left(\frac{z_{\min}}{z}\right)^{\alpha_z}$ doctors with ability at least z in each region and by definition, a share $\varphi(z)$ of those move from region B to region A. Since each doctor can provide services to λ patients, after doctors have relocated the total supply over a quality z in region A is given by the right-hand side of (21). 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:

$$\left(x_{\min}^B / m(z)\right)^{\alpha_x^B} = \lambda \mu_d \left(1 - \varphi(z)\right) \left(z_{\min}/z\right)^{\alpha_z}. \tag{22}$$

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 (21) and (22) for $z = z_c$, we obtain $z_c = (\lambda \mu_d)^{\frac{1}{\alpha_z}} z_{min}$, as in the baseline model.

Similarly, combining (21) and (22) for any z, we obtain

$$x_{\min}^A \left(1 + \varphi(z)\right)^{-\frac{1}{\alpha_x^A}} = x_{\min}^B \left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^B} - \frac{\alpha_z}{\alpha_x^A}} \left(1 - \varphi(z)\right)^{-\frac{1}{\alpha_x^B}}. \tag{23}$$

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 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 (21) 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:

$$w(z) \to \frac{\lambda \beta \alpha_x^A 2^{-\frac{1}{\alpha_x^A}}}{\alpha_z (1-\beta) + \beta \alpha_x^A} \left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^A}} \tag{24}$$

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 $2\mu_d \left(\frac{z_{\min}}{z}\right)^{\alpha_z}$ doctors eventually located in region A. Then, as in the baseline model, doctor's income is asymptotically Pareto distributed with coefficient α_x^A in A. Further, using (23), we get:

$$1-\varphi\left(z\right)\rightarrow 2^{\alpha_x^B/\alpha_x^A}\left(x_{\min}^B/x_{\min}^A\right)^{\alpha_x^B}\left(z/z_c\right)^{\alpha_z\left(1-\alpha_x^B/\alpha_x^A\right)}. \tag{25}$$

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:

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

where w above denotes the wage function. Indeed, there are initially $\mu_d P\left(Z>w^{-1}\left(\widetilde{w}\right)\right)$ doctors in region B with a talent sufficient to earn $\widetilde{w}$. A share of $1-\varphi\left(w^{-1}\left(\widetilde{w}\right)\right)$ of these doctors stay in B. Moreover, the total mass of active doctors in region B is given by $\mu_d P\left(Z>z_c\right)$, since overall there is no net movement of actual doctors. Using (24), we get:

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

Using this expression and (25) we get that:

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

Therefore, doctors' income in region B is Pareto distributed with shape parameter α_B as in the baseline model. This establishes Proposition 3.

A.5 Generalizing the model to different utility functions and asymptotically Pareto distributions

We now consider a generalized version of the model. There is a mass 1 of patients and a mass μ_d 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 $1-\mu_d$) or not (the mass of widget makers is 1). The technology for health services is the same as before and we keep $\lambda>\mu_d^{-1}$. 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.

The income of patients is asymptotically Pareto distributed.[47] Therefore, we get

$$P_x(X > x) = \overline{G}_x(\overline{x})\,\overline{G}_{x,\overline{x}}(x),$$

where $\overline{G}_{x,\overline{x}}(x)$ is the conditional counter-cumulative distribution above $\overline{x}$ and $\overline{G}_x(\overline{x})$ is the unconditional counter-cumulative distribution. For $\overline{x}$ large enough, we have

$$\overline{G}_x(x, \overline{x}) \approx (\overline{x}/x)^{\alpha_x} \text{ with } \alpha_x > 1.$$

The ability distribution of potential doctors is also asymptotically Pareto distributed.

We assume that the utility of patient features positive assortative matching (and put more structure in the following subsections). As a result, 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 $z_c = \overline{G}_z^{-1}\left(\frac{1}{\lambda\mu_d}\right)$, which is independent of α_x. As a result, 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 are approximately Pareto distributed, which allows us to rewrite the previous equation as:

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

which gives

$$m(z) = Bz^{\frac{\alpha_z}{\alpha_x}} + o\left(z^{\frac{\alpha_z}{\alpha_x}}\right) \text{ with } B = m(\overline{z})\left(\frac{\overline{G}_x(m(\overline{z}))}{\overline{G}_{\overline{z},z_c}(\overline{z})\,\overline{z}^{\alpha_z}}\right)^{\frac{1}{\alpha_x}}. \tag{26}$$

A.5.1 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 (26) in (2) gives:

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

Up to some 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.

[47] 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.

Proof. We can rewrite (2) as

$$w'(z)\,z=\frac{\beta_z}{1-\beta_z}\left(\lambda Bz^{\frac{\alpha_z}{\alpha_x}}-w(z)\right)+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right). \tag{27}$$

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

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

This gives

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

Integrating we obtain:

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

for some constant K, therefore $\widetilde{w}(z)$ is negligible in front of $\overline{w}(z)$.

This ensures that

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

From this we get that for $\overline{w}_d$ large enough, doctors' income is distributed according to

$$P\left(W_d>w_d|w_d>\overline{w}_d\right)\approx\left(\overline{w}_d/w_d\right)^{\alpha_x}. \tag{29}$$

That is, 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.5.2 CES case and Proof of Proposition 4

We now assume that patients' utility is CES (10) with $\varepsilon\neq1$. The first order condition for the patient's problem can be written as:

$$\frac{\partial u}{\partial z}=\omega'(z)\,\frac{\partial u}{\partial c}. \tag{30}$$

Using (10) 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

$$w'(z)=\lambda^{\frac{\varepsilon-1}{\varepsilon}}\frac{\beta_z}{\beta_c}z^{-\frac{1}{\varepsilon}}\left(\lambda Bz^{\frac{\alpha_z}{\alpha_x}}-w(z)\right)^{\frac{1}{\varepsilon}}\left(1+o(1)\right). \tag{31}$$

The asymptotic distribution of doctors' wages is given either by Proposition 4 or by the following Proposition, which studies the remaining cases.

Proposition 6. *1) If either (i) $\varepsilon > 1$ and $\alpha_x > \alpha_z$, or (ii) $\varepsilon < 1$ and $\alpha_x < \alpha_z$, 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 $\alpha_x > \frac{\alpha_z}{1-\varepsilon}$, doctors' wages doctors are bounded. For $\alpha_x = \frac{\alpha_z}{1-\varepsilon}$, doctors' wages are asymptotically exponentially distributed. In both cases, the elasticity of health expenditures with respect to income tend to 0, $\frac{\ln h(x)}{\ln x} \to 0$.

In the first case, of Proposition 6, demand for health care services at the top is sufficiently strong to generate a Pareto distribution of income for potential doctors. In the second case, it is too weak to generate a Pareto distribution.

Proof. We now establish Propositions 4 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

$$w'(z) = \lambda \frac{\beta_z}{\beta_c} B^{\frac{1}{\varepsilon}} z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}} + o\left(z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}\right). \tag{32}$$

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

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

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

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

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 $\alpha_w = \frac{\alpha_z}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}$, which is increasing in α_x (and we have $\alpha_w > \alpha_x$).

If $\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1=0$, then $\alpha_z = \alpha_x(1-\varepsilon)$, and integrating (32), one obtains

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

Therefore

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

In that case, w is distributed exponentially.

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

$$w(z) = Az^{\frac{\alpha_z}{\alpha_x}} + o\left(z^{\frac{\alpha_z}{\alpha_x}}\right) \tag{33}$$

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 (33) in (31), we get:

$$A\frac{\alpha_z}{\alpha_x}z^{\frac{\alpha_z}{\alpha_x}-1} + o\left(z^{\frac{\alpha_z}{\alpha_x}-1}\right) = \lambda^{\frac{\varepsilon-1}{\varepsilon}}\frac{\beta_z}{\beta_c}(\lambda B - A)^{\frac{1}{\varepsilon}}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}} + o\left((\lambda B - A)^{\frac{1}{\varepsilon}}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}\right). \tag{34}$$

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

Consider now that $\alpha_z \neq \alpha_x$. If $\lambda B \neq A$ then (34) is impossible when $\varepsilon \neq 1$, therefore 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

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

satisfies (31) provided that 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 4 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. $\square$

A.5.3 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 (26), we can then write (30) as

$$w'(z) = \lambda \frac{\frac{\partial u}{\partial z}}{\frac{\partial u}{\partial c}} \equiv \lambda f\left(Bz^{\frac{\alpha_z}{\alpha_x}-1} - \frac{w(z)}{z\lambda}\right). \tag{35}$$

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 ($k=0$):

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

where β is a constant. In these two cases, we can then rewrite (35) as:

$$w'(z) = \lambda^{\frac{\varepsilon_k-1}{\varepsilon_k}}\beta\left(\lambda Bz^{\frac{\alpha_z}{\alpha_x}-1} - \frac{w(z)}{z}\right)^{\frac{1}{\varepsilon_k}}(1+o(1)), \tag{36}$$

which is the same expression as (31) in the CES case (except that there are two potential values for ε_k). We then obtain

Proposition 7. *Propositions 4 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 it is immediate that the logic of Propositions 4 and 6 apply 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 4, 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 (36) 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 $g'(z) \to \frac{\alpha_z}{\alpha_x}\lambda B z^{\frac{\alpha_z}{\alpha_x}-1}$, 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. $\square$

A.6 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 $\lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z}$ 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.5.1. 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 $\widetilde{G}(z) = \overline{G}_z(z)/\overline{G}_z(z_c) = \lambda\mu_d\overline{G}_z(z)$, 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 $\lim\frac{z\widetilde{g}(z)}{\widetilde{G}} = \lim\frac{zg_z(z)}{\overline{G}_z}$. 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:

$$w'(z)z = \frac{\beta_z}{1-\beta_z}\left(\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}} - w(z)\right). \tag{37}$$

instead of (5). We then establish the following Proposition which generalizes Proposition 1:

Proposition 8. *If the ability distribution has a tail at least as fat as Pareto ($\lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z}$ exists and is finite), doctor's income is asymptotically Pareto distributed with shape parameter α_x. If the ability distribution has a tail thinner than Pareto ($\lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z} = \infty$), 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, $\frac{\ln(P(W>w))}{\ln w} = -\alpha_w$. Therefore the statement that $\frac{\ln(P(W>w))}{\ln w}$ decreases with α_x for high w directly generalizes the notion that top income inequality spills over from the general distribution to the doctors' income distribution to the case where 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: $w(z) \to A\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$. In that case,

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

so that the income distribution of doctors 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 (37), one gets:

$$A'(z)z + \frac{1}{\alpha_x}\frac{z\widetilde{g}(z)}{\widetilde{G}(z)}A(z) = \frac{\beta_z(1-A(z))}{1-\beta_z}. \tag{38}$$

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 consistent with equation (38) 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, (37) leads to

$$w'(z) = \frac{\beta_z}{1-\beta_z}\lambda x_{\min}\frac{1}{z}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}(1+o(1)). \tag{39}$$

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

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

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, with the derivative of $\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$ being $z\widetilde{g}(z)\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}-1}$, we necessarily get that the derivative of $\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$ is larger than that of $w(z)$ when $\lim \frac{zg_z(z)}{\widetilde{G}_z(z)} = \infty$, which justifies the assumption that $w(z) = o\left(\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}\right)$.

Integrating (39), we can write

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

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:

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

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

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

By definition, we can rewrite

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

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.7 Scalability in health care

We start from the baseline model of Section 3.1 except that doctors can now choose how many patients to treat. Specifically, doctors pay effort costs $k_1\lambda^{k_2}/k_2$ where λ is the number of patients treated and $k_2 > 1$ with $k_1 > 0$ (results are identical if doctors pay monetary costs). Doctors' utility maximization problem immediately implies that their scale depends on the income they receive for their services as

$$\lambda(z) = (\omega(z)/k_1)^{\varepsilon^S} \text{ where } \varepsilon^S \equiv 1/(k_2 - 1). \tag{40}$$

ε^S is the supply elasticity of health services. Widget makers' consumption problem remains the same, so there is still positive assortative matching and equation (2) holds. Health care market clearing now takes into account that doctors serve different number of patients. With Pareto distributions for doctors' ability and widget makers' income, we get:

$$(m_z/x_{\min})^{-\alpha_x} = \int_z^{\infty} \lambda(\zeta)\, \alpha_z \frac{1}{\zeta} (\zeta/z_c)^{-\alpha_z} d\zeta, \tag{41}$$

with z_c the least able doctor. The equilibrium is defined by (2), (40) and (41) and we get:

Proposition 9. *Doctors' income is asymptotically Pareto distributed with shape parameter* $\alpha_w = \frac{\alpha_x+\varepsilon^S}{1+\varepsilon^S}$*, and an increase in income inequality for widget makers leads to an increase in income inequality for doctors.*

Proposition 9 implies that the central spillover result from Proposition 1 generalizes to this setup. The elasticity of doctors' income inequality with respect to widget makers' income inequality is $\left(1 + \frac{\varepsilon^S}{\alpha_x}\right)^{-1}$. This expression decreases with the supply elasticity of health services, and in the limit when $\varepsilon^S = 0$ (i.e. health care services are not scalable), we recover the exact result of Proposition 1. Intuitively, more elastic supply from each doctor leads to increased supply of healthcare quality, especially at the top. This reduces the pass-through from widget makers' inequality into price inequality and thus into doctors' income inequality—despite the increase in supply at the top.

Proof. Combining (2), (40) and (41), we obtain the differential equation:

$$\omega'(z)\, z + \frac{\beta}{1-\beta}\omega(z) = \frac{\beta}{1-\beta} x_{\min} \left(\int_z^{\infty} \left(\frac{\omega(\zeta)}{k_1}\right)^{\varepsilon^S} \alpha_z \frac{1}{\zeta} \left(\frac{\zeta}{z_c}\right)^{-\alpha_z} d\zeta \right)^{-\frac{1}{\alpha_x}}. \tag{42}$$

We verify that a solution to the problem of the form $\omega(z) = C_1 z^{\psi}$ exists. Plugging this expression in (42), 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 (42) 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 shape parameter $\alpha_w = \frac{\alpha_x+\varepsilon^S}{1+\varepsilon^S}$, so $\frac{\partial \ln \alpha_w}{\partial \ln \alpha_x} = \left(1+\frac{\varepsilon^S}{\alpha_x}\right)^{-1}$. □

A.8 Partly tradable health care

Consider the set-up of section 3.2.4. 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.5.1) 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 the sample of Census/ACS respondents to those who are age 25 or older. We further restrict the sample to individuals that either (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 labour 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.

Construction of Figure 1 For both 1980 and 2012 we calculate two sets of statistics. First, the percentiles of average occupational log wage income where occupations are weighted by occupation size. Then within each of these percentile groups, we rank individual observations based on income and calculate 500 bins of the distribution of the difference between average log earnings in that bin and the corresponding occupation-percentile. The green series (with circles) shows the actual change from 1980 to 2012. The "Between-Occupation Effects Only" series (with red triangles) shows the counter-factual distribution if occupational percentiles' wage 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 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.

Weighing All constructed variables use individual census weights ("perwt").

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

1. Among *uncensored* observations, calculate the income at that percentile in that region-year among all persons regardless of occupation. Drop all 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, correcting for censoring.

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

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

1. Among *uncensored* observations, calculate the income at that percentile in that region-year among all people regardless of occupation. Drop all observations with income below that income level.
2. Then, keep only observations from the outcome occupation of interest (*i.e.* doctors).
3. Then, calculate the inverse Pareto parameter correcting for censoring.

Instrument Construction in the Regression Analysis To construct the inequality measures entering the shift-share instrument, we first 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 separately for each occupation. To calculate the weights (or shares) in the Bartik, for each region s, we identify the 20 most common occupations in the upper tail of that region in 1980, where upper tail is defined in the same way as described above. For each of the top 20 occupations, we calculate the fraction of the upper tail population that is employed in each occupation in 1980. These population shares are the weights in the Bartik, as described in the main text. When we estimate the log-log form of our regressions, these weights (among the top 20 occupations) are normalized to sum to 1 by dividing each by their collective sum. When the regressions are estimated in level-level form, the weights are not normalized, and thus sum to less than one. Retirees are excluded from the instrument.

Minimum required number of observations The asymptotic variance of our maximum likelihood estimate of $\hat{\alpha}$ is α^2/N, which is decreasing in N such that few observations imply a higher estimated variance. In addition, the variance estimate suffers from small-sample bias. A ratio of the standard error to the estimate of the maximum likelihood estimator, $\hat{\sigma}/\hat{\alpha}$, of less than 20% would require 25 observations. Monte Carlo simulations also show that the average absolute deviation of the estimated variance with a bootstrapped estimate exceeds 20% when the number of observations is less than around 25 (results not shown). These two results guide our choice of 25 as the cutoff but we perform robustness checks for other values.

Regressions 1. Weight: Regressions are weighed by the number of observations in the outcome occupation above the cutoff in that region-year.

2. Sample: Only LMAs that have at least 25 outcome occupation observations above the cutoff in *all years* are included so that the regression sample is balanced.

3. We estimate the regressions with the Stata add-on commands reghdfe and ivreghdfe.

4. Standard errors are clustered at the LMA level.

B.2 Crosswalks

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 all engineering occupations into one, all managers (excluding those working in real estate) into one, and combine primary and secondary school teachers together.

Geography: Labor Market Areas We compute local inequality within 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. 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

B.3.1 O*NET

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 5.

B.3.2 Blinder

We rely on the measure of offshorability in Blinder (2009) to act 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 (at least partially offshorable). When two or more O*NET occupation codes correspond to a single SOC code, and those ONET occupations are deemed to be substantially different in their offshorability, he keeps the ONET occupation codes separate, as opposed to aggregating them to the SOC code

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

Occupation	98th to 90th percentile ratio 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

level. In the few cases in which that occurs, he only reports the ONET 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 ONET 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 (offshorability index of 75), and does not list which of the three O*NET sub-types it represents, just the higher-level SOC code.

16 of our 28 occupations are not offshorable (score less than 25). We map this to the IV spillovers in two ways: First, we rescale the score for our 28 occupations (including a 0 for those that are not offshorable) to be in percentile rankings, where each occupation is assigned their percentile ranking in the occupation distribution of offshorability (weighted by occupation size). This measure underlies Panel 5a. Second, we group occupations into those deemed non-offshorable and those that are offshorable to some extent. These two groups have statistically distinct average spillover estimates as discussed in footnote 42.

C Empirical Appendix

C.1 Additional tables of descriptive statistics

Table B.1 shows the 98^{th} to 90^{th} percentile for selected occupations in 1980 and 2012 – including physicians, dentists and real estate agents, as well as for all for the overall population. Table C.1 gives descriptive statistics for the largest occupations in the top of the income distribution for the year 2000.

Table C.1: Descriptive Statistics for Top Occupations in 2000. Wage Income

Occupation	Mean income $1000	Occupations share in: 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: This table shows basic descriptive statistics 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. 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.

C.2 Additional empirical results

We next show the results discussed in section 6.2. In Table C.2, we first replace the left hand side with earned income as the sum of wage and business income. Column (1) shows the OLS regression and Column (2) the IV. Alternatively, we control for the specialties of physicians. Columns (3) and (4) report OLS and IV coefficients, respectively.

Table C.3 considers a linear specification for physicians, dentists and real estate agents, respectively. Coefficients show little change. The coefficient on the IV regression for physicians differs from zero with $p = 0.105$.

Table C.5 shows regressions for physicians wherein we exclude, one at a time, the occupations with the highest Rotemberg weights from the instrument. The Rotemberg weights are computed for a linear specification which includes a constant set of occupations. For each LMA in 1980, we find the top 12 occupations in that LMA in the upper tail. We then build our Bartik instrument with the union of those top 12 across all LMAs in the regression sample. The weights are reported in Table C.4.

Table C.6 changes cutoffs in three dimensions. Column (1) gives the result where we only rely on the top 5% of the local population when calculating our inverse Pareto parameter (instead of 10%). Columns (2) and (3) uses LMAs with at least 40 or 15 observations, respectively. In our baseline regressions we use 25. Columns (4) and (5) change the number of occupations in

Table C.2: Robustness checks for Physicians

	Earned income		Specialities	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
$\ln(\alpha_{-o}^{-1})$	0.09	1.57**	0.17**	1.12**
	(0.08)	(0.76)	(0.07)	(0.55)
ln(Average Income)	−0.43***	−0.67***	−0.42***	−0.57***
	(0.10)	(0.15)	(0.08)	(0.11)
ln(Population)	0.01	0.11	−0.01	0.05
	(0.03)	(0.08)	(0.02)	(0.05)
Sh. neurosurgeons			3.53	2.60
			(2.47)	(3.00)
Sh. higher earning specialties			2.28***	2.17**
			(0.77)	(0.95)
Sh. lower earning specialties			−0.80***	−0.38***
			(0.24)	(0.35)
Sh. unequal earning specialties			−1.21***	−0.60***
			(1.26)	(1.74)
LMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	750	750	750	750
F-Statistic		7.59		8.48

Notes: This table includes OLS and IV regressions for physicians including controls. Column (1) is the OLS regression where the dependent variable is calculated using earned income (instead of wage income). The remaining variables are unchanged. Column (2) is the corresponding IV regression. Column (3) is the OLS where dependent variable is wage income, but four controls for specialties of physicians are included (see text for details). Column (4) shows the corresponding IV regression. N is the number of observations rounded to the nearest integer divisible by 50. *p<0.1, **: p<0.05, ***: p<0.01.

Table C.3: Linear specification

	Physicians		Dentists		Real estate	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
α_{-o}^{-1}	0.44***	1.85	1.16***	2.40***	1.37***	1.65***
	(0.13)	(1.14)	(0.36)	(0.81)	(0.21)	(0.55)
ln(Average Income)	−0.32***	−0.45***	0.10	0.02	0.12	0.09
	(0.08)	(0.12)	(0.14)	(0.14)	(0.09)	(0.08)
ln(Population)	−0.03	0.01	0.02	0.05	0.00	0.01
	(0.03)	(0.04)	(0.05)	(0.07)	(0.03)	(0.03)
Share upper tail not included in Bartik*Y	Yes	Yes	Yes	Yes	Yes	Yes
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	750	750	150	150	200	200
F-Statistic		7.13		13.80		7.38

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. Regressions are run using a linear specification instead of log-on-log, but all other details remain the same. Regressions include a control for the share of employment of the local employment not captured by the occupations in the instrument, interacted with a year dummy. Columns (1), (3) and (5) show the OLS regressions and columns (2), (4) and (6) show the IV regressions. *p<0.1, **: p<0.05, ***: p<0.01.

Table C.4: Occupations with the largest Rotemberg weights for physician regressions

Occupation	Weights	Occupation	Weights
Financial service sales occupations	0.34	Managers, Excl. Real Estate	0.06
Financial managers	0.22	Real estate sales occupations	0.05
Lawyers and judges	0.19	Computer systems analysts and	
Production supervisors or foremen	0.16	computer scientists	-0.06
Airplane pilots and navigators	0.16	Tool and die makers and die setters	-0.06
Other financial specialists	0.13	Driver/sales workers and truck Drivers	-0.08
Sales occupations and sales representatives	0.10	Engineers	-0.10
Accountants and auditors	0.06	Subject instructors, college	-0.13

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

Table C.5: Physician regressions: excluding occupations with the highest Rotemberg weights

Excluding:	**Financial services sales occupations**	**Financial managers**	**Lawyers and judges**	**Production supervisors and foremen**	**Airplane pilots and navigators**
	(1)	(2)	(3)	(4)	(5)
$\ln(\alpha_{-o}^{-1})$	1.54**	1.61**	1.54**	1.63**	1.56**
	(0.73)	(0.72)	(0.70)	(0.74)	(0.74)
ln(Average Income)	−0.61***	−0.62***	−0.61***	−0.63***	−0.62***
	(0.15)	(0.15)	(0.14)	(0.15)	(0.14)
ln(Population)	0.07	0.08	0.07	0.08	0.07
	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
LMA FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	750	750	750	750	750
F-Statistic	6.90	7.92	6.91	6.73	8.10

Notes: This table shows IV regressions for physicians when we in turn exclude the 5 occupations with the highest Rotemberg weights from the instrument. These occupations are Financial services sales occupations (0.34), Financial managers (0.22), Lawyers and judges (0.19), Production supervisors or foremen (.16), and Airplane pilots and navigators (0.16). The dependent variable is logarithm of inverse Pareto parameter, $\log(\alpha_o^{-1})$, for physicians in all regressions. All other details are as in Table 3. *p<0.1, **: p<0.05, ***: p<0.01.

the instrument to 30 or 15. Panel A shows results for physicians, Panel B for dentists and Panel C for real estate agents.

C.3 Shift-share robustness checks

We follow the framework of Goldsmith-Pinkham et al. (2020) and assume that the occupational shares are not correlated with changes in the outcome variable other than through their effect on local top income inequality. We, however, deviate from the standard setting in two ways. First, our preferred specification uses log of income inequality instead of levels. As shown in Appendix Table C.3, the linear specification somewhat reduces the significance of our results for physicians, but with little effect in the point estimate. The results for dentists and real estate agents are, if anything, even stronger.

Second, to improve the power of the instrument and to take into account that the set of dominant occupations varies across LMA, we let the set of (20) occupations in the instruments differ across LMAs. If we use the same set of occupations, the first stage regression is considerably weaker. In Appendix Table C.7, we show that this is driven by smaller LMAs. We do so by considering the same set of 20 occupations for all LMAs (and include a control for the share of the population not included in the instrument, as suggested by Goldsmith-Pinkham et al., 2020), but with a focus on only bigger LMAs. We do so by restricting attention to those with at least 120 physicians. We get a number of LMA×year observations similar to those of dentists and real estate agents and we obtain a much stronger first stage F-statistic of 15.36. Presumably, our mechanism is more relevant in larger areas with a broader range of medical services. This specification also permits us to implement the Adão et al. (2019) standard errors. Results for physicians and dentists remain significant.

C.4 Occupational characteristics

Table C.8 shows the regressions results associated with the scatter plots of Figure 5. All coefficients show the expected sign (the p-value for the importance of customer service is 0.11.)

Table C.6: Alternative cutoffs

	Top 5 pct. (1)	**Only ≥40 obs.** (2)	**Only ≥15 obs** (3)	**Using 30 occ.** (4)	**Using 15 occ.** (5)
		Panel A: Physicians			
$\ln(\alpha_{-o}^{-1})$	1.34***	1.78**	1.38**	1.77**	1.53**
	(0.43)	(0.86)	(0.68)	(0.86)	(0.62)
ln(Average Income)	−0.80***	−0.63***	−0.58***	−0.65***	−0.61***
	(0.20)	(0.16)	(0.13)	(0.17)	(0.15)
ln(Population)	0.10	0.08	0.06	0.09	0.07
	(0.06)	(0.09)	(0.06)	(0.08)	(0.07)
LMA FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	750	550	1100	750	750
F-Statistic	9.36	5.60	7.11	6.98	7.22
		Panel B: Dentists			
$\ln(\alpha_{-o}^{-1})$	0.67	3.18	1.67**	1.84*	2.34*
	(0.51)	(2.98)	(0.82)	(1.00)	(1.25)
ln(Average Income)	−0.10	−0.04	−0.09	0.00	−0.04
	(0.30)	(0.43)	(0.21)	(0.22)	(0.25)
ln(Population)	0.02	0.41	0.05	0.05	0.07
	(0.11)	(0.46)	(0.11)	(0.14)	(0.18)
LMA FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	150	100	300	150	150
F-Statistic	10.32	2.48	7.19	4.54	4.72
		Panel C: Real Estate Agents			
$\ln(\alpha_{-o}^{-1})$	0.66	1.34**	1.19**	1.34*	1.50***
	(0.41)	(0.57)	(0.59)	(0.80)	(0.57)
ln(Average Income)	0.17	0.04	0.15	0.10	0.08
	(0.23)	(0.17)	(0.16)	(0.18)	(0.18)
ln(Population)	0.01	−0.02	0.00	0.01	0.01
	(0.07)	(0.06)	(0.05)	(0.06)	(0.06)
LMA FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
N	200	150	200	200	200
F-Statistic	10.22	7.15	6.80	6.80	7.05

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) shows a regression where we use the top 5% (instead of 10%) of the income distribution (for the dependent variable, the independent variable and the construction of the IV). This regression keeps the same number of LMAs so that number of observations underlying inequality measures is lower. Columns (2) and (3) use a cutoff of 40 and 15, respectively, for the selection of LMAs (instead of 20). Columns (4) and (5) use 30 and 15 occupations in the construction of the instrument (instead of 20). *p:<0.1, **: p<0.05, ***: p<0.01.

Table C.7: Regressions with Adão et al. (2019) standard errors

	Physicians		Dentists		Real estate	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
α_{-o}^{-1}	1.67**	1.89*	1.60*	2.09*	1.11*	1.33
	(0.68)	(0.57)	(0.88)	(0.75)	(0.64)	(0.65)
		[1.09]		[1.11]		[1.2]
ln(Average Income)	−0.63***	−0.51***	0.02	0.05	0.15	0.07
	(0.15)	(0.12)	(0.21)	(0.12)	(0.18)	(0.08)
ln(Population)	0.06	0.02	0.04	0.07	0.00	0.00
	(0.09)	(0.04)	(0.12)	(0.07)	(0.06)	(0.03)
lag Excluded share*Y	No	Yes	No	Yes	No	Yes
LMA FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	200	200	150	150	200	200
F-Statistic	5.56	15.36	7.11	14.62	9.75	8.45

Notes: OLS and IV regressions for selected occupations. The specification is linear and includes controls for average income, population and the share of general population that is not included in the instrument interacted with a year dummy. Standard errors in squared brackets are Adão et. al. (2019) standard errors. Columns (1), (3) and (5) are OLS regressions and columns (2), (4) and (6) are IV regressions. Columns (1) and (2) are for physicians, columns (3) and (4) are for dentists and (5) and (6) are for real estate occupations. F-stats for IV regressions refer to the first stage regressions. *p<0.1, **: p<0.05, ***: p<0.01 using Adão et. al. (2019) standard errors.

Table C.8: Spillover t-stats and occupational characteristics

	(1)	(2)	(3)	(4)	(5)
Offshore (Blinder)	−1.36*				
	(0.54)				
Customer service - level		2.11*			
		(0.78)			
Customer service - importance			1.53		
			(0.79)		
Working with public - level				1.76*	
				(0.76)	
Working with public - importance					2.11**
					(0.73)
N	28	28	28	28	28

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 occupations. These 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 measures are rescaled as percentiles. We use 28 occupations which are those amongst the biggest 30 occupations in the top 10% with at least 20 LMAs for the IV regressions. *: p<0.1, **: p<0.05, ***: p<0.01.

Additional Reference

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.

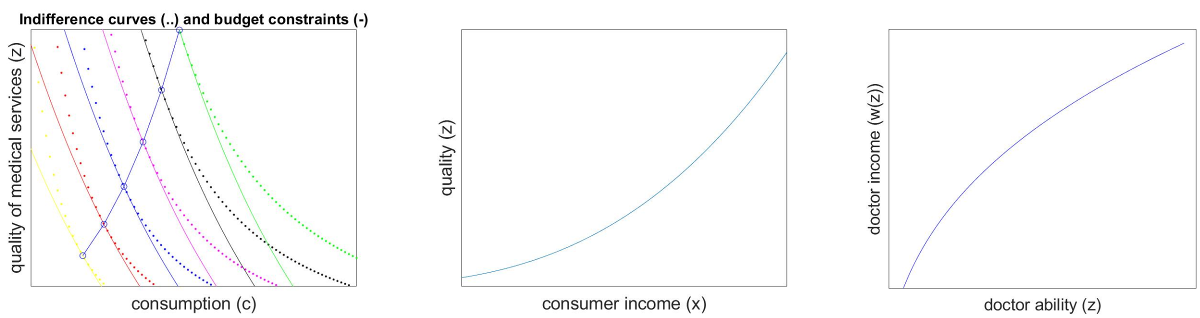
D Supplementary Material for: "The Spillover Effects of Top Income Inequality"

Joshua D. Gottlieb, David Hémous, Jeffrey Hicks, and Morten Olsen
June 2023

D.1 Graphical representation of the model

Figure D.1: Equilibrium in the baseline model when $\alpha_z < \alpha_x$

(a) Offer curve **(b)** Matching Function **(c)** Doctor ability and pay



Notes: This figure illustrates the matching mechanism in the model when $\alpha_z < \alpha_x$. Panel (a) shows the budget sets and indifference curves for six different consumers, along with the matching function that this equilibrium generates. The horizontal axis shows consumption c of the homogeneous good, and the vertical axis shows the quality of physician z that the consumer obtains. The dotted curves represent the indifference curves, and the solid curves the budget constraints. The budget constraints are curved because there is not a constant price per unit of quality; in this example, additional units of quality have decreasing cost. So, for any given budget constraint, the constraint flattens as we move to the left. Because there is much more skill dispersion among doctors than among consumer (the ability distribution of doctors has a fatter tail), the higher-income consumers have a increasingly abundant medical services. Consequently, the quality of physician is convex in consumer income (Panel (b)). Panel (c) shows the income of physicians.

We illustrate equilibrium of the baseline model in Figure D.1. Panel D.1a shows the budget sets and indifference curves for six consumers with different income levels. For this illustration we choose $\alpha_z < \alpha_x$; that is, skill inequality among physicians is higher than among widget makers. For each consumer, the horizontal axis shows consumption c of the homogeneous good, and the vertical axis shows the quality of the physician z that the consumer obtains. The dotted curves represent the indifference curves, and the solid curves the budget constraints. With Cobb-Douglas utility, the higher indifference curves are proportionally scaled versions of lower indifference curves—the slopes are constant on any ray out from the origin. But the budget constraints behave differently from those with constant relative prices: they are curved because there is not a constant price per unit of quality. In this example, additional units of quality have decreasing cost, due to the relative abundance of skill for physicians in the top. So, for any given budget constraint, the budget line becomes steeper as we move to the left. Income differences lead to parallel shifts left or right in the budget constraint. As a result, the slopes at which the indifference curves are tangent to the budget constraint change for different budget constraints. Consequently, the curve that traces out optimal bundles for increasing income is not a straight line from the origin, as would be the case for Cobb-Douglas utility and constant prices, but instead a convex function. Panel D.1b shows the matching function: Since $\alpha_z < \alpha_x$ it must match widget makers with increasingly high-skill physicians, implying a convex matching function. Finally, Panel D.1c shows the income of physicians with a given ability. With divisible services, Panel D.1c would be a straight line.

D.2 Disentangling supply side and demand side effects

As described in the text of Section 3.2.1, in the model with occupational mobility, doctors and widget makers interact through both demand and outside option effects. To disentangle these two effects, we now build a model where doctors have an outside option positively correlated with their ability but where patients are a separate group. Formally, there are two types of agents: a mass 1 of consumers, with income x distributed with the Pareto distribution $P(X > x) = (x_{\min}/x)^{\alpha_x}$ and a mass M of potential doctors. Consumers consume the homogeneous good and health care services according to the utility function (1). Potential doctors only consume the homogeneous good. As in section 3.2.1, they are ranked in descending order of ability and we denote i their rank. Agent i can choose between being a doctor providing health 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:

$$w(z) = y_{\min}(z/z_{\min})^{\alpha_z/\alpha_y}. \tag{43}$$

Market clearing for health care services above z implies:

$$\left(\frac{x_{\min}}{m(z)}\right)^{\alpha_x} = \lambda M \int_z^\infty \mu(\zeta)\, g_z(\zeta)\, d\zeta, \tag{44}$$

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 (43), we obtain:

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

Taking the derivative with respect to z, we get:

$$\mu(z) = \frac{\alpha_x}{\alpha_y} \frac{1}{\lambda M} \left(\frac{\frac{\beta}{1-\beta} \lambda x_{\min}}{\left(\frac{\alpha_z}{\alpha_y} + \frac{\beta}{1-\beta} \right) y_{\min}} \right)^{\alpha_x} \left(\frac{z}{z_{\min}} \right)^{\alpha_z \left(1 - \frac{\alpha_x}{\alpha_y}\right)}. \tag{46}$$

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 $\alpha_y = \alpha_x$, $\frac{\alpha_x}{\alpha_y} \frac{1}{\lambda M} \left(\frac{\alpha_y \beta \lambda x_{\min}}{(\alpha_z(1-\beta) + \beta \alpha_y) y_{\min}} \right)^{\alpha_x} \leq 1$). We then obtain that doctors' income distribution obeys (for w high enough):

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

Therefore doctors' income is distributed like the patients' income and not according to doctors' outside option.

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. This result, however, relies on the Cobb-Douglas assumption and does not generalize to a CES case.