SUPPLEMENT TO "THE SPILLOVER EFFECTS OF TOP INCOME INEQUALITY"

JOSHUA D. GOTTLIEB
Harris School of Public Policy, University of Chicago and NBER

DAVID HÉMOUS
Department of Economics, University of Zürich and CEPR

JEFFREY HICKS
Department of Management, University of Toronto

MORTEN OLSEN
Department of Economics, University of Copenhagen

References to numbered sections, equations, propositions, tables, and figures without an appendix letter refer to the main article.

APPENDIX A: THEORY APPENDIX

A.1. *Positive assortative matching in equilibrium*

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

LEMMA 1: *The equilibrium features positive assortative matching between the income of the patient and the skill of the doctor if the utility function has a positive cross-partial derivative and weakly diminishing marginal utility of consumption.*

PROOF: We prove the result by contradiction. Consider two individuals, 1 and 2, with income $x_1 < x_2$ whose consumption bundles are such that $z_1 > z_2$ and $c_1 < c_2$. Utility depends on z and the remaining disposable income $x - \omega(z)$. Because utility is increasing in both arguments, the equilibrium price schedule must be nondecreasing: otherwise, a higher-quality doctor charging a lower price would strictly dominate a lower-quality doctor. Since widget maker 1 chooses a doctor of quality z_1, it must be that $u(z_1, x_1 - \omega(z_1)) \geq u(z_2, x_1 - \omega(z_2))$. Further, we have:

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

For all $x \in (x_1, x_2)$, the term under the integral is positive: $\frac{\partial u}{\partial c}(z_1, x - \omega(z_1)) > \frac{\partial u}{\partial c}(z_2, x - \omega(z_1))$ because u has positive cross partial derivatives (and $z_1 > z_2$), and $\frac{\partial u}{\partial c}(z_2, x - \omega(z_1)) \geq \frac{\partial u}{\partial c}(z_2, x - \omega(z_2))$ because u has weakly diminishing marginal utility of consumption and $\omega(z_1) \geq \omega(z_2)$. Since $x_2 > x_1$, the first term is positive. The second term is weakly positive, so $u(z_1, x_2 - \omega(z_1)) > u(z_2, x_2 - \omega(z_2))$. In other words, widget maker 2 would rather pick a doctor of ability z_1. This is a contradiction and it must be that $z_1 \leq z_2$.

Assume now that $z_1 = z_2 = z^*$. Then, by weak monotonicity, all individuals with income in (x_1, x_2) would choose the same quality z^*. This would match a positive mass of patients with an atomistic mass of doctors, violating market clearing. Therefore, we must have $z_1 < z_2$, which establishes the lemma. *Q.E.D.*

A.2. *Proofs for the baseline model*

A.2.1. *Solving equation (5)*

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

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

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

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

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

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

A.2.2. *Proof of Proposition 2*

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

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

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

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

Therefore $eq(x)$ obeys

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

which implies that for x large enough:

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

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

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

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

A.3. *Generalization to other utility functions*

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

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

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

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

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

A.3.1. *Cobb-Douglas case*

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

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

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

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

PROOF: We can rewrite (2) as

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

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

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

This gives

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

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

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

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

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

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

A.3.2. *CES case and Proof of Proposition 3*

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

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

Using (9), (20), and $w(z) = \lambda\omega(z)$, we find that for high levels of z the wage function obeys a differential equation given by

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

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

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

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

PROOF: We now establish Propositions 3 and 6. Since consumption of the homogeneous good must remain positive, $\lambda B z^{\frac{\alpha_z}{\alpha_x}} - w(z) \geq 0$ for all sufficiently large z. Hence $w(z)$ 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 $w(z) \propto z^{\frac{\alpha_z}{\alpha_x}}$.[2]

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

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

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

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

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

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

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

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

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

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

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

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

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

Consequently,

$$\left(f^w\right)^{-1}(w) = \exp\left(\frac{1-\beta}{\lambda\beta B^{1/\varepsilon}} w + o(w)\right).$$

[2]One can establish that $A(z) \equiv \frac{w(z)}{z^{\frac{\alpha_z}{\alpha_x}}}$ must have a limit by studying the differential equation that $A(z)$ satisfies and using barrier arguments—proof omitted.

Using the asymptotically Pareto tail of doctors' ability, we obtain

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

Equivalently,

$$\lim_{w\to\infty} \frac{\ln \Pr(W > w)}{w} = -\frac{\alpha_z(1-\beta)}{\lambda\beta B^{1/\varepsilon}}.$$

Thus, doctors' wages have an asymptotically exponential upper tail.

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

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

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

$$\Pr(W > w) = \Pr\left(Z > \left(\left(\frac{w}{A}\right)^{\frac{\alpha_x}{\alpha_z}} + o\left(w^{\frac{\alpha_x}{\alpha_z}}\right)\right)\right) = \Pr(W > \overline{w})\left(\frac{\overline{w}}{w}\right)^{\alpha_x}(1 + o(1))$$

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

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

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

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

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

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

A.3.3. *Homothetic utility function*

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

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

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

$$\lim_{z/c\to\infty} -\frac{d\ln\left(\frac{\partial u}{\partial z}/\frac{\partial u}{\partial c}\right)}{d\ln(z/c)} = \frac{1}{\varepsilon_\infty} \text{ and } \lim_{z/c\to 0} -\frac{d\ln\left(\frac{\partial u}{\partial z}/\frac{\partial u}{\partial c}\right)}{d\ln(z/c)} = \frac{1}{\varepsilon_0}, \tag{30}$$

where $\varepsilon_k \in (0,\infty)$ for $k \in \{0,\infty\}$. In fact, we make the slightly stronger assumption that for z/c arbitrarily large ($k=\infty$) or small ($k=0$):

$$\frac{\partial u}{\partial z} \Big/ \frac{\partial u}{\partial c} = \kappa_k \left(\frac{z}{c}\right)^{-\frac{1}{\varepsilon_k}} (1+o(1)), \tag{31}$$

where κ_k is a constant (but may differ for $k=0$ and $k=\infty$).[3] In these two cases, we can then rewrite (29) as:

$$w'(z) = \lambda^{\frac{\varepsilon_k - 1}{\varepsilon_k}} \kappa_k \left(\lambda \frac{m(z)}{z} - \frac{w(z)}{z}\right)^{\frac{1}{\varepsilon_k}} (1+o(1)), \tag{32}$$

which is similar to (25) in the CES case. We obtain:

PROPOSITION 7: *Propositions 3 and 6 apply to any homothetic utility function that admits positive and finite local elasticities of substitution as the ratio z/c tends to 0 or infinity with the regularity condition (31). The relevant elasticity is ε_0 when $\alpha_z > \alpha_x$ and ε_∞ when $\alpha_z < \alpha_x$ (Proposition 1 applies when $\varepsilon_k = 1$).*

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

Consider now the case $\alpha_z > \alpha_x$. We can use the same logic as in Propositions 3 and 6 with ε_0, but need to verify that $z/c \to 0$ in all cases. If $\varepsilon_0 = 1$, wages are Pareto distributed and health expenditures are an interior share of total income. If $\varepsilon_0 > 1$, then, following Proposition 3, health expenditures become a negligible share of total income. In both cases, using (20), we get that $c/z \sim \lambda B z^{\frac{\alpha_z}{\alpha_x}-1} - \frac{w(z)}{\lambda z} \to \infty$.

If $\varepsilon_0 < 1$, then, following Proposition 6, health expenditures are asymptotically equal to total income. Therefore, $c(z) = o(m(z))$, and using (20), $c(z) = o\left(z^{\frac{\alpha_z}{\alpha_x}}\right)$. Substituting $w(z) = \lambda m(z) - \lambda c(z)$ into (32) gives:

$$m'(z) - c'(z) = \kappa_0 \left(\frac{c(z)}{z}\right)^{\frac{1}{\varepsilon_0}} (1+o(1)).$$

If $c(z)/z$ were bounded, this equation and $m'(z) \sim (\alpha_z/\alpha_x) B z^{\alpha_z/\alpha_x - 1}$ would imply $c'(z) \sim m'(z)$, contradicting $c(z) = o(m(z))$. Thus $c(z)/z$ is unbounded. At any fixed bounded value of $c(z)/z$, the same equation implies $d(c(z)/z)/dz > 0$ for all sufficiently large z. Hence $c(z)/z \to \infty$, as required. *Q.E.D.*

A.4. *Generalized ability distribution*

We consider the set-up of the baseline model with a Cobb-Douglas utility function but generalize the ability distribution to any unbounded distribution with a counter-CDF denoted $\overline{G}_z$. (We keep the widget makers' income distribution Pareto but this could be generalized as well.) We assume that $\lim\limits_{z\to\infty} \frac{z g_z(z)}{\overline{G}_z(z)}$ exists—an asymptotically Pareto distribution corresponds to the

[3](30) does not directly imply (31), because a slowly varying κ_k in (31) is consistent with (30). Assuming (31) directly simplifies the proof of Proposition 7. Nevertheless, Proposition 7 can still be further generalized when (31) is violated (except for the knife-edge case $\alpha_x^{-1} = (1-\varepsilon_\infty)\alpha_z^{-1}$).

case where that limit is positive and finite. As before, there is a cut-off value z_c above which all potential doctors choose to be doctors. We then define $\widetilde{G}(z) = \overline{G}_z(z)/\overline{G}_z(z_c) = \lambda\mu\overline{G}_z(z)$, which is the counter-cumulative ability distribution of individuals who actually choose to be doctors (and $\widetilde{g}$ is the corresponding conditional PDF). We have $\lim_{z\to\infty}\frac{z\widetilde{g}(z)}{\widetilde{G}(z)} = \lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z(z)}$. Equation (4) is then replaced by $m(z) = x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$, which allows us to derive the differential equation for the wage function as:

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

instead of (5). We then establish:

PROPOSITION 8: *If the ability distribution has a tail at least as fat as Pareto ($\lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z(z)}$ exists and is finite), doctors' income is asymptotically Pareto distributed with shape parameter α_x. If the ability distribution has a tail thinner than Pareto ($\lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z(z)} = \infty$), doctors' income need not have a tail asymptotically proportional to a Pareto tail but $\frac{\ln(P(W>w))}{\ln w}$ decreases with α_x for $w > 1$.*

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

PROOF: Define $A(z) \equiv \frac{w(z)}{\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}}$. Budget feasibility implies that $0 < A(z) \leq 1$. Using (33), we get that $A(z)$ satisfies:

$$A'(z)\,z + \frac{1}{\alpha_x}\frac{z\widetilde{g}(z)}{\widetilde{G}(z)}A(z) = \frac{\beta(1-A(z))}{1-\beta}. \tag{34}$$

Case 1: Suppose first that $\lim_{z\to\infty}\frac{zg_z(z)}{\overline{G}_z}$ is finite and equal to ε_g. Define $\tilde{A}(z) = A(z) - \frac{\beta\alpha_x}{\beta\alpha_x+(1-\beta)\varepsilon_g}$. Then (34) implies:

$$\tilde{A}'(z)\,z + \left(\frac{\varepsilon_g}{\alpha_x} + \frac{\beta}{1-\beta}\right)\tilde{A}(z) = \frac{1}{\alpha_x}\left(\varepsilon_g - \frac{z\widetilde{g}(z)}{\widetilde{G}(z)}\right)A(z).$$

Knowing that $A(z)$ is bounded above and integrating, we get:

$$z^{\left(\frac{\varepsilon_g}{\alpha_x}+\frac{\beta}{1-\beta}\right)}\tilde{A}(z) = K + o\left(z^{\left(\frac{\varepsilon_g}{\alpha_x}+\frac{\beta}{1-\beta}\right)}\right),$$

for some constant K. Therefore, $\tilde{A}(z) \to 0$, so that $A(z)$ tends toward $A \equiv \frac{\beta\alpha_x}{\beta\alpha_x+(1-\beta)\varepsilon_g}$ (which is equal to 1 if $\varepsilon_g = 0$). Therefore, $w(z) \sim A\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}$, and

$$P(W > w) \sim \left(\frac{w}{A\lambda x_{\min}}\right)^{-\alpha_x},$$

so that the doctors' income distribution is Pareto distributed with shape parameter α_x.

Case 2: Suppose now that $\lim_{z\to\infty} \frac{z g_z(z)}{\widetilde{G}_z} = \infty$. Assume that $A(z)$ does not tend toward 0 and consider some $\epsilon > 0$. If $A(z) \geq \epsilon$, then for all sufficiently large z, (34) implies that $zA'(z)$ is negative and bounded away from zero. Thus $A(z)$ eventually falls below ϵ. Moreover, at $A(z) = \epsilon$ the derivative is negative, so $A(z)$ cannot subsequently cross that level upward. Then for z, large enough we must have $A(z) < \epsilon$. As this holds for every $\epsilon > 0$, $A(z) \to 0$, that is $w(z) = o\left(\lambda x_{\min}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}\right)$. Then, (33) leads to

$$w'(z) = \frac{\beta}{1-\beta}\lambda x_{\min}\frac{1}{z}\left(\widetilde{G}(z)\right)^{-\frac{1}{\alpha_x}}(1+o(1)). \tag{35}$$

Since $\widetilde{G}(z)$ decreases faster than any power of z, this implies that $w(z) \to \infty$. Integrating (33) and using that $w(z_c) = x_{\min}$, we can write:

$$w(z,\alpha_x) = z^{-\frac{\beta}{1-\beta}}\left[x_{\min} z_c^{\frac{\beta}{1-\beta}} + \frac{\beta}{1-\beta}\lambda x_{\min}\int_{z_c}^{z} s^{\frac{\beta}{1-\beta}-1}\left(\widetilde{G}(s)\right)^{-\alpha_x^{-1}} ds\right],$$

where we explicitly note the dependence of the solution on the parameter α_x. As $x_{\min}$ and z_c are independent of α_x and since $\left(\widetilde{G}(s)\right)^{-\alpha_x^{-1}}$ is decreasing in α_x (recall that $\widetilde{G}(s) < 1$), we get that $w(z,\alpha_x)$ is decreasing in α_x. We then immediately get that, for a given w, $P(W > w)$ decreases with α_x, and so does $\frac{\ln(P(W>w))}{\ln w}$ when $w > 1$. *Q.E.D.*

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

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

$$(m(z)/x_{\min})^{-\alpha_x} = \int_z^{\infty} \lambda(\zeta)\,\alpha_z \frac{1}{\zeta}(\zeta/z_c)^{-\alpha_z} d\zeta, \tag{36}$$

Cobb-Douglas case. In the Cobb-Douglas case, we combine (2), (11) and (36), and we obtain the differential equation:

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

We verify that a solution to the problem of the form $\omega(z) = C_1 z^{\psi}$ exists. Plugging this expression in (37), we obtain a solution with $\psi = \frac{\alpha_z}{\alpha_x+\varepsilon^S}$ and some constant C_1. In fact, the solution must asymptotically behave like $C_1 z^{\psi}$, otherwise the left-hand and right-hand sides of (37) cannot be of the same order. Doctors' income can then be written as

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

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

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

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

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

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

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

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

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

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

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

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

$$\omega'(z) \approx \frac{\beta}{1-\beta} z^{-\frac{1}{\varepsilon}} m(z)^{\frac{1}{\varepsilon}}. \tag{38}$$

We guess and verify that the solution features $m(z) = m_0 z^{m_1} + o(z^{m_1})$ with $m_1 > 0$ and $m_1 \neq 1-\varepsilon$.[5] Integrating (38) gives

$$\omega(z) \approx \frac{\beta}{1-\beta} \frac{1}{\frac{m_1-1}{\varepsilon}+1} m_0^{\frac{1}{\varepsilon}} z^{\frac{m_1-1}{\varepsilon}+1} + K + o\left(z^{\frac{m_1-1}{\varepsilon}+1}\right), \tag{39}$$

for some constant K. Plugging that expression in (36) and using (11) gives

$$(m(z))^{-\alpha_x} x_{\min}^{\alpha_x} = \frac{\alpha_z z_c^{\alpha_z}}{k^{\varepsilon^S}} \int_z^\infty \left(\frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}} \zeta^{\frac{m_1-1}{\varepsilon}+1}}{\frac{m_1-1}{\varepsilon}+1} + K + o\left(\zeta^{\frac{m_1-1}{\varepsilon}+1}\right)\right)^{\varepsilon^S} \zeta^{-\alpha_z-1} d\zeta. \tag{40}$$

[5]For $m_1 = 1-\varepsilon$, we would get $\omega(z) \sim \frac{\beta}{1-\beta} m_0^{1/\varepsilon} (\ln(z))$, and therefore $\lambda(z) \sim C\left(\ln(z)^{\varepsilon^S}\right)$ for some constant $C > 0$. Plugging this expression in (36) leads to a contradiction with the assumption that $(1-\varepsilon)\alpha_x \neq \alpha_z$.

We need to consider three cases. i) First, assume that $\frac{m_1-1}{\varepsilon}+1>0$. Then K is negligible relative to $\zeta^{\frac{m_1-1}{\varepsilon}+1}$ for sufficiently large ζ, and we can rewrite (40) as

$$\left(m_0 z^{m_1}\right)^{-\alpha_x} x_{\min}^{\alpha_x} \approx \frac{\alpha_z z_c^{\alpha_z}}{k^{\varepsilon^S}} \left(\frac{\beta}{1-\beta} \frac{m_0^{\frac{1}{\varepsilon}}}{\frac{m_1-1}{\varepsilon}+1} \right)^{\varepsilon^S} \frac{z^{\left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S-\alpha_z}}{\alpha_z-\left(\frac{m_1-1}{\varepsilon}+1\right)\varepsilon^S}. \tag{41}$$

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

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

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

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

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

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

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

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

We have that

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

since we have assumed that $\alpha_z<\varepsilon^S+\alpha_x$ when $\varepsilon<1$ and $\alpha_z>\varepsilon^S+\alpha_x$ when $\varepsilon>1$. Therefore health care expenditures rise less than proportionately with income and become a negligible income share asymptotically, as assumed initially. This case is therefore internally consistent.

ii) Assume that $\frac{m_1-1}{\varepsilon}+1<0$ and $K\neq 0$. Then (40) leads to

$$m_0^{-\alpha_x}z^{-m_1\alpha_x}x_{\min}^{\alpha_x}\approx\frac{z_c^{\alpha_z}}{k^{\varepsilon^S}}K^{\varepsilon^S}z^{-\alpha_z}. \tag{42}$$

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

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

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

iii) Assume now that $\frac{m_1-1}{\varepsilon}+1<0$ and $K=0$. Then (39) implies that $\omega(z)<0$ for large z, which is impossible.

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

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

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

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

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

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

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

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

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

A.6. *Occupational mobility*

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

Assume that below a certain rank, some individuals choose to be widget makers and some doctors. This holds in equilibrium under a condition specified below. Then, individuals must be indifferent between the two occupations, so that for i low enough, we have $w\left(z\left(i\right)\right)=x\left(i\right)$.

Therefore the wage function must satisfy $w(z) = \overline{G}_x^{-1}\left(\overline{G}_z(z)\right)$ for z high enough. As both ability distributions are Pareto, we get:

$$w(z) = x_{\min}\left(z/z_{\min}\right)^{\alpha_z/\alpha_x}. \tag{43}$$

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

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

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

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

The first-order condition for health care consumption in equation (2) still applies. For sufficiently high z, μ is interior, and (43) holds. Combining these expressions with (44), we obtain:

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

Differentiating with respect to z, we find that μ is a constant: $\mu = \lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta}+1\right)^{-\alpha_x}$. Intuitively, with a constant μ, doctors' wages grow proportionately with patients' incomes, in line with the Cobb-Douglas assumption. To be consistent with our assumption of an interior equilibrium, we must have $\lambda^{\alpha_x-1}\left(\frac{\alpha_z}{\alpha_x}\frac{1-\beta}{\beta}+1\right)^{-\alpha_x} < 1$.[6]

With a constant share μ of individuals choosing to be doctors (above a threshold) and a mass $1/\lambda$ of doctors in equilibrium, we get that $P_{doc}(W_d > w_d) = \lambda\mu P\left(Z > w^{-1}(w_d)\right)$ for w_d high enough so that the observed distribution for doctor wages is Pareto with a shape parameter α_x. Therefore, Proposition 1 still applies:[7]

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

Therefore the models with and without occupational mobility are observationally equivalent for top income inequality: doctors' top income inequality perfectly traces that of widget makers. Finally, note that with occupational mobility, doctors and widget makers interact through two channels: a demand side and an outside option side. Appendix A.6.1 below presents an additional model that separates the two. It highlights that the demand effect drives the result. Intuitively, if top income inequality increases for the outside option, higher-ability doctors will move to the outside option. This generates an increase in the relative pay of the remaining high-ability doctors, which, under Cobb-Douglas preferences, exactly compensates for the change in ability distribution of active doctors, leaving the observed income distribution unchanged.[8]

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

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

[8]This intuition does not generalize to the CES case, in which part of the effect of a rise in income inequality on doctors' income inequality results from the outside-option effect rather than purely from the demand-side effect.

A.6.1. *Disentangling supply-side and demand-side effects*

To disentangle demand-side and outside-option effects, we now assume that doctors have an outside option positively correlated with their ability, whereas patients form a separate group. There are two types of agents: a mass 1 of consumers and a mass M of potential doctors. Consumers have the same utility as in equation (1), and their income, x, is Pareto distributed with shape parameter α_x. Potential doctors only consume the homogeneous good. They are ranked in descending order of ability and agent i can choose between being a doctor providing health services of quality $z(i)$ and earning $w(z(i))$ or working in the homogeneous good sector earning $y(i)$. y and z are distributed according to the counter-cumulative distributions:

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

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

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

$$w(z) = y_{\min}(z/z_{\min})^{\alpha_z/\alpha_y}. \tag{45}$$

Market clearing for health care services above z implies:

$$\left(\frac{x_{\min}}{m(z)}\right)^{\alpha_x} = \lambda M \int_z^\infty \mu(\zeta)\, g_z(\zeta)\, d\zeta, \tag{46}$$

where $\mu(\zeta)$ denotes the share of potential doctors who choose to be doctors. Plugging this expression into the first-order condition in equation (2), together with (45), we obtain:

$$\int_z^\infty \mu(\zeta)\, g_z(\zeta)\, d\zeta = \frac{1}{\lambda M}\left(\frac{\frac{\beta}{1-\beta}\lambda x_{\min}}{\left(\frac{\alpha_z}{\alpha_y} + \frac{\beta}{1-\beta}\right) y_{\min}}\right)^{\alpha_x}\left(\frac{z}{z_{\min}}\right)^{-\alpha_x\frac{\alpha_z}{\alpha_y}}. \tag{47}$$

Taking the derivative with respect to z, we get:

$$\mu(z) = \frac{\alpha_x}{\alpha_y}\frac{1}{\lambda M}\left(\frac{\frac{\beta}{1-\beta}\lambda x_{\min}}{\left(\frac{\alpha_z}{\alpha_y} + \frac{\beta}{1-\beta}\right) y_{\min}}\right)^{\alpha_x}\left(\frac{z}{z_{\min}}\right)^{\alpha_z\left(1-\frac{\alpha_x}{\alpha_y}\right)}. \tag{48}$$

Since $\mu(z) \in (0,1)$, this case is possible only if $\alpha_y < \alpha_x$ (or $\alpha_y = \alpha_x$ and $\frac{\alpha_x}{\alpha_y}\frac{1}{\lambda M}\left(\frac{\alpha_y\beta\lambda x_{\min}}{(\alpha_z(1-\beta)+\beta\alpha_y)y_{\min}}\right)^{\alpha_x} < 1$), that is, consumers' income distribution has a tail no fatter than the potential doctors' outside-option distribution. Then, for w high enough, doctors' income distribution obeys:

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

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

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

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

A.7. *Doctors moving: Proof of Proposition 5*

With no trade in goods between the two regions, we can normalize the price of the homogeneous good to 1 in both. As doctors only consume the homogeneous good, doctors' nominal wages must be equalized in the two regions and the price of health care of quality z must be the same in both regions. From the first-order condition for health care consumption, the matching function is also the same: doctors of quality z provide health care to widget makers of income $m(z)$ regardless of their region of origin (because the income supports differ across regions, for low z, all widget makers with income $m(z)$ live in the same region). Moreover, the least able potential doctor who decides to become a doctor must have the same ability z_c in both regions. [9]

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 (note that $\varphi(z)$ may be negative on part of the support). Labor market clearing in A implies that, for $z \geq z_c$,

$$\left(x^A_{\min}/m(z)\right)^{\alpha^A_x} = \lambda\mu\left(1+\varphi(z)\right)\left(z_{\min}/z\right)^{\alpha_z}. \tag{49}$$

There are initially $\mu\left(\frac{z_{\min}}{z}\right)^{\alpha_z}$ doctors with ability at least z in each region and a share $\varphi(z)$ of those move from region B to region A. As each doctor serves λ patients, after doctors have relocated the total supply above a quality z in region A is given by the right-hand side of (49). Total demand corresponds to region A patients with an income higher than $m(z)$, of which there are $P(X > m(z))$. Since $\alpha^A_x < \alpha^B_x$, the assumption that average incomes are equal in the two regions implies that $x^A_{\min} < x^B_{\min}$. Consequently, doctors of a sufficiently low quality serve patients only in region A. Hence, there exists a threshold z_B, such that doctors born in B with ability $z \in (z_c, z_B)$, move to A. For $z > z_B$, the same equation as (49), replacing $\varphi(z)$ by $-\varphi(z)$, holds in region B:

$$\left(x^B_{\min}/m(z)\right)^{\alpha^B_x} = \lambda\mu\left(1-\varphi(z)\right)\left(z_{\min}/z\right)^{\alpha_z}. \tag{50}$$

Since the two regions are of equal size, total demand for health services must be the same and on net, no doctors move: $\varphi(z_c) = 0$. Using (49) for $z = z_c$, we obtain $z_c = (\lambda\mu)^{\frac{1}{\alpha_z}} z_{\min}$, as in the baseline model.

Similarly, combining (49) and (50) for any $z > z_B$, we obtain

$$x^A_{\min}\left(1+\varphi(z)\right)^{-\frac{1}{\alpha^A_x}} = x^B_{\min}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha^B_x}-\frac{\alpha_z}{\alpha^A_x}}\left(1-\varphi(z)\right)^{-\frac{1}{\alpha^B_x}}. \tag{51}$$

Since $\alpha^B_x > \alpha^A_x$, we find that $\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha^B_x}-\frac{\alpha_z}{\alpha^A_x}}$ tends towards 0. Because $\varphi(z)$ is a net share, $\varphi(z) \in [-1, 1]$. If $1+\varphi(z)$ is not bounded away from 0, then the left-hand side is arbitrarily

[9] Potential doctors who decide to work in the homogeneous good sector earn the same outside option $\widehat{x} < \lambda x^A_{\min}$ in the two regions.

large on some interval while the right-hand side is arbitrarily close to 0, which is a contradiction. Therefore $1+\varphi(z)$ must be bounded away from zero, which ensures that the left-hand side is bounded above and bounded away from 0. If $\varphi(z) \not\to 1$, then the right-hand side is arbitrarily close to 0 on some interval, which is also a contradiction. Therefore asymptotically, we must have that $\varphi(z) \to 1$: nearly all the best doctors move to the most unequal region.

Plugging (49) in (2), we get that in region A:

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

Therefore, asymptotically:

$$w(z) \sim \frac{\lambda\beta\alpha_x^A 2^{-\frac{1}{\alpha_x^A}} x_{\min}^A}{\alpha_z(1-\beta)+\beta\alpha_x^A}\left(\frac{z}{z_c}\right)^{\frac{\alpha_z}{\alpha_x^A}} \tag{52}$$

As $\varphi(z) \to 1$, doctors' ex-post ability distribution in region A is asymptotically Pareto with shape parameter α_z. For high z, the mass of doctors with ability at least z who are eventually located in A is asymptotic to $2\mu\left(\frac{z_{\min}}{z}\right)^{\alpha_z}$. Then, as in the baseline model, doctors' income is asymptotically Pareto distributed with coefficient α_x^A in A. Further, using (51), we get:

$$1-\varphi(z) \sim 2^{\alpha_x^B/\alpha_x^A}\left(x_{\min}^B/x_{\min}^A\right)^{\alpha_x^B}\left(z/z_c\right)^{\alpha_z\left(1-\alpha_x^B/\alpha_x^A\right)}. \tag{53}$$

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

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

where w denotes the wage function. Indeed, there are initially $\mu P\left(Z>w^{-1}(\widetilde{w})\right)$ doctors in region B with a talent sufficient to earn $\widetilde{w}$. A share of $1-\varphi\left(w^{-1}(\widetilde{w})\right)$ of these doctors stay in B. Moreover, the total mass of active doctors in region B is given by $\mu P(Z>z_c)$, since overall there is no net movement of actual doctors. Using (52), we get:

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

Using this expression and (53) we get that:

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

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

A.8. *Quality-quantity tradeoff*

We consider the model of Section 3.3.4.[10] With $\omega(z)$ the price per unit of health care of quality z, the budget constraint faced by a consumer with income x is:

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

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

PROPOSITION 10: *1) In the Cobb-Douglas case, $\theta = 1$, there is no positive assortative matching. Doctors' income is Pareto distributed with inverse Pareto parameter $\frac{\gamma}{1-\gamma}\alpha_z^{-1}$, independent of widget makers' top income inequality.*

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

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

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

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

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

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

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

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

$$\max_z H(z,h) = \left((1-\gamma)^{\frac{1}{\theta}} \left(\frac{h}{\omega(z)}\right)^{\frac{\theta-1}{\theta}} + \gamma^{\frac{1}{\theta}} z^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}. \tag{54}$$

We obtain the first-order condition:

$$\frac{\partial H}{\partial z} = \left(\gamma^{\frac{1}{\theta}} z^{-\frac{1}{\theta}} - (1-\gamma)^{\frac{1}{\theta}} h^{\frac{\theta-1}{\theta}} \omega(z)^{\frac{1-2\theta}{\theta}} \omega'(z)\right) H^{\frac{1}{\theta}} = 0 \tag{55}$$

which leads to the optimal z satisfying:

$$z^* = \frac{\gamma}{1-\gamma} h^{1-\theta} \omega(z^*)^{2\theta-1} \omega'(z^*)^{-\theta}. \tag{56}$$

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

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

$$\frac{\partial^2 H}{(\partial z)^2}(z^*, h) = -\gamma^{\frac{1}{\theta}} g(z^*) \frac{\omega'(z^*)}{\omega(z^*)} (z^*)^{-\frac{1}{\theta}} H^{\frac{1}{\theta}}$$

$$\text{with } g(z^*) \equiv \frac{1}{\theta} \frac{1}{\varepsilon_\omega(z^*)} + \frac{1-2\theta}{\theta} + \frac{\omega(z^*)\omega''(z^*)}{(\omega'(z^*))^2} \quad (57)$$

where we defined $\varepsilon_\omega(z) \equiv \frac{z\omega'(z)}{\omega(z)}$ as the elasticity of the price function. Therefore at an interior optimum, we must have $g(z^*) > 0$. Differentiating (55), we obtain:

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

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

$$H(z) = z\gamma^{\frac{1}{\theta-1}} \left(\frac{1}{\varepsilon_\omega(z)} + 1\right)^{\frac{\theta}{\theta-1}}, \quad (58)$$

$$\text{with } \frac{dH}{dz} = \frac{\partial H}{\partial z} + \frac{\partial H}{\partial h}\frac{dh}{dz} = \frac{\partial H}{\partial h}\frac{1}{\dfrac{dz}{dh}} = \gamma^{\frac{1}{\theta}} g(z^*)(z^*)^{-\frac{1}{\theta}} \frac{H^{\frac{1}{\theta}}\theta}{1-\theta} > 0. \quad (59)$$

In addition, the quantity of health care demanded by a consumer consuming quality z is then given by

$$q(z) = z\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}} (\varepsilon_\omega(z))^{\frac{\theta}{1-\theta}}. \quad (60)$$

Using (56), we can then rewrite the original problem as

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

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

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

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

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

$$\beta\frac{dH}{dz}\left(m(z)-\left(\frac{1-\gamma}{\gamma}z\omega(z)^{1-2\theta}\omega'(z)^{\theta}\right)^{\frac{1}{1-\theta}}\right) \tag{61}$$

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

We note that

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

Substituting this expression and (59) into (61), and noting that $g(z)>0$ at the optimum, we get

$$\beta z^{-\frac{1}{\theta}}\left(m(z)-\left(\frac{1-\gamma}{\gamma}z\omega(z)^{1-2\theta}\omega'(z)^{\theta}\right)^{\frac{1}{1-\theta}}\right)$$

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

Then using (58), we obtain:

$$\beta m(z)=\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}z\omega(z)\left(\varepsilon_{\omega}(z)\right)^{\frac{\theta}{1-\theta}}\left[1+\varepsilon_{\omega}(z)(1-\beta)\right]. \tag{62}$$

This differential equation characterizes the optimal behavior of consumers.

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

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

Differentiating with respect to z, we obtain:

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

Plugging in (60), we get a second differential equation that characterizes market clearing:

$$z^{-\alpha_z-2}\alpha_z\lambda\mu z_{\min}^{\alpha_z}=m'(z)\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\theta}}\left(\varepsilon_{\omega}(z)\right)^{\frac{\theta}{1-\theta}}m(z)^{-(\alpha_x+1)}\alpha_x x_{\min}^{\alpha_x}. \tag{63}$$

Together with initial conditions determining the cutoff z_c and ensuring that $w(z_c)=x_{\min}$, equations (62) and (63) fully characterize the equilibrium.

We are interested in the asymptotic behavior of the solution. Multiplying equations (62) and (63) with each other, we get

$$\left[\omega(z)z^{-\alpha_z-1}+(1-\beta)z^{-\alpha_z}\omega'(z)\right]\alpha_z\lambda\mu z_{\min}^{\alpha_z}=m'(z)m(z)^{-\alpha_x}\beta\alpha_x x_{\min}^{\alpha_x}. \tag{64}$$

We assume that the solution behaves regularly in the sense that $\varepsilon_\omega(z)$ admits a limit in $[0,\infty]$ (we know that $\varepsilon_\omega(z)\geq 0$) and, whenever this limit is finite and positive, $\omega(z)$ and its first two derivatives admit the power-function asymptotic expansions used below. We consider in turn the three possible cases of an infinite limit, a finite but positive limit and a null limit.

Case 1: $\varepsilon_\omega(z)\to\infty$. Then, $\omega(z)$ must grow faster than any power function and $\omega(z)\, z^{-\alpha_z-1}$ becomes negligible relative to $z^{-\alpha_z}\omega'(z)$. Using (64), we can write:

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

Integrating, we get

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

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

Case 2: $\varepsilon_\omega(z)\to\omega_1$ with $\omega_1>0$ and finite. Then, by assumption, we can write $\omega(z)=\omega_0 z^{\omega_1}+o(z^{\omega_1})$ and $\omega'(z)=\omega_0\omega_1 z^{\omega_1-1}+o(z^{\omega_1-1})$. Using (64), we can write

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

Integrating, we obtain[11]

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

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

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

Plugging that expression in (62), we get

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

This implies that we must have

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

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

[11]This assumes $\omega_1\neq\alpha_z$. For $\omega_1=\alpha_z$, integration leads to a $\ln$ on the left-hand side and a finite tail integral on the right-hand side (as $\alpha_x>1$ and $m(z)\to\infty$), which is impossible

We still need to check that $g(z) > 0$ to satisfy the second-order condition. To do so, we first compute $\omega''(z)$. Log-differentiating (62), we get

$$\frac{zm'(z)}{m(z)} = \frac{1}{1-\theta} + \frac{1-2\theta}{1-\theta}\varepsilon_\omega(z) + \frac{\theta}{1-\theta}\frac{z\omega''(z)}{\omega'(z)} + \frac{(1-\beta)\varepsilon_\omega(z)\left(1+\frac{z\omega''(z)}{\omega'(z)} - \varepsilon_\omega(z)\right)}{1+\varepsilon_\omega(z)(1-\beta)} \tag{65}$$

From (63), we get that $m'(z)$ is asymptotically a power function and that $\frac{zm'(z)}{m(z)}$ must tend toward a constant. Then (65) implies that $\frac{z\omega''(z)}{\omega'(z)}$ also tends toward a constant, which must be $\omega_1 - 1$, so that we can write $\omega''(z) = \omega_0\omega_1(\omega_1 - 1)z^{\omega_1-2} + o(z^{\omega_1-2})$. Using (57), we get

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

so that the SOC is satisfied.

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

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

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

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

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

Case 3: $\varepsilon_\omega(z) \to 0$. Then $\omega(z)$ grows more slowly than any power function, and $\omega'(z)$ is negligible relative to $\frac{\omega(z)}{z}$. Using (64), we can write:

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

Integrating, we get

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

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

$$m(z) \sim \left[\frac{z^{\alpha_z}}{\omega(z)}\frac{\beta\alpha_x}{\alpha_x - 1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1}{\alpha_x-1}}. \tag{66}$$

Plugging this in (62), we obtain

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

Rearranging terms, we get

$$\left[\frac{\beta^{\alpha_x}\alpha_x}{\alpha_x - 1}\frac{x_{\min}^{\alpha_x}}{\lambda\mu z_{\min}^{\alpha_z}}\right]^{\frac{1-\theta}{(\alpha_x-1)\theta}} z^{\frac{\alpha_z-\alpha_x+1}{\alpha_x-1}\frac{1-\theta}{\theta}-1} \sim \omega'(z)\,\omega(z)^{\frac{\alpha_x}{\alpha_x-1}\frac{1-\theta}{\theta}-1}\left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{\theta}}. \tag{67}$$

Integrating (and recalling that we are ignoring the case $\alpha_z = \alpha_x - 1$ in this proposition) we get

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

for some constant K. The right-hand side is increasing in $\omega(z)$, while the left-hand side is increasing in z only if $\alpha_x - 1 > \alpha_z$. This condition is therefore necessary for the equilibrium to fall into case 3. Assuming that this is the case, we get $\omega(z) = \Omega - \omega_0 z^{-\omega_1} + o(z^{-\omega_1})$, with $\Omega, \omega_0 > 0$ and $\omega_1 = \frac{(\alpha_x - 1 - \alpha_z)(1-\theta)}{(\alpha_x - 1)\theta} > 0$. Therefore the price of health care quality is bounded above and so is doctors' income.

We then need to check that the SOC holds. From (67), we get

$$\omega'(z) \sim \Omega^{1 - \frac{\alpha_x}{\alpha_x - 1}\frac{1-\theta}{\theta}} \left(\frac{\gamma}{1-\gamma} \right)^{\frac{1}{\theta}} \left[\frac{\beta^{\alpha_x} \alpha_x}{\alpha_x - 1} \frac{x_{\min}^{\alpha_x}}{\lambda \mu z_{\min}^{\alpha_z}} \right]^{\frac{1-\theta}{(\alpha_x - 1)\theta}} z^{\frac{\alpha_z}{\alpha_x - 1}\frac{1-\theta}{\theta} - \frac{1}{\theta}}. \tag{68}$$

Equation (65) still holds and plugging (63) in it, we can write:

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

Using (66) and (68), we can derive

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

Plugging this expression and (68) in (57), we obtain

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

which ensures that the second-order condition is asymptotically satisfied.

Bringing together cases 1), 2) and 3) establishes parts 2 and 3 of Proposition 10.

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

$$\max_z H(z,h) = h^{1-\gamma} \frac{z^\gamma}{(\omega(z))^{1-\gamma}}. \tag{69}$$

The solution of this problem is independent of h. If consumers were not indifferent across qualities, they would all choose the same health care quality. This cannot be an equilibrium, so consumers must be indifferent across health care qualities, which requires $\omega(z) \propto z^{\frac{\gamma}{1-\gamma}}$. Hence, there is a common quality-adjusted price of health care, which doctors take as given. Doctors' income is Pareto distributed with shape parameter $\alpha_z \frac{1-\gamma}{\gamma}$. *Q.E.D.*

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

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

A.9. *Partly tradable health care*

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

APPENDIX B: DATA APPENDIX

B.1. *Details of data construction and Figure 1*

Sample Selection. We restrict to Census/ACS respondents (U.S. Census Bureau, 2000, 2014) who are age 25 or older, and (1) have positive income and are categorized as "employed, at work" according to the variable ESR (employment status recoded) or (2) have positive income, are not in the labor force, and are age 65 or older. The latter group approximates retirees. For the positive income restriction, income refers to whichever definition is being used; typically wage income, but in robustness checks we also use all earned income. All constructed variables use Census weights (perwt).

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

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

1. Among *uncensored* observations, calculate the income at that percentile in that LMA-year among all persons regardless of occupation. Drop observations below that income level.
2. Then, drop the occupation of interest, that is, the occupation in the dependent variable.
3. Then, calculate the inverse Pareto parameter as described in the main text. We adjust equation (13) to account for a few censored observations. Consider a sample of draws of a random variable $\tilde{X}$ which follows a Pareto distribution $P(\tilde{X} > \tilde{x}) = (\tilde{x}/x_{\min})^{-\alpha}$. With censoring, the observed value is $x = \min\{\tilde{x}, \bar{x}\}$ for some known censoring point $\bar{x} > x_{\min}$. Denote by $\mathcal{N}_{cen}$ and $\mathcal{N}_{unc}$ the sets of censored and uncensored observations, respectively. Since all estimates use Census person weights, the frequency-weighted maximum likelihood estimator of α^{-1} is

$$\widehat{\alpha}^{-1} = \frac{\sum_{i \in \mathcal{N}_{unc}} \text{perwt}_i \ln\left(\frac{x_i}{x_{\min}}\right) + \sum_{i \in \mathcal{N}_{cen}} \text{perwt}_i \ln\left(\frac{\bar{x}}{x_{\min}}\right)}{\sum_{i \in \mathcal{N}_{unc}} \text{perwt}_i}.$$

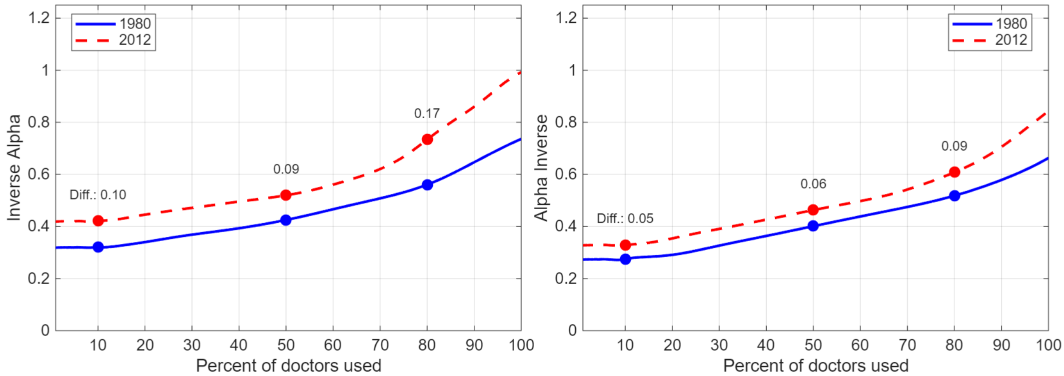
Armour, Burkhauser, and Larrimore (2016) use this method with Current Population Survey data (March supplement) to show that income inequality trends match those found by Kopczuk, Saez, and Song (2010) using Social Security data.

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

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

Instrument Construction in the Regression Analysis. For each occupation of interest o, we construct our shift-share instrument as follows. We first identify the 10 most common occupations (excluding o) in the upper tail of each LMA in 1980, where the upper tail corresponds to the 90th percentile of the local uncensored observations. We define the set of shift-share occupations, K_{-o}, as the union of all these occupations, which corresponds to around 30 occupations. Then, for each occupation κ in K_{-o}, we calculate the income at the percentile cutoff defining the upper tail (*e.g.*, the 90th percentile) among uncensored observations *nationwide*. Observations with income below that level are dropped. Then, for each region s, we drop respondents residing in that region and calculate the nationwide inverse Pareto parameter for each occupation ($\alpha^{-1}_{\kappa,t,-s}$). We then calculate the weights (or shares) for occupation $\kappa \in K_{-o}$

FIGURE B.1.—Estimated α^{-1} for doctors for various cutoffs.



(a) New York State (b) National

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

in the instrument as the fraction of the upper tail population that is employed in κ in 1980. The weights are not normalized, and thus sum to less than one. Retirees are excluded from the instrument.

Measures of α^{-1} and sample selection. In our main specifications, for each occupation we estimate α^{-1} using members of that occupation whose incomes lie in the top 10% of the general population. For New York State and the national sample, we illustrate how the estimated α^{-1} for doctors depends on the size of the sample (we use New York State and not New York LMA for disclosure reasons). Our data here are the same as those used for the calibration exercise in Section 6. We defer a detailed discussion of data and data cleaning until then.

Figure B.1, panel (a), shows the α^{-1} estimated for 1980 and 2012 on increasing fractions of doctors in New York State once we restrict attention to doctors with incomes higher than those in medical residency. When using the highest-earning 10% of doctors we get a difference between 1980 and 2012 of 0.101. When we use half the observations we get a difference of 0.09. The fraction of (non-resident) doctors who are in the top 10% of the general population is around 80% in 2012. When we use this fraction we get 0.173.[12] Panel (b) performs the same calculation for the nation as a whole with the same points highlighted.

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

B.2. *Occupation and LMA definitions*

Occupations. We use the occupation classification constructed by Deming (2017), which ensures consistent occupational groups throughout our sample. We create some additional

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

groupings: We combine (1) all engineering occupations into one, (2) all managers (excluding those working in real estate) into one, (3) primary and secondary school teachers into one, (4) respiratory, occupational, physical, speech and not-elsewhere-classified therapists into one, (5) hairdressers and barbers into one, and (6) waiters and bartenders into one.

Geography. We use Labor Market Areas (LMAs) defined by Tolbert and Sizer (1996). These are aggregates of the 741 Commuting Zones (CZs) popularized by Dorn (2009). Both commuting zones and labor market areas are defined based on the commuting patterns between counties. CZs are unrestricted in size, whereas LMAs aggregate CZs to ensure a population of at least 100,000. LMAs are constructed such that they can be driven through in a matter of a few hours, *e.g.* Los Angeles or New York. Given that our estimation strategy relies on a relatively high number of observations of a particular occupation, LMAs are a more natural choice.

B.3. *Occupational characteristics*

We use the Occupational Information Network (O*NET) 10.0 June 2006 release of occupation characteristics (National Center for O*NET Development, 2006), drawing on two traits in particular: "Customer and Personal Service" from the list of "Knowledge" traits, and "Performing for or Working Directly with the Public" from the list of "Work Activities". Each trait is scored on a scale from 0 to 7 based on the "level" of skill in that trait required for the occupation. We crosswalk from O*NET-SOC codes to 2000 Census occupation groups (occ2000) using the SOC-to-occ2000 weights used by Acemoglu and Autor (2011) (available at https://economics.mit.edu/people/faculty/david-h-autor/data-archive). We then collapse from occ2000 to the occupation definition used in this paper, which was developed by Deming (2017) and which we call occ1990dd, as described in the main text. For this collapse, we take the weighted average of each O*NET trait at the `occ1990dd` level, where the weights correspond to the fraction of each occ1990dd population derived from each occ2000 category. We calculate these weights using the 2000 public-use Census sample (RFSB+, 2025). Finally, once we have O*NET traits at the occ1990dd level, we normalize the traits to be in percentile rankings (from 0 to 1) rather than on the 0 to 7 scale. Each occupation is assigned its percentile ranking in the occupation distribution of the trait (weighted by occupation size). These two normalized traits are the bases of Figure 3.

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

APPENDIX C: ASSORTATIVE MATCHING IN HEALTHCARE AND PHYSICIANS' SCALE

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

C.1. *Spending and price data*

We use medical spending data from the Medical Expenditure Panel Survey (MEPS), accessed through IPUMS Health Surveys (Blewett, Rivera Drew, Griffin, and Williams, 2019). MEPS is a nationally representative survey of families' health insurance coverage and medical spending conducted annually by the Agency for Healthcare Research and Quality. We use the survey waves 2010-2014 to match the five-year ACS data used in our main results. For each family in each year, we calculate total medical spending and total dental spending. For families present in the survey for multiple years, we average their annual spending and annual family income. We then take logs of spending and income to estimate the income elasticity of spending.

We measure service provider *prices* using the Colorado All-Payer Claims Data (APCD-CO) (Center for Improving Value in Health Care, 2015), described in Clemens, Gottlieb, and Molnár (2017). These data provide details on patient visits for medical care, including the service provided (a 5-digit code established by the Healthcare Common Procedure Coding System (HCPCS)) and the identity of the treating provider. They cover "the majority of covered lives in the state across commercial health insurance plans, Medicare (Fee-for-Service and Advantage), and Health First Colorado (Colorado's Medicaid program)". To match the MEPS years and avoid confounding price changes during a calendar year (Clemens, Gottlieb, and Molnár, 2017), we include visits in January through June of each year from 2010 to 2014. Crucially, APCD-CO indicates the patient's residential zip code and the amount the provider was paid for each service (whether by an insurer, directly by the patient, or both).[13]

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

$$\ln r_{g,j,i,t} = \varphi_g + \varphi_j + \varphi_t + \varepsilon_{g,j,i,t} \tag{70}$$

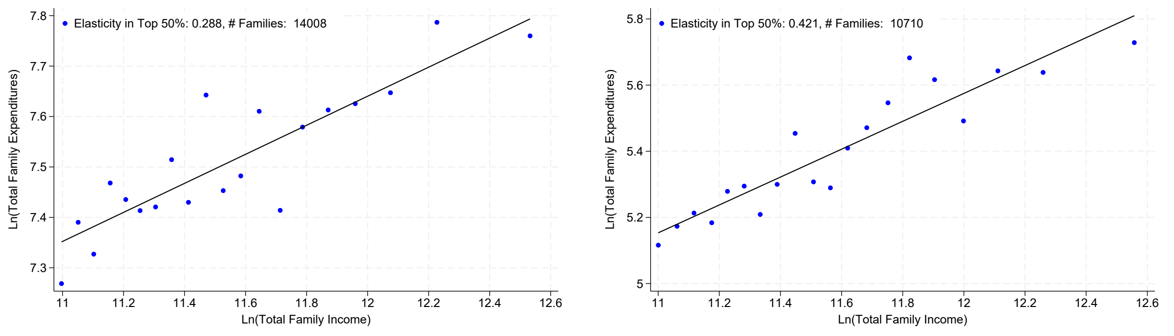
where $\hat{\varphi}_j$ are fixed effects for HCPCS procedure codes, $\hat{\varphi}_t$ are fixed effects for each year, and $\hat{\varphi}_g$ are fixed effects for provider g that reflect each provider's average mark-up.

We approximate patients' income using the median family income in their residential zip code.[14] We compute the mean of provider markups, $\hat{\varphi}_g$, among all provider visits from patients in a given zip code z: $\overline{\varphi_z} = \frac{1}{N_{visit}} \sum_{visit \in z} \hat{\varphi}_g$. We then estimate $\overline{\varphi_z} = \mu_0 + \mu_1 \ln(\text{income})_z + \varepsilon_z$ while weighting observations by the number of underlying patient claims in z.

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

[14]We obtain data on median family income in each Zip Code Tabulation Area (areas that closely approximate zip codes) from IPUMS NHGIS (Manson, Schroeder, Riper, and Ruggles, 2017).

FIGURE C.1.—Engel Curve for Medical Spending



(a) Total Medical Spending (b) Dental Spending

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

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

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

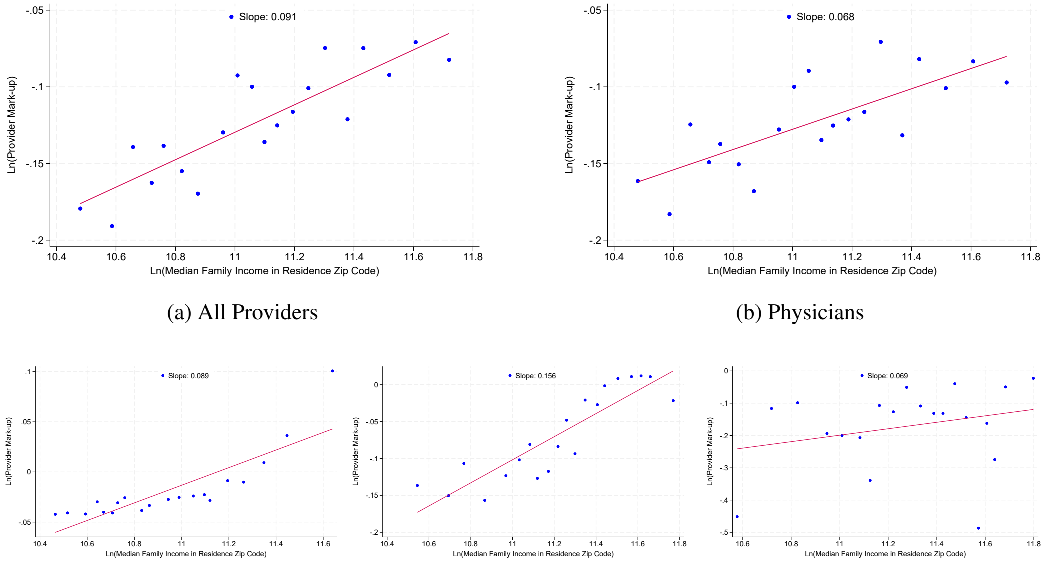
Figure C.2 shows Engel curves for medical provider "prices" using the APCD-CO data described above. The first panel shows results using all medical providers, while subsequent panels show curves for different types of providers: physicians, dentists, physical and occupational therapists, and psychologists. The elasticity among all providers combined is 0.091. The gradients are statistically significant and positive for each occupation individually, with the highest gradient among therapists at 0.156. The estimate of the price elasticity is likely to be biased down because we use median family income in a zip code whereas an unbiased estimate would require the average of log income—or, better yet, a link to exact family incomes. This introduces a downward bias when zip codes vary in their local log income inequality. The positive relationship is evident throughout the income distribution and not solely at the top. These results, based on a cross-section, support our assortative matching prediction.

C.3. *Some facts about physician scale*

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

To provide some context on whether the physician scale (λ in our model) may have changed differently across doctors, we access data from the National Ambulatory Medical Care Survey (NAMCS) for 1991–2011 (National Center for Health Statistics, 2011). NAMCS is a sample of office-based physicians. Physicians are surveyed for a week and report detailed information

FIGURE C.2.—Engel Curves for Provider Prices



(a) All Providers (b) Physicians

(c) Dentists (d) Therapists (e) Psychologists

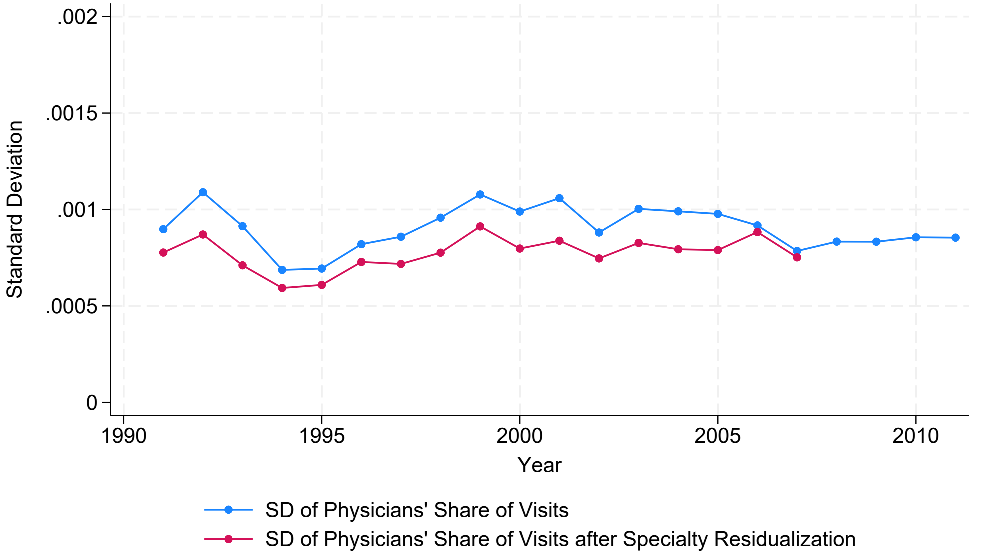
Notes: Each panel plots the relationship between log markups (normalized to mean zero within the relevant provider sample) charged by the medical provider who treats a patient and family income (as proxied by the median income in the patient’s residential zip code). The calculation of the markups is described in the text above. Each panel shows a binned scatter plot with 20 vigintiles; the slope of the regression line is the elasticity shown in that panel. Panel (a) shows the relationship for all medical providers; panel (b) for physicians; panel (c) for dentists; panel (d) for therapists; and panel (e) for psychologists.

on their patient visits. For example, in 2000 the NAMCS had 1,200 reporting physicians and recorded 27,369 visits to those physicians during the sampling week. We use this information to examine time trends in the variance of per-physician quantity of care.

In the first step of the survey, physician j estimates their visit volume V_j for the upcoming survey week. We do not observe V_j. Based on V_j, the surveyors then tell the physicians what fraction of their visits to record in the survey, F_j, which is a decreasing function of V_j. After the survey is conducted, the surveyor constructs patient visit sampling weights $W_{i(j)}$ (where a visit is indexed by i) that sum to the total estimated nationwide visits. Based on this, in a given year, for physician j we calculate $\hat{V}_j = \sum_i W_{i(j)}$ and treat this as the quantity of medical care provided. The share of all sampled visits performed by physician j is $s_j = \hat{V}_j / \left(\sum_j \sum_i W_{i(j)}\right)$.

Figure C.3 shows the raw standard deviation of s_j in each year from 1991 to 2011 and the specialty-adjusted standard deviation from 1991 to 2007. (The survey sampling and enumeration method changed in 2012, and specialty coding changed in 2008). This illustrates whether, on a nationwide basis, some physicians have captured larger shares of overall patient visits. The blue line shows the raw standard deviation and the red line the standard deviation after controlling for practitioner specialty. Both series were fairly stable over the periods for which they are available, suggesting that there were no large secular trends in the ability of some doctors to capture disproportionate shares of the market.

FIGURE C.3.—Standard Deviation in Physicians' Share of Total Patient Visits.



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

TABLE D.1

RATIO OF 98^{th} TO 90^{th} PERCENTILE OF WAGE INCOME FOR SELECTED OCCUPATIONS

Occupation	Ratio of 98th to 90th Percentile 1980	2012	Change
Chief executives, public administrators, and legislators	1.63	2.42	0.80
Financial managers	1.62	2.38	0.75
Managers, Excl. Real Estate	1.90	1.79	-0.11
Engineers	1.37	1.46	0.09
Physicians	1.50	1.72	0.22
Dentists	1.54	1.74	0.20
Registered nurses	1.29	1.48	0.20
Primary/Secondary School Teachers	1.26	1.33	0.07
Lawyers and judges	1.89	2.31	0.42
Real estate sales occupations	1.94	2.17	0.22
Financial service sales occupations	1.79	2.81	1.01
All Occupations Combined	1.70	1.99	0.29

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

APPENDIX D: EMPIRICAL APPENDIX

D.1. *Additional tables of descriptive statistics*

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

TABLE D.2

DESCRIPTIVE STATISTICS FOR TOP OCCUPATIONS IN 2000. WAGE INCOME

Occupation	Mean Income $1,000	Occupation's share in: Top 10%	Top 5%	Top 1%
Managers, Excl. Real Estate	61	0.225	0.234	0.183
Chief executives, public administrators, and legislators	120	0.062	0.094	0.152
Computer systems analysts and computer scientists	57	0.053	0.040	0.015
Engineers	63	0.050	0.035	0.011
Sales supervisors and proprietors	45	0.040	0.044	0.043
Sales occupations and sales representatives	53	0.039	0.042	0.032
Physicians	136	0.038	0.068	0.129
Lawyers and judges	98	0.037	0.051	0.065
Financial managers	68	0.025	0.028	0.032
Accountants and auditors	47	0.022	0.021	0.017
Other financial specialists	61	0.015	0.019	0.028
Subject instructors, college	43	0.015	0.012	0.004
Sales workers	29	0.014	0.015	0.012
Computer programmers	57	0.014	0.009	0.003
Primary/Secondary School Teachers	35	0.013	0.004	0.002
Registered nurses	40	0.012	0.007	0.004
Production supervisors or foremen	43	0.012	0.007	0.005
Financial service sales occupations	102	0.011	0.017	0.037
Real estate sales occupations	53	0.011	0.014	0.016
Office supervisors	37	0.010	0.007	0.005

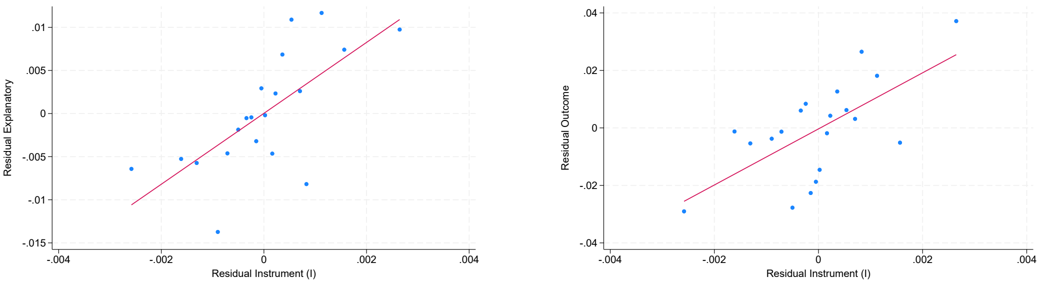
Notes: For the top twenty occupations in the top ten percent of the national income distribution in 2000, column (1) reports mean wage income (for the whole population), where the (very few) censored values have been replaced with the state-level mean income among those above the censoring point, and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

D.2. *Additional empirical results*

Figure D.1 represents the IV results of Table III in two binned scatter plots. Figure D.1, panel (a), shows the relationship between the shift-share instrument and non-physician inequality, *i.e.*, the first-stage regression. Both of these are residuals based on regressions with year and LMA dummies and the controls of column (6) in Table III. For disclosure reasons, we bin our LMA $\times$ year observations into 20 equal-frequency bins. We plot the average value of the residuals in each bin. Figure D.1, panel (b), shows the relationship between the instrument and physicians' inequality, *i.e.*, the reduced-form regression. In both cases, we see strong upward-sloping relationships. The results are not driven by outliers.

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

FIGURE D.1.—Graphical Representation of the First Stage and Reduced Form: Residual Top-Income Inequality



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

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

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

TABLE D.3

SPILLOVER t-STATISTICS AND OCCUPATIONAL CHARACTERISTICS

	(1)	(2)	(3)	(4)	(5)	(6)
At Least Partly Offshoreable	-0.79 (0.40)					
Offshore (Blinder)		-0.93 (0.51)				
Customer service - level			2.11 (0.73)			
Customer service - importance				1.31 (0.73)		
Working with public - level					1.10 (0.67)	
Working with public - importance						1.52 (0.65)
Constant	1.36 (0.34)	1.33 (0.33)	-0.10 (0.40)	0.31 (0.40)	0.40 (0.37)	0.13 (0.37)
Observations	30	30	30	30	30	30

Notes: This table shows the relationship between the t-statistics of the spillover coefficients from the IV regressions (from Table VI) and five characteristics of the 30 most common occupations in the top 10%. These characteristics are a measure of offshorability from Blinder (2009) and four measures from O*NET: the level and importance of "Customer and Personal Service" from Knowledge Requirements, and the level and importance of "Performing for or Working Directly with the Public" from Work Activities. O*NET and offshorability measures are rescaled as percentiles. "At Least Partly Offshoreable" indicates that Blinder (2009) categorized the occupation as having at least some offshoring potential—12 of the 30 occupations in our list.

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

Table D.5 shows robustness checks for Physicians and Dentists. Columns (1)-(3) consider variations of the baseline regressions on physicians: column (1) restricts to physicians who are older than 35, column (2) to those who have not moved within the past 5 years (other mobil-

TABLE D.4

EARNED INCOME

	Doctors		Dentists		Real Estate Agents	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
α_{-o}^{-1}	0.23	1.73	0.41	2.16	0.71	0.94
	(0.22)	(0.57)	(0.36)	(0.86)	(0.15)	(0.50)
		[0.77, 2.97]		[0.69, 4.25]		[0.01, 2.02]
Ln(Average Income)	-0.30	-0.43	0.06	-0.06	-0.03	-0.05
	(0.11)	(0.11)	(0.14)	(0.13)	(0.06)	(0.06)
Ln(Population)	-0.02	0.03	0.04	0.09	0.02	0.02
	(0.03)	(0.03)	(0.04)	(0.07)	(0.03)	(0.03)
N	200	200	200	200	200	200
F-Statistic		13.38		13.35		7.06

Notes: The table contains OLS and IV estimates exactly as in the baseline specification, except the outcome inequality measure is constructed using earned income (instead of wage and salary income). The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

ity questions are not available throughout our sample), and column (3) controls for specialty composition. We use data from the Area Resource File on the composition of specialties across LMAs (we use numbers from 1985 for year 1980) and data from Medical Group Management Association (MGMA) (2009) on average and standard deviation of income by specialty in 2008. We build four control variables: (1) the share of physicians in neurosurgery, which is the specialty with the highest mean income and the largest standard deviation; (2) the share of physicians in the 8 highest earning specialties (excluding neurosurgeons); (3) the share of physicians in the 7 specialties with the lowest income; and (4) the share of physicians in the 4 specialties with the largest standard deviation in income (excluding neurosurgeons). The rationale behind (2) and (3) is that these specialties share similar average incomes while the next specialty (down or up) in the ranking has a substantially different average income. Column (4) looks at the baseline regression for dentists and adds a control for the ratio of dental hygienists to dentists in the LMA-year to proxy for the increased specialization of dentists and the potential resulting increase in their operating scale.

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

While we follow Goldsmith-Pinkham, Sorkin, and Swift (2020) and assume that identification in our exercise comes from the exogeneity of the shares (the occupation weights), an alternative identification strategy in shift-share instruments relies on the exogeneity of the shifts (here nationwide occupational inequality). In that case, Adão, Kolesár, and Morales (2019) emphasize that correlated errors for (in our case) LMAs with similar occupational compositions can lead researchers to reject the null too often, and suggest a formula for standard errors that takes this issue into account. In Table D.12, we report their standard errors for our three focal occupations and our three main placebo occupations. For our focal occupations, standard errors only increase by a bit, and the spillover coefficients remain significant (the p-value for real estate agents with controls is 5.1%). Table D.12 also reports the confidence intervals from LMMP+ (2025) which are valid under weak instruments, shown in curly brackets.

TABLE D.5

ROBUSTNESS CHECKS FOR PHYSICIANS AND DENTISTS

	Physicians		Dentists	
	Age > 35	No Recent Move	Specialty Controls	Specialization Control
	(1)	(2)	(3)	(4)
α_{-o}^{-1}	3.07	3.32	2.00	2.13
	(1.37)	(1.38)	(0.72)	(0.87)
	[0.47, 6.81]	[0.76, 7.01]	[0.83, 3.70]	[0.66, 4.15]
Ln(Avg. Income)	-0.11	-0.05	-0.49	0.02
	(0.22)	(0.22)	(0.09)	(0.14)
Ln(Population)	0.03	0.02	-0.00	0.08
	(0.10)	(0.12)	(0.04)	(0.07)
Sh. neurosurgeons			-1.43	
			(5.13)	
Sh. high earning specialties			3.19	
			(1.20)	
Sh. low earning specialties			0.11	
			(0.48)	
Sh. unequal earning specialties			2.12	
			(3.82)	
# hygienist / # dentists				0.04
				(0.03)
N	200	200	200	200
F-Statistic	12.26	12.62	11.03	13.75

Notes: The table contains robustness checks on the baseline IV estimates. The first three columns are for physicians, and the last column is for dentists. Column (1) restricts the sample of physicians to those older than 35. Column (2) restricts the sample of physicians to those who have not recently moved (within the past 5 years). Column (3) uses the baseline sample of physicians but adds four controls for specialties of physicians (see text for details). Column (4) (dentists) controls for the ratio of the number of hygienists to the number of dentists in the area, as a measure of task specialization. The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

TABLE D.6

CONTROLLING AND TESTING FOR ENTRY EFFECTS

	Physicians		Dentists		Real Estate Agents	
Dependent Variable:	E_o/E	α_o^{-1}	E_o/E	α_o^{-1}	E_o/E	α_o^{-1}
	(1)	(2)	(3)	(4)	(5)	(6)
α_{-o}^{-1}	0.028	2.11	0.003	2.44	0.116	1.61
	(0.013)	(0.66)	(0.002)	(0.94)	(0.043)	(0.99)
	[0.006, 0.054]	[0.99, 3.58]	[0.000, 0.006]	[0.85, 4.72]	[0.042, 0.273]	[-0.17, 3.91]
Ln(Avg. Income)	-0.001	-0.47	-0.001	-0.03	-0.002	-0.01
	(0.002)	(0.12)	(0.000)	(0.15)	(0.008)	(0.09)
Ln(Pop.)	-0.001	0.04	0.000	0.09	-0.002	0.03
	(0.001)	(0.04)	(0.000)	(0.08)	(0.003)	(0.03)
E_o/E		6.52		-62.78		0.83
		(11.41)		(49.24)		(4.75)
N	200	200	200	200	200	200
F-Statistic	13.210	14.22	13.360	12.96	7.124	6.29

Notes: E_o is the number of people working in outcome occupation o (*e.g.* physicians) in a given year and LMA. E is the total employed population in the same year and LMA. When E_o/E is included as an independent variable, its purpose is to check that inequality spillovers are not driven by entry. When E_o/E is the outcome variable, we are testing whether the instrument-induced variation in local inequality causes entry. All results are IV estimates with the same specification described for our baseline results. The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

TABLE D.7

CONTROLLING FOR THE OCCUPATIONAL SHARE IN THE TOP 10%

	Main Occupations		
	Physicians	Dentists	Real Estate Agents
	(1)	(2)	(3)
α_{-o}^{-1}	2.27	2.27	1.70
	(0.63)	(0.93)	(0.67)
	[1.22, 3.69]	[0.71, 4.46]	[0.57, 3.38]
Outcome Occupation Region-Specific Percentile	0.43	0.67	-0.08
	(0.61)	(0.55)	(1.02)
Ln(Avg. Income)	-0.47	-0.03	-0.01
	(0.12)	(0.14)	(0.09)
Ln(Population)	0.03	0.07	0.02
	(0.04)	(0.08)	(0.03)
N	200	200	200
F-Statistic	12.78	13.33	6.15
	Placebo Occupations		
	Financial Managers	Managers Excl. Real Estate	Engineers
	(1)	(2)	(3)
α_{-o}^{-1}	0.62	-0.03	-0.10
	(0.78)	(0.15)	(0.30)
	[-1.14, 1.93]	[-0.32, 0.23]	[-0.64, 0.43]
Outcome Occupation Region-Specific Percentile	4.50	-0.11	-0.23
	(1.23)	(0.67)	(0.41)
Ln(Avg. Income)	0.17	0.06	-0.02
	(0.11)	(0.03)	(0.05)
Ln(Population)	-0.10	0.03	0.05
	(0.05)	(0.02)	(0.02)
N	200	200	200
F-Statistic	12.64	16.15	28.91

Notes: IV regressions for selected occupations with an additional control for the "Outcome Occupation Region-Specific Percentile," which is the fraction of workers in the outcome occupation whose income is in the top 10% of the overall LMA-year income distribution. This fraction will be greater than 10% when the outcome occupation is disproportionately high income (such as physicians). The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

TABLE D.8

USING 10% OF DOCTORS FOR THE DEPENDENT VARIABLE

	50 LMAs (Baseline)		30 LMAs	
	(1)	(2)	(3)	(4)
α_{-o}^{-1}	1.03	1.00	1.24	1.23
	(0.55)	(0.64)	(0.58)	(0.64)
	[0.10, 2.28]	[-0.08, 2.50]	[0.24, 2.57]	[0.17, 2.79]
Ln(Avg. Income)		0.01		-0.01
		(0.12)		(0.12)
Ln(Population)		-0.01		-0.01
		(0.04)		(0.05)
N	200	200	100	100
F Statistic	13.36	13.11	14.61	12.63

Notes: This table shows OLS and IV regression results for physicians using a different approach to calculating physician income inequality. Rather than taking all physicians with income above the LMA's general 90th income percentile, we take all physicians above the physician-specific LMA 90th percentile income. Due to physicians' high average earnings, this method results in calculating income inequality using far fewer doctors. N is the number of observations rounded to the nearest integer divisible by 50 as required by disclosure rules. The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

TABLE D.9

ROBUSTNESS TO CUTOFFS

	Panel (a): Physicians				
	Top 5 pct.	70 LMAs	30 LMAs	Using 13 Occs	Using 7 Occs
α_{-o}^{-1}	2.31	2.16	1.74	2.27	2.28
	(0.95)	(0.87)	(0.51)	(0.64)	(1.37)
	[0.92, 6.44]	[0.75, 4.51]	[0.90, 2.67]	[1.22, 3.88]	[., .]
Ln(Average Income)	-0.60	-0.49	-0.40	-0.47	-0.48
	(0.19)	(0.10)	(0.13)	(0.12)	(0.14)
Ln(Population)	0.07	0.04	0.02	0.03	0.04
	(0.07)	(0.04)	(0.03)	(0.04)	(0.06)
N	200	300	100	200	200
F Statistic	5.32	9.37	12.34	11.17	1.89
	Panel (b): Dentists				
	Top 5 pct.	70 LMAs	30 LMAs	Using 13 Occs	Using 7 Occs
α_{-o}^{-1}	1.33	2.15	2.32	2.27	2.94
	(0.83)	(0.98)	(1.14)	(0.86)	(1.42)
	[0.01, 3.51]	[0.54, 4.49]	[0.51, 5.37]	[0.81, 4.35]	[0.95, 10.42]
Ln(Average Income)	-0.18	-0.08	0.03	0.01	-0.07
	(0.20)	(0.13)	(0.15)	(0.13)	(0.15)
Ln(Population)	0.07	0.03	0.09	0.08	0.10
	(0.08)	(0.07)	(0.10)	(0.07)	(0.10)
N	200	300	100	200	200
F Statistic	8.71	10.46	9.16	13.14	4.46
	Panel (c): Real Estate Agents				
	Top 5 pct.	70 LMAs	30 LMAs	Using 13 Occs	Using 7 Occs
α_{-o}^{-1}	0.64	1.49	1.22	1.32	2.43
	(0.44)	(0.64)	(0.55)	(0.59)	(1.15)
	[-0.12, 1.40]	[0.18, 2.66]	[0.16, 2.43]	[0.18, 2.43]	[0.80, 7.91]
Ln(Average Income)	0.09	0.03	0.02	0.04	-0.08
	(0.12)	(0.08)	(0.08)	(0.08)	(0.14)
Ln(Population)	-0.00	0.02	0.02	0.02	0.03
	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)
N	200	300	100	200	200
F Statistic	18.37	7.96	5.63	8.22	4.58

Notes: This table shows the IV regressions for physicians (Panel (a)), dentists (Panel (b)), and real estate agents (Panel (c)) for 5 different specifications. Column (1) uses the top 5% (instead of 10%) of the income distribution (for the dependent variable, the independent variable, and the IV). Columns (2) and (3) use 70 and 30 LMAs respectively, instead of 50. Columns (4) and (5) construct the instrument differently: rather than finding the top 10 occupations in each LMA in 1980 then taking the union, column (4) uses the top 13 before taking the union while column (5) uses the top 7 before taking the union. N is the number of observations rounded to the nearest integer divisible by 50 as required by disclosure rules. The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

TABLE D.10

INSTRUMENT OCCUPATIONS WITH LARGEST ROTEMBERG WEIGHTS

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

Notes: The table reports the occupations with the sixteen largest Rotemberg weights in absolute value in the instrument for physician regressions.

TABLE D.11

PHYSICIAN REGRESSIONS: EXCLUDING OCCUPATIONS WITH THE HIGHEST ROTEMBERG WEIGHTS

	Outcome Occupation: Physicians				
Excl. from Instrument:	Financial Managers	Other financial Specialists	Engineers	Airplane pilots and navigators	Financial service sales occupations
α_{-o}^{-1}	2.57	2.39	2.36	2.63	2.54
	(0.80)	(0.73)	(0.60)	(0.84)	(0.92)
	[1.27, 4.69]	[1.17, 4.08]	[1.33, 3.76]	[1.23, 4.64]	[1.08, 5.00]
Ln(Average Income)	-0.50	-0.48	-0.47	-0.50	-0.49
	(0.14)	(0.13)	(0.13)	(0.14)	(0.13)
Ln(Population)	0.04	0.03	0.03	0.04	0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
N	200	200	200	200	200
F Statistic	9.75	12.66	15.45	12.47	8.56

Notes: This table shows the baseline IV regressions for physicians when we exclude in turn each of the five occupations with the highest Rotemberg weights from the instrument (Rotemberg weights are shown in Table D.10). The column heading denotes the occupation excluded from the instrument. The square brackets contain 95% confidence intervals that are robust to weak instruments following LMMP+ (2025).

TABLE D.12

ALTERNATIVE INFERENCE APPROACHES FOR THE IV ESTIMATOR

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

Notes: IV regressions for selected occupations, with additional inference approaches. The first column for each outcome occupation is the IV regression without population and average income controls. The second column adds these controls. The standard errors in parentheses are from standard cluster-robust inference. The standard errors in square brackets are Adão, Kolesár, and Morales (2019) standard errors. The curly brackets contain the 95% confidence bounds from LMMP+ (2025).

APPENDIX E: CALIBRATION APPENDIX

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

E.1. *Data*

Disclosure restrictions allow us to export 50 bins from each of four distributions in New York State: physicians and consumers (non-physicians) for 1980 and 2012. Each bin represents the average log income within 2 percentiles. The underlying sample is the same as that used for the regression. The consumer data have a rather long left tail, and we drop the bottom 10% (those below \$7,080 and \$8,791 in 1980 and 2012, respectively, with all values deflated to 2000 dollars using the CPI). We fit a kernel to the bottom 90% of each distribution, and impose a Pareto distribution on the top 10%, where α^{-1} is estimated using the top of the income distribution. We then choose the scale parameter of the Pareto distribution to ensure a continuous CDF.[15] Panels (b) and (d) in Figure E.1 show histograms drawn from the fitted kernel, with the lightly shaded part representing the 10% of observations not included. The figure also shows histograms drawn from the kernels of the doctors. These underlying data are bimodal, reflecting a sizable mass of doctors who are still in their medical residencies. Given that wages for medical residents are not driven by consumers' demand for physician skill (Chandra, Khullar, and Wilensky, 2014), we exclude them from the analysis by dropping observations with values lower than the 90th percentile of doctors aged 35 and younger (\$52,052 and \$56,387 in 1980 and 2012, respectively). As can be seen from the histograms this is practically equivalent to only including observations after the first "peak" of the distributions. For the analysis, we scale all incomes by the lowest consumer income in 1980.

E.2. *Calibration*

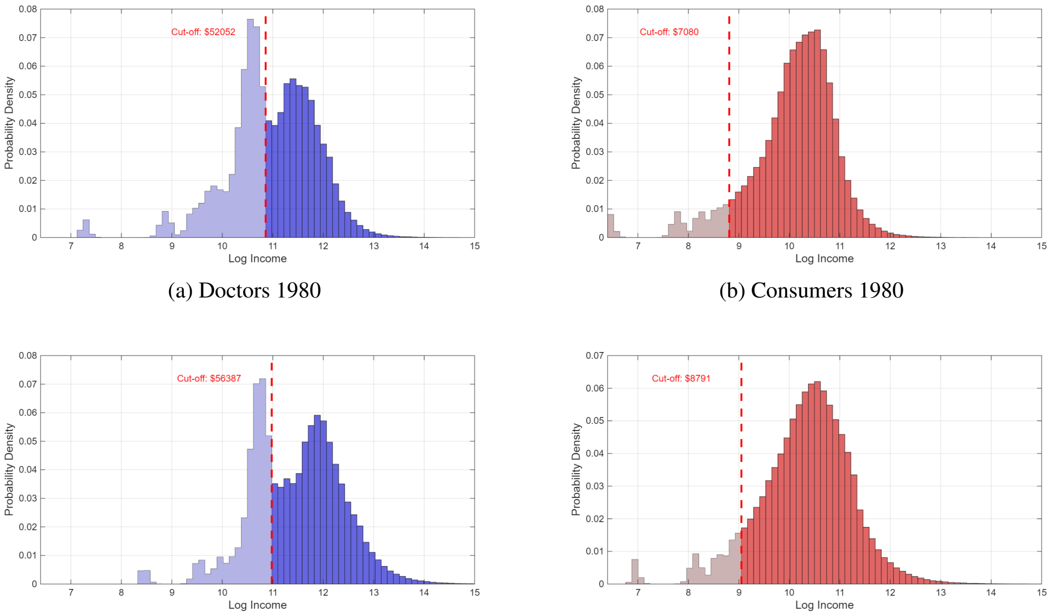
We calibrate our model for exogenously given λ and ε. The calculation of λ is based on the number of doctors as described in the text and we describe how we choose ε to match the IV coefficients in subsection E.2.1 below. Though the underlying ability distribution of potential doctors is given by $F_z(z)$, we can only calibrate the conditional distribution for active doctors, $\hat{F}_z(z)$. We normalize the lowest active doctor to have ability 1. We parameterize this distribution using 12 points along the distribution and interpolate between them.[16] With the tail value α_z and preference parameter β, this gives us 14 parameters to calibrate for 2012. We minimize the squared deviation of log wages between the empirical kernel and the predicted wage distribution from the model. This gives us the parameters in Table IX. The calibrated $\hat{F}_z(z)$ is shown in Appendix Figure E.2, panel (a). For this figure we also ask the following question: By how much should the ability distribution have changed between 1980 and 2012 to exactly fit the wage distribution in both 1980 and 2012, holding the other parameters, λ, β, and ε constant? The result is shown in the same panel. For the top 70%, the distribution shows a clear shift to the right, as well as a higher α_z^{-1} in 2012. This increase in ability inequality is commensurate with the change in α^{-1} for consumers. Panel (b) shows the same two distributions in a Pareto plot, where a Pareto distribution would imply a straight line.

E.2.1. *The choice of ε and the measure of α^{-1}*

Proposition 3 shows that the spillover coefficient equals $1/\varepsilon$, but this result only holds exactly in the tail of the distribution. Here we calculate the spillover coefficient for the model-predicted

[15] Our procedure does not require the CDF to be differentiable at the first point of the Pareto distribution, but it is nearly so because of the good fit of the Pareto distribution in the upper tail. We confirm that applying the censoring

FIGURE E.1.—Histograms Drawn from the CDFs of Consumers and Doctors



(a) Doctors 1980 (b) Consumers 1980

(c) Doctors 2012 (d) Consumers 2012

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

distributions when using a larger share of the population to calculate α^{-1}. We further show how we use these calculations to choose $1/\varepsilon$ to match the IV coefficient from Table III.

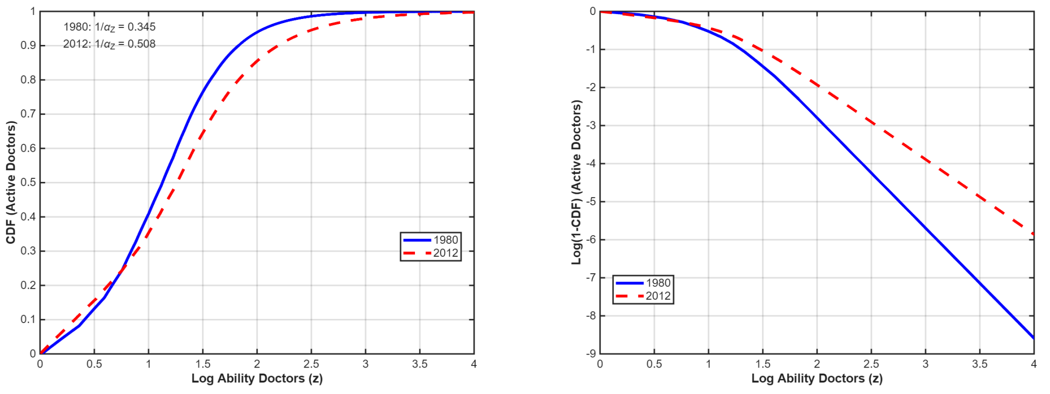
Throughout most of the paper, we calculate α^{-1} for an occupation of interest o using observations from that occupation in the top 10% of the general population. For the empirical distribution, Appendix Figure B.1 demonstrates that although the value of α^{-1} is sensitive to the choice of cutoff, the change in α^{-1} is much less so. Figure E.3 performs an analogous analysis on the model-predicted distributions, which by design fit the data in 2012, but not in 1980. For reference, slightly more than 80% of (non-resident) doctors are in the top 10% of the overall distribution in 2012, and the difference in α^{-1} at this cutoff is 0.193, while using the sample of the top 10% of doctors gives a similar value of 0.215. This implies that the model-predicted spillover estimates are also not very sensitive to the choice of cutoff.

The IV coefficient corresponds to the ratio of changes in α^{-1} for doctors and consumers. Employing the change for consumers of 0.14 (Table IX), we get a spillover coefficient from only changing the consumer distribution of 1.38 $(= 0.193/0.14)$ and 1.56, for the two cutoffs

procedure on data drawn from our CDF recreates binned observations that are indistinguishable from the actual data (not shown).

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

FIGURE E.2.—Calibrated doctor ability distribution

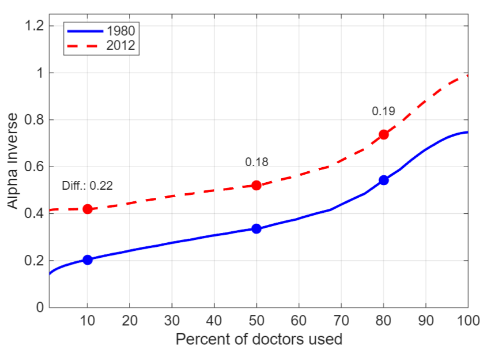


(a) CDF

(b) Pareto plot

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

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



Notes: The figure reports α^{-1} calculated using different fractions of the doctors' distribution for both 1980 and 2012, with the fraction increasing from left to right. The numbers report the difference between 1980 and 2012 for the selected fractions.

respectively. We chose ε to ensure that these are both in line with our target of 1.5, the IV spillover coefficient without controls in Table III.

E.2.2. *Welfare and spillovers.*

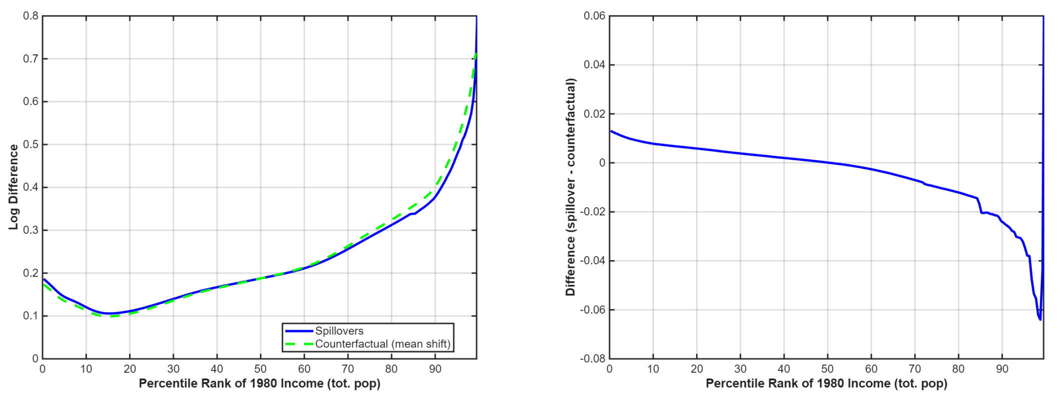
For the analyses behind Table X, we consider two different scenarios. For the spillover exercise, we consider the model counterfactual change to the doctors' income underlying Figure 5 – that is, keeping doctors' ability distribution fixed to its 2012 value. For the alternative counterfactual mean shift we calculate the mean of log income in both 1980 and 2012 in the model, and we shift the 2012 distribution down by the difference between the 2012 and 1980 means, preserving the shape. For the exercise in Table X we calculate the various inequality measures in the table for 2012 and the two 1980 distributions and consider differences.

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

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

For lower-income consumers, the mean-shift scenario implies higher increases in medical service costs. This implies that low-income consumers benefit from the spillover scenario. The reverse is true for higher-income consumers, who pay higher prices with spillovers. However, because spillover occupations are concentrated at the top of the income distribution, most consumers use services provided by those who earn more than themselves. Consequently, for the highest earners (top 1%), the spillover scenario features higher welfare gains than the mean-shift scenario. Appendix Figure E.4 Panel (b) shows the difference between the welfare gains in these two scenarios by income percentile.[17]

FIGURE E.4.—Differences in Welfare for the Full Income Distribution (with and without Spillovers)



(a) Welfare differences - spillovers and counterfactual (b) Difference between spillovers and counterfactual

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

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

APPENDIX: REFERENCES

Acemoglu, Daron and David H. Autor (2011), "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics* (David Card and Orley Ashenfelter, eds.), Vol. 4, Part B, chapter 12, 1043–1171, Elsevier, URL https://doi.org/10.1016/S0169-7218(11)02410-5. [26]

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales (2019), "Shift-share designs: Theory and inference." *The Quarterly Journal of Economics*, 134 (4), 1949–2010. [33, 37]

Armour, Philip, Richard Burkhauser, and Jeff Larrimore (2016), "Using the pareto distribution to improve estimates of topcoded earnings." *Economic Inquiry*, 54, 1263–1273. [24]

Becker, Gary and Gregg Lewis (1973), "On the interaction between the quantity and quality of children." *Journal of Political Economy*, 81 (2), 279–288. [17]

Blewett, Lynn A., Julia A. Rivera Drew, Risa Griffin, and Kari C. W. Williams (2019), "IPUMS Health Surveys: Medical Expenditure Panel Survey, Version 1.1 [dataset]." Minneapolis, MN: IPUMS, URL https://doi.org/10.18128/D071.V1.1. [27]

Blinder, Alan S. (2009), "How many US jobs might be offshorable?" *World Economics*, 10 (2), 41–78. [26, 32]

Center for Improving Value in Health Care (2015), "Colorado All Payer Claims Database (CO APCD) Research Data Set." Denver, CO, URL https://civhc.org/get-data/non-public-data/. Data licensed from CIVHC. Vintages: 2010–2014. [27]

Chandra, Amitabh, Dhruv Khullar, and Gail R. Wilensky (2014), "The economics of graduate medical education." *The New England Journal of Medicine*, 370 (25), 2357–2360, URL https://doi.org/10.1056/NEJMp1402468. [38]

Clemens, Jeffrey and Joshua D. Gottlieb (2017), "In the shadow of a giant: Medicare's influence on private physician payments." *Journal of Political Economy*, 125 (1), 1–39. [27]

Clemens, Jeffrey, Joshua D. Gottlieb, and Tímea Laura Molnár (2017), "Do health insurers innovate? evidence from the anatomy of physician payments." *Journal of Health Economics*, 55, 153–167. [27]

Deming, David J. (2017), "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics*, 132 (4), 1593–1640. [25, 26]

Dorn, David (2009), "Essays on inequality, spatial interaction, and the demand for skills, ph.d. dissertation, university of st. gallen." [26]

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020), "Bartik instruments: What, when, why, and how." *American Economic Review*, 110 (8), 2586–2624. [33]

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song (2010), "Earnings inequality and mobility in the united states: Evidence from social security data since 1937." *Quarterly Journal of Economics*, 125 (1), 91–128. [24]

[LMMP+] Lee, David S., Justin McCrary, Marcelo J. Moreira, Jack R. Porter, and Luther Yap (2025), "What to do when you can't use '1.96' confidence intervals for IV." Working Paper 31893, National Bureau of Economic Research. [33, 34, 35, 36, 37]

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles (2017), *IPUMS National Historical Geographic Information System: Version 12.0 [Database]*, Minneapolis: University of Minnesota. [27]

Medical Group Management Association (MGMA) (2009), "Physician compensation and production survey: 2009 report based on 2008 data." [33]

Mogstad, Magne and Matthew Wiswall (2016), "Testing the quantity-quality model of fertility: Estimation using unrestricted family size models." *Quantitative Economics*, 7 (1), 157–192. [17]

National Center for Health Statistics (2011), "National Ambulatory Medical Care Survey, 1991–2011." Public-use data file and documentation, Hyattsville, MD, URL https://www.cdc.gov/nchs/namcs/documentation/index.html. [28]

National Center for O*NET Development (2006), *O*NET 10.0 Database*, U.S. Department of Labor, Employment and Training Administration (USDOL/ETA). Retrieved from https://www.onetcenter.org/db_releases.html. [26]

Rosen, Sherwin (1974), "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy*, 82 (1), 34–55. [17]

[RFSB+] Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C. W. Williams (2025), "IPUMS USA: Version 16.0 [Extract: 2000 Census 5% Sample]." Minneapolis, MN: IPUMS, URL https://doi.org/10.18128/D010.V16.0. [26]

Tolbert, Charles M. and Molly Sizer (1996), "U.s. commuting zones and labor market areas. a 1990 update." *Economic Research Service Staff Paper No. 9614*. [26]

U.S. Census Bureau (2000), "Decennial Census (SAP 690) [Restricted-Use Data]." URL https://www.researchdatagov.org/product/690. Restricted-use data accessed via Federal Statistical Research Data Centers: 1960, 1970, 1980, 1990, and 2000. [23]

U.S. Census Bureau (2014), "American Community Survey (SAP 691) [Restricted-Use Data]." URL https://www.researchdatagov.org/product/691. Restricted-use data accessed via Federal Statistical Research Data Centers: Vintages: 2010–2014 (5-year ACS sample). [23]