The Spillover Effects of Top Income Inequality*

Jeffrey Clemens† Joshua D. Gottlieb‡ David Hémous§
Morten Olsen¶

November 2016

Abstract

Since the 1980s top income inequality within occupations as diverse as bankers, managers, doctors, lawyers and scientists has increased considerably. Such a broad pattern has led the literature to search for a common explanation. In this paper, however, we argue that increases in income inequality originating within a few occupations can "spill over" into others creating broader changes in income inequality. In particular, we study an assignment model where *generalists* with heterogeneous income buy the services of *doctors* with heterogeneous ability. In equilibrium the highest earning generalists match with the highest quality doctors and increases in income inequality among the generalists feed directly into the income inequality of doctors. We use data from the Decennial Census as well as the American Community Survey from 1980 to 2014 to test our theory. Specifically, we identify occupations for which our consumption-driven theory predicts spill-overs and occupations for which it does not and show that patterns align with the predictions of our model. In particular, using a Bartik-style instrument, we show that an increase in general income inequality causes higher income inequality for doctors, dentists and real estate agents; and in fact accounts for most of the increase of inequality in these occupations.

*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 thanks the Social Sciences and Humanities Research Council of Canada (430-2016-00030). We thank Innessa Colaiacovo and Andrew Vogt for excellent research assistance.

†University of California, San Diego and NBER

‡University of British Columbia and NBER

§University of Zürich and CEPR

¶IESE Business School, University of Navarra.

JEL: D31; J24; J31; O15

KEYWORDS: Income inequality, Assignment model, Occupational inequality, Decennial Census

1 Introduction

Since the 1980s the share of total earnings going to the top of the income distribution has increased considerably. At the same time income inequality within the top has also increased with a higher share of top earnings going to the very high earners. Moreover this pattern holds within high-earning occupations so that the overall growth of top income inequality is not simply due to the growth within particular occupations (Bakija, Cole, and Heim, 2012). As argued by Kaplan and Rauh (2013), this broad pattern suggests that a plausible explanation whether it be globalization, deregulation, changes to the tax structure, or technological change, would have to apply to occupations as diverse as financial managers, doctors, and CEOs. In this paper we argue that this need not be the case by showing that exogenous increases in income inequality within one occupation "spill over" into others through the former's consumption, driving up income inequality for a broader set of occupations.

This paper provides a model where changes in within-occupation income inequality propagates to other occupations through consumption rather than competition for skill in the broader labor market. We study an assignment model where *generalists* with heterogeneous income buy the services of *doctors* with heterogeneous ability. In equilibrium the highest earning generalists match with the highest ability doctors and increases in income inequality among the generalists feed directly into the income inequality of doctors. Two conditions on the services provided by doctors are necessary for the equilibrium to feature an assignment mechanism and thereby income inequality spillovers: *heterogeneity* and *non-divisibility* in output (one high-ability doctor is not the same as two decent-ability doctors). We focus on physicians, dentists and real estate agents who meet these conditions and contrast them with occupations who do not. Using data from the Decennial Census and the American Community Survey, we find that an increase in general income inequality causes an increase in inequality for these occupations, with a spillover elasticity ranging from 0.5 to 2.7. These occupations are important within the top 1%, in fact Physicians are the most common Census-occupation in the top 1% in 2014.

We first present our baseline model and demonstrate that for occupations of heterogeneous ability where production is not scalable (that is, no mechanism exists that would allow the more talented to scale up output) and consumption is non-divisible, the income distribution is tightly linked to that of the general population. Specifically, generalists of heterogeneous ability produce a homogeneous product in quantity proportional to their skill level. Besides this homogeneous product, each generalist consumes the services of one doctor. Doctors also have heterogeneous ability but their ability translates proportionately into the quality of the services they provide and not the quantity. Instead, all doctors service the same number of patients. The abilities of both generalists and doctors are Pareto distributed but with different parameters. The result is an assignment model with positive assortative matching in which the highest ability generalists match with the highest ability doctors. An exogenous mean-preserving spread in the income inequality of generalists increases the number of high-earning generalists, increases the demand for the best doctors and increases top income inequality among doctors as well. In fact, in the special case of Cobb-Douglas utility, top income inequality of doctors is entirely driven by the earnings distribution of generalists and is independent of the underlying ability distribution of doctors.

We extend the model in three directions. First, we allow for occupational mobility at the top: high-ability doctors can choose to be high-ability generalists and vice-versa. Since changes in top income inequality for generalists completely translate into changes in top income inequality for doctors when there is no occupational mobility, allowing doctors to switch occupation or not has no impact on doctors' inequality and the two settings are observationally equivalent. Second, we consider two regions in one nation that differ only in the top income inequality of generalists and allow patients to import their medical services. We show that top income inequality among doctors for each region must follow generalist top income for the most unequal region. This distinction will be important for the empirical test: When a service is local, i.e. non-tradable, spill-over effects will happen at the local level, whereas for tradable services the spill-over effect will happen at the national level. Finally, we let doctors move across regions and show that the most unequal region will attract the most able doctors, but, as in the baseline model, doctors' inequality is determined by general inequality in the region where they eventually live. Hence the observed top income inequality of doctors is the same whether they can move or not.

To test our model we take as a starting point the fact that top income inequality has increased broadly across occupations. The solid line of Figure 1 shows that the relative income of those in the top 0.1% relative to the 1% of the income distribution has risen from 3.1 in 1979 to 4.3 in 2005. The pattern is similar for occupations as diverse as doctors, real estate agents, and scientists and in fact holds for a number of occupations with incomes mostly below the top 1% such as college professors and secretaries (this is not a result of large changes in the occupational distribution in the top 1% and top 0.1%. Except for financial professionals whose weight in the top has increased substantially, the distribution of the largest occupations in the top has remained relatively constant from 1979 to 2005). We test our theory using a combination of the Decennial Census and the American Community Survey for every decade since 1980 and we focus on labor market areas (an aggregation of commuting zones, Dorn; 2009) as the unit of analysis.[1] We construct measures of top income inequality that are specific to the year/occupation/labor market area, but since the income data are top-coded for around 0.5 per cent of observations — and therefore significantly more for high-earning occupations — we impose an assumption of a Pareto distributed right tail of the income distribution and use the exponential parameter of the estimated Pareto distribution as an (inverse) measure of income inequality. Our theory predicts which occupations will feature spill-overs: those with non-divisibility in output. Furthermore, since we will focus on geographical variation across the United States, our estimation methodology will only pick up spill-over effects if they are local, that is if workers mostly service local clients. We classify occupations into two groups: Those that meet these conditions (such as physicians, dentists and real estate agents) and those that do not (such as financial managers, college professors and secretaries.) Using panel data, in OLS regressions, we find that an increase in general income inequality (excluding the occupation of interest) is positively correlated with an increase in inequality for occupations in the first group, but mixed results for the other occupations. Naturally state-specific changes in regulation, labor demand or taxes might cause both occupation-specific income inequality to increase at the same time as general income inequality. To establish a causal link, we use a Bartik (1991)-style instrument. We construct a weighted average of nationwide inequality for the 20 occupations that are the most represented in the top 5% nationwide (excluding the occupation of interest). The weights correspond to the labor market area-specific relative importance of each occupation in the beginning of our sample. In other words, we only exploit the changes in labor market income inequality that arises from the occupational distribution in 1980 combined with the nationwide trends in occupational inequality. This weighted average serves as our instrument for general inequality in the area in question.

Using this instrument, we find a very clear distinction between the first group and other occupations. An increase in general income inequality at the local level causes an

[1] At the moment we use the publicly available data downloaded from IPUMS (Ruggles et al., 2015). We are in the process of obtaining access to the uncensored data from the Decennial Census as well as the American Community Survey.

increase in inequality for physicians, dentists and real estate agents, who operate in local markets. The parameter estimates are consistent with the majority of the increase in income inequality for these occupations being explained by increases in general income inequality. On the other hand, we find that *local* general income inequality does *not* spill over to financial managers, college professors and secretaries. This contrasts our theory of consumption-driven spill-overs of income inequality with a theory driven by broad increases in demand for skills: Such a theory would predict spill-over effects across a broad set of occupations, whereas we predict such changes only for local non-divisible occupations, in line with the empirical evidence.

The increase in top income inequality has inspired substantial scholarship (see notably Piketty and Saez, 2003 and 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, 2015, for some of the most recent evidence, and Pareto, 1896, for the earliest). Further, Jones and Kim (2014) show that the increase in top income inequality is linked with a fattening of the right tail of the income distribution, which corresponds to a decrease in the shape parameter of the Pareto distribution. This literature is related, but distinct from the large literature on skill-biased technological change and income inequality which seeks to explain changes in income inequality throughout the income distribution and primarily *across* occupations (Goldin and Katz, 2010; Acemoglu and Autor, 2011).

More specifically, our paper builds on the literature on "superstars", which originated with Rosen (1981), who explains how small differences in talent may lead to large differences in income. The key element in his model is an indivisibility of consumption result 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. Specifically, if $w(z)$ denotes the income of an individual of talent z, $p(z)$ denotes the average price for his services, and $q(z)$ is the quantity provided, such that $w(z) = p(z)q(z)$, "superstars" are mostly characterized by very large markets (a large $q(z)$). This makes such a framework poorly suited for occupations where output is not easily scalable, such as doctors. In contrast we focus on such occupations and study

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

an assignment model that is "one-to-one" (or more accurately "a constant-to-one") where superstars are characterized by a large price for their services $p(z)$. This makes our paper closer to Gabaix and Landier (2008) who build a "one-to-one" assignment model to study CEO's compensation. They argue that since executives' talent increases the overall productivity of firms, the best CEO's are assigned to the largest firms, and show empirically that the increase in CEO's compensation can be fully attributed to the increase in firms' market size (Grossman (2007) builds a model with similar results). Along the same line, Määttänen and Terviö (2014) build an assignment model to study house price dispersion and income inequality, they calibrate their model to six US metropolitan areas and find that the increase in inequality has led to an increase in house price dispersion. Gabaix, Lasry, Lions and Moll (2015) argue that the fast rise in both the share of income held by the top earners and income inequality among these earners requires aggregate shocks to the return of high income earners ("superstars shocks"). Our analysis suggests that even if such shocks only directly affect some occupations they will spill over into other occupations. The original shock may arise from technological change in occupations where span of control features are pervasive as suggested by Geerolf (2015).[3] In addition, several papers have looked at how globalization can increase the share of income going to the top earners and also increase inequality among these earners (see Manasse and Turini, 2001; Kukharskyy, 2012; Gesbach and Schmutzel, 2014 and Ma, 2015[4]).

Beyond "superstars" effects, the economic literature has investigated several possible explanations for the rise in top income inequality: for instance, Jones and Kim (2014) and Aghion, Akcigit, Bergeaud, Blundell and Hémous (2015) look at the role played by innovation;[5] Piketty (2014) argues that top income inequality has increased because of the high returns on capital that a concentrated class of capitalists enjoy; Piketty, Saez and Stantcheva (2014) argue that low marginal income tax rates divert manager's

[3]Geerolf (2015) builds a span of control model to micro-found the fact that firms' size distribution follows Zipf's law. His model naturally leads to "superstars" effects and a bounded distribution of talents can lead to an unbounded distribution of income. Similarly, Garicano and Hubbard (2012) build a span of control model which features positive assortative matching as the most skilled individuals become the most skilled managers who manage large firms which employ the most skilled workers. They use data from the 1992 Census of services on law offices to find support for their model. Yet, span of control issues do not seem *directly* relevant for doctors or real estate agents.

[4]Yet, none of these papers are able to generate a change in the shape parameter of the Pareto distribution of top incomes through globalization.

[5]Jones and Kim (2014) build a model close to the superstars literature where the distribution of income for top earners is Pareto and results from two forces: the efforts of incumbents to increase their market share and the innovations of entrants who can replace incumbents. Using a panel analysis of US states, Aghion et al. (2015) show empirically that an increase in innovation leads to more top income inequality.

compensation from perks to wages and increase their incentive to bargain for higher wages and Philippon and Reshef (2012) emphasize the role played by the financial sector.

Section 2 presents the theoretical model, Section 3 describes our empirical strategy and data, Section 4 gives our empirical results and Section 5 concludes.

2 Theory

We first present our baseline model and demonstrate that for occupations of heterogeneous ability where production is not scalable (that is, no mechanism exists that would allow the more talented to scale up output) and consumption is non-divisible, the income distribution is tightly linked to that of the general population. To help guide our empirical analysis and demonstrate when we would expect to see spill-over effects in the data, we then relax a number of assumptions. Throughout "doctors" will represent occupations where the most skilled workers can produce a good of higher quality but cannot serve more customers than the less skilled, and where customers cannot divide their consumption across several producers: one high-ability doctor is not the same as two low-ability doctors. Besides doctors, prominent examples are dentists, college professors, and real estate agents.

2.1 The baseline model

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

Production. Generalists produce a homogeneous good, x, which serves as the numeraire. They differ in their ability to produce such that a generalist of ability x can produce x units of the homogeneous good. The ability distribution is Pareto such that a generalist is of ability $X > x$ with probability:

$$P(X > x) = \left(\frac{x_{\min}}{x}\right)^{\alpha_x},$$

with lower bound $x_{\min} = \frac{\alpha_x - 1}{\alpha_x}\widehat{x}$ and shape $\alpha_x > 1$, which keeps the mean fixed at $\widehat{x}$ when α_x changes. The parameter α_x is an (inverse) measure of the spread of abilities. We will keep α_x exogenous throughout and will capture a general increase in top income inequality by a reduction in α_x. Doctors produce health services and can each serve λ customers, where we impose $\lambda \geq \max(1, 1/\mu_d)$ such that there are enough doctors to serve everyone. Potential doctors differ in their ability z, according to a Pareto distribution of shape α_z such that they will have ability $Z > z$ with probability:

$$P(Z > z) = \left(\frac{z_{min}}{z}\right)^{\alpha_z},$$

where all potential doctors can alternatively work as generalist and produce the homogeneous good with ability x_{min} (see Section 2.2 for a model where doctors' and generalists' abilities are perfectly correlated). Though the ability of a doctor does not change how many patients she can take care of, it increases the utility benefit that patients get from the health services that are provided.

Consumption. Generalists consume the two goods according to the Cobb-Douglas utility function

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

where c is the consumption of homogeneous good and z is the quality of the health care (equal to the ability of the doctor providing it).[6] The notion that medical services are not divisible is captured by the assumption that each generalist needs to consume the services of exactly one doctor. This implies that there will not exist a common price per unit of quality-adjusted medical services. For simplicity, doctors only consume the homogeneous good, an assumption that can easily be generalized (see section 2.4.1 below).

2.1.1 Equilibrium

Generalists. Since a generalist of ability x produces x units of the consumption good, their income must be distributed like their ability. The consumption problem of generalist of ability x can then be written as:

$$\max_{z,c} u(z, c) = z^{\beta_z} c^{1-\beta_z},$$

$$\text{st } \omega(z) + c = x, \tag{2}$$

where $\omega(z)$ is the price of one unit of medical services by a doctor of ability z.

Taking first order conditions with respect to the quality of the health services con-

[6]For our purpose, one should think of z as the perceived quality of health care by the consumers at the time when they decide on a doctor. Whether this is the "true" quality of the health care provided or not does not matter for the predictions of the model.

sumed and the homogeneous good gives:

$$\omega'(z)\,z = \frac{\beta_z}{1-\beta_z}\left(x - \omega(z)\right). \tag{3}$$

Since no generalist spends all her income on health, this equation immediately implies that in equilibrium, $\omega(z)$ must be increasing such that doctors of higher ability earn more per unit of medical services. Importantly, the non-divisibility of medical services implies that doctors are "local monopolists" in that they are in direct competition only with the doctors with slightly higher or lower ability. As a consequence, doctors do not take prices as given implying that $\omega(z)$ will in general not be a linear function of z.

As a result, the equilibrium involves positive assortative matching between generalists' income and doctors' ability. We denote by $m(z)$ the matching function such that a doctor of ability z will be hired by a generalist whose income is $x = m(z)$ and $m(z)$ is an increasing function (see Appendix A.1 for a proof).

Doctors. Since there are (weakly) more doctors than needed the least able doctors will choose to work as generalists. We denote by z_c the ability level of the least able doctor who decides to provide health services so that $m(z)$ is defined over (z_c, ∞) and $m(z_c) = x_{\min}$ (the worst doctor is hired by a generalist with income $x_{\min}$). Then, market clearing at all quality levels implies that

$$P\left(X > m(z)\right) = \lambda\mu_d P\left(Z > z\right),\ \forall z \geq z_{min} \tag{4}$$

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 $m(z)$. If $\lambda\mu_d > 1$ then $z_c > z_{\min}$ and if $\lambda\mu_d = 1$ then $z_c = z_{\min}$.

Using the assumption that abilities are Pareto distributed, we can use (4) to obtain the matching function as:

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

Intuitively if $\alpha_z > \alpha_x$, so that top talent is 'scarcer' among doctors than generalists then the matching function is convex because it must assign increasingly relatively productive generalists to doctors. As $m(z_c) = x_{\min}$, we obtain the ability of the least able doctor working as a doctor: $z_c = (\lambda\mu_d)^{\frac{1}{\alpha_z}} z_{min}$, which is independent of the generalists' income distribution.

We denote by $w(z)$ 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. Furthermore as a potential doctor of ability z_c is indifferent between working as a doctor and in the homogeneous good sector earning a wage equal to x_{min}, we must have $w(z_c) = x_{\min}$. Now plugging the matching function in (3), we obtain the following differential equation which must be satisfied by the wage function $w(z)$:

$$w'(z)\, z + \frac{\beta_z}{1-\beta_z} w(z) = \frac{\beta_z}{1-\beta_z} x_{\min} \left(\frac{\lambda^{\alpha_x - 1}}{\mu_d}\right)^{\frac{1}{\alpha_x}} \left(\frac{z}{z_{min}}\right)^{\frac{\alpha_z}{\alpha_x}}. \tag{6}$$

using the boundary condition at $z = z_c$, we obtain a single solution for the wage profile of doctors which obeys (see Appendix A.2):

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

One can show that $w(z)$ is increasing in doctor's ability z as expected, with $w(z_c) = x_{\min}$. Intuitively, equation (7), consists of two parts: The first term which dominates for large z/z_c and ensures an asymptotic Pareto distribution and the second term which fulfills the indifference condition for the least able active doctor. Hence, for z/z_c large, we get that

$$w(z) \approx x_{\min}\frac{\lambda\beta_z\alpha_x}{\alpha_z(1-\beta_z)+\beta_z\alpha_x}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x}}. \tag{8}$$

Therefore, the wage schedule must be convex in z if $\alpha_z > \alpha_x$. To understand the intuition, consider again the case where top-talented doctors are scarce ($\alpha_z > \alpha_x$). This implies a fatter tail among generalists than doctors, such that a generalist of twice the income does not have a doctor of twice the ability. Hence, a linear schedule $\omega(z) \propto z$ cannot be an equilibrium as the Cobb-Douglas utility function would require a constant share spent on medical services, which would imply double the payment to a doctor that is not twice as good. For the same reason, the schedule cannot be concave when $\alpha_z > \alpha_x$.

We define $P_{doc}(W_d > w_d)$ as the probability that the wage of an actual doctor is higher than w (that is we only take into account the potential doctors who actually choose to work as doctors). We get that $P_{doc}(W_d > w_d) = \left(z_c/w^{-1}(w_d)\right)^{\alpha_z}$, so that using (8), for w_d large enough:

$$P_{doc}(W_d > w_d) \approx \left(\frac{x_{\min} \lambda \beta_z \alpha_x}{\alpha_z (1-\beta_z) + \beta_z \alpha_x} \frac{1}{w_d} \right)^{\alpha_x}. \tag{9}$$

That is, the income of (actual) doctors is distributed in a Pareto fashion at the top, with a shape parameter inherited from the generalists, independent of the spread of doctor ability, α_z. Similarly, the income distribution of potential doctors (denoted P_{pot_doc}) must then obey for w_d large enough $P_{pot\_doc}(W_d > w_d) \approx \frac{1}{\lambda \mu_d} \left(\frac{x_{\min} \lambda \beta_z \alpha_x}{\alpha_z(1-\beta_z)+\beta_z \alpha_x} \frac{1}{w_d} \right)^{\alpha_x}$. 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 generalists leads to an increase in inequality among doctors. In other words, the increase in top income inequality spills over from one occupation (the generalists) 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.[7] Formally:

Proposition 1. *Doctors' incomes are asymptotically Pareto distributed with the same shape parameter as the generalists'. In particular an increase in top income inequality for generalists increases top income inequality for doctors.*

Further, a decrease in the mass of potential doctors μ_d (or an increase in the mass of generalists which here has been normalized to 1) does not affect inequality among doctors at the top but it increases the mass of active doctors (z_c decreases) and their wages (as $w(z)$ increases if z_c decreases).

Health expenditures. Health care prices increase sharply at the top, in fact, thanks to the Cobb-Douglas assumption, we obtain that rich generalists spend close to a constant fraction of their income on health. Formally, a generalist with income x spend $w(m^{-1}(x))/\lambda$ in health services. Using (5), we obtain that his health spending of $h(x)$ must obey:

$$h(x) = \frac{\beta_z \alpha_x}{\alpha_z (1-\beta_z) + \beta_z \alpha_x} x + \frac{1}{\lambda} \frac{\alpha_z (1-\beta_z) + \beta_z \alpha_x (1-\lambda)}{\alpha_z (1-\beta_z) + \beta_z \alpha_x} x_{\min} \left(\frac{x}{x_{\min}} \right)^{-\frac{\alpha_x}{\alpha_z} \frac{\beta_z}{1-\beta_z}}. \tag{10}$$

Note that health care is a necessity if $\alpha_z (1-\beta_z) + \beta_z \alpha_x (1-\lambda) > 0$. This follows from

[7] 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 generalist income, then all doctors would have benefited.

the price gradient consumers face (equation 7) which follow from the prices low-quality doctors charge — determined by the indifference condition for the lowest doctor z_c — and the pricing the high-quality doctors charged — which is determined purely by the parameters of the utility function and the ability distributions. Specially, consider the case in which $\alpha_z (1-\beta_z)+\beta_z\alpha_x (1-\lambda) > 0$: a doctor in the right end of the tail servicing a patient of income, x, earns $\lambda\beta_z\alpha_x/(\alpha_z(1-\beta_z)+\beta_z\alpha_x)x$. If the lowest quality doctor were to charge the same, she would earn $\lambda\beta_z\alpha_x/(\alpha_z(1-\beta_z)+\beta_z\alpha_x)x_{min} < x_{min}$, which would be insufficient to compensate her for not working as a generalist and earning x_{min}. Consequently, she must be charging a larger share of patient income, and since everybody consumes the services of exactly one doctor, medical services are a necessity. This is more likely to be the case when the number of patients a doctor can service, λ, is low or when $\alpha_z > \alpha_x$ such that generalists have fatter tails than doctors and doctors can charge a smaller part of patient income.

Welfare inequality. The lack of a uniform quality-adjusted price implies that prices vary along the income distribution. Deaton (1998) argued that heterogeneity in consumption patterns implies that people at different points of the income distribution can face price changes that are different from that faced by a "Consumer Price Index-representative household". In the following we show that taking this into account implies that a given increase in income inequality translates into a lower increase in welfare inequality. The assignment mechanism implies that as inequality increases, the rich generalists cannot obtain better health services (in fact they pay more for health services of the same quality), this mechanism limits the welfare increase in inequality. 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.

To assess this formally, we use as a consumption-based measure of welfare the homogeneous good consumption $eq\,(x)$ which, when combined with a fixed level of health quality (namely z_c) gives the same utility to the generalist as what she gets in the market. That is we define $eq\,(x)$ through $u\,(z_c, eq\,(x)) = u\,(z\,(x)\,, c\,(x))$. We then obtain (see Appendix A.3):

Remark 1. *For x large enough, the welfare measure eq is Pareto-distributed with shape parameter $\alpha_{eq} \equiv \frac{\alpha_x}{1+\frac{\alpha_x}{\alpha_z}\frac{\beta_z}{1-\beta_z}}$ so that $\frac{d\ln\alpha_{eq}}{d\ln\alpha_x} = \frac{1}{1+\frac{\alpha_x}{\alpha_z}\frac{\beta_z}{1-\beta_z}}$, implying that an increase in inequality for generalists' income translates into a less than proportional increase in their welfare inequality. The mitigation is stronger when health services matter more (high β_z) or when doctors' abilities are more unequal (low α_z).*

Taking stock. Proposition 1 establishes the central theoretical result of our paper. However, for the empirical analysis it is important to establish what assumptions are necessary for the "spill-over" result and which are not. Consequently, in the following, we first establish the necessity of the "non-divisibility" assumption by introducing brewers who produce a divisible good, beer, and show that income inequality of brewers is independent of that of generalists. Second, we show that the predictions of our model are unchanged if we allow mobility across occupations such that high-earning doctors can work as high-earning generalists. Third, in preparation for our empirical analysis, we introduce a multi-region model. Naturally, without trade or migration between regions top income inequality among doctors must be determined by local generalist income inequality. However, this remains true even if we allow doctors to move across regions. If we instead, allow for the cross-region trade of medical services income inequality for doctors will be the same for all regions. This distinction between "local" services that cannot be traded across regions and "non-local" services that can will be important for the empirical section which is guided by local variation in general income inequality. Finally, we allow doctors to consume medical services and use a more general utility function and ability distributions that are only Pareto in the tail and show that the spill-over effect survive, although it demonstrates that the prediction of a spill-over elasticity of 1 from Proposition 1 does not generalize.

2.1.2 The role of the assortative matching mechanism

To highlight the specificity of our mechanism, we add "brewers" to the system. Potential brewers can produce the divisible good, beer. They differ in their ability such that a brewer of ability y can produce quantity y of (quality-adjusted) beer. Their ability distribution is Pareto with shape α_y, that is a brewer is of ability $Y > y$ with probability

$$P(Y > y) = \left(\frac{y_{min}}{y}\right)^{\alpha_y},$$

and α_y is kept constant. If potential brewers do not produce beer they produce x_{min} units of the homogeneous good. We modify the utility function such that $u(z, c, y) = z^{\beta_z} c^{1-\beta_z-\beta_y} y^{\beta_y}$. The first order condition for beer consumption together with a market clearing equation determine the price of beers. As beer is divisible, the beer price p must be taken as given by each producer and their income will simply be given by py. As a result, the income of active beer producers is Pareto distributed with a shape parameter α_y. A change in inequality among generalists can only affect active producers proportionately.[8] Moreover since beers are divisible, the distribution of the "real" income inequality is unaffected by the presence of beer and the difference between nominal and real income is only driven by the presence of doctors: Remark 1 still applies and α_{eq} does not change. Consequently, divisibility is essential for spill-overs through consumption.

2.2 Occupational mobility

Above we assumed that a potential doctor working as a generalist makes the minimum amount possible as a generalist: x_{min}. In reality it is quite plausible that those succeeding as doctors would have succeeded in other occupations as well. To capture this we assume that there is perfect correlation between abilities as a doctor and as a generalist. We keep the model as before, except we assume that there is a mass 1 of agents who decide whether they want to be doctors or generalists. We rank agents in descending order of ability and use i to denote their rank (so that the most able agent has rank 0 and the least able has rank 1). For two agents i and i' with $i < i'$, i will be better both as a generalist and as a doctor than i'. We assume that both ability distributions are Pareto with parameters (x_{min}, α_x) for generalist and (z_{min}, α_z) for doctors. An agent i can choose between becoming a generalist 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 (1- the cumulative distribution function) for x and z obey:

$$\overline{G}_x(x(i)) = \overline{G}_z(z(i)) = i.$$

For individuals of a sufficiently high level of ability, some will choose to be doctors and others to be generalists. That is for i low enough, agents must be indifferent between becoming a doctor or a generalist: $w(z(i)) = x(i)$, which directly implies that, for z high enough, the wage function must satisfy:

$$w(z) = \overline{G}_x^{-1}\left(\overline{G}_z(z)\right),$$

[8] As showed in Appendix A.4, a decrease in α_x increases p for parameters where all potential brewers are actively producing beer. If the extensive margin of brewers is operative the (mean-preserving) increase in income inequality will lower x_{min} and encourage a supply increase of brewers. As a consequence the effect on beer prices, p, from a decrease in α_x is ambiguous.

which — since both ability distributions are Pareto — can be written as:

$$w(z) = x_{\min} \left(\frac{z}{z_{\min}} \right)^{\frac{\alpha_z}{\alpha_x}}. \tag{11}$$

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

Let $\mu(z) \in (0,1)$ denotes the share of individuals able to provide heath services of quality z who are doctors. the share of agents with ability z that work as doctors. For z sufficiently high that individuals of rank $\overline{G}_z(z)$ and below and their patient work both as generalist and doctors, market clearing implies

$$\left(\frac{x_{\min}}{m(z)} \right)^{\alpha_x} = \int_z^{\infty} \lambda \mu(\zeta)\, g_z(\zeta)\, d\zeta, \tag{12}$$

where $m(z)$ denotes the income (earned either as a generalist or a doctor) of the patient of a doctor of quality z.

The first order condition on health care consumption (3) still applies, and together with (11) and (12) it implies that:

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

Differentiating with respect to z, we obtain that μ is a constant given by $\mu = \lambda^{\alpha_x - 1} \left(\frac{\alpha_z}{\alpha_x} + \frac{\beta_z}{1-\beta_z} \right)^{-\alpha_x}$. Intuitively, with a constant μ, doctors' wages grow proportionately with the patient's income, which is in line with the Cobb-Douglas assumption.

Therefore, we have 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: Proposition 1 still applies (in fact the distribution is now exactly Pareto above a threshold). In Appendix A.7, we solve the full model and show that individuals of a sufficiently low ability will all choose one occupation, while above a given threshold, they work in both occupation but with a constant μ. Further, if the distributions of x and z are only asymptotically Pareto, then our results remain true asymptotically, so that Proposition 1 applies.

Note, that in terms of observed top income inequality the model where agents can move and the one where agents cannot are observationally equivalent: doctor top income inequality perfectly traces that of the generalists. This is so because even when doctors are not allowed to shift occupation, the relative reward to the very best doctors adjusts correspondingly with the shift for generalists.

2.3 Mobility and open Economy

So far we assumed a closed economy. With our empirical analysis driven by local variation in income inequality, we here consider an economy with more than one region and analyze a case in which medical services can be traded between regions and a case in which doctors can move across regions.

2.3.1 Tradable health care

We consider the baseline model of section 2.1. We now assume that there are several regions, $s = 1, .., S$ and we allow some patients (a positive share of generalists in all regions) to purchase their medical services across regions to get health care services. The distribution of potential doctors' ability is the same in all regions (and so is the parameter λ). The other parameters, and in particular the Pareto shape parameter of generalists' income α_x^s is allowed to differ across regions. The cost of health care services must be the same everywhere, otherwise, the generalists who can travel would go to the country with the cheapest health care. Since top talented potential doctors work as doctors (instead of being generalists with income $x_{\min}^s$), they must all earn the same wage. In all regions, the income distribution of patients is asymptotically Pareto with parameter $\min_s \alpha_x^s$, because at the very top, overall income income inequality follows the income inequality of the most unequal region. Using a logic identical to that in section 2.2, we get that in all regions, doctors' income is asymptotically Pareto with shape parameter $\min_s \alpha_x^s$. In other words, income inequality for generalists in the *most unequal* region spills over to doctors in *all* regions.

Empirically, whether the service provided is "local", i.e. non-tradable or "non-local", i.e. tradable, will depend on the occupations of interest and we will use the results of this section and the previous ones to guide our empirical analysis.

2.3.2 Doctors moving

Here as well, we consider the baseline model of section 2.1, but we now assume that there are 2 regions A and B (although the results can be generalized) and that doctors can move across regions, but that medical services are non-tradable and patients cannot move (when doctors are mobile and medical services tradable the geographic location of agents is undetermined and we have no empirical predictions). The two regions are identical except for the ability distribution of generalists, which is Pareto in both but with possibly different means and shape parameters.[9] Without loss of generality, we assume that $\alpha_x^A < \alpha_x^B$, that is region A is more unequal than region B.

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, which from the first order condition on health care consumption, implies that the matching function is the same: doctors of quality z provide health care to generalists 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.

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. Then, labor market clearing in region A, implies that for $z \geq z_c$,

$$\left(x_{\min}^A/m(z)\right)^{\alpha_x^A} = \lambda\mu_d\left(1+\varphi(z)\right)\left(z_{\min}/z\right)^{\alpha_z}. \tag{13}$$

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

$$\left(x_{\min}^B/m(z)\right)^{\alpha_x^B} = \lambda\mu_d\left(1-\varphi(z)\right)\left(z_{\min}/z\right)^{\alpha_z}. \tag{14}$$

Since the two regions are of equal size, total demand for health services must be the same and on net, no doctors move: $\varphi(z_c) = 0$. On the other hand, most rich patients are in region A (as $\alpha_x^A > \alpha_x^B$), as doctor's income increases with the income of their patient, then nearly all the most talented doctors will eventually locate in region A, that is: $\lim_{z\to\infty}\varphi(z) = 1$. Therefore, we obtain that in region A, the distribution of doctors' ability after relocation is asymptotically Pareto, so that as in the baseline model, doctors' income will be asymptotically distributed in a Pareto way with a shape parameter equal

[9] Our results directly generalize to a case where the two regions do not have the same mass of potential doctors and generalists.

to α_x^A.

In region B, doctors of a given quality level earn the same as in region A. That is the income of doctors initially in region B is still distributed in a Pareto way with coefficient α_x^A. However, after the move, the share of doctors that stay in region B decreases with their quality. Using (13) and (14), we get that $1-\varphi(z) \propto z^{\alpha_z\left(1-\alpha_x^B/\alpha_x^A\right)}$. Therefore, the ex-post talent distribution of doctors in region B is still Pareto but with a coefficient α_z' now given by $\alpha_z' = \alpha_z\alpha_x^B/\alpha_x^A$. As in the baseline model, the distribution of income for doctors who stay must be asymptotically Pareto with a shape Parameter α_x^B.[10] We obtain (formal proof in Appendix A.7.1):

Proposition 2. *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.*

Consequently, whether doctors can move or not does not alter the *observable* local income distribution, although it does matter considerably for the unobservable local *ability* distribution. Consequently, for our empirical analysis we need not take a stand on which assumption is the most reasonable and we cannot empirically distinguish between them using data on income inequality.

Finally, we show that our analysis does not depend on the simplifying assumptions made in the baseline model of Section 2.1, although the result of Proposition 1 that the spill-over must have an elasticity of 1 does not carry through.

2.4 Utility Function and Ability Distribution

2.4.1 Doctors consume medical services and ability distribution is only Pareto distributed in the tail

We now alter the model so there is a mass 1 of agents a fraction, μ_d, of which are potential doctors. The technology for health services is the same as before (and we now assume that $\lambda > 1/\mu_d$). Agents not working as doctors produce a composite good which we take as the numeraire. Contrary to the baseline model above all agents have the same utility function (1).

[10] To see that there is no contradiction, note that the baseline model predicts that the income of individual z, $w(z) \propto z^{\frac{\alpha_z'}{\alpha_x^B}}$ but $\frac{\alpha_z'}{\alpha_x^B} = \frac{\alpha_z}{\alpha_x^A}$, so we also have $w(z) \propto z^{\frac{\alpha_z}{\alpha_x^A}}$ and doctors do indeed earn the same in both regions.

The equilibrium results in a wage distribution. We assume that this distribution and also the distributions of skills for potential doctors are asymptotically Pareto. Therefore we can write

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

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

$$\overline{G}_x\left(x,\overline{x}\right) \approx \left(\overline{x}/x\right)^{\alpha_x},$$

with $\alpha_x > 1$. The same holds for doctors' talents z (moreover, potential doctors can work as generalists with the lowest productivity $x_{\min}$ as an alternative).

As before solving for the consumer problem leads to the differential equation (3). Furthermore since health care services are not divisible, the equilibrium also features assortative matching and we still denote the matching function $m\left(z\right)$. Market clearing at every z can still be written as (4). The least able potential doctor who actually works as a doctor will have ability $z_c = \overline{G}_z^{-1}\left(1/\lambda\mu_d\right)$. Therefore z_c is independent of α_x. As a result we get that (4) implies that $m\left(z\right)$ is defined by $m\left(z\right) = \overline{G}_x^{-1}\left(\overline{G}_{z,z_c}\left(z\right)\right)$.

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:

$$\overline{G}_x\left(m\left(\overline{z}\right)\right)\left(m\left(\overline{z}\right)/\left(m\left(z\right)\right)\right)^{\alpha_x} \approx \overline{G}_{\overline{z},z_c}\left(\overline{z}\right)\left(\overline{z}/z\right)^{\alpha_z},$$

which gives

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

Plugging this in (3) we can rewrite the differential equation as:

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

Therefore for z large enough, we must have (see Appendix A.5 for a rigorous derivation):

$$w\left(z\right) \approx \frac{\beta_z\alpha_x}{\alpha_z\left(1-\beta_z\right)+\beta_z\alpha_x}\lambda Bz^{\frac{\alpha_z}{\alpha_x}}. \tag{15}$$

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

$$P\left(W_d > w_d | w_d > \overline{w}_d\right) \approx \left(\frac{\overline{w}_d}{w_d}\right)^{\alpha_x}, \tag{16}$$

that is doctors' income follows a Pareto distribution with shape parameter α_x. Proposition 1 still applies: a decrease in α_x will directly translate into an increase in top income inequality among doctors.

2.4.2 The role of the Cobb-Douglas utility function

We keep the same model as just introduced, but we replace the utility function of equation (1) with:

$$u(z,c) = \left(\beta_z z^{\frac{\varepsilon-1}{\varepsilon}} + \beta_c c^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{17}$$

with $\varepsilon \neq 1$. As before, the first order conditions gives the differential equation:

$$\partial u / \partial z = \omega'(z)\, \partial u / \partial c. \tag{18}$$

Since CES exhibit positive cross partial derivatives, we know that the equilibrium features positive assortative matching. Therefore, with income and ability asymptotically Pareto, the matching function still obeys (5). Using (17), combining (18) and (5), and using that $w(z) = \lambda \omega(z)$, we find that for high levels of z the wage function obeys a differential equation given by

$$w'(z) \approx \lambda^{\frac{\varepsilon-1}{\varepsilon}} \frac{\beta_z}{\beta_c} z^{-\frac{1}{\varepsilon}} \left(\lambda B z^{\frac{\alpha_z}{\alpha_x}} - w(z)\right)^{\frac{1}{\varepsilon}}. \tag{19}$$

We solve for this equation in Appendix A.6 and we prove:

Proposition 3. *i) Assume that $\varepsilon > 1$. Then for $\alpha_x \geq \alpha_z$, wages of doctors are asymptotically Pareto distributed with exponential parameter $\alpha_w = \alpha_x$. For $\alpha_x < \alpha_z$, wages of doctors are asymptotically Pareto distributed with $\alpha_w = \frac{\alpha_z}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}$.*
ii) Assume that $\varepsilon < 1$. Then for $\alpha_x > \frac{\alpha_z}{1-\varepsilon}$, wages of doctors are bounded. For $\alpha_x = \frac{\alpha_z}{1-\varepsilon}$, wages of doctors are asymptotically exponentially distributed. For $\alpha_z < \alpha_x < \frac{\alpha_z}{1-\varepsilon}$, they are asymptotically distributed with $\alpha_w = \frac{\alpha_z}{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}+1}$. For $\alpha_x \leq \alpha_z$, they are asymptotically Pareto distributed with $\alpha_w = \alpha_x$

Therefore, when doctors' income distribution is Pareto, we still obtain that a reduction in α_x leads to a reduction in α_w (that is an increase in general top income inequality increases top income inequality among doctors), although the elasticity may now be lower than 1 (it can naturally not be asymptotically higher than 1 since that would imply high-paying generalists spending more than their income on medical services). Further, a decrease in α_x also reduces the size of the parameter space for which doctors' wage distribution is bounded (which corresponds to a situation where top income inequality for doctors is very low).

To understand intuitively the results of Proposition 3, consider first the case where $\alpha_z > \alpha_x$: that is where the tail of the ability distribution of generalists is fatter than that of doctors, implying a shortage of doctors at the top. This must mean a convex pricing schedule for medical services. If $\varepsilon > 1$, health services and the homogeneous good are substitute, so that the expenditure share on health services declines with income. As a result, $w(z)$ cannot grow as fast as the income of the generalist who pays the services of doctor z, namely $m(z)$, which grows as $z^{\frac{\alpha_z}{\alpha_x}}$, implying less income inequality among the top doctors than the top generalists (a higher Pareto exponential parameter). On the contrary if $\varepsilon < 1$, then richer generalists are forced to spend an increasing amount (eventually all their resources) on health services, $m(z)$ and $w(z)$ grow at the same rate, so that doctor's income is Pareto distributed with coefficient α_x. The reverse holds when doctors are relatively abundant at the top (i.e. when $\alpha_z < \alpha_x$), except that with $\varepsilon < 1$, doctors' income can even be bounded.

2.5 Empirical prediction

To summarize, our model makes the following predictions:

1. An increase in general inequality will lead to an increase in inequality for doctors if they service the general population directly and their services are non-divisible.

2. This is true, whether doctors can move across regions or not, and whether doctors' ability is positively correlated with the income they would get working in alternative occupations or not.

3. If patients can easily travel, doctors' income in each region does not depend on local income inequality.

3 Empirical Strategy and Data

3.1 Empirical strategy

We are centrally interested in the causal effect of general top income inequality in a region s on the top income inequality of a particular subgroup i in region s. Since our data are top censored, we make the distributional assumption that the right tail of the income distribution is Pareto distributed: $P(X > x) = (x/x_{min})^{-\alpha}$ above some cut-off x_{min} and use $1/\alpha$ as a our measure of income inequality. Specially, for such a distribution the relative income of somebody at the 99th percentile relative to somebody at the 95th percentile is $5^{1/\alpha}$ and the Gini-coefficient is $(2\alpha - 1)^{-1}$. Guvenen, Karahan, Ozkan, and Song (2015) and Jones and Kim (2014) also employ $1/\alpha$ as a measure of income inequality.

Using this, the regression of interest is:

$$log(1/\alpha_{o,t,s}) = \gamma_s + \gamma_t + \beta log(1/\alpha_{-o,t,s}) + X_{t,s}\delta + \epsilon_{o,t,s}, \tag{20}$$

where $1/\alpha_{o,t,s}$ is top income inequality for occupation i at time t for geographical area s and $1/\alpha_{-o,t,s}$ is the corresponding value for the general population except i at time t in geographical area s, γ_s is a dummy for the geographical area, γ_t is a time dummy, and $X_{t,s}$ are a vector of controls, in particular population of labor market area and average income. We are centrally interested in β which measures the elasticity of top income inequality for our occupation of interest with respect to the general income inequality. We will focus on labor market areas, which are aggregations of commuting zones (Dorn, 2009) and can generally be driven through in a matter of a few hours, i.e. Los Angeles or New York City. Central results carry through if we instead use States as the unit of analysis.

We focus on a number of occupations, but are limited by the fact that our analysis requires a relatively high number of observations. We split these occupations into two groups: First, we focus on occupations whose output is non-divisible and who primarily operate in local markets: physicians, dentists and real estate agents. Admittedly, some patients do travel for special medical treatment. To the extent that this creates an integrated market, this would create a downward bias in our estimate of the spill-over effects. We contrast these with other occupations who have also seen increases in income inequality, but that do not satisfy these conditions: college professors who, although they do provide a non-divisible, do not operate in a local market (at least not in the right tail of the distribution), secretaries, who although they do operate in local markets do not service the general population directly, and financial managers. According to the *Standard Occupational Classification scheme* financial managers "Plan, direct, or coordinate accounting, investing, banking, insurance, securities, and other financial activities of a branch, office, or department of an establishment." Hence, such fall into two categories: those who manage financial matters for corporations and who are unlikely to be affected by higher local income inequality, and those that manage the financial means of private individuals who are likely to operate in more integrated markets. In either case we expect to see no local spillover effect. Figure 2 shows the increase in the 99th / 90th percentile for selected occupations and demonstrates that the increase in within-occupation top income inequality is a trend outside of the very top of the income distribution as well.

We will use estimate (20) using the Publicly available Decennial Census and American Community Survey to estimate both $1/\alpha_{o,t,s}$ and $1/\alpha_{-o,t,s}$. This allows us to examine the time period 1980 to 2014 and consider a relatively broad set of occupations. Due to possible concerns about endogeneity we will use an instrumental variable approach, using a "shift-share" instrument (following Bartik, 1991) based on the occupational distribution across geographical areas in 1980. We first show some summary statistics in Section 3.3. We then perform our regression analysis on the income data in Section 4. A number of plots are contained in the Data Appendix D

3.2 Income data

Our central data set is a combination of the Decennial Census for 1980, 1990 and 2000 and add the IPUMS Combined American Community Survey (ACS) for 2010-2014 as one year, which we will henceforth refer to as 2014 (Ruggles et al. 2015).[11] This leaves us with between 5.4 million observations in 1980 and 7.4 million observations in 2014 with positive wage income. We use 2010-2014 as opposed to the perhaps more natural 2008-2012 to avoid the immediate aftermath of the Great Recession which had large impact on top income. IPUMS does have data from farther back but it is a substantially smaller sample and we exclude it from the analysis. Throughout we use the 1990 census occupa-

[11]The Decennial Censuses are each 5 per cent of the population, whereas the ACS each are 1 per cent. Combining the years 2010-2014 creates an 'artificial' sample of 5 per cent for 2014. The IPUMS inflates all numbers to 2014 using the consumer price index.

tional classification from IPUMS which consistently assigns occupations throughout the 1980-2014 period. The publicly available income data are censored, generally at a level at around top 0.5 per cent of the whole income distribution, which complicates our estimation of the parameter of the Pareto distribution.[12] In particular, suppose $\tilde{X}$ follows a Pareto distribution $P(\tilde{X} > \tilde{x}) = (\tilde{x}/x_{min})^{-\alpha}$, but the observed wage is $x = max\{\tilde{x}, \bar{x}\}$ for some censoring point, $\bar{x}$. Then we can write the maximum likelihood function as: $\Pi_{i \in \mathcal{N}_{unc}} \alpha x^{-(\alpha+1)} x_i^{\alpha} (\bar{x}/x_{min})^{-\alpha N_{cen}}$, where $\mathcal{N}_{unc}$ is the set of uncensored observations and N_{cen} is the number of censored observations. Armour, Burkhauser and Larrimore (2014) use the same methodology on the Current Population Survey (March edition) to show that trends in income inequality match those found by Kopczuk, Saez and Song (2010) using uncensored social security data. The resulting maximum likelihood estimate is

$$\frac{1}{\hat{\alpha}} = \frac{\sum_{i \in \mathcal{N}_{unc}} log(x_i/x_{min}) + N_{cen} log(\bar{x}/x_{min})}{N_{unc}}, \tag{21}$$

where N_{unc} is number of uncensored observations. Note, that even without the assumption of a Pareto distribution, equation (21) is a measure of income inequality: It is the average log-difference from the minimum possible observation for the uncensored observations plus the product of the relative number of censored observations times the log distance of the censoring point to the minimum. This will be our measure of income inequality throughout.

In estimating equation (20), we will (for some specifications) control for average income and population.

3.3 Summary statistics

We will be focusing on the combined pre-tax wage and salary income throughout.[13] Table 1 below shows the mean, median, 90th, 95th and 98th percentile for for positive wage income for each year all in 2014 dollars (we choose 98th as the censoring doesn't allow for the calculations of 99th for all years). As discussed in the introduction, income inequality measured as the ratio of the 98th to 95th percentile or the ratio of the 95th

[12]Specially, the censoring takes place at 75,000 for 1980, 140,000 for 1990, 175,000 for 2000 and at the 99.5 pct ratio at the state level for each individual year 2010-2014. What information is given about the censored variables varies from year to year.

[13]The census includes other measures of income, in particular business income which would be relevant for some occupations. Unfortunately, since wage income and business income are censored separately estimating a joint distribution for the two would be substantially more complicated. We are in process of getting access to the full uncensored data which would allow us to use total income.

to the median has increased during the period. The table also shows the estimate of $1/\alpha$ on the top 10 per cent of observations with positive wage income for each year and the 98/95th and 95/90th ratios from the estimated Pareto Distribution in parentheses. There is a high level of agreement between the predicted and the actual ratios consistent with a good fit of the Pareto distribution.

Although the censoring point is sufficiently high to allow standard measures of top income inequality to be calculated for most occupations, the high average income of some occupations leads to a larger share being censored. Whereas the censoring has little impact on the overall distribution, slightly more than 26 per cent of Physicians with positive income are censored in 2000. This implies that we cannot calculate measures of income inequality using high percentiles. But we can still calculate $1/\alpha$ using the assumption of a Pareto distribution. Table 2 shows the result using the top 65 per cent of the uncensored observations.[14] Consistent with Figures 1, $1/\alpha$ has increased for most occupations in the top during this period. Table D.1 in the Appendix shows the calculated measures of income inequality (using top 10 per cent of the population) for a number of other occupations along with the fraction of observations with positive income that are censored. The table shows the same general trend, but with some notable exceptions. In particular there has been little upward trend in top income inequality for truck drivers, sales people, and computer software developers, but substantial increases for financial managers and chief executives.[15] Table D.2 shows which occupations were in the top 1, 5 and 10% for 1980 and 2014. What is particularly noteworthy is that Physicians are increasingly important in the high end of the income distribution and that this importance has grown from 1980 to 2014. In fact, Physicians are the most common (Census) occupation in the top 1% in 2014.

Each observation in the data is associated with a particular geographical area (for 1980 'County groups' from 1990 onward 'Public Use Microdata Area'). As argued in Dorn (2009) these are statistical areas created to ensure confidentiality and have little economic meaning. Alternatively, one could use states, but some local economies, say greater New York City or Washington D.C. span several states, some states are too

[14] Throughout the paper we follow the following rule of thumb when calculating occupation, year, labor market specific measures of income inequality: If there are very few censored observations —say for secretaries— we use the top 10 per cent of the distribution. For occupations that are heavily censored — physicians and dentists — we move the cut-off until we have around twice as many uncensored observations as censored for all labor market areas we use. For Physicians that is the top 65 per cent, for dentists it is top 50 per cent and for Real Estate agents it is top 20 per cent.

[15] Table D.2 in the appendix shows the 10 most prominent occupations in the top 1, 5 and 10% of the population for 1980 and 2014.

large to meaningfully capture a local economy and some states are too small to have sufficient number of observations. Dorn instead uses commuting zones of which there are 741. We take a similar approach but use labor market areas. Both commuting zones and labor market areas are defined based on the commuting patterns between counties (Tolbert and Sizer, 1996). But whereas commuting zones are unrestricted in size, labor market areas aggregate commuting zones to ensure a population of at least 100,000. Given that our estimation strategy relies on a relatively high number of observations of a particular occupation, labor market areas are a more natural choice. Table 3 shows the size distribution of number of total observations with positive wage income and physicians with positive wage income across labor market areas.

To asses the fit of the Pareto distribution at the labor market area, year, occupation level, we use the fact that a Pareto distribution implies a linear relationship between value and frequency. Figure 3 shows this relationship for the biggest labor market area for both all occupations (Los Angeles) and physicians specifically (New York City). The line shows the predicted number of observations in each bin whereas the orange dots give the actual number of observations in each bin, both on the left hand side. The right hand axis gives the corresponding values for the censored values, scaled to ensure that the predicted censored values are on the same line as the uncensored predicted values.[16]

Figure 3.a below uses the biggest labor market area (Los Angeles) for the year 2000 and bins the income interval between that of the 90th percentile and the censoring point of 175,000 into 20 evenly sized bins and plots the (linear) predicted number of observations from the associated Pareto distribution with the observed number of observations in each bin (the choice of bins in the figure does not influence any estimation results). The figure further shows the actual and predicted number of censored observations on the right hand side scaled to fit a linear line. The fit for the general population is very close to a straight line and therefore a Pareto distribution. We perform an analogous analysis for the physicians (where New York City is the biggest labor market), although with the lower number of observations overall and the much higher number of censored observations we use the top 65 per cent of the positive uncensored observations.[17] The

[16]Formally, we use the fact that for a dataset with N observations on wages drawn from a Pareto distribution $P(X > x) = (x/x_{min})^{-\alpha}$ with a corresponding pdf of $f(x) = \alpha x^{-(\alpha+1)} x_{min}^{\alpha}$, the expected number of observations that have wage income in the interval $[x' - \frac{\Delta}{2}, x' + \frac{\Delta}{2}]$ is $N_{x'} = N \int_{x'-\frac{\Delta}{2}}^{x'+\frac{\Delta}{2}} f(x)dx \simeq N\Delta\alpha x^{-(\alpha+1)} x_{min}^{\alpha}$, giving a negative linear relationship between $logN_{x'}$ and $logx$. The predicted number of censored observations is $P(X > \bar{x}) = (\bar{x}/x_{min})^{-\alpha}$ to which we (arbitrarily) assign the value $\bar{x} + \Delta/2$ and scale to fit on the same predicted line.

[17]Though we carry out the main analysis using the top 65% of observations, Table B.5 in the Appendix

fewer observations implies a fit that is less tight, but there are no systematic deviations from the straight line. Figures D.1 and D.2 in the Appendix give equivalent figures for the 20 biggest labor markets in the United States.

3.4 Instrument

One might worry about endogeneity when estimating equation (20). In particular, even controlling for labor market area 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 and not reflect a causal effect from general income inequality to inequality for the occupation of interest. To address this issue, we use a Bartik (1991)-style instrument. Namely we define:

$$logI_{-o,t,s} = log\left[\sum_{\kappa \in K_{-o}} \omega_{\kappa,1980,s}(1/\alpha_{\kappa,t})\right], \, for\, t = 1980,\, 1990,\, 2000,\, 2014.$$

K_{-o} is the set of the 20 most important occupations in the top 5 per cent of the income distribution nationwide in 1980 (excluding occupation o). $\omega_{\kappa,1980,s}$ is the share of individuals in occupation κ among individuals in an occupation belonging to K_{-o} in 1980 in LMA s. In other words our instrument is a weighted average of nationwide occupational inequalities for the most important occupations in the top of the income distribution, where the weights correspond to the initial occupational composition of each LMA. The instrument has strong predictive power: the correlation between $logI_{-o,t,s}$ and $log(1/\alpha_{-o,t,s})$ for Physicians is between 0.39 and 0.55 for each year (depending on the number of LMAs considered). It is practically the same for all occupations as each occupation represents a relatively small share of total top income holders.[18] In other words, in the IV regression we only exploit the changes in labor market income inequality that arises from the occupational distribution in 1980 combined with the nationwide trends in occupational inequality. Furthermore, by using nationwide trends, our instrument is more likely to capture the effects of globalization, technological change or deregulation, which affect local inequality but are exogenous to the LMA, which is in line with a decrease in α_x in our theoretical model.

shows that the parameter estimate is relatively insensitive to the choice of cut-off.

[18] The qualitative conclusions of our analysis remain unchanged by using a different number of top occupations than 20, although the point estimate of β is somewhat sensitive.

4 Empirical Analysis

4.1 Testing the model for occupations with positive predicted spill-overs

Having estimated the Pareto distributions described above we next estimate equation (20). We start out by conducting the analysis for physicians. Table 4 presents summary statistics for the regressors of interest. We restrict ourselves to the biggest 253 labor market areas — those with at least 8 observations of physicians in 1980 — for a total of 1,012 observations. Table B.6 moves this cut-off between the top 100 LMAs and using all. The parameter estimate, β, remains significant and between 0.8 and 1.7.

The result is shown in Table 5.[19] The first column shows an OLS regression of physicians' income inequality on general income inequality including year and LMA fixed effects and shows an elasticity of around 1/3. To take account of the fact that the variables of the estimation are themselves estimated, confidence intervals are calculated using bootstrap sampling, stratified at the occupational-labor market area-year level using 300 replications.[20] This estimate remains unchanged in column (2), where we include controls for labor market population and the average wage income among those with positive wage income. Neither control has a significant impact on physicians' income inequality. Column (3) shows the first stage of the instrumental variable regression using the instrument as constructed in (21). The instrument has a strong predictive power and along with the time trends accounts for 82 percent of the variation in the variation for general income inequality (R^2 is computed excluding the LMA fixed effects). The F-statistic for the first stage is 40 for our preferred specification and higher than 20 for all regressions in this paper. The fourth and fifth columns give the second stage IV results, which show point estimates of the coefficient of interest of 1.19 or 1.33 depending on controls. This is strongly significantly different from 0, and not significantly different from the predicted value of 1 from the simplest model of Section 2.1. With our measure of general top income inequality increasing by 27 percent and top income inequality for physicians increasing by 31 percent (Tables 1 and 2) an elasticity of 1 suggests that a

[19] One can show that the inverse of the variance of the MLE estimator of equation 21 is proportional to the number of uncensored observations and we correspondingly weigh the equation by the number of uncensored observations of physicians.

[20] That is for each draw we resample person-observations, recalculate the α's and reestimate the regressions. For computational reasons we do restrict attention to the top end of the redistribution, i.e. we only resample from top 10 per cent of the income distribution in a given labor market area to calculate general income inequality.

large share of the rise of income inequality among doctors can be explained by the general increase in income inequality, although the exact fraction is measured with uncertainty. Moreover, note that neither the controls nor the year fixed effects are significant in the IV regressions, which is consistent with our mechanism explaining most of the changes in income inequality for doctors.

The medical industry in the United States is not perfectly described by the simple free market model of Section 2.1: the government plays a substantial role through Medicare and Medicaid, the insurance sector has an important role as an intermediary, there is substantial asymmetric information between patients and doctors and patients are often willing to travel to seek medical attention. However, these features are unlikely to substantially affect our analysis. The government sets administrative prices for those whose care it pays for directly, but providers' negotiations with private insurers generally lead to higher prices (Clemens and Gottlieb, forthcoming) . Even in the presence of asymmetric information, patients often have clear beliefs about who the "best" local doctor in a special field is (even though these beliefs may be incorrect). And although patients occasionally travel for care, a patient in Dallas is vastly more likely to seek medical care in Dallas than Boston. Furthermore, our empirical strategy more heavily weights large metropolitan areas, which are more likely to have a full portfolio of medical specialties implying less need to travel. Finally, the extent to which the medical industry is described by a national market would bias our parameter estimate downwards, but we still find a significantly positive effect.

Nevertheless, we analyze two other occupations that are much less regulated and more local: dentistry and real estate. We perform an analogous examination of dentists in table 6 and reach broadly similar conclusions. Again we focus on labor market areas with at least 8 observations in 1980, which severely reduces the number of labor market areas from 253 to 40. Yet, we see a pattern broadly similar to that of physicians, albeit with less precision (the OLS coefficients have p-values of 12 per cent). Both OLS and IV point estimates are around twice as high for dentists as for physicians. Though this might reflect the fact that dentistry is more local and prices are less regulated, the point estimates are not significantly distinct and we cannot rule out a difference purely due to sampling error. With a spill-over elasticity of 2.8 and a rise in income inequality for dentists that has mirrored that of the general population we substantially "over-explain" the rise in income inequality for dentists, though with this few observations there is substantial imprecision in the estimate.[21]

Finally, we use an occupation outside the medical industry: real estate agents. The fee structure in real estate is often proportional to housing prices (Miceli, T., Pancak, K. and Sirmans, C., 2007) and the increase in the spread of housing prices is consistent with the increase in income inequality (Määttänen and Terviö, 2014). Real estate is a difficult business to scale up, as each house still needs to be shown individually and each transaction negotiated separately. Consequently, one would expect to see spill-over effects from general income inequality to real estate agents. Table 7 shows that this is indeed the case. Though the OLS estimates are somewhat lower than for the physicians, the IV estimates are very close. Income inequality for Real Estate agents has increased from 0.45 to 0.69, an increase of 50%. With general income inequality increasing by around 27% the IV estimate suggests that more than half the increase in agents' income inequality can be attributed to the general increase in income inequality.

4.2 Testing the model for occupations with zero predicted spill-overs

Whereas our theory predicts local spill-over effects from general income inequality to the income inequality for occupations such as physicians, dentists and real agents it predicts no such spill-overs for other occupations. We perform analogous regressions for financial managers, who, as argued above, do not fit the conditions required for local spill-overs. Table 8 shows that this is the case. Though the OLS estimate is positive, the IV estimate is close to zero (and in fact the point estimate is now negative). This also shows that spurious correlation between general inequality and occupational inequality at the local level is likely but that our instrument can address this concern.

Finally, we perform the analysis for two other occupations with substantial increases in top income inequality from Figure 2 but where our model predicts to spill-overs: College Professors and Secretaries.[22] Universities with faculty in the top 10% operate in a very national market and secretaries are not hired by private citizens. In either case

[21] We also perform the analogous analysis on nurses for whom top inequality has grown as well (See Figure 2) though it is less clear that our model would apply to this occupation. Whereas the OLS estimates are similar to physicians, the IV estimate is lower and estimated with more imprecision and we cannot clearly establish spill-overs as a cause for the increase in income inequality for nurses. See Table B.1 in Appendix B.1 for details.

[22] There is a data break in the IPUMS data: For 1980 to 1990 Post-Secondary teachers (those teaching at higher level than high-school) are partly categorized by subject of instruction (code 113-154). From 2000 onward they are not. We collapse all codes 113-154 into 154 for 1980 to 1990.

we see no effect. See Tables B.2 and B.3.

Interestingly income inequality for secretaries correlates strongly with income inequality for "Chief Executives and Public Administrators" in an OLS regression. Though we cannot establish causality this is consistent with a theory analogous to the one presented in 2.1 where CEOs compete for the most skilled secretaries. At slightly more than 10% the implied elasticity is substantially lower than for other occupations considered here. See Table B.4 in the Appendix.

It is important to distinguish the predictions of our theory of consumption-driven spill-overs with alternative theories of spill-overs through increased demand for skill in the local labor market. Even if these were driven by the original distribution of occupations, and therefore captured by our instrument, one would expect income inequality to spill over broadly into other occupations. In contrast, our theory predicts that this will only happen for a subset of occupations, in particular those that provide non-divisible local services, in line with the empirical evidence.

5 Conclusion

In this paper, we established that an increase in income inequality in one occupation can spill over through consumption to other occupations, such as physicians, dentists and real estate agents, that provide non-divisible services directly to customers. We show that changes in general income inequality at the level of the local labor market area do indeed spill-over into these occupations. We distinguish this with other occupations that have seen rises in top income inequality, but that either do not fit the conditions or operate in a national labor market such as financial managers and college professors and show that there are no spill-over effects. This contrasts our consumption-driven theory with different theories of changes in the relative demand for skills which would predict broader spill-overs of income inequality.

The magnitude of the estimate suggests that this effect may explain most of the increase in income inequality for occupations such as doctors, dentists and real estate agents. As a result, the increase in top income inequality across most occupations observed in the last 40 years may not require a common explanation: in particular, increases in inequality for, say, financial managers or CEOs because of deregulation or globalization may have spilled over to other occupations in the top causing a broader increase in top income inequality.

Further, even though our analysis has been purely positive, it clearly has normative implications, that we plan on exploring in future work. In particular, our analysis could be relevant to the study of top income taxation (see for instance, Scheuer and Werning, 2015).

References

Acemoglu, D. and Autor, D. (2011). *Handbook of Labor Economics*, 40:1043–1171.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., and Hémous, D. (2015). Innovation and top income inequality. Nber working paper no. 21247, National Bureau of Economic Research. NBER wp 21247.

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

Armour, P., Burkhauser, R., and Larrimore, J. (2016). *Economic Inquiry*, 54:1263–1273.

Bakija, J., Cole, A., and Heim, B. (2010). Jobs and income growth of top earners and the causes of changing income inequality: Evidence from u.s. tax return data.

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

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

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

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

Gabaix, X., Lasry, J.-M., Lions, P.-L., and Moll, B. (2015). The dynamics of inequality. NBER wp 21363.

Garicano, L. and Hubbard, T. (2012). Learning about the nature of production from equilibrium assignment patterns. *Journal of Economic Behavior {&* Organization, 84:136–153.

Geerolf, F. (2014). A static and microfounded theory of zipf's law for firms and of the top labor income distribution.

Gesbach, H. and Schmutzler, A. (2014). Does globalization create superstars? a simple theory of managerial wages. *European Economic Review*, 71:34–51.

Goldin, C. and Katz, L. (2010). *The Race between Education and Technology.*

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

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2015). What do data on millions of u.s. workers reveal about life-cycle earnings risk. NBER working paper 20913.

Jones, C. and Kim, J. (2014). A schumpeterian model of top income inequality. NBER working paper 20637.

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

Kopczuk, W., Saez, E., and Song, J. (2010). *Quarterly Journal of Economics*, 125(1):91–128.

Kukharskyy, B. (2012). Trade, superstars, and welfare. BGPE Discussion Paper No 120.

Ma, L. (2015). Globalization and top income shares.

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

Manasse, P. and Turrini, A. (2001). *Journal of International Economics*, 54:97–117.

Miceli, T., Pancak, K., and Sirmans, C. (2007). Is the compensation model for real estate brokers obsolete? 35(7).

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

Pareto, V. (1896). *Cours d'Economie Politique*.

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

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

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

Piketty, T. and Saez, E. (2003). Income inequality in the united states: 1913-1998 , 1, 1-39. *Quarterly Journal of Economics*, 118 (1):1–41.

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

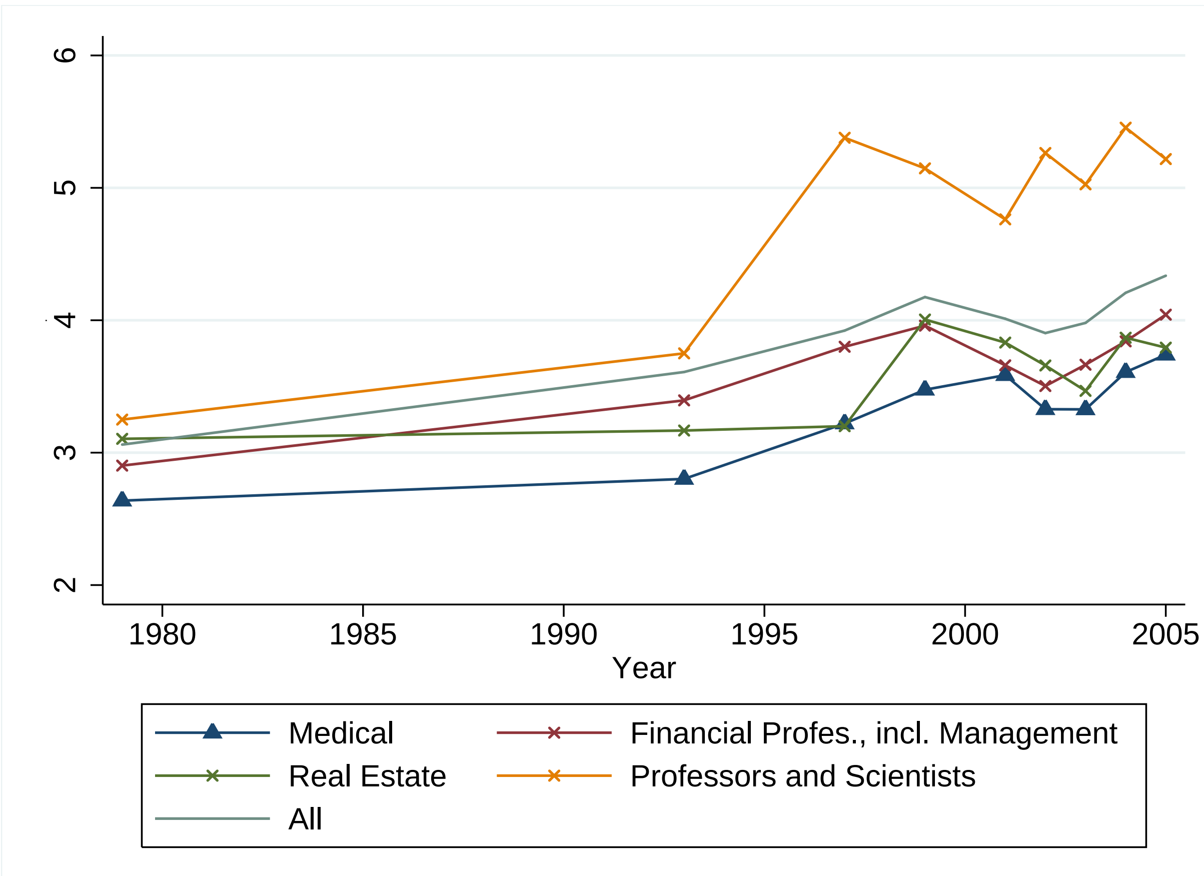
Rosen, S. (1981). The economics of superstars. *American Economic Review*, 71:485–858.

Ruggles, S., Goeken, R., Grover, J., and Sobek, M. (2015). Integrated public use microdata series: Version 6.0 [machine-readable database]. *Minneapolis: University of Minnesota*.

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

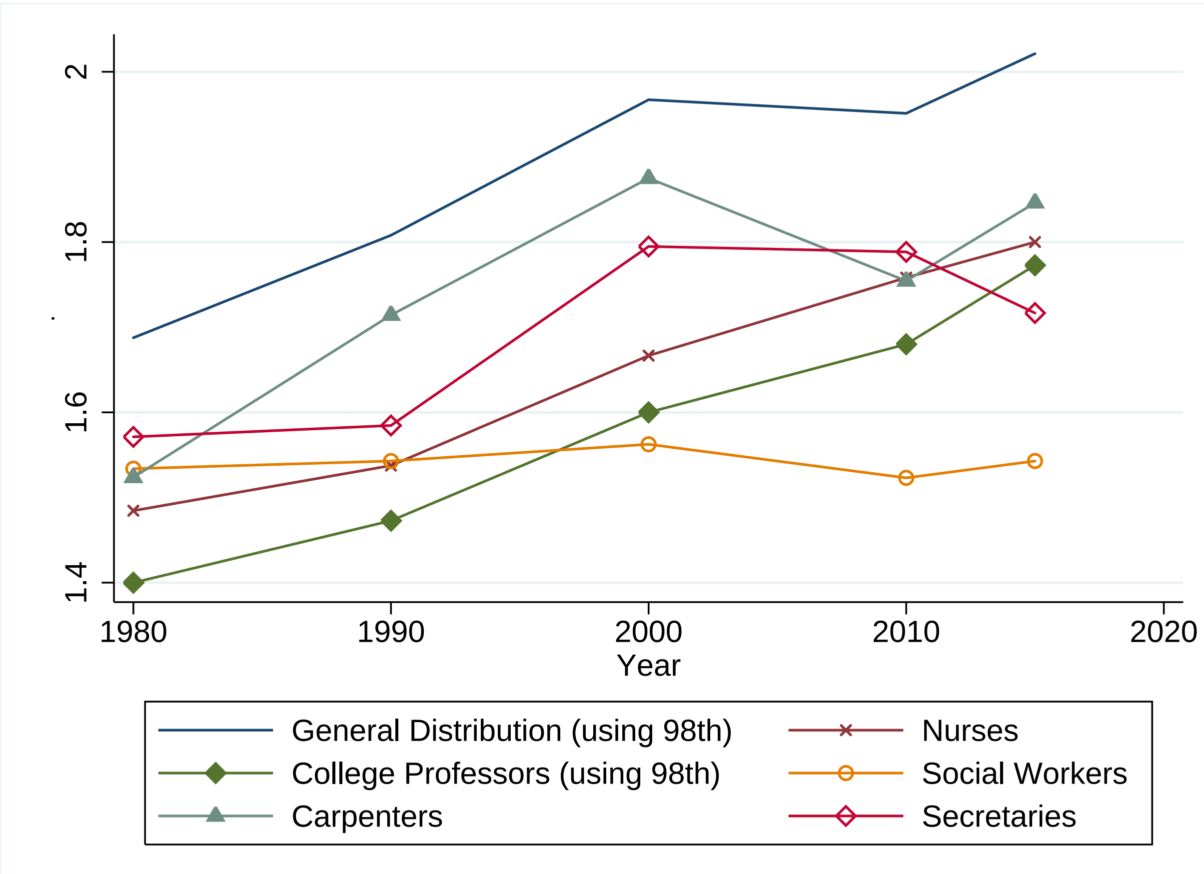
Scheuer, F. and Werning, I. (2015). The taxation of superstars. NBER wp 21323.

Figure 1: Relative income of top 0.1% to top 1% of income distribution for selected occupations



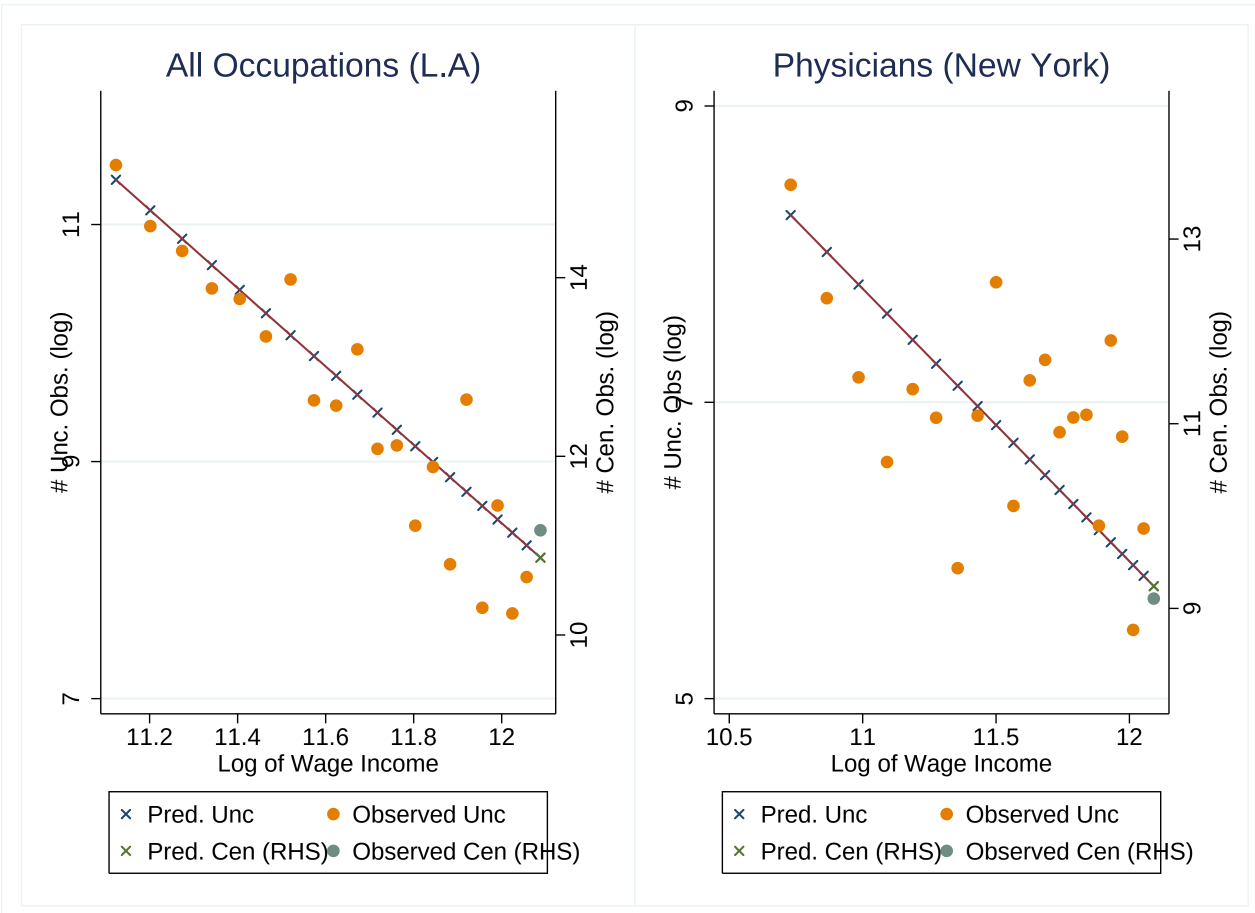
Notes: Relative earnings of those in the top 0.1% compared with those in top 1% across occupations for selected occupations "All" refers to all in the top 1% or top 0.1%, not just the occupations shown here. Source: Bakija, Cole and Heim (2012)

Figure 2: The ratio of the 99th to 90th percentile for selection occupations



Notes: Using strictly positive wage income. Source: Decennial Census and American Community Survey. Authors own calculations. Top-censoring prevents the calculation of 99th percentile for the general distribution as well as for college professors.

Figure 3: Fit of the Pareto Distribution for the year 2000



Notes: Predicted (line) and observed (dots) number of observations. Uncensored observations on left hand side, Censored observations on right hand side

Table 1: Wage income 1980-2014 for general population

Year	Median	percentile p90	p95	p98	p95/p90 (predicted)	p98/p95 (predicted)	$1/\alpha$
1980	25.9	68.1	85.2	114.9	1.25 (1.26)	1.35 (1.35)	0.33
1990	29.0	76.1	96.8	137.6	1.27 (1.30)	1.42 (1.42)	0.38
2000	33.0	83.9	111.4	165.0	1.33 (1.32)	1.48 (1.44)	0.40
2014	30.5	90.0	120.0	177.9	1.33 (1.34)	1.48 (1.47)	0.42

Notes: Real wage income for observations with positive income (1000s of 2014 dollars using CPI). p95/p90 is the relative income of top 5 and top 10 per cent (predicted values in parenthese). Censoring prevents the calculation of 99th percentile wages

Table 2: Wage income 1980-2014 for Physicians

Year	median	$1/\alpha$	p95/p90 (pred)
1980	100.6	1.29	1.71
1990	126.8	1.49	1.59
2000	137.5	1.58	1.55
2014	160.9	1.70	1.50

Notes: Real Wage Income (1000s of 2014 dollars using CPI) calculation of α uses top 65 per cent of non-censored positive income

Table 3: Number of observations across labor market areas

	All				Physicians			
Year	Mean	25th	Median	75th	Mean	25th	Median	75th
1980	39584	17500	24560	37269	70	17	28	54
1990	43772	21207	29393	43286	89	22	37	70
2000	47541	21829	31786	46297	105	26	42	87
2014	51495	22789	34212	49489	138	29	56	115

Notes: Number of observations for labor market areas (all and physicians)

Table 4: Summary Table For Regression Variables

Variable	Obs.	Mean	Std. dev.	Min	Median	Max
Physicians						
$\log(1/\alpha(o))$	1,012	0.28	0.50	-2.17	0.33	2.34
$\log(1/\alpha(-o))$	1,012	-1.11	0.14	-1.60	-1.10	-0.70
log(I)	1,012	-1.00	0.06	-1.25	-1.00	-0.84
Dentists						
$\log(1/\alpha(o))$	160	-0.22	0.33	-1.11	-0.22	0.91
$\log(1/\alpha(-o))$	160	-0.98	0.14	-1.34	-0.96	-0.67
log(I)	160	-0.96	0.06	-1.14	-0.95	-0.81
College Professors						
$\log(1/\alpha(o))$	703	-1.67	0.55	-7.10	-1.60	-0.60
$\log(1/\alpha(-o))$	704	-1.05	0.14	-1.49	-1.04	-0.66
log(I)	704	-0.96	0.07	-1.24	-0.96	-0.72
Financial managers						
$\log(1/\alpha(o))$	1,421	-1.76	1.81	-16.46	-1.41	0.72
$\log(1/\alpha(-o))$	1,432	-1.08	0.15	-1.64	-1.07	-0.58
log(I)	1,432	-0.98	0.07	-1.26	-0.98	-0.79
Real estate sales occupations						
$\log(1/\alpha(o))$	1,448	-0.41	0.32	-3.32	-0.38	0.72
$\log(1/\alpha(-o))$	1,448	-1.08	0.15	-1.58	-1.08	-0.60
log(I)	1,448	-1.00	0.07	-1.25	-1.00	-0.79
Secretaries						
$\log(1/\alpha(o))$	1,576	-1.55	0.27	-2.60	-1.55	-0.49
$\log(1/\alpha(-o))$	1,576	-1.08	0.15	-1.63	-1.07	-0.58
log(I)	1,576	-0.98	0.08	-1.26	-0.98	-0.77

Notes: For labor market areas with at least 8 uncensored observations for MLE estimation.

Table 5: Regression Table for Physicians

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.34***	0.32**		1.19**	1.33**
	[0.06, 0.50]	[0.05, 0.49]		[0.35, 2.20]	[0.38, 2.46]
Instrument			1.06***		
			[0.91, 1.23]		
Log of Population		-0.02	-0.09***		0.10
		[-0.13, 0.08]	[-0.11,-0.08]		[-0.06, 0.23]
Log of Income		0.12	0.06***		0.06
		[-0.04, 0.24]	[0.04, 0.09]		[-0.13, 0.17]
1990	0.09***	0.02	-0.03***	-0.01	-0.07
	[0.07, 0.14]	[-0.06, 0.15]	[-0.06,-0.01]	[-0.12, 0.11]	[-0.20, 0.09]
2000	0.10***	-0.01	0.10***	-0.09	-0.20
	[0.07, 0.17]	[-0.13, 0.19]	[0.07, 0.12]	[-0.32, 0.11]	[-0.44, 0.08]
2014	0.21***	0.06	0.06**	-0.00	-0.15
	[0.17, 0.28]	[-0.09, 0.31]	[0.01, 0.09]	[-0.24, 0.20]	[-0.41, 0.20]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2 (ex. LMA FE)	0.21	0.21	0.82	.	.
Observations	1,012	1,012	1,012	1,012	1,012

Bootstrapped standard errors based on 300 draws, stratefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table 6: Regression Table for Dentists

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.60	0.54		2.17*	2.77*
	[-0.12, 1.21]	[-0.16, 1.20]		[-0.11, 3.79]	[-0.19, 5.20]
Instrument			1.66***		
			[1.20, 2.04]		
Log of Population		-0.04	-0.12***		0.30
		[-0.35, 0.35]	[-0.14,-0.08]		[-0.21, 0.83]
Log of Income		0.42	0.03		0.38
		[-0.15, 0.85]	[-0.01, 0.08]		[-0.20, 0.87]
1990	-0.20**	-0.47**	-0.10**	-0.42**	1.56**
	[-0.30,-0.04]	[-0.80,-0.05]	[-0.17,-0.02]	[-0.66,-0.07]	[0.36, 2.58]
2000	-0.17	-0.60	0.11***	-0.56*	0.77**
	[-0.34, 0.04]	[-1.11, 0.07]	[0.03, 0.17]	[-0.97, 0.03]	[0.19, 1.30]
2014	-0.38***	-0.92**	0.04	-0.79**	0.38***
	[-0.55,-0.15]	[-1.57,-0.09]	[-0.06, 0.13]	[-1.24,-0.19]	[0.15, 0.58]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2(ex. LMA FE)	0.19	0.20	0.87	.	.
Observations	160	160	160	160	160

Bootstrapped standard errors based on 300 draws, statefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table 7: IV Regressions for Real Estate Agents (top 20 per cent)

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.17*	0.17*		1.02**	1.32**
	[-0.03, 0.30]	[-0.03, 0.30]		[0.20, 2.09]	[0.29, 2.56]
Instrument			0.64***		
			[0.51, 0.76]		
Log of Population		0.05*	-0.04***		0.11***
		[-0.00, 0.10]	[-0.05,-0.03]		[0.04, 0.20]
Log of Income		0.23***	0.03**		0.19**
		[0.13, 0.33]	[0.00, 0.05]		[0.08, 0.31]
1990	0.01	-0.14***	0.03***	-0.09	0.57***
	[-0.01, 0.04]	[-0.21,-0.08]	[0.01, 0.05]	[-0.21, 0.01]	[0.34, 0.90]
2000	0.06***	-0.20***	0.18***	-0.13	0.31***
	[0.01, 0.10]	[-0.30,-0.10]	[0.15, 0.20]	[-0.38, 0.04]	[0.19, 0.50]
2014	0.01	-0.30***	0.17***	-0.20**	0.14***
	[-0.03, 0.06]	[-0.43,-0.19]	[0.14, 0.20]	[-0.48,-0.01]	[0.10, 0.19]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2(ex. LMA FE)	0.05	0.05	0.82	.	.
Observations	1,448	1,448	1,448	1,448	1,448

Bootstrapped standard errors based on 300 draws, statefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table 8: Regression Table for Financial Managers (top 10 per cent)

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.27***	0.27***		-0.25	-0.29
	[0.10, 0.37]	[0.08, 0.36]		[-1.13, 0.57]	[-1.23, 0.58]
Instrument			0.78***		
			[0.64, 0.90]		
Log of Population		0.00	-0.06***		-0.04
		[-0.07, 0.06]	[-0.08,-0.05]		[-0.11, 0.06]
Log of Income		0.04	0.03***		0.05
		[-0.06, 0.13]	[0.01, 0.05]		[-0.05, 0.13]
1990	0.05***	0.03	0.01	0.12**	0.09
	[0.03, 0.08]	[-0.03, 0.10]	[-0.02, 0.03]	[0.02, 0.21]	[-0.03, 0.20]
2000	0.14***	0.11**	0.15***	0.26**	0.22**
	[0.12, 0.19]	[0.01, 0.22]	[0.13, 0.18]	[0.08, 0.44]	[0.00, 0.40]
2014	0.18***	0.13**	0.13***	0.31***	0.26**
	[0.15, 0.22]	[0.01, 0.28]	[0.10, 0.16]	[0.11, 0.51]	[0.01, 0.46]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2(ex. LMA FE)	0.31	0.31	0.80	.	.
Observations	1,432	1,432	1,432	1,432	1,432

Bootstrapped standard errors based on 300 draws, stratefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

A Appendix: Theory

A.1 Positive assortative matching in equilibrium

Here we show that the equilibrium must feature positive assortative matching between the income of the patient and the skill of the doctor. To do so, we assume that there are 2 individuals 1 and 2 with income $x_1 < x_2$ whose consumption bundles are so that $z_1 > z_2$ and $c_1 < c_2$. For simplicity we write the utility function as a function of health services and the income left for other goods $(x - \omega(z))$.

Note that since consumer 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:

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

If the utility function has a positive cross-partial (which is the case for a Cobb-Douglas), then the first term is positive as $z_1 > z_2$. Since the second term is also weakly positive, then it must be the case that $u(z_1, x_2 - \omega(z_1)) > u(z_2, x_2 - \omega(z_2))$, in other words, consumer 2 would rather pick a doctor of ability z_1. Therefore there is a contradiction and it must be the case that $z_1 < z_2$.

A.2 Solving (6)

We look for a specific solution to(6) of the type $w(z) = K_1 z^{\frac{\alpha_z}{\alpha_x}}$. We find that such a K_1 must satisfy

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

As the solutions to the differential equation $w'(z)\,z+\frac{\beta_z}{1-\beta_z}w(z)=0$ are given by $Kz^{-\frac{\beta_z}{1-\beta_z}}$ for any constant K. We get that all solutions to (6) take the form:

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

We then obtain (7) by using that $w(z_c)=x_{\min}$ which fixes

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

A.3 Proof of remark 1

Using (1), (2), (5) and (10), we get that the utility of a generalist with income x is given by

$$\begin{aligned}u(x)&=(x-h(x))^{1-\beta_z}\left(m^{-1}(x)\right)^{\beta_z}\\&=\left(\frac{\alpha_z(1-\beta_z)}{\alpha_z(1-\beta_z)+\beta_z\alpha_x}x-\frac{1}{\lambda}\frac{\alpha_z(1-\beta_z)+\beta_z\alpha_x(1-\lambda)}{\alpha_z(1-\beta_z)+\beta_z\alpha_x}x_{\min}\left(\frac{x}{x_{\min}}\right)^{-\frac{\alpha_x}{\alpha_z}\frac{\beta_z}{1-\beta_z}}\right)^{1-\beta_z}\left(z_c\left(\frac{x}{x_{\min}}\right)^{\frac{\alpha_x}{\alpha_z}}\right)\end{aligned}$$

Therefore $eq(x)$ obeys

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

which implies that for x large enough

$$eq(x)\approx\frac{\alpha_z(1-\beta_z)\,x_{\min}^{-\frac{\alpha_x}{\alpha_z}\frac{\beta_z}{1-\beta_z}}}{\alpha_z(1-\beta_z)+\beta_z\alpha_x}x^{1+\frac{\alpha_x}{\alpha_z}\frac{\beta_z}{1-\beta_z}}.$$

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

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

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

A.4 Brewers' case

Taking first order conditions with respect to c and y, we obtain that expenditures on beers and on the homogeneous good are related by

$$py = \frac{\beta_y}{1 - \beta_y - \beta_z} c. \tag{22}$$

The first order condition with respect to the quality of the health services consumed and the homogeneous good similarly imply

$$\omega'(z)\, z = \frac{\beta_z}{1 - \beta_y - \beta_z} c. \tag{23}$$

Together with the budget constraint equation

$$\omega(z) + py + c = x,$$

(22) and (23) give (3) so that all results concerning $w(z)$ including (10) still apply, and

$$y(x) = \frac{1}{p} \frac{\beta_y}{1 - \beta_z} (x - h(x)). \tag{24}$$

Market clearing imposes

$$\int_{x_{\min}}^{\infty} y(x)\, dG_x(x) = \mu_m \int_{y_c}^{\infty} y dG_y(y), \tag{25}$$

where $y(x)$ denotes the consumption of beer by a generalist of income x and G_a the cdf of variable a. Plugging (24) in (25) we obtain:

$$py_c = \frac{\psi}{\mu_m} \left(\frac{y_c}{y_{\min}} \right)^{\alpha_y},$$
$$\text{with } \psi \equiv \frac{\alpha_y - 1}{\alpha_y} \frac{\alpha_z \beta_y \left(\frac{1}{\alpha_x} + \lambda - 1 \right)}{\lambda \left(\beta_z + \alpha_z (1 - \beta_z) \right)} \widehat{x}.$$

This implies that there are two possible scenarios. If $\psi \geq \mu_m x_{\min}$ then $py_{\min} \geq x_{\min}$ so that all possible brewers end up working as brewers. We then have

$$p = \frac{\psi}{\mu_m y_{\min}}.$$

Since ψ is decreasing in α_x, a decrease in the shape parameter of generalist income is associated with a proportional increase in brewer's income.

Note that

$$\frac{\psi}{x_{\min}} = \frac{\alpha_z \beta_y (\alpha_y - 1)}{\lambda (\beta_z + \alpha_z (1 - \beta_z)) \alpha_y} \frac{1 + (\lambda - 1) \alpha_x}{\alpha_x - 1}$$

is decreasing in α_x. Therefore as α_x decreases then this situation becomes more and more likely.

Otherwise, $y_c > y_{\min}$ with

$$y_c = y_{\min} \left(\mu_m \frac{x_{\min}}{\psi} \right)^{\frac{1}{\alpha_y}},$$

so that as α_x decreases (and consequently x_{min} to keep mean income of generalists constant), y_c decreases and more and more potential brewers decide to become brewers. This leads to

$$\begin{aligned} p &= \left(\frac{\psi}{\mu_m} \right)^{\frac{1}{\alpha_y}} \left(\frac{\alpha_x - 1}{\alpha_x} \widehat{x} \right)^{\frac{\alpha_y - 1}{\alpha_y}} \frac{1}{y_{\min}} \\ &= \left(\frac{1}{\mu_m} \frac{\alpha_z \beta_y (\alpha_y - 1)}{\lambda (\beta_z + \alpha_z (1 - \beta_z)) \alpha_y} \right)^{\frac{1}{\alpha_y}} \left(\frac{1 + (\lambda - 1) \alpha_x}{\alpha_x - 1} \right)^{\frac{1}{\alpha_y}} \frac{\alpha_x - 1}{\alpha_x} \frac{\widehat{x}}{y_{\min}}. \end{aligned}$$

Note that

$$\begin{aligned} &\frac{d}{d\alpha_x} (1 + (\lambda - 1) \alpha_x)^{\frac{1}{\alpha_y}} \frac{(\alpha_x - 1)^{\frac{\alpha_y - 1}{\alpha_y}}}{\alpha_x} \\ &= [(1 + (\lambda - 1) \alpha_x)(\alpha_y - 1) - (\alpha_x - 1)] \frac{(1 + (\lambda - 1) \alpha_x)^{\frac{1}{\alpha_y} - 1} (\alpha_x - 1)^{\frac{-1}{\alpha_y}}}{\alpha_y \alpha_x^2}, \end{aligned}$$

the sign of which is ambiguous since λ can be close to 1 and we may have $\alpha_x > \alpha_y$. Therefore in this case, a decrease in α_x increases the supply of beers but as a result the impact on brewers' income is ambiguous.

For any price level $\widetilde{p}$, we can define the real welfare measure similarly as the income which gives the same utility in the market and when the agent is forced to consume (for free) z_c while having y prices at $\widetilde{p}$. That is we now have:

$$u \left(z_c, \frac{1 - \beta_z - \beta_y}{1 - \beta_z} eq(x), \frac{\beta_y}{1 - \beta_z} \frac{eq(x)}{\widetilde{p}} \right) = u(z(x), c(x), y(x)).$$

We then obtain:

$$eq\left(x\right)=\left(\frac{\widetilde{p}}{p}\right)^{\frac{\beta_y}{1-\beta_z}}\left(\frac{\alpha_z\left(1-\beta_z\right)}{\alpha_z\left(1-\beta_z\right)+\beta_z\alpha_x}x-\frac{1}{\lambda}\frac{\alpha_z\left(1-\beta_z\right)+\beta_z\alpha_x\left(1-\lambda\right)}{\alpha_z\left(1-\beta_z\right)+\beta_z\alpha_x}x_{\min}\left(\frac{x}{x_{\min}}\right)^{-\frac{\alpha_x}{\alpha_z}\frac{\beta_z}{1-\beta_z}}\right)\left(\frac{x}{x_{\min}}\right)^{\frac{\alpha_x}{\alpha_z}}$$

Therefore the analysis of real income inequality is the same whether $\beta_y=0$ or not.

A.5 Deriving (15)

Using that both doctors talents and income are approximately Pareto distributed, we can rewrite (3) as:

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

From this we get that $m\left(z\right)$ is of the order of $z^{\frac{\alpha_z}{\alpha_x}}$ and therefore

$$m\left(z\right)=Bz^{\frac{\alpha_z}{\alpha_x}}+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right)$$

with B defined as in the text. We can then rewrite (3) as

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

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

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

This gives

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

Integrating we obtain:

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

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

A.6 Appendix: proof of Proposition 3

We rewrite (19) more precisely as:

$$w'(z) = \lambda^{\frac{\varepsilon-1}{\varepsilon}} \frac{\beta_z}{\beta_c} z^{-\frac{1}{\varepsilon}} \left(\lambda B z^{\frac{\alpha_z}{\alpha_x}} - w(z)\right)^{\frac{1}{\varepsilon}} + o\left(\frac{\lambda B z^{\frac{\alpha_z}{\alpha_x}} - w(z)}{z}\right)^{\frac{1}{\varepsilon}}. \quad (27)$$

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

$$w'(z) = \lambda \frac{\beta_z}{\beta_c} B^{\frac{1}{\varepsilon}} z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}} + o\left(z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}\right). \quad (28)$$

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

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

where the notation f^w is introduced to help notation. 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 $\alpha_z=\alpha_x\left(1-\varepsilon\right)$, and integrating (28), one obtains

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

Therefore

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

In that case, w is distributed exponentially.

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

$$w\left(z\right)=Az^{\frac{\alpha_z}{\alpha_x}}+o\left(z^{\frac{\alpha_z}{\alpha_x}}\right) \tag{29}$$

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

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

That is w is Pareto distributed with coefficient α_x.

Plugging (29) in (27), we get:

$$A\frac{\alpha_z}{\alpha_x}z^{\frac{\alpha_z}{\alpha_x}-1}+o\left(z^{\frac{\alpha_z}{\alpha_x}-1}\right)=\lambda^{\frac{\varepsilon-1}{\varepsilon}}\frac{\beta_z}{\beta_c}\left(\lambda B-A\right)^{\frac{1}{\varepsilon}}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}+o\left(\left(\lambda B-A\right)^{\frac{1}{\varepsilon}}z^{\left(\frac{\alpha_z}{\alpha_x}-1\right)\frac{1}{\varepsilon}}\right). \tag{30}$$

First, assume that $\alpha_z=\alpha_x$, then we get that the solution is characterized by $A=\lambda^{\frac{\varepsilon-1}{\varepsilon}}\frac{\beta_z}{\beta_c}\left(\lambda B-A\right)^{\frac{1}{\varepsilon}}$.

Consider now that $\alpha_z\neq\alpha_x$. If $\lambda B\neq A$ then (30) is impossible when $\varepsilon\neq1$, therefore we must have that $\lambda B=A$. This equation then requires that

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

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

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

satisfies (27) 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 proposition 3.

A.7 Appendix: different adjustment margin

In this appendix we fully solve the model described in section 2.2. We obtained that when individuals of a certain rank choose both careers then the share that does so must be constant. Therefore, we guess and verify that the equilibrium takes the following form: individuals above a certain rank choose only one occupation and those below that rank choose both.

Case 1. Consider first the case where there exists a z_c such that individuals of rank higher than $\overline{G}_z(z_c)$ all choose to be generalists. Then (12) applies for $z > z_c$ and we know that for $z \geq z_c$, $\mu = \frac{\lambda^{\alpha_x - 1}}{\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta_z}{\beta_z}+1\right)^{\alpha_x}}$. Since $m(z_c) = x_{\min}$, we obtain:

$$z_c = z_{\min} \left(\frac{\lambda}{\frac{\alpha_z}{\alpha_x}\frac{1-\beta_z}{\beta_z} + 1} \right)^{\frac{\alpha_x}{\alpha_z}}, \tag{31}$$

which is only possible if $\lambda \geq \frac{\alpha_z}{\alpha_x}\frac{1-\beta_z}{\beta_z} + 1$.

Case 2. Consider now the opposite case. Individuals ranked above $\overline{G}_z(z_m)$ all choose to be doctors, those ranked below all choose to be generalists. Since $\lambda > 1$, the supply of health services by agents ranked higher than $\overline{G}_z(z_m)$ is enough to cover their own demand for health services. Therefore, if one denotes by $r(z)$ the rank of the patient of a doctor of quality z, we obtain that there exists a $z_p < z_m$, such that $r(z_p) = z_m$: doctors with ability lower than z_p only provide health services to doctors and those with ability above z provide health services to both doctors and generalists. Since $z_m > z_p$, we have that for $z \geq z_m$, (12) applies which directly leads to $\mu = \frac{\lambda^{\alpha_x - 1}}{\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta_z}{\beta_z}+1\right)^{\alpha_x}}$ for $z \geq z_m$.

We then get to further write for $z \leq z_m$:

$$r(z) = \int_z^{z_m} \lambda g_z(\zeta)\, d\zeta + \int_{z_m}^{\infty} \lambda \mu(\zeta)\, g_z(\zeta)\, d\zeta = \lambda \left(\left(\frac{z_{\min}}{z}\right)^{\alpha_z} - (1-\mu)\left(\frac{z_{\min}}{z_m}\right)^{\alpha_z} \right). \tag{32}$$

For $z \geq z_p$, $m(z) = \overline{G}_x^{-1}(r(z))$, so that (32) implies

$$m(z) = x_{\min} \lambda^{-\frac{1}{\alpha_x}} \left(\left(\frac{z_{\min}}{z}\right)^{\alpha_z} - (1-\mu)\left(\frac{z_{\min}}{z_m}\right)^{\alpha_z} \right)^{-\frac{1}{\alpha_x}} \text{ for } z \in (z_p, z_m).$$

(3) still applies and now gives the differential equation:

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

Using that $w(z_m) = x_{\min} \left(\frac{z_m}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}$, the solution to this differential equation is then given by:

$$w(z) = z^{-\frac{\beta_z}{1-\beta_z}} x_{\min} (z_{\min})^{-\frac{\alpha_z}{\alpha_x}} \left(z_m^{\frac{\alpha_z}{\alpha_x} + \frac{\beta_z}{1-\beta_z}} - \frac{\beta_z}{1-\beta_z} \lambda^{\frac{\alpha_x - 1}{\alpha_x}} \int_z^{z_m} \zeta^{-\frac{1-2\beta_z}{1-\beta_z}} \left(\zeta^{-\alpha_z} - (1-\mu)\, z_m^{-\alpha_z} \right)^{-\frac{1}{\alpha_x}} d\zeta \right).$$

For this to be an equilibrium, we need to check that $w(z) \geq x_{\min} \left(\frac{z}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}$, which is the income that a doctor of rank $\overline{G}_z(z)$ would obtain as a generalist. We can rewrite:

$$w(z) - x_{\min} \left(\frac{z}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}} = x_{\min} (z_{\min})^{-\frac{\alpha_z}{\alpha_x}} z^{-\frac{\beta_z}{1-\beta_z}} T(z)$$

with

$$T(z) \equiv z_m^{\frac{\alpha_z}{\alpha_x} + \frac{\beta_z}{1-\beta_z}} - z^{\frac{\alpha_z}{\alpha_x} + \frac{\beta_z}{1-\beta_z}} - \frac{\beta_z}{1-\beta_z} \lambda^{\frac{\alpha_x - 1}{\alpha_x}} \int_z^{z_m} \zeta^{-\frac{1-2\beta_z}{1-\beta_z}} \left(\zeta^{-\alpha_z} - (1-\mu)\, z_m^{-\alpha_z} \right)^{-\frac{1}{\alpha_x}} d\zeta.$$

We get

$$T'(z) = \left(1 - \left(\frac{z^{-\alpha_z} - (1-\mu)\, z_m^{-\alpha_z}}{\mu z^{-\alpha_z}} \right)^{\frac{1}{\alpha_x}} \right) \frac{\beta_z \lambda^{\frac{\alpha_x - 1}{\alpha_x}}}{1-\beta_z} z^{-\frac{1-2\beta_z}{1-\beta_z}} \left(z^{-\alpha_z} - (1-\mu)\, z_m^{-\alpha_z} \right)^{-\frac{1}{\alpha_x}}.$$

where we used that

$$\frac{\alpha_z}{\alpha_x} \frac{1-\beta_z}{\beta_z} + 1 = \lambda (\mu \lambda)^{-\frac{1}{\alpha_x}}. \tag{33}$$

Further for $z < z_m$, we get that $z^{-\alpha_z} - (1-\mu) z_m^{-\alpha_z} > \mu z^{-\alpha_z}$, so that $T'(z) < 0$. Since $T(z_m) = 0$, then we get that $T(z) > 0$ for $z < z_m$, which ensures that $w(z) > x_{\min} \left(\frac{z}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}$ for $z_p \leq z < z_m$.

Finally, we consider what happens for $z < z_p$. Denote by $d(z)$ the doctor's ability of the individual of rank $r(z)$, then using (32) we get:

$$d(z) = \lambda^{-\frac{1}{\alpha_z}} \left(z^{-\alpha_z} - (1-\mu) z_m^{-\alpha_z}\right)^{-\frac{1}{\alpha_z}}. \tag{34}$$

To close the market, it must be that $d(z_{\min}) = z_{\min}$, which implies that

$$z_m = z_{\min} \left(\frac{1-\mu}{1-\frac{1}{\lambda}}\right)^{\frac{1}{\alpha_z}}. \tag{35}$$

Therefore $z_m > z_{\min}$ is only possible if $\mu < 1/\lambda$, which corresponds to $\lambda < \frac{\alpha_z}{\alpha_x} \frac{1-\beta_z}{\beta_z} + 1$ (the opposite from case 1). Therefore, as long as the equilibrium exists in this case for $\mu < 1/\lambda$, we will have found the overall equilibrium for all possible parameters. We can henceforth assume that in case 2, $\mu < 1/\lambda$.

Further, by definition again, we must have $d(z_p) = z_m$, so that:

$$z_p = \frac{z_m}{\left(1 + \frac{1}{\lambda} - \mu\right)^{\frac{1}{\alpha_z}}} = z_{\min} \left(\frac{1-\mu}{\left(1-\frac{1}{\lambda}\right)\left(1+\frac{1}{\lambda}-\mu\right)}\right)^{\frac{1}{\alpha_z}}. \tag{36}$$

It is direct to verify that for $\mu < 1/\lambda$, $z_{\min} < z_p < z_m$.

Now the patient of the doctor of quality z will have an income given by $w(d(z))$. Therefore (3) gives that for $z \leq z_p$, $w(z)$ must satisfy:

$$w'(z) z = \frac{\beta_z}{1-\beta_z} \left(\lambda w(d(z)) - w(z)\right).$$

Multiply this equation by $z^{\frac{\beta_z}{1-\beta_z}-1}$ and integrate over (z, z_p) to obtain that the solution must satisfy:

$$w(z) = \left(w(z_p) z_p^{\frac{\beta_z}{1-\beta_z}} - \int_z^{z_p} \frac{\beta_z}{1-\beta_z} \zeta^{\frac{2\beta_z - 1}{1-\beta_z}} \lambda w(d(\zeta)) d\zeta\right) z^{-\frac{\beta_z}{1-\beta_z}} \text{ for } z \leq z_p.$$

Once again, we need to verify that $w(z) \geq x_{\min} \left(\frac{z}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}$ for $z < z_p$. Taking the difference we can write:

$$
\begin{aligned}
&w\left(z\right)-x_{\min}\left(\frac{z}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}\\
&=\left(\left(w\left(z_p\right)-x_{\min}\left(\frac{z_p}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}\right)z_p^{\frac{\beta_z}{1-\beta_z}}+\frac{x_{\min}}{z_{\min}^{\frac{\alpha_z}{\alpha_x}}}\left(z_p^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}}-z^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}}\right)-\int_z^{z_p}\frac{\lambda\beta_z\zeta^{\frac{2\beta_z-1}{1-\beta_z}}}{1-\beta_z}w\left(d\left(\zeta\right)\right)d\zeta\right)z^{-1
\end{aligned}
$$

We already know that $w\left(z_p\right)>x_{\min}z_{\min}^{-\frac{\alpha_z}{\alpha_x}}z_p^{\frac{\alpha_z}{\alpha_x}}$. Moreover for $\zeta\in\left(z,z_p\right)$, $d\left(\zeta\right)<z_m$, since $w\left(z\right)$ is increasing we get

$$
w\left(d\left(\zeta\right)\right)\leq w\left(z_m\right)=x_{\min}\left(\frac{z_m}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}.
$$

Therefore, we get:

$$
w\left(z\right)-x_{\min}\left(\frac{z}{z_{\min}}\right)^{\frac{\alpha_z}{\alpha_x}}>\frac{x_{\min}z^{-\frac{\beta_z}{1-\beta_z}}}{z_{\min}^{\frac{\alpha_z}{\alpha_x}}}T_2\left(z\right).
$$

with

$$
T_2\left(z\right)=\left(z_p^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}}-z^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}}\right)-\lambda z_m^{\frac{\alpha_z}{\alpha_x}}\left(z_p^{\frac{\beta_z}{1-\beta_z}}-z^{\frac{\beta_z}{1-\beta_z}}\right).
$$

Differentiating, we get:

$$
T_2'\left(z\right)=\left(\lambda\frac{\beta_z}{1-\beta_z}z_m^{\frac{\alpha_z}{\alpha_x}}-\left(\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}\right)z^{\frac{\alpha_z}{\alpha_x}}\right)z^{\frac{\beta_z}{1-\beta_z}-1}.
$$

Therefore $T_2'\left(z\right)$ has the sign of $\lambda\frac{\beta_z}{1-\beta_z}z_m^{\frac{\alpha_z}{\alpha_x}}-\left(\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}\right)z^{\frac{\alpha_z}{\alpha_x}}$, which is more likely to be negative for a higher z and can change sign at most once on $\left(z_{\min},z_p\right)$. Using (36) and (33) we get that

$$
T_2'\left(z_p\right)=\frac{\beta_z\lambda^{\frac{\alpha_x-1}{\alpha_x}}\mu^{-\frac{1}{\alpha_x}}}{1-\beta_z}z^{\frac{\beta_z}{1-\beta_z}-1}z_p^{\frac{\alpha_z}{\alpha_x}}\left(\left(\left(1+\frac{1}{\lambda}-\mu\right)\lambda\mu\right)^{\frac{1}{\alpha_x}}-1\right).
$$

Note that $\left(1+\frac{1}{\lambda}-\mu\right)\lambda\mu=1-\left(1-\mu\right)\left(1-\lambda\mu\right)$, since $\lambda\mu<1$ and $\lambda>1$ (which implies $\mu<1$), then we get $\left(1+\frac{1}{\lambda}-\mu\right)\lambda\mu<1$. Therefore $T_2'\left(z_p\right)<0$, so that over $\left(z_{\min},z_p\right)$ either T_2 is everywhere decreasing or T_2 is initially increasing and afterwards decreasing. In the former case since $T_2\left(z_p\right)>0$, we directly get that $T_2\left(z\right)>0$ for $z\in\left(z_{\min},z_p\right)$. In the latter case, a necessary and sufficient condition to get $T_2\left(z\right)>0$ over the interval

$(z_{\min}, z_p)$ is that $T_2(z_{\min}) > 0$.

Using (35) and (36), we now compute

$$T_2(z_{\min}) = z_{\min}^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}} \left[\lambda \left(\frac{1-\mu}{1-\frac{1}{\lambda}} \right)^{\frac{1}{\alpha_x}} - 1 - \left(\lambda - \left(\frac{1}{1+\frac{1}{\lambda}-\mu} \right)^{\frac{1}{\alpha_x}} \right) \left(\frac{1-\mu}{1-\frac{1}{\lambda}} \right)^{\frac{1}{\alpha_x}} \left(\frac{1-\mu}{\left(1-\frac{1}{\lambda}\right)\left(1+\frac{1}{\lambda}-\mu\right)} \right)$$

Note that $\lambda - \left(\frac{1}{1+\frac{1}{\lambda}-\mu} \right)^{\frac{1}{\alpha_x}} > 0$ since $\frac{1}{\lambda} > \mu$ and that $\frac{1-\mu}{\left(1-\frac{1}{\lambda}\right)\left(1+\frac{1}{\lambda}-\mu\right)} > 1$ so that $\left(\frac{1-\mu}{\left(1-\frac{1}{\lambda}\right)\left(1+\frac{1}{\lambda}-\mu\right)} \right)^{\frac{1}{\alpha_z}\frac{\beta_z}{1-\beta_z}} > 1$, therefore:

$$\begin{aligned}
T_2(z_{\min}) &> z_{\min}^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}} \left[\lambda \left(\frac{1-\mu}{1-\frac{1}{\lambda}} \right)^{\frac{1}{\alpha_x}} - 1 - \left(\lambda - \left(\frac{1}{1+\frac{1}{\lambda}-\mu} \right)^{\frac{1}{\alpha_x}} \right) \left(\frac{1-\mu}{1-\frac{1}{\lambda}} \right)^{\frac{1}{\alpha_x}} \right] \\
&\geq z_{\min}^{\frac{\alpha_z}{\alpha_x}+\frac{\beta_z}{1-\beta_z}} \left[\left(\frac{1-\mu}{\left(1-\frac{1}{\lambda}\right)\left(1+\frac{1}{\lambda}-\mu\right)} \right)^{\frac{1}{\alpha_x}} - 1 \right] \\
&> 0,
\end{aligned}$$

since $\frac{1-\mu}{\left(1-\frac{1}{\lambda}\right)\left(1+\frac{1}{\lambda}-\mu\right)} > 1$. This guarantees that we always have $T_2(z) > 0$ over $(z_{\min}, z_p)$, so that we obtain $w(z) > x_{\min} \left(\frac{z}{z_{\min}} \right)^{\frac{\alpha_z}{\alpha_x}}$ for $z \in (z_{\min}, z_m)$, which ensures that we do have an equilibrium: no doctor of rank higher than $\overline{G}_z(z_m)$ would like to switch and be a generalist.

A.7.1 Appendix: Proof of Proposition 2

Since $\omega(z)$ is equalized between the two regions, then the threshold z_c of the least able potential doctor must also be the same in the two regions.[23] Summing up the market clearing equations (13) and (14), we obtain that as in the baseline model, $z_c = (\lambda \mu_d)^{\frac{1}{\alpha_z}} z_{min}$. Next combining (13) and (14), we get that

$$x_{\min}^A (1+\varphi(z))^{-\frac{1}{\alpha_x^A}} = x_{\min}^B \left(\frac{z}{z_c} \right)^{\frac{\alpha_z}{\alpha_x^B}-\frac{\alpha_z}{\alpha_x^A}} (1-\varphi(z))^{-\frac{1}{\alpha_x^B}}. \tag{37}$$

[23]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.

Since $\alpha_x^B > \alpha_x^A$, we get 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 $\varphi(z) \to -1$, then the left-hand side would tend 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 would be asymptotically 0, this 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 (13) in (3), we get that in region A:

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

Therefore, asymptotically:

$$w(z) \to \frac{\lambda \beta_z \alpha_x^A 2^{-\frac{1}{\alpha_x^A}}}{\alpha_z(1-\beta_z)+\beta_z\alpha_x^A}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^A}} \tag{38}$$

Since $\varphi(z) \to 1$, after the location decision, doctors' talent is asymptotically distributed with Pareto coefficient α_z in region A: for z high enough, there are $2\mu_d\left(z_{\min}/z\right)^{\alpha_z}$ doctors eventually located in region A. We then directly get that doctor's income distribution is asymptotically Pareto distributed with coefficient α_x^A.

From (37), we get that:

$$\begin{aligned} 1-\varphi(z) &= \left(x_{\min}^B/x_{\min}^A\right)^{\alpha_x^B}\left(1+\varphi(z)\right)^{\alpha_x^B/\alpha_x^A}\left(z/z_c\right)^{\alpha_z\left(1-\alpha_x^B/\alpha_x^A\right)} \\ &\to 2^{\alpha_x^B/\alpha_x^A}\left(x_{\min}^B/x_{\min}^A\right)^{\alpha_x^B}\left(z/z_c\right)^{\alpha_z\left(1-\alpha_x^B/\alpha_x^A\right)}. \end{aligned} \tag{39}$$

Then we can write that in region B, the probability that a doctor earns at least $\widetilde{w}$ is given by:

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

where w above denotes the wage function. Indeed, there are originally $\mu_d P\left(Z > w^{-1}(\widetilde{w})\right)$ doctors present in region B with a talent sufficient to earn $\widetilde{w}$. Out of these doctors, $1-\varphi\left(w^{-1}(\widetilde{w})\right)$ stay in region 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 (38) we get that,

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

Using this expression and (39) we get that:

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

This establishes Proposition 2.

B Empirical Appendix

B.1 Additional Regressions for other occupations

We perform an analysis like that of Tables 5 and 6 for nurses, College professors and Real Estate agents (occupation code 254). Real Estate agents are censored at around top 7 per cent and we use top 20 per cent uncensored observations.

And finally, we show that income inequality for chief executives and public administrators positively predict the income inequality for secretaries in Table B.4.

B.2 Robustness Checks for Physicians

We perform robustness checks for the the regression in Table 5. In particular, Table B.5 shows the regression for different cut-offs. The parameter estimate is generally not far from 1 and remains significant at the 10% level throughout the regressions. Table B.6 shows that the choice of how many LMAs to include does not affect the parameter estimate much.

C Construction of Data on Labor Market Areas

The publicly available data from IPUMS gives information on "country group" in 1980 and "Public Use Microdata Area" (PUMA) for 1990 and onward. We wish to assign these to labor market areas. Dorn (2009) uses a probabilistic approach using the aggregate correspondence between county groups/PUMAs and counties and counties and

Table B.1: IV Regressions for Nurses (top 10 per cent)

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.46***	0.49***		0.86	1.04
	[0.10, 0.59]	[0.11, 0.64]		[-0.61, 2.27]	[-0.72, 2.61]
Instrument			0.82***		
			[0.63, 0.96]		
Log of Population		0.09	-0.07***		0.14
		[-0.04, 0.21]	[-0.08,-0.06]		[-0.03, 0.30]
Log of Income		0.28**	-0.01		0.28**
		[0.04, 0.43]	[-0.02, 0.02]		[0.04, 0.44]
1990	-0.07**	-0.26***	0.03***	-0.12	-0.34**
	[-0.10,-0.00]	[-0.32,-0.06]	[0.01, 0.05]	[-0.33, 0.07]	[-0.58,-0.06]
2000	0.11***	-0.21	0.19***	0.01	-0.35
	[0.07, 0.21]	[-0.33, 0.09]	[0.17, 0.22]	[-0.35, 0.36]	[-0.77, 0.11]
2014	0.16***	-0.24	0.19***	0.06	-0.41
	[0.13, 0.28]	[-0.40, 0.12]	[0.16, 0.23]	[-0.37, 0.46]	[-0.90, 0.14]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2(ex. LMA FE)	0.27	0.28	0.82	.	.
Observations	1,176	1,176	1,176	1,176	1,176

Bootstrapped standard errors based on 300 draws, statefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table B.2: IV Regressions for College Professors (top 10 per cent)

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.35	0.42		-0.82	-0.70
	[-0.18, 0.63]	[-0.12, 0.71]		[-2.64, 1.07]	[-2.59, 1.31]
Instrument			0.95***		
			[0.70, 1.15]		
Log of Population		0.25***	-0.05***		0.19**
		[0.10, 0.41]	[-0.07,-0.03]		[0.01, 0.41]
Log of Income		-0.12	0.03**		-0.10
		[-0.39, 0.21]	[0.01, 0.06]		[-0.37, 0.26]
1990	0.03	0.07	-0.02	0.17	-0.70**
	[-0.02, 0.11]	[-0.14, 0.26]	[-0.06, 0.01]	[-0.06, 0.39]	[-1.27,-0.07]
2000	0.21***	0.25	0.14***	0.47**	-0.50***
	[0.13, 0.35]	[-0.07, 0.58]	[0.10, 0.17]	[0.04, 0.88]	[-0.81,-0.17]
2014	0.38***	0.43**	0.11***	0.67***	-0.20***
	[0.31, 0.53]	[0.03, 0.83]	[0.05, 0.15]	[0.20, 1.13]	[-0.29,-0.10]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2(ex. LMA FE)	0.43	0.44	0.83	.	.
Observations	703	703	704	703	703

Bootstrapped standard errors based on 300 draws, stratefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table B.3: IV Regressions for Secretaries (top 10 per cent)

	(1) OLS $\log(1/\alpha(o))$	(2) OLS $\log(1/\alpha(o))$	(3) 1st Stage $\log(1/\alpha(-o))$	(4) 2SLS $\log(1/\alpha(o))$	(5) 2SLS $\log(1/\alpha(o))$
$\log(1/\alpha(-o))$	0.03	0.06		0.39	0.41
	[-0.11, 0.21]	[-0.09, 0.24]		[-0.69, 1.38]	[-0.68, 1.30]
Instrument			0.69***		
			[0.57, 0.81]		
Log of Population		0.13***	-0.06***		0.16***
		[0.05, 0.22]	[-0.07,-0.04]		[0.04, 0.27]
Log of Income		0.16***	0.02***		0.15**
		[0.03, 0.31]	[0.00, 0.04]		[0.01, 0.31]
1990	-0.05***	-0.17***	0.02*	-0.09	0.28**
	[-0.08,-0.02]	[-0.27,-0.09]	[-0.00, 0.04]	[-0.21, 0.04]	[0.01, 0.57]
2000	0.12***	-0.09	0.17***	0.04	0.08
	[0.06, 0.15]	[-0.25, 0.01]	[0.14, 0.19]	[-0.20, 0.26]	[-0.08, 0.22]
2014	0.07**	-0.20**	0.16***	-0.02	0.12***
	[0.01, 0.12]	[-0.39,-0.06]	[0.13, 0.19]	[-0.27, 0.23]	[0.07, 0.16]
LMA FE	Yes	Yes	Yes	Yes	Yes
R^2(ex. LMA FE)	0.13	0.14	0.80	.	.
Observations	1,576	1,576	1,576	1,576	1,576

Bootstrapped standard errors based on 300 draws, stratefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table B.4: OLS regressions for secretaries on Chief executives and public administrators for 2000 and 2014

	(1) Secretaries	(2) Secretaries	(3) Secretaries	(4) Secretaries
Chief executives and public administrators	0.178***	0.177***	0.136**	0.136**
	(4.24)	(4.28)	(2.02)	(2.00)
2014		-0.0953***	-0.0968***	-0.0892
		(-4.50)	(-4.64)	(-1.21)
Log of Inc.				-0.0446
				(-0.18)
Log of Pop.				0.0391
				(0.19)
Observations	769	769	769	769
Adjusted R^2	0.022	0.046	0.095	0.090

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Regressions limited to 2000 and 2014 due to insufficient information on CEOs in 1980 and 1990. Weighted by number of secretaries by LMA. For 8 observations or more. Column (I) is univariate OLS, Column (II) includes time dummy for 2014, Column (III) further includes labor market area fixed effects and Column (IV) controls for average wage income as well as population.

Table B.5: IV Regressions for Physicians for different cut-offs of Pareto Distribution

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
cut-off	35	40	45	55	65	75
$\log(1/\alpha(-o))$	1.33***	1.48**	1.41*	1.07*	1.17*	1.82*
	[0.25, 2.42]	[0.06, 2.23]	[-0.35, 2.86]	[-0.10, 2.00]	[-0.12, 2.41]	[-0.14, 3.67]
Log of Population	0.10	0.13	0.17*	0.17*	0.20**	0.31**
	[-0.09, 0.22]	[-0.11, 0.26]	[-0.02, 0.34]	[-0.02, 0.35]	[0.02, 0.47]	[0.02, 0.48]
Log of Income	0.06	-0.03	-0.25***	-0.31***	-0.32***	-0.25**
	[-0.11, 0.17]	[-0.23, 0.12]	[-0.45,-0.05]	[-0.56,-0.17]	[-0.66,-0.16]	[-0.71,-0.06]
1990	-0.07	-0.04	-0.20	-0.38***	-0.41***	0.20**
	[-0.19, 0.11]	[-0.14, 0.15]	[-0.70, 0.03]	[-0.92,-0.21]	[-0.97,-0.23]	[0.08, 0.60]
2000	-0.20	-0.17	-0.08	-0.17**	-0.16**	0.00
	[-0.40, 0.12]	[-0.33, 0.19]	[-0.36, 0.06]	[-0.45,-0.08]	[-0.45,-0.04]	[-0.17, 0.66]
2014	-0.15	-0.09	-0.13***	-0.21***	-0.25***	0.18*
	[-0.37, 0.22]	[-0.27, 0.31]	[-0.19,-0.10]	[-0.28,-0.18]	[-0.34,-0.21]	[-0.03, 0.94]
Observations	1,012	1,012	1,012	1,012	1,011	1,011

Bootstrapped standard errors based on 100 draws, statefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Table B.6: IV Regressions for Physicians for different number of LMAs

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
LMAs	100	150	200	253	300	All
$\log(1/\alpha(-o))$	1.70***	1.35**	0.83*	1.33***	1.21**	1.11**
	[0.76, 2.45]	[0.27, 2.01]	[-0.16, 1.57]	[0.25, 2.42]	[0.09, 2.10]	[0.05, 1.87]
Log of Population	0.17	0.07	0.02	0.10	0.12	0.09
	[-0.03, 0.33]	[-0.12, 0.20]	[-0.14, 0.17]	[-0.09, 0.22]	[-0.05, 0.21]	[-0.05, 0.21]
Log of Income	-0.04	0.03	0.09	0.06	0.07	0.08
	[-0.22, 0.15]	[-0.16, 0.17]	[-0.08, 0.19]	[-0.11, 0.17]	[-0.09, 0.22]	[-0.11, 0.24]
1990	0.16	0.13	0.05	-0.07	0.14	0.11
	[-0.22, 0.39]	[-0.25, 0.26]	[-0.30, 0.24]	[-0.19, 0.11]	[-0.26, 0.32]	[-0.24, 0.26]
2000	0.09	0.07	0.02	-0.20	0.07	0.05
	[-0.11, 0.22]	[-0.13, 0.14]	[-0.16, 0.11]	[-0.40, 0.12]	[-0.14, 0.18]	[-0.12, 0.15]
2014	-0.02	-0.04**	-0.05***	-0.15	-0.06***	-0.08***
	[-0.10, 0.03]	[-0.10,-0.01]	[-0.11,-0.02]	[-0.37, 0.22]	[-0.13,-0.03]	[-0.13,-0.04]
Observations	400	600	800	1,012	1,200	1,573

Bootstrapped standard errors based on 100 draws, stratefied at the occupation/year/labor market level. 95 pct confidence interval in brackets. Income is average wage income for those with positive income. $\log(1/\alpha(o))$ refers to income inequality for the occupation of interest. $\log(1/\alpha(-o))$ refers to all occupations, except the occupation of interest. $1/\alpha$ calculated using MLE * $p <= 0.10$, ** $p <= 0.05$, *** $p <= 0.01$

Figure C.1: Labor Market Areas as defined for 1990

commuting zones and creates a "crosswalk" assigning weights for each country group in 1980 to 1990 commuting zones and for each PUMA to 1990 commuting zones. If a given county group or PUMA is assigned to multiple commuting zones we "split" all individuals in the county group or PUMA and give each weights from the crosswalk. The IPUMS data from 2012 onward uses the PUMA2010 (updated from the 2010 federal census) and we construct a new crosswalk along the same lines as Dorn (2009). Counties are very stable across town and we manually correct for county changes between 2000 and 2010. Finally, since our unit of analysis is labor market areas we use Missouri Census Data Center (http://mcdc.missouri.edu/websas/geocorr2k.html) to aggregate commuting zones into labor market areas. Each commuting zone is uniquely assigned to a labor market area. If a single individual had been split into two commuting zones within the same labor market area using Dorn's algorithm we combine the two into one observation aggregating their weights. Figure C.1 shows the labor market areas for 1990.

Table D.1: Top occupations and income inequality ($1/\alpha$)

Occupation	1980 (pred. 95/90)	1990	2000	2014 (pred. 95/90)
	$1/\alpha$			
Chief executives and public administrators	0.24 (1.18)	0.34	0.65	0.57 (1.48)
Financial managers	0.32 (1.25)	0.43	0.48	0.52 (1.44)
Managers and specialists in marketing, advertising, and public relations	0.30 (1.23)	0.33	0.36	0.37 (1.29)
Managers in education and related fields	0.19 (1.14)	0.24	0.23	0.29 (1.22)
Managers and administrators, n.e.c.	0.43 (1.34)	0.45	0.36	0.38 (1.30)
Accountants and auditors	0.27 (1.21)	0.32	0.38	0.44 (1.35)
Computer systems analysts and computer scientists	0.16 (1.12)	0.21	0.25	0.25 (1.19)
Physicians	0.47 (1.39)	0.78	0.55	0.62 (1.54)
Registered nurses	0.17 (1.13)	0.17	0.20	0.23 (1.17)
Subject instructors (HS/college)	0.20 (1.14)	0.24	0.28	0.33 (1.26)
Lawyers	0.42 (1.34)	0.53	0.58	0.58 (1.49)
Computer software developers	0.18 (1.13)	0.19	0.23	0.24 (1.18)
Supervisors and proprietors of sales jobs	0.40 (1.32)	0.45	0.44	0.44 (1.36)
Insurance sales occupations	0.42 (1.34)	0.50	0.52	0.58 (1.49)
Real estate sales occupations	0.45 (1.37)	0.59	0.67	0.69 (1.61)
Salespersons, n.e.c.	0.35 (1.27)	0.40	0.39	0.42 (1.34)
Supervisors of construction work	0.30 (1.23)	0.33	0.29	0.30 (1.23)
Production supervisors or foremen	0.20 (1.15)	0.20	0.26	0.29 (1.23)
Truck, delivery, and tractor drivers	0.20 (1.15)	0.22	0.24	0.26 (1.20)
Military	0.28 (1.21)	0.25	0.28	0.25 (1.19)

Notes: Estimates of $1/\alpha$ for top 20 occupations using top 10 per cent of population

D Pareto Fit and Tables for Top Occupation s

Table D.1 gives the change in α for the top occupations. The top occupations for 1980 and 2014 are given in Table D.2.

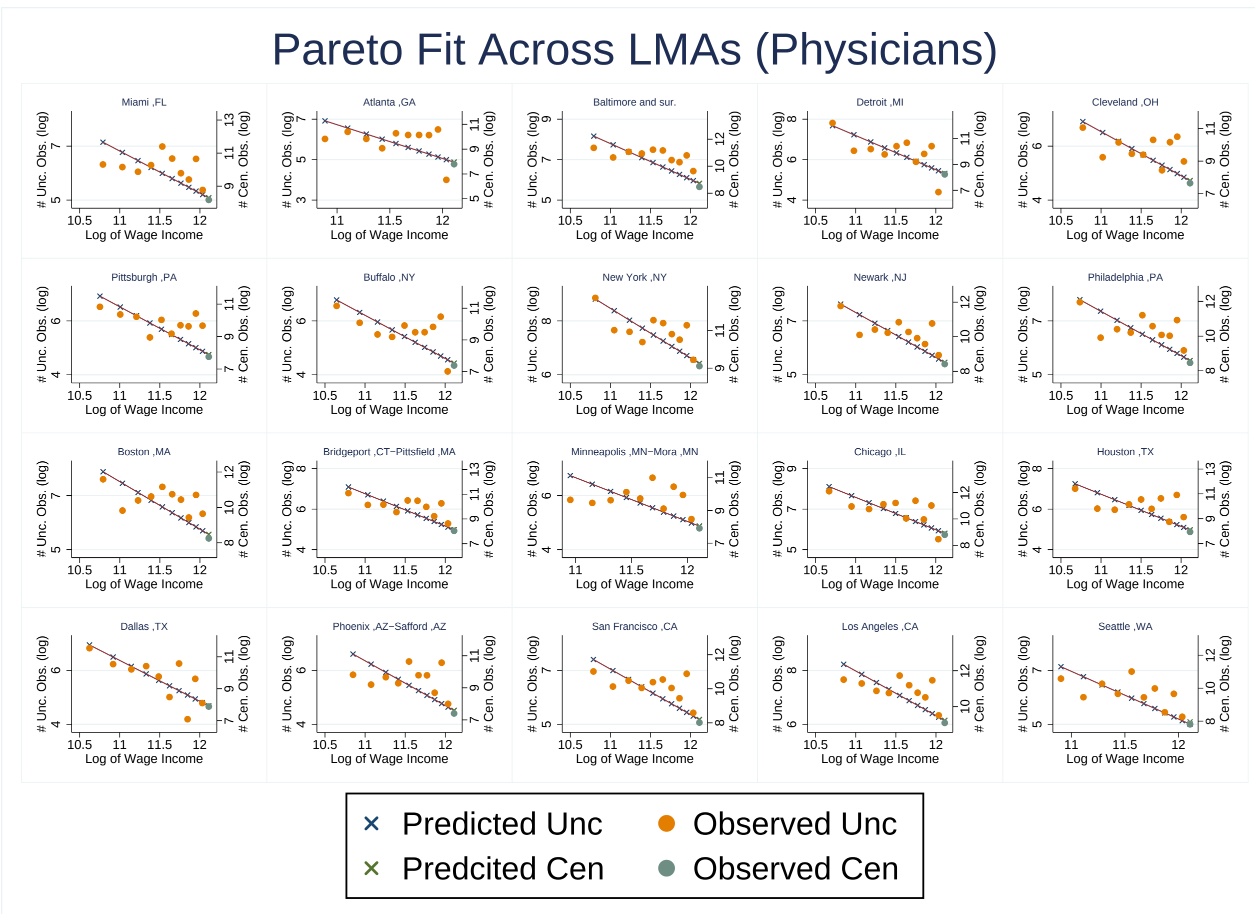
Table D.2: Top occupations in 1%, 5%, 10%, for year 1980 and 2014

	1980			2014		
rank	top 10 pct	top 5 pct	top 1 pct	top 10 pct	top 5 pct	top 1 pct
1	Managers and administrators, n.e.c.	Managers and administrators, n.e.c.	Managers and administrators, n.e.c.	Managers and administrators, n.e.c.	Managers and administrators, n.e.c.	Physicians
2	Salespersons, n.e.c.	Salespersons, n.e.c.	Physicians	Chief executives and public administrators	Physicians	Managers and administrators, n.e.c.
3	Production supervisors or foremen	Production supervisors or foremen	Real estate sales occupations	Computer software developers	Chief executives and public administrators	Chief executives and public administrators
4	Truck, delivery, and tractor drivers	Managers and specialists in marketing, advertising, and public relations	Salespersons, n.e.c.	Physicians	Lawyers	Lawyers
5	Managers and specialists in marketing, advertising, and public relations	Real estate sales occupations	Managers and specialists in marketing, advertising, and public relations	Lawyers	Salespersons, n.e.c.	Real estate sales occupations
6	Real estate sales occupations	Physicians	Lawyers	Salespersons, n.e.c.	Supervisors and proprietors of sales jobs	Supervisors and proprietors of sales jobs
7	Supervisors of construction work	Supervisors and proprietors of sales jobs	Supervisors and proprietors of sales jobs	Supervisors and proprietors of sales jobs	Computer software developers	Salespersons, n.e.c.
8	Supervisors and proprietors of sales jobs	Truck, delivery, and tractor drivers	Production supervisors or foremen	Computer systems analysts and computer scientists	Managers and specialists in marketing, advertising, and public relations	Other financial specialists
9	Physicians	Lawyers	Airplane pilots and navigators	Registered nurses	Financial managers	Financial managers
10	Primary school teachers	Supervisors of construction work	Dentists	Accountants and auditors	Real estate sales occupations	Managers and specialists in marketing, advertising, and public relations

Notes:

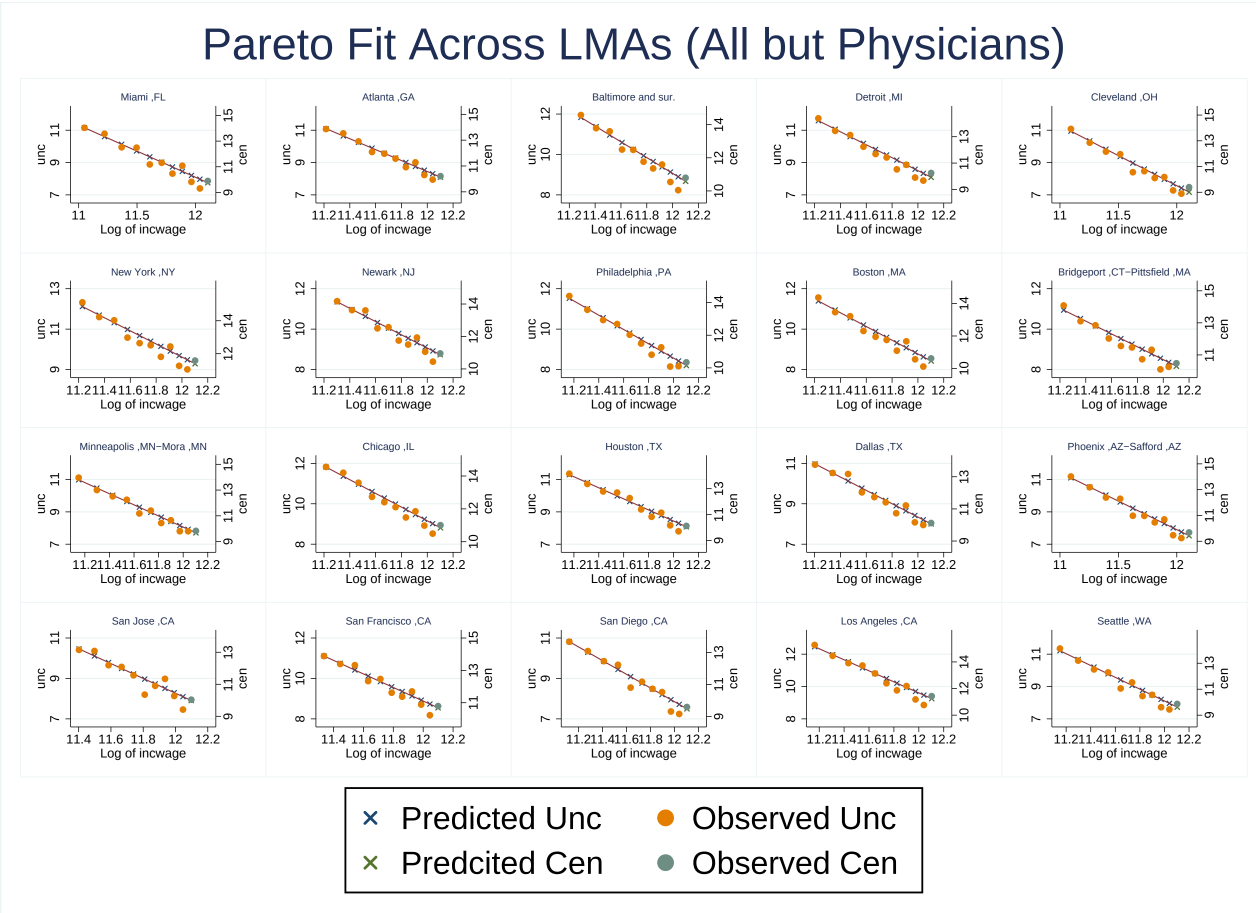
The paper uses the assumption of Pareto for physicians on the LMA-year-occupation level, for LMA-year for the general population and for occupation-year level for the top 20 occupations. Figure 3 in the main text shows the fit with Pareto distribution for the biggest LMA for the whole distribution and for physicians specifically. Figures D.1 and D.2 show analogous figures for the 20 biggest labor market areas for physicians and for the all other occupations than physicians both for the year 2000. Whereas the general population fits the Pareto assumption remarkably well, there is more noise around the line for the physicians, though no systematic deviation.

Figure D.1: Fit to the Pareto Distribution for general income distribution for Physicians for 20 biggest labor market areas for 2000 (using top 65 per cent of uncensored observations)



Notes: Using top 65 per cent of uncensored observations

Figure D.2: Fit to the Pareto Distribution for general income distribution excluding Physicians for 20 biggest labor market areas for 2000 (using top 10 per cent of uncensored observations)



Notes: Using top 10 per cent of uncensored observations