What's Driving the Decline in Entrepreneurship? Appendix

For Online Publication

Contents

A Empircs (Section 2)

A.1 Discussion of sample and variable definitions

The sample period of 1987–2015 has been chosen to ensure that self-employment can be measured consistently over time. The CPS does have data prior to 1987 on self-employment, but for this period the BLS only reported people as self-employed if their business was not incorporated. From 1987 onward people with incorporated businesses have been counted as self-employed as well. The exclusion of people with incorporated businesses from self-employment prior to 1987 is likely to downwardly bias the trend in self-employment since people have been increasingly likely to incorporate their businesses over time. Since the share of people who are self-employed is a critical moment for the analysis, I exclude the pre-1987 data.[1] One additional point regarding the consistency of the data over time is that in 1994 the CPS questionnaire and data collection methods were updated (see Polivka and Miller, 1998). For the moments that I consider, this redesign had no systematic impact, so no adjustments are made for this.

The entrepreneur and self-employed shares are calculated with respect to the labor force throughout the analysis. The labor force is defined to include all people who worked for profit, pay, or as an unpaid family worker for at least 8 weeks during the relevant year. A week is counted even if a person only worked for a few hours, or was on paid time off (for vacation or illness). The weeks requirement is intended to omit people with very low labor force participation, while maintaining a broad sample. 2.3% of the sample are excluded from the labor force due to this criteria.

A.2 Source of owners per firm estimate

The estimates of the number of owners per business in various size categories is from the 1992 Characteristics of Business Owners Survey from the Census Bureau. This data provides the number of sole proprietorships, partnerships and S corporations, and the number of owners of these businesses, by firm size. I use 1992 data since this is the closest year to 1997 with this information (the survey was discontinued after 1992). C corporations are omitted from this dataset so I am assuming that they account for a negligible number of the businesses with less than 100 employees owned by self-employed people.

[1] In their analysis of entrepreneurs Levine and Rubinstein (2017) distinguish between people with incorporated and unincorporated businesses arguing that incorporation is a signal of entrepreneurial quality. In this paper I don't do analysis dividing the sample by the legal form of businesses since I am focusing on trends over time and the data shows that there is a trend towards incorporation over time so that this division is not stable.

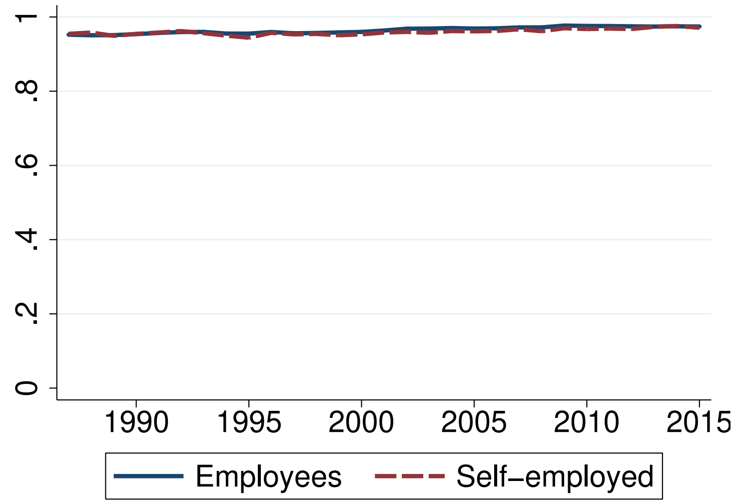


Figure A1: **Average share of income from longest job for employees and the self-employed.** Each line is the average value of income from the longest job as a share of total income from self-employment and dependent employment. For each person the share is winsorized to be in $[0, 1]$.

A.3 Income share of main job

The March supplement of the CPS asks respondents for information about their "longest" job in the previous calendar year. The data shows that on average, a person's longest job accounts for nearly all of their income from employment. To see this in detail, Figure A1 plots a person's income from their longest job as a share of their total income from self-employment and dependent employment.[2] This share is plotted separately for those whose longest jobs were self-employment and dependent employment. The figure shows that throughout the period of analysis the share of income from the main job was above 95% for both groups.

This feature of the data also supports the way that occupational choice is modeled. In the model agents who are working have to choose between being being an entrepreneur and an employee. They can't split their time between the two. While this is a simplification of reality, the data shows that it is a reasonable approximation.

A.4 Composition effects on the entrepreneur share

In this section I will show that the decline in entrepreneurship is not driven by changes in the composition of the population over time. To evaluate whether changes in composition are driving the result I compute the entrepreneur share holding the composition of the economy fixed along several dimensions. Specifically, the entrepreneur share in year t

[2]Due to negative income from self-employment, it is possible for this share to fall outside $[0, 1]$. For the purposes of these calculations I bound the shares by 0 and 1. This affects a small number of observations, for example, 0.56% for 1991 and 0.01% for 2015.

can be written as

$$e_t = \sum_{g \in \mathcal{G}} \omega_{g,t} e_{g,t}$$

where $\mathcal{G}$ is a partition of the labor force, $\omega_{g,t}$ is the share of the sample in subset $g \in \mathcal{G}$ and $e_{g,t}$ is the share of that subset who are entrepreneurs. Holding the composition fixed over time with respect to partition $\mathcal{G}$ the entrepreneur share in year t is

$$e_{\mathcal{G},t} \equiv \sum_{g \in \mathcal{G}} \omega_{g,1991} e_{g,t}. \tag{A1}$$

This equation keeps the share of each subset of the economy fixed while allowing the entrepreneur share within each subset to vary.

I perform this exercise to control for composition along six dimensions individually and also do the exercise controlling for several of these dimensions jointly. These dimensions are the sector, age, education, gender, geographic and metropolitan/non-metropolitan distributions. To control for the sector distribution $\mathcal{G}$ is composed of the 11 major non-agricultural non-government sectors from the 1990 Census Industrial Classification System;[3] for age $\mathcal{G}$ has four categories: 25–35, 36–45, 46–55 and 56–65; for education $\mathcal{G}$ is composed of five categories for the highest level of education a person has completed: less than high school, high school, some college education but less than a bachelor's degree, a bachelor's degree and more education than a bachelor's degree; for gender $\mathcal{G}$ is male and female; for geographic distribution $\mathcal{G}$ is the nine Census divisions; and to control for the metropolitan and non-metropolitan shares of the labor force $\mathcal{G}$ has these two categories.

The results for $e_{\mathcal{G},t}$ for each of these composition controls are presented in Figure A2. They show that the decrease in the entrepreneur share is either virtually unchanged or larger when each of these composition controls is imposed. This implies that changes in composition are not causing the decrease in the entrepreneur share and, in fact, the decrease in the entrepreneur share would be larger without changes in composition. Due to sample size limitations I can not control for all of the changes in composition jointly, but I have taken the three dimensions that matter most (age, sector and education) and controlled for these jointly. To ensure that cell sizes are large enough for this exercise I use three sectors (mining, manufacturing, construction, and utilities; wholesale and retail trade; and finance, insurance, real estate, and services), two education groups (less than college and at least college) and all four age categories. $\mathcal{G}$ is the set of all possible intersection of these sets.[4] The resulting $e_{\mathcal{G},t}$ series is presented in Figure A2 and labeled *Sect, age, ed*. The decrease in the entrepreneur share is larger again under these joint controls, em-

[3]These sectors are mining; construction; manufacturing; transportation, communication and public utilities; wholesale trade; retail; finance, insurance and real estate; business and repair services; personal services; entertainment and recreation services; and professional services.

[4]An example of an element of $\mathcal{G}$ for this case is all people in the sample aged 25–35 with less than a college education working in mining, manufacturing, construction or utilities.

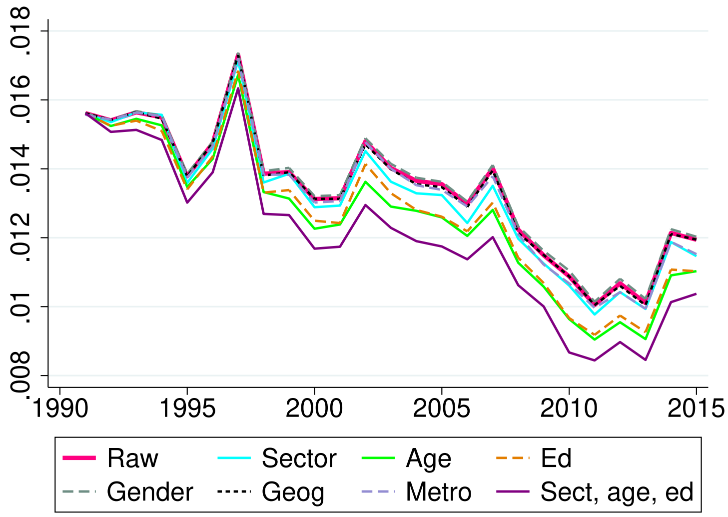


Figure A2: **Entrepreneur share with composition controls.** The *Raw* line is the entrepreneur share without any composition control. For the remaining lines the composition of the labor force along various dimensions is held fixed at its 1991 distribution, per equation (A1). The subsets of the labor force that are used for each of the lines are as follows. *Sector:* 11 major non-agricultural non-government sectors from the 1990 Census Industrial Classification System. *Age:* age groups 25–35, 36–45, 46–55 and 56–65. *Ed:* less than a high school education, completed high school, some college, completed college and more than college. *Gender:* male and female. *Geog:* nine Census divisions. *Metro:* metropolitan and non-metropolitan areas. *Sect, age, ed:* Cartesian product of three sectoral groups (see text for details), four age groups (25–35, 36–45, 46–55 and 56–65) and two education groups (less than college and at least college).

phasizing that composition changes not are causing this decline, they are working against it.

This exercise has been replicated for the self-employed share, instead of the entrepreneur share, with the results in Figure A3. The main message is the same.

A.5 Additional details on composition changes

In the previous section I showed that changes in the composition of the economy have generally worked against the decrease in the entrepreneur share. In this section I provide additional details for the composition changes that have had the largest effect on the entrepreneur share: changes in the sectoral, education and age compositions.

Figure A4(a) shows how the sectoral distribution has evolved over time. The main change is that the share of employed people who are in services has been steadily increasing while the share in manufacturing has been decreasing. This has worked against the decrease in the entrepreneur share since, as panel (b) shows, the share of people in the services sector who are entrepreneurs is larger than the share in manufacturing.[5]

Panels (c) and (d) illustrate the effects of changes in the education distribution. Over time the share of people with a college or more than a college education has increased,

[5]The entrepreneur shares are shown for 1991 as an illustration. The ranking of entrepreneur shares across sectors, and also across education and age, are stable over time.

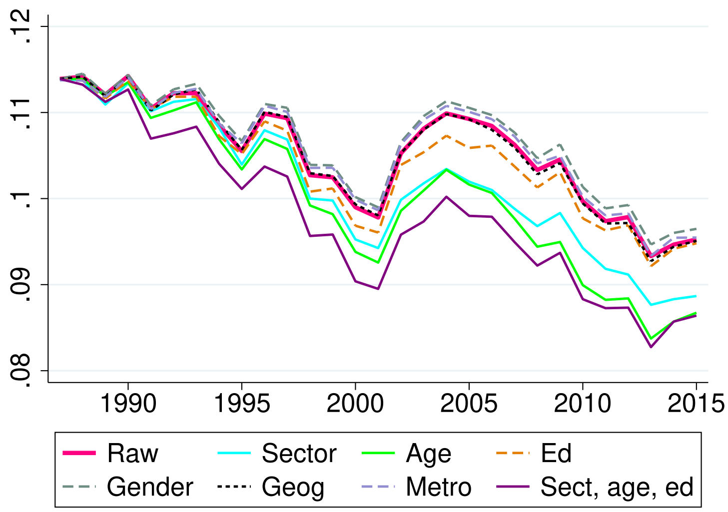


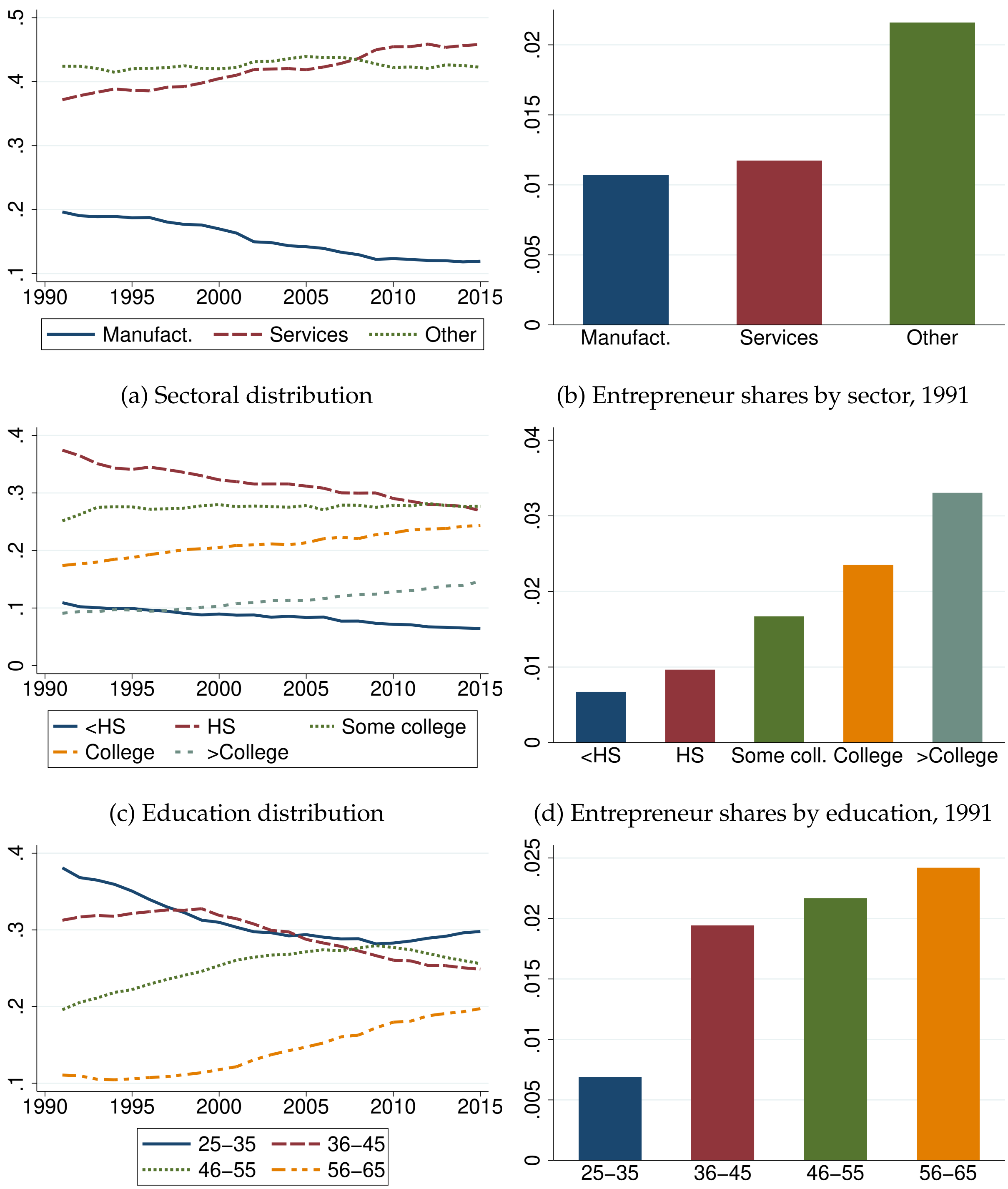
Figure A3: **Self-employed share with composition controls.** This figure has exactly the same setup as Figure A2. The difference is that the results are for the self-employment rate instead of the entrepreneurship rate. See notes of Figure A2 for details.

while the shares in all lower education categories have decreased. Since more educated people have higher entrepreneur shares—see panel (d)—this change has pushed the entrepreneur share up.

The effects of the changes in the age distribution are demonstrated by Figures A4(e) and (f). While the change in the share of the labor force in each age category has not been monotone in age, in general there has been an aging of the population. This has pushed the entrepreneur share upwards since the entrepreneur share is increasing in age. Note that the entrepreneur share is increasing in age rather than having the familiar hump shape because I use the labor force as the denominator. If we looked at the share of *people* in age groups who are entrepreneurs then we would see a hump shape in age.

A.6 Decline in entrepreneurship by sector

To establish that the decline in the entrepreneur share is not driven by one sector Table A1 presents details of the change in the entrepreneur share by sector and the contribution of each sector to the aggregate change. To increase cell sizes I group the mining, construction and transportation, communication and public utilities sectors together, and the business and repair services, personal services, and entertainment and recreation services sectors. I also add the wholesale trade sector to the retail sector, since the former is relatively small and the entrepreneur shares have very similar trends in the two sectors. To smooth out year-to-year volatility in the data I take averages of the entrepreneur share in the first four and last four years of the sample. The table shows that there was a large decline in the entrepreneur share in all sectors except other services, for which the decline was more modest. The last column of the table presents the share of the decrease in the aggregate entrepreneur share that each sector accounts for when the sectoral composition of the



(a) Sectoral distribution

(b) Entrepreneur shares by sector, 1991

(c) Education distribution

(d) Entrepreneur shares by education, 1991

(e) Age distribution

(f) Entrepreneur shares by age, 1991

Figure A4: **Details of sectoral, education and age composition changes** The sectoral distribution is the share of the labor force in manufacturing, services (including business and repair services, personal services, entertainment and recreation services, and professional and related services) and all other sectors. The education and age distributions are the share of the labor force in each education and age group, respectively. The entrepreneur shares are the share of the labor force who are entrepreneurs within each group.

Sector	1991 share	Entrepreneur share '91–'94	'12–'15	% change	% of total change
Mining, Construction and TCU	15.8	1.7	1.4	−17.2	10.2
Manufacturing	19.6	1.0	0.7	−28.8	12.6
Wholesale and retail trade	19.3	2.5	1.5	−39.0	40.9
FIRE	7.2	2.1	1.2	−43.0	14.5
Professional services	26.3	1.1	0.7	−33.5	20.5
Other services	10.9	1.5	1.5	−3.4	1.3

Table A1: **Entrepreneur share by sector.** The columns contain: (1) share of the labor force in each sector in 1991; (2)–(3) the average share of the labor force in each sector who are entrepreneurs for 1991–94 and 2012–15, respectively; (4) percentage change in these rates from 1991–94 to 2012–15; (5) each sector's share of the total change in the entrepreneur share when the sector distribution is held fixed at 1991.TCU stands for the transportation, communication, and public utilities sector.

economy is held fixed. For sector g this is

$$\frac{\omega_{g,1991}(\bar{e}_{g,2015} - \bar{e}_{g,1994})}{\bar{e}_{\mathcal{G},2015} - \bar{e}_{\mathcal{G},1994}}$$

where the partition $\mathcal{G}$ is the set of sectors being used and $\bar{x}_t \equiv (x_t + x_{t-1} + x_{t-2} + x_{t-3})/4$ for any variable x_t. The results show that all sectors contribute to the decline, with the largest contributions coming from wholesale and retail trade, and professional services, with other services only making a small contribution.

A.7 Broader definitions of an entrepreneur

The questions in the CPS dictate how an entrepreneur can be defined. There are a number of types of people who one might want to include, that are omitted by this definition. This section discusses a number of these and explains why the data suggests that these omissions are unlikely to reverse the trend decline in the entrepreneur share.

The first relevant class of people are those who own and manage a business, but are not classified as self-employed. This could be because they do not work the majority of their hours in the business or because the ownership or legal structure of the business is such that they consider themselves to be an employee rather than self-employed. If the share of people in this category has increased over time, then it could explain some of the decline in the entrepreneur share. One way to assess this is to use an alternative dataset on businesses that doesn't rely on employment status of the manager of the firm for its classification. One such dataset is the BDS from the Census Bureau. Using this, we can compute the ratio of employer firms in the economy, relative to the number of working people in the economy.[6] This ratio is presented in Figure A5 and shows a decline over time.

[6]Specifically I use the number of non-agricultural firms from the BDS and estimate the number of employees and self-employed in the non-farm private sector using the CPS.

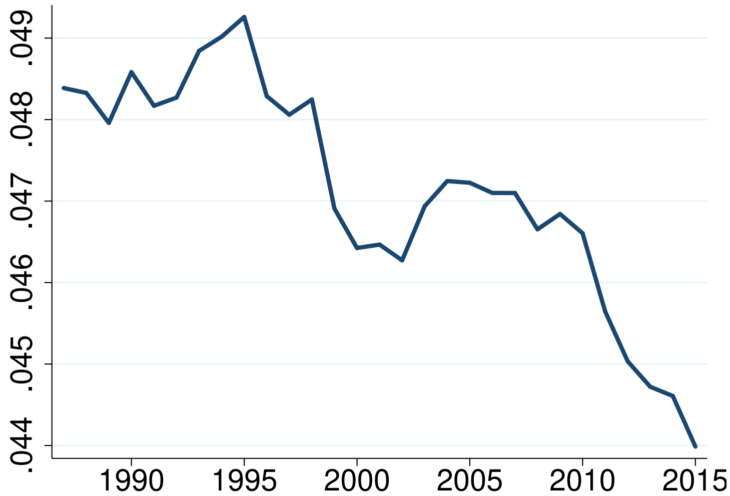


Figure A5: **Ratio of employer firms to the labor force.** The number of employer firms is from the BDS. The labor force is estimated, using the CPS, as the number of people in the civilian non-institutional population aged 16 and over who worked in the private non-farm sector in the relevant calendar year.

A second class of people missed by the definition is people who were self-employed in the previous year, but self-employment was not their main job. One way of looking at this is to use Figure A5 again, since the measure of entrepreneurship in that figure does not depend on whether someone manages a firm that they own as their primary job. Another approach is to look at whether there is evidence that people have earned an increasing share of their income from secondary occupations over time. This would be consistent with an increasing share of people running businesses as a supplementary source of income. With the CPS we can measure a person's income from their main job in a year, as a share of all of their income from working as an employee and from self-employment.[7] In Figure A6 I plot the mean, and several percentiles from the left tail, of the distribution of this share for people who work as employees in their main job. The data show that on average secondary income sources make up a very small share of income ($< 5\%$ in all years), and this share has actually decreased over time, rather than increasing. Looking at the mean only could hide a decrease in the share of income from the main job for people in the left tail of the distribution for this variable. However, the figure clearly shows that that left tail has increased in value as well.

A third note on measurement issues is that the definition of an entrepreneur is likely to omit some people whose businesses have merged with others, or been acquired. If merger and acquisition activity has increased over time then this could be contributing to the decline in entrepreneurship. The relevance of this can be assessed using the BDS data. This dataset provides information, *inter alia*, on the number of firms each year, the number of new firms, and the number of firm exits. A firm exit occurs when all

[7]For income from self-employment I include farm and non-farm income.

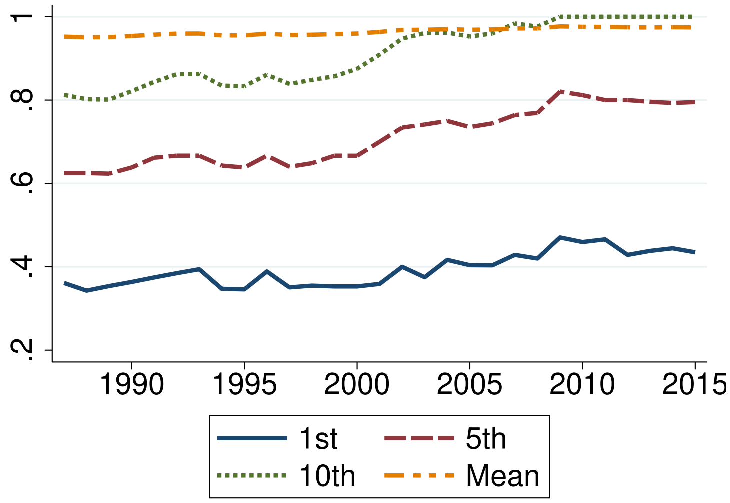


Figure A6: **Share of income from main job for people working as employees.** This figure presents the 1^{st}, 5^{th} and 10^{th} percentiles, and the mean, of the distribution of the share of income that employees earn from their main job.

establishments of a firm close down, so that mergers and acquisitions (M&A) that keep at least one establishment operating are not included. The number of firms in the dataset can also change due to firms splitting into multiple firms. The net M&A rate (the rate of M&A less splits) can be computed as:

$$\text{M\&A rate}(t) = \frac{\text{firms}(t) - \text{deaths}(t) + \text{entrants}(t+1) - \text{firms}(t+1)}{\text{firms}(t)}, \tag{A2}$$

where firms(t), entrants(t) and deaths(t) are the total number of firms, the number of entrants and the number of firms that die in year t, respectively. This measure of M&A is plotted in Figure A7 and shows that there is not an upward trend over time.

A.8 Evidence of declining entrepreneurship from the Survey of Income and Program Participation

To provide additional evidence of the decline in the entrepreneur share, including showing that it holds for a different period that excludes the Great Recession, I have computed the change in the entrepreneur share from 1983 to 1995 using the Survey of Income and Program Participation (SIPP) from the Census Bureau.

The SIPP is a nationally representative survey of US households that started in late 1983 and has been conducted regularly since. Using weights that are provided a nationally representative sample of individuals can be constructed. For my analysis I use the interviews conducted in October 1983–January 1984 and October 1995–January 1996. I will refer to these as the 1983 and 1995 data. There is SIPP data after 1996, however the survey changed and it is not possible to construct a consistent measure of entrepreneurship across this change. Note that 1983 is the year after a recession trough while 1995

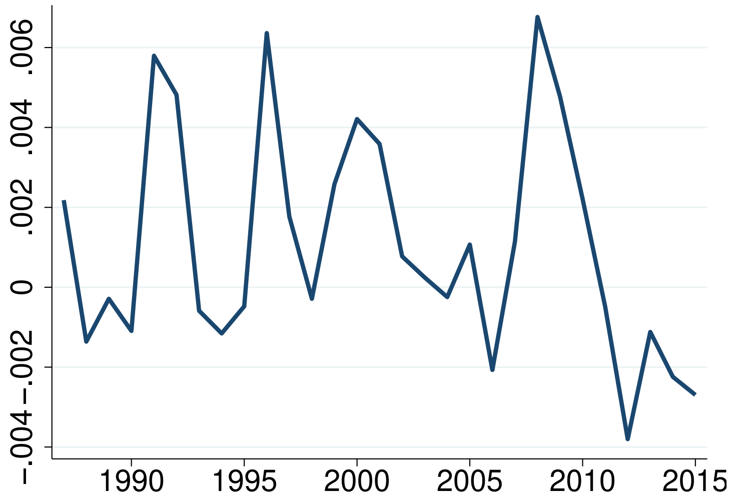


Figure A7: **Net merger and acquisition rate.** This is the implied net merger and acquisition rate from the BDS, computed using equation (A2). It is a net rate because it measures M&A less firm splits. A negative value implies that there were more firms splitting than merging or being acquired.

is four years after a recession trough, so the cyclicality of the entrepreneur share should work against any decline over this period.

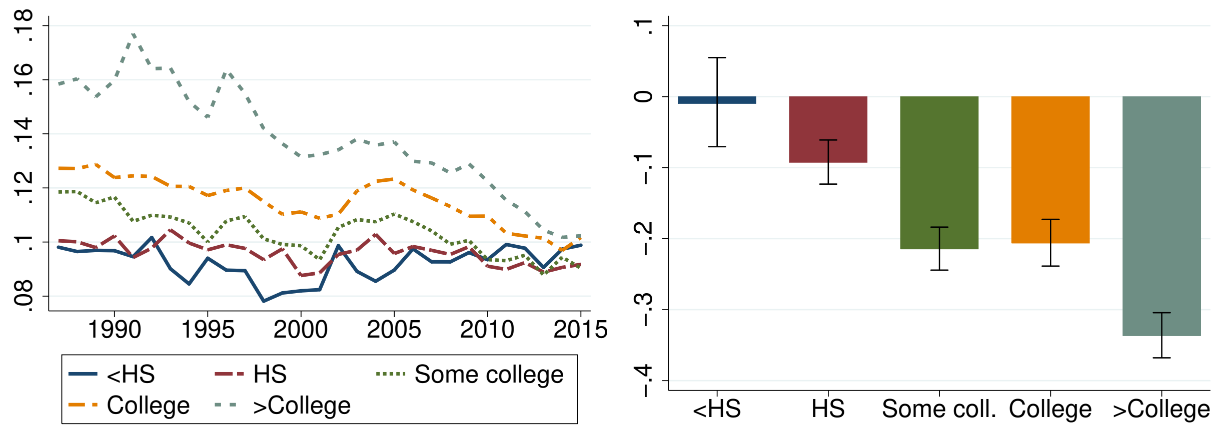
For the analysis of the entrepreneur share I have used two samples. Men and women aged at least 18, and men aged 24–65 who are not in education. I define an entrepreneur as a person who works at least 15 hours per week in self-employment, expects their business to generate at least $1,000 in revenue in the next 12 months and has at least one employee other than the owner and co-owners in the same household. For the first sample I find that the entrepreneur share (share of the labor force who are entrepreneurs) decreases from 5.38% in 1983 to 4.62% in 1995, a decrease of 14%. For the second sample I find a decrease from 9.40% to 7.67%, a decrease of 18.4%.

A.9 Robustness exercises for the change in entrepreneurship by education

This section contains two robustness exercises for the result that the decline in entrepreneurship has been larger for higher education groups.

Figure 2 in the main text shows how the entrepreneur share has changed over time for each education group. This analysis is reproduced for the self-employed share in Figure A8 to show that the results hold for this broader measure of entrepreneurship as well. This analysis goes back to 1987, rather than starting in 1991, since it does not require firm size information.

A second potential concern is that the larger decline in entrepreneurship for more educated people could be driven by changes in the structure of specific industries. Specifically, there are a number of professional services industries such as legal services, account-



(a) Self-employmed share, 1987–2015 (b) Self-employment, 1987/90–2012/15

Figure A8: **Self-employed share by education and percentage change.** Panel (a) is the share of the labor force for each education level who are self-employed. Panel (b) is the the relative change in the self-employed share from 1987–90 (pooled data) to 2012–15 for each education group (i.e. -0.1 is a decline of 10%). The whiskers are 95% confidence intervals estimated by Poisson regression. The education categories are people who did not finish high school (<HS), finished high school (HS), have some college education less than a bachelor's degree (some college), completed a bachelor's degree (college), and have more education than a bachelor's degree (>College).

ing, financial consulting and medical services, that seem to have shifted over time from small practices to larger companies containing many professionals. To assess whether this change is driving the result, I redo the analysis for the change in the entrepreneur share by education group, excluding the professional services and FIRE sectors. This changes the sample significantly. For example, the total sample for the analysis shrinks by 38%, and for people with more than a college education the decline is 75%.[8] Despite this, the point estimates in Figure A9 are very similar to the main results. Confidence intervals are, of course, wider due to the smaller sample size.

B Model and proofs

B.1 Optimal input choices and profit function for entrepreneurs

The Γ functions for the optimal input choices and the profit function for entrepreneurs are:

$$\begin{aligned}
\Gamma_{k_o} &= \left[\left(\frac{\eta}{r_o}\right)^{1-\alpha} D_3^{\alpha}\right]^{\frac{1}{1-\eta-\alpha}} \left(\phi + (1-\phi) D_1^{\frac{\gamma}{1-\gamma}} D_2^{\frac{\gamma(1-\tau)}{\tau(1-\gamma)}}\right)^{\frac{\alpha(1-\alpha)}{\gamma(1-\eta-\alpha)}}, \\
\Gamma_{\ell_h} &= D_3^{\frac{1}{1-\alpha}} \Gamma_{k_o}^{\frac{\eta}{1-\alpha}}, \\
\Gamma_{\ell_l} &= \left(D_1 D_2^{\frac{\gamma-\tau}{\tau}}\right)^{\frac{1}{1-\gamma}} \Gamma_{\ell_h},
\end{aligned}$$

[8]The full sample decrease from 521 to 324 thousand, and for people with more than a college education it is from 62 to 16 thousand.

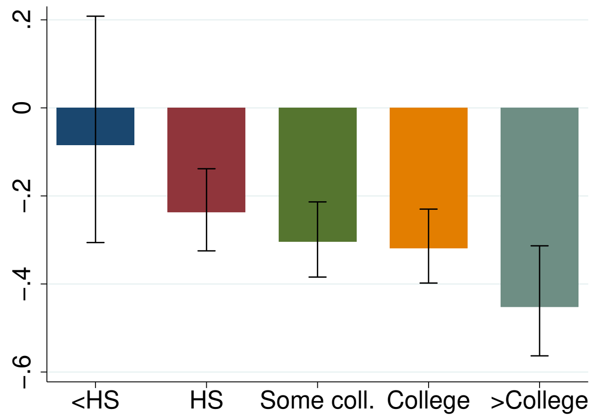


Figure A9: **1991–2015 percentage change in entrepreneur share by education, omitting professional services and FIRE.** This figure presents the relative change in the entrepreneur share from 1991–94 (pooled date) to 2012–15 for each education group with the professional services and FIRE sectors omitted (i.e. -0.1 is a decline of 10%). The whiskers are 95% confidence intervals estimated by Poisson regression. The education categories are the same as in Figure A8.

$$
\begin{aligned}
\Gamma_{k_i} &= \left[\left(\frac{\lambda}{1-\lambda}\right)\left(\frac{w_l}{r_i}\right)\right]^{\frac{1}{1-\tau}}\Gamma_{\ell_l}, \\
\Gamma_{\pi} &= \Gamma_{k_o}^{\eta}\left[\phi\Gamma_{\ell_h}^{\gamma} + (1-\phi)\Big(\lambda(\Gamma_{k_i})^{\tau} + (1-\lambda)\Gamma_{\ell_l}^{\tau}\Big)^{\frac{\gamma}{\tau}}\right]^{\frac{\alpha}{\gamma}} \\
&\quad - \Gamma_{k_o}r_o - \Gamma_{k_i}r_i - \Gamma_{\ell_h}w_h - \Gamma_{\ell_l}w_l,
\end{aligned}
$$

where

$$
\begin{aligned}
D_1 &= \left(\frac{1-\phi}{\phi}\right)\left(\frac{w_h}{w_l}\right)(1-\lambda), \\
D_2 &= \lambda\left[\left(\frac{\lambda}{1-\lambda}\right)\left(\frac{w_l}{r_i}\right)\right]^{\frac{\tau}{1-\tau}} + 1 - \lambda, \\
D_3 &= \frac{\alpha\phi}{w_h}\left(\phi + (1-\phi)D_1^{\frac{\gamma}{1-\gamma}}D_2^{\frac{\gamma(1-\tau)}{\tau(1-\gamma)}}\right)^{\frac{\alpha-\gamma}{\gamma}}.
\end{aligned}
$$

B.2 Stationary distribution

The state space is the Cartesian product $\mathbb{R}_+^3 \times \{0,1\}$, denoted by $\mathbb{Z}$. Let the σ-algebra $\Sigma_{\mathbb{Z}}$ be defined as $B_{\mathbb{R}_+^3} \otimes P(\{0,1\})$, where $B_{\mathbb{R}_+^3}$ is the Borel σ-algebra of $\mathbb{R}_+^3$ and $P(\{0,1\})$ is the power set of $\{0,1\}$. Let the typical subset of $\Sigma_{\mathbb{Z}}$ be denoted by $\mathcal{Z} \times \mathcal{E}$. With this notation, the transition function for the distribution of agents, $q : \mathbb{Z} \times \Sigma_{\mathbb{Z}} \to [0,1]$, can be expressed as:

$$
q((\mathbf{z},\epsilon), \mathcal{Z}\times\mathcal{E}) = (1-\delta)\big[\big(1-\mathbb{1}_{o}(e)\big)\mathbb{1}_{\mathcal{E}}(0) + \mathbb{1}_{o}(e)\mathbb{1}_{\mathcal{E}}(1)\big]\int_{\mathcal{Z}} g(\mathbf{z}'|\mathbf{z})d\mathbf{z}' + \delta\mathbb{1}_{\mathcal{E}}(0)\int_{\mathcal{Z}} g(\mathbf{z}')d\mathbf{z}',
$$

where $g(\mathbf{z}'|\mathbf{z})$ and $g(\mathbf{z})$ are the probability density functions of $G(\mathbf{z}'|\mathbf{z})$ and $G(\mathbf{z})$ respectively. The indicator function for the set $\mathcal{E}$, $\mathbb{1}_{\mathcal{E}}(x)$, indicates whether element x is in set

$\mathcal{E}$. To understand this formula, recall that with probability $1-\delta$ an agent survives to the next period. If they are not an entrepreneur this period ($o \neq e$) then $\epsilon' = 0$, and if they are then $\epsilon' = 1$. Their productivity vector evolves according to $G(\mathbf{z}'|\mathbf{z})$. With probability δ an agent will die. In this case they will be replaced by a new agent next period who will have $\epsilon = 0$ and will draw her productivities from $G(\mathbf{z})$. A stationary distribution of agents is a function $Q : \Sigma_{\mathbb{Z}} \to [0,1]$, such that for all $\mathcal{Z} \times \mathcal{E} \in \Sigma_{\mathbb{Z}}$

$$Q(\mathcal{Z} \times \mathcal{E}) = \int_{\mathbb{Z}} q((\mathbf{z}, \epsilon), \mathcal{Z} \times \mathcal{E}) dQ(\mathbf{z}, \epsilon). \tag{B3}$$

B.3 Proofs of propositions

Proposition 1

(a) First consider the derivative with respect to r_i, holding wages fixed. $\underline{z}_e^s(z_s, \epsilon)$ is a piecewise function with two parts. When wages are held fixed, r_i only enters both parts through Γ_π. $\underline{z}_e^s(z_s, \epsilon)$ is strictly increasing in r_i if Γ_π is strictly decreasing in it. $\partial \Gamma_\pi / \partial r_i < 0$ can be proved by contradiction. Take any entrepreneur with productivity $z_e > 0$ and any rental rate of IT capital, $r_{i,1} > 0$. Let the entrepreneur's profit maximizing input choice be $(k_{o,1}^*, k_{i,1}^*, \ell_{l,1}^*, \ell_{h,1}^*)$ and its profit (before fixed and entry costs) be π_1^*. Now consider any $r_{i,2} > r_{i,1}$. Denote the optimal input choices and the resulting profit in the same way as for $r_{i,1}$, but with subscript 2 this time. Suppose that $\pi_2^* > \pi_1^*$. Then if, for $r_i = r_{i,1}$, the firm chose inputs $(k_{o,2}^*, k_{i,2}^*, \ell_{l,2}^*, \ell_{h,2}^*)$ instead of $(k_{o,1}^*, k_{i,1}^*, \ell_{l,1}^*, \ell_{h,1}^*)$ it would achieve a profit strictly great than π_2^*, and therefore strictly greater than π_1^*. This contradicts $(k_{o,1}^*, k_{i,1}^*, \ell_{l,1}^*, \ell_{h,1}^*)$ being the optimal input choice for $r_i = r_{i,1}$.
The proof for the derivative with respect to w_h (holding w_l fixed) follows the same logic. By the same argument just outlined, $\partial \Gamma_\pi / \partial w_h < 0$, which causes $\underline{z}_e^s(z_s, \epsilon)$ to increase in w_h. For $s = h$, w_h also enters in the numerator of the expression for $\underline{z}_e^s(z_s, \epsilon)$ for $z_h > \underline{z}_h$. This also causes $\underline{z}_e^h(z_h, \epsilon)$ to increase in w_h.

(b) This part of the proposition restricts attention to $z_s > \underline{z}_h$, so the relevant expression for the entrepreneur thresholds is:

$$\underline{z}_e^s(z_s, \epsilon) = \left(\frac{z_s w_s + \psi + \mathbb{1}_\epsilon(0)\psi_e}{\Gamma_\pi} \right)^{1-\alpha-\eta}.$$

From part (a) it is established that $\partial \underline{z}_e^s(z_s, \epsilon)/\partial r_i|_{\mathbf{w}} > 0$. This derivative is larger for $s = h$ because the wage w_s is in the numerator and $w_h > w_l$. The derivative of $\underline{z}_e^s(z_s, \epsilon)$ with respect to w_h is also positive from part (a). It is larger for $s = h$ because (i) $w_h > w_l$, which increases the value of the numerator, and (ii) w_s increases in the numerator for the case of $s = h$, while it does not for $s = l$.

(c) This part of the proposition follows from equation (11) and the proof of part (a) of this proposition. From equation (11), $\underline{z}_e^s(z_s, 0) - \underline{z}_e^s(z_s, 1) > 0$ since $\psi_e > 0$. The proof of part (a) established that $\partial \Gamma_\pi / \partial r_i|_{\mathbf{w}} < 0$. The inequality in this part of the proposition follows from this.

Proposition 2

(a) Since $\partial w_s / \partial z_f > 0$ and $\partial \Gamma_\pi / \partial w_s < 0$ for all $s \in \{l, h\}$,[9] it follows that $\partial \underline{z}_e^s(z_s, \epsilon)/\partial z_f > 0$ for $z_s \in (0, \underline{z}_s]$. For $z_s > \underline{z}_s$, the decrease in Γ_π also causes $\underline{z}_e^s(z_s, \epsilon)$ to increase. The increases in w_l and w_h cause this function to increase further through the w_s term in the numerator.

(b) The first inequality comes from taking the derivative of the function specified in equation (10) with respect to ψ.
For the second inequality, it follows from equation (11) that, for $z_s > \underline{z}_s$,

$$\left.\frac{\partial[\underline{z}_e^s(z_s,0) - \underline{z}_e^s(z_s,1)]}{\partial \psi}\right|_{\mathbf{w}} = (1-\alpha-\eta)\left(\frac{1}{\Gamma_\pi}\right)^{1-\alpha-\eta}\left(\frac{1}{(z_s w_s + \psi + \psi_e)^{\alpha+\eta}} - \frac{1}{(z_s w_s + \psi)^{\alpha+\eta}}\right).$$

This derivative is strictly negative since $z_s w_s + \psi + \psi_e > z_s w_s + \psi > 0$.
For $z_s \in (0, \underline{z}_s]$, the analysis is identical, with $z_s w_s$ replaced by b in the previous equation.

(c) The first inequality comes from taking the derivative of the function specified in equation (10), with ϵ set equal to zero, with respect to ψ_e.
For the second inequality, since $\partial w_l / \partial \psi_e < 0$ and $\partial w_h / \partial \psi_e < 0$, $\partial \Gamma_\pi / \partial \psi_e > 0$. Using equation (10), it follows that $\partial \underline{z}_e^s(z_s, 1)/\partial \psi_e < 0$.

Proposition 3 Using equation (6) for $\ell_s(z)$, inequality (13) can be expressed as

$$\frac{\int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_o(e|\mathcal{P}_\psi) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z}, \epsilon|\mathcal{P}_\psi)}{z_f(\mathcal{P}_\psi)^{\frac{1}{1-\alpha-\eta}}} > \frac{\int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_o(e|\mathcal{P}_{z_f}) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z}, \epsilon|\mathcal{P}_{z_f})}{z_f(\mathcal{P}_{z_f})^{\frac{1}{1-\alpha-\eta}}}.$$

Since $z_f(\mathcal{P}_{z_f}) > z_f(\mathcal{P}_\psi)$, a sufficient condition for this is that

$$\int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_o(e|\mathcal{P}_\psi) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z}, \epsilon|\mathcal{P}_\psi) > \int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_o(e|\mathcal{P}_{z_f}) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z}, \epsilon|\mathcal{P}_{z_f}). \quad \text{(B4)}$$

To condense notation, let

$$\tilde{Z}_e^{s'}(\mathcal{P}) \equiv \int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_o(e|\mathcal{P}) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z}, \epsilon|\mathcal{P}),$$

so that (B4) can be expressed as $\tilde{Z}_e^{s'}(\mathcal{P}_\psi) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) > 0$.

Since $\partial w_s / \partial z_f > 0$ and $\partial w_s / \partial \psi > 0$ for all $s \in \{l, h\}$, $\partial \Gamma_\pi / \partial z_f < 0$ and $\partial \Gamma_\pi / \partial \psi > 0$. It

[9] $\partial \Gamma_\pi / \partial w_s < 0$ can be proved in the same way that $\partial \Gamma_\pi / \partial r_i < 0$ is proved for Proposition 1(a).

follows that, for $z_{s'} > \underline{z}_{s'}(\mathcal{P}_{z_f})$,[10]

$$\frac{\partial \underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{z_f})}{\partial z_{s'}} > \frac{\partial \underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi})}{\partial z_{s'}}.$$

Since the share of agents of skill type s' who are entrepreneurs must be equal under $\mathcal{P}_{z_f}$ and $\mathcal{P}_{\psi}$ (equation 12), the previous inequality implies that there are thresholds $z^*_{s'}(\epsilon)$, for $\epsilon \in \{0,1\}$, such that

$$\begin{aligned}
\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi}) &> \underline{z}_e^{s'}(z_s,\epsilon|\mathcal{P}_{z_f}) \text{ for } z_{s'} < z^*_{s'}(\epsilon),\\
\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi}) &= \underline{z}_e^{s'}(z_s,\epsilon|\mathcal{P}_{z_f}) \text{ for } z_{s'} = z^*_{s'}(\epsilon),\\
\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi}) &< \underline{z}_e^{s'}(z_s,\epsilon|\mathcal{P}_{z_f}) \text{ for } z_{s'} > z^*_{s'}(\epsilon).
\end{aligned}$$

Using this mapping of the entrepreneur thresholds for the two sets of parameter values, $\tilde{Z}_e^{s'}(\mathcal{P}_{\psi}) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f})$ can be expressed as

$$\begin{aligned}
\tilde{Z}_e^{s'}(\mathcal{P}_{\psi}) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) = \sum_{\epsilon\in\{0,1\}} \int_{z^*_{s'}(\epsilon)}^{\infty} \int_{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi})}^{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{z_f})} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} Q(\mathbf{z},\epsilon) dz_e dz_{s'}\\
- \int_0^{z^*_{s'}(\epsilon)} \int_{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{z_f})}^{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi})} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} Q(\mathbf{z},\epsilon) dz_e dz_{s'},
\end{aligned} \quad \text{(B5)}$$

where $Q(\mathbf{z})$ is the marginal distribution of $\mathbf{z}$. Taking the limit as $\psi_e \to 0$,

$$\begin{aligned}
\lim_{\psi_e\to 0} \tilde{Z}_e^{s'}(\mathcal{P}_{\psi}) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) = \int_{z^*_{s'}(0)}^{\infty} \int_{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{\psi})}^{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{z_f})} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} g(\mathbf{z}) dz_e dz_{s'}\\
- \int_0^{z^*_{s'}(0)} \int_{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{z_f})}^{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{\psi})} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} g(\mathbf{z}) dz_e dz_{s'}.
\end{aligned} \quad \text{(B6)}$$

Observe that every value of z_e in the range $(\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{\psi}), \underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{z_f}))$ for $z_{s'} > z^*_{s'}(0)$ is greater than every value of z_e in the range $(\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{z_f}), \underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{\psi}))$ for $z_{s'} < z^*_{s'}(0)$. Since equation (12) ensures that the weights placed on these two sets of values of z_e are equal, therefore $\lim_{\psi_e\to 0} \tilde{Z}_e^{s'}(\mathcal{P}_{\psi}) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) > 0$.

C Model-data mapping and calibration moments

C.1 Additional details for mapping model to data

Data The main dataset that is used for the calibration is the CPS March supplement, which was introduced in Section 2. The sample is the same as the main sample for the analysis in that section: people aged 25–65 not working in the agriculture or government

[10]The one exception to this is at $z_{s'} = \underline{z}_{s'}(\mathcal{P}_{\psi})$ because $\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{\psi})$ is not differentiable at this point; but that is not material for the proof.

sectors. The main moments that are used are from the occupation distribution and the income distribution. Full details of how these moments are computed are below. Wherever other datasets are used, this is specified.

Skills The occupation classification scheme from Acemoglu and Autor (2011) divides occupations into four categories according what types of tasks each occupation is most intensive in: non-routine cognitive, routine cognitive, routine manual or non-routine manual tasks.[11] For a detailed discussion of these categories see Autor et al. (2003) and Acemoglu and Autor (2011). Briefly, routine tasks are repetitive tasks that could be summarized by a set of instructions that a machine could follow. They are cognitive if they require mostly mental effort (e.g. book-keeping) while they are manual if they require mostly physical effort (e.g. production line assembly). Non-routine tasks are difficult to get a machine to do with a set of instructions. Cognitive non-routine tasks include research, marketing activities and managerial tasks. Manual non-routine tasks include many low-skill service jobs. In terms of relative wages, non-routine manual occupations earn the lowest wages, followed by routine occupations and then non-routine cognitive occupations. I therefore use non-routine cognitive occupations as high-skill occupations and the rest as low-skill occupations.

There is a line of research on routine-biased technical change that distinguishes between non-routine manual occupations and routine occupations (e.g. Acemoglu and Autor, 2011; Autor and Dorn, 2013; Goos et al., 2014; Jaimovich and Siu, 2020; vom Lehn, 2015; Cortes et al., 2017; Lee and Shin, 2016). The rationale for this is that employment and wages in non-routine manual jobs has increased relative to that of routine manual jobs in recent decades, although much less than the relative wages of non-routine cognitive occupations have increased. The present paper abstracts from the difference between non-routine manual and routine occupations by grouping them together since the key force under my theory is the increase in demand for high-skill employees as technology changes, rather than the differential effects among low-skill workers who are all worse off relative to the high-skilled. Adding an additional employee type would clutter the analysis without adding much.

C.2 Entrepreneur share

In the model an entrepreneur is a person who owns and manages a business with employees. In the data I define these people to be the self-employed with employees. This creates a challenge for the data. The size information provided in the CPS does not separate self-employed people with businesses with no employees from those that have a small number of employees. For 1991–2015 the smallest size category is $<$10 employees and for 1988–91 it is $<$25 employees.

[11] Under this classification managerial, professional and technical occupations are non-routine cognitive; sales, clerical and administrative support occupations are routine cognitive; production, craft, repair and operative occupations are routine manual; and service occupations are non-routine manual.

Figure C10: **Numbers of self-employed people with <10 employees and firms with 1–9 employees (millions).** The *self-employed* series is the number of people aged 16+ in the US who are self-employed and whose businesses have <10 employees. The *firms* series is the number of firms in the US with 1–9 employees. Agriculture and public-administration sectors are excluded

To estimate the share of the self-employed in the <10 category who have employees I take the following approach. For 1991–2014 there are two steps. First, data from the BDS provides information on the number of firms in various size categories on an annual basis up to 2014, including establishments with 1–9 employees.[12] Since these are small firms I assume that they each are owned and run by one person, so that they are each associated with one self-employed person.[13] This gives me an estimate of the number of self-employed people with businesses with 1–9 employees each year. I exclude the agriculture sector from the data, just as I did in the empirical analysis in Section 2.

Second, using the CPS data I estimate the number of people in the population who are self-employed with non-agricultural businesses in a range of size categories.[14] The population for this analysis is the civilian non-institutional population aged 16 years and over, rather than the restricted population that I used for the empirical analysis, since the self-employment estimates need to be for the whole population to be comparable to the BDS data. The estimate for the number of people in the US who are self-employed

[12]This is an annual dataset going back to 1977 that provides information on the *population* of private sector firms in the US which have at least one employee. The information includes the number of firms in a range of size bins, with size measured with the number of employees. When I compute the number of firms with 1–9 employees I omit those in the agriculture sector since I don't count self-employed people in agriculture when I measure entrepreneurship in the CPS data.

[13]Some supporting evidence for this that for firms in the next size category up, 10–99 employees, the average ratio of the number of self-employed people, estimated from the CPS, to the number of firms in the BDS is 0.96.

[14]The CPS data provides estimates of the share of the population who are self-employed with businesses in a number of size categories and I multiply these by the size of the population that the weighted CPS sample represents to estimate the number of self-employed people with businesses in each size category in the US. The size of the population that the CPS sample represents come from the BLS.

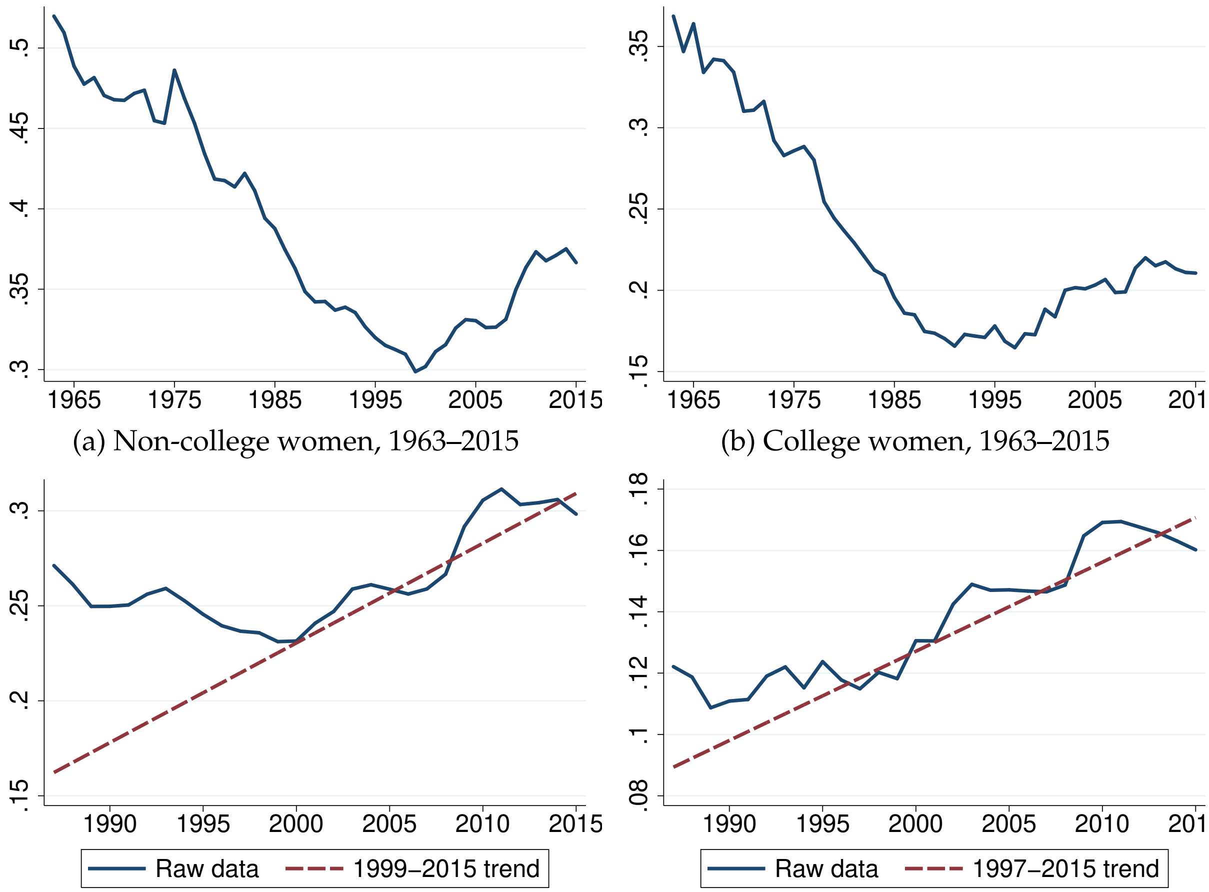
with less than 10 employees and the number of firms with 1–9 employees are presented in Figure C10. Both series grow steadily over time and the ratio of the number of firms to self-employed people is fairly stable, starting at 0.42 and ending at 0.40. I use the estimate of the number of self-employed people with 1–9 employees from the BDS data to divide the number of self-employed people with <10 employees in the CPS data into those with 0 employees and those with 1–9 employees. This provides the information necessary to compute the share of self-employed people with <10 employees who have at least one employee. Finally I assume that this share also holds for the restricted sample that I am studying (ages 25–65) and for both of the education levels I use.[15] This allows me to then compute the number of entrepreneurs in the data for each education level, and thereby the entrepreneur shares.

For 1987–90 the size categories for small firms in the BDS and CPS don't match up. Since the size distribution of self-employed businesses is quite stable over time (see Figure 1(b)) I estimate the share of people who are self-employed with at least one employee for each education level by taking the share who are self-employed each year and multiplying it by the average share of the self-employed who have employees for 1991–1993 for the relevant education level. For 2015 BDS data on the number of firms with 1–9 employees is not yet available. I assume that the share of the self-employed with less than 10 employees who have at least one employee equals to the average of this moment for 2012–14.

C.3 Out of labor force share

A second challenge with matching up the occupation distributions in the model and data arises because of changes in female labor force participation over time. As is well known, there was a strong and steady increase in the female labor force participation rate throughout at least the second half of the last century and this rate leveled off in the mid to late 1990s. Since my analysis starts in 1987 and I do not model gender this creates a disjunction between the model and the data. I deal with this by making adjustments to the data so that the two are comparable. The approach is as follows. I start with the out of labor force shares for women in my sample with non-college and college educations. For each education level there is a strong downward trend from when the CPS starts in the early 1960s until the late 1990s when both out of labor force shares start to rise. For non-college women the turning point is 1999, while for college educated women it is 1997 (see panels a and b of Figure C11). I assume that after these turning points the force generating the long run increase in female labor participation has ended. I therefore interpret the data after the turning points as representing the effect of other forces operating in the economy. To estimate what the data would have looked like prior to the turning points without the trend increase in female labor force participation I take the series for men and women

[15] Ideally I would compute this share for each education group separately, but the data does not provide the information necessary to do this.



(a) Non-college women, 1963–2015

(b) College women, 1963–2015

(c) Non-college men and women, 1987–2015

(d) College men and women, 1987–2015

Figure C11: **Out of labor force share by education level** Panels (a) and (b) present the out of labor force share for women with the two education levels for 1963–2015. Panels (c) and (d) present the out of labor force shares for men and women for the two education levels for 1987–2015. These panels also show linear trends for 1999–2015 and 1997–2015, respectively, extrapolated back to 1987.

combined for each education level, estimate the trend in the out of labor force share from the turning point (1999 for non-college and 1997 for college) to 2015, and then extrapolate the trend back to 1987. For both education groups the out of labor force share is approximately linear after the turning points, so I use a linear trend. See panels (c) and (d) of Figure C11.

C.4 Occupation distribution

To complete the occupational distribution for each education level I also need estimates of the shares of low and high-skill employees. The low and high-skill employee shares can be measured directly from the CPS data. Since I don't have unemployed people in the model I treat them as employees and use the occupation of their last job to determine their skill type.[16] This gives me estimates of the occupation distribution for each edu-

[16]There is a small number of unemployed people who don't have an occupation reported in the CPS. To deal with this I scale up the shares of low and high-skill employees in the data so that their relative sizes are

	Non-college		College	
	1987	2015	1987	2015
Out of labor force	16.8	29.8	9.2	16.0
Low-skill	65.6	53.2	23.5	21.2
High-skill	13.1	13.8	60.0	58.4
Entrepreneur	4.5	3.2	7.3	4.4

Table C2: **Occupation distributions from data.** These are the occupation distributions for college and non-college agents for 1987 and 2015 after I adjust the out of labor force shares to remove the effect of increasing female labor force participation prior to 1999 and remove self-employed people without employees from the data.

cation level, consisting of the shares of people who are out of the labor force, low-skill employees, high-skill employees and entrepreneurs.[17] To compute the aggregate occupation distribution I sum the two distributions conditional on education, weighting them by the shares of people with and without a college education. The final empirical occupation distributions that are used in the paper are presented in Table C2.

C.5 1987 income moments

The calibration moments require computing the mean and coefficient of variation of income for low-skill people, high-skill people and entrepreneurs, within each education group. These moments are computed using the March CPS, which provides data on income earned in the previous calendar year. To ensure a clean sample that is analogous to the model, I restrict the sample to people who worked full time in the previous year (at least 50 weeks and an average of at least 40 hours per week), earned nearly all of their income (at least 99%) from their main job, and did not make a loss on a business. Since the model does not allow for variation in hours worked, I use average hourly income rather than total income. To compute each person's average hourly income I take their income earned from their main job and divide it by the number of weeks he or she worked multiplied by his or her usual hours worked per week. Once the average hourly income is constructed for each person, it is straightforward to compute means and coefficients of variation for each relevant subsample. For the rest of this section "income" should be taken to refer to average hourly income.

There are three additional issues with the income data that are addressed. First is top coding. While there is income top coding in the CPS data, replacement values are available to maintain the top of the income distribution while protecting the anonymity

constant and these two shares sum to the share of people who are employed and unemployed in the data.

[17] Putting together the shares of people in each education group who are out of labor force, low-skill employees, high-skill employees and entrepreneurs does not produce a distribution that sums to one since I have estimated the out of labor force share and dropped self-employed people without employees from the data. To correct this I scale up the low-skill employee, high-skill employee and entrepreneur shares so that their relative sizes are constant and the total share of people who are working equals one minus the out of labor force share.

of respondents. The replacement values for the 1988 March CPS have been taken from the CPS IPUMS website[18]. Second, there is evidence that self-employed people underreport their income in the Panel Study of Income Dynamics, another income survey in the US. Hurst et al. (2014) estimate an underreporting rate of 25%. To adjust for this I scale up the income of entrepreneurs by a factor of $1/0.75$.

The third issue arises because the CPS does not provide information on the exact number of employees of each self-employed person. Thus it is necessary to estimate moments for the group of people who are defined as entrepreneurs in the model—self-employed people with at least one employee. I use the CPS data for 1991 for this purpose since, as described in Section 2 of the paper, it has more detailed information on the size of small firms than the data for 1987. I combine the 1991 estimates with information for 1987 to get estimates for that year, as I describe in detail below. For the coefficient of variation I use the data for 1991 to compute this moment of entrepreneur income for the two education groups for all self-employed people, and self-employed people with at least 10 employees. These moments are very similar, so the exact employment threshold doesn't appear to affect this moment very much. Therefore to estimate the 1987 coefficient of variation for entrepreneur income, I just use the value of this moment for all self-employed.

For average entrepreneur income the general approach is to use the data to estimate upper and lower bounds for this moment for each education group, and use this range to guide the choice of value. The details of the procedure are, using the data for 1991 unless stated otherwise:

- Compute average income of the self-employed, for each of the two education groups, conditional on three employment levels: any number of employees, < 10 employees and ≥ 10 employees.
- Take the estimate of the share of people who are self-employed with 0–10 employees who have at least one employee from the work done to estimate the share of people who are entrepreneurs (see discussion above). This value is 42.03% for 1991.
- Construct a lower bound for the average income of self-employed people with at least one employee, conditional on education, using a weighted average of the average income of the self-employed with < 10 employees and the average income of the self-employed with at least 10 employees:

$$\frac{(0.4203 \times shr^{\xi}_{<10})inc^{\xi}_{<10} + shr^{\xi}_{\geq 10}inc^{\xi}_{\geq 10}}{(0.4203 \times shr^{\xi}_{<10}) + shr^{\xi}_{\geq 10}}$$

  where shr^{ξ}_{x} is the share of the self-employed with education level $\xi \in \{N, C\}$ in size category x and inc^{ξ}_{x} is the average income of self-employed in this education-size category.
- Construct an upper bound for the average income of self-employed people with at

[18] https://cps.ipums.org/cps/income_cell_means.html, accessed 4 May 2020

least one employee, conditional on education, in a similar way:

$$\frac{(0.4203 \times shr^{\xi}_{<10})inc^{\xi}_{10\text{–}24} + shr^{\xi}_{\geq 10}inc^{\xi}_{\geq 10}}{(0.4203 \times shr^{\xi}_{<10}) + shr^{\xi}_{\geq 10}}.$$

The difference for the upper bound is that the average income of the self-employed with 10–24 employees is being used to put an upper bound on the income of the self-employed with 1–10 employees.

- The last step is to use these lower and upper bounds for 1991 to estimate such bounds for 1987. To do this I compute the ratio of 1987 to 1991 average self-employed income, and scale the lower and upper bounds by this factor, all conditional on education.

The resulting estimated ranges for mean (hourly) income of the self-employed in 1987 are: $\$14.19$–$20.47$ for non-college educated people and $\$27.62$–$32.43$ for the college educated. For calibration purposes I use the midpoints of these ranges.

The moments of entrepreneur income abstract from the asset value of entrepreneurs' businesses due to data limitations. To the extent that these businesses are a savings vehicle for entrepreneurs, this should not significantly affect results since I am also abstracting from savings for other agents.[19] The more important omission is the sale value of intangible capital accumulated by businesses. For recent work on measuring this, see Bhandari and McGrattan (2021).

C.6 1987–2015 income growth

Two of the key moments for the calibration are the growth of average real income for low and high-skill agents from 1987 to 2015. A limitation of using the CPS data on its own for these estimates is that it does not include non-wage compensation, the growth of which has differed across skill levels over time. To adjust for this, data from the BLS' Employer Costs of Employee Compensation (ECEC) survey is used. This dataset provides information going back to 1986 on compensation costs for employers by employee occupation and breaks the cost of compensation down into different components.[20] Particularly relevant for the purposes of this paper is that it separates wage and salary costs (which I'll call wages for brevity) from other forms of compensation. The data is annual up to 2001 and uses payroll data that includes March 12$^{\text{th}}$ each year. From 2002 onward the data is quarterly and I use the observation for the first quarter of each year to match up with the timing of the annual data.

The approach to adjusting the growth in the average income for each skill level from 1987 to 2015 from the CPS data to account for growth in non-wage compensation follows three steps.

[19] This assumes that the return on savings invested in private businesses and elsewhere are similar.

[20] The data used in this paper come from ECEC Table 9 for 1987–2003 and Table 15 for 2004–15.

1. Using the CPS compute the average hourly income for low and high-skill workers for 1987 and 2015. The sample for this is the main sample for the calibration, described above. Put the 2015 values in 1987 dollars using the Personal Consumption Expenditures Index from the BEA. The ratio of 2015 to 1987 average hourly wages for low and high-skill workers are 1.1303 and 1.3362, respectively.
2. For each skill level use the ECEC data to compute the ratio of 2015 to 1987 average hourly wages and average hourly total compensation, for the two skill levels. These ratios are presented in Table C3.
3. Use the ratio of compensation growth to wage growth to scale up the wage growth numbers from the CPS to account for non-wage compensation. For example, the estimated ratio of 2015 to 1987 average hourly total compensation for low-skilled employees is $1.1303\times(2.080/2.017) = 1.166$. This procedure assumes that the growth of compensation relative to wages is the same for my CPS sample as for the ECEC sample.

The one detail that has been omitted so far is how to compute the growth in average wages and total compensation for each skill level in step two. The ECEC data is by occupation so start by allocating each occupation to a skill level using the division described in Section C.1.[21] There is a change in the occupation classification system that the data uses from 2003 to 2004 so there is discontinuity in the data between these years. Next compute the average wage and average total compensation for each skill for 1987, 2003, 2004 and 2015. This requires aggregating the data across occupations. To do this weight each occupation by the share of the CPS sample in that occupation in the relevant year. In doing this use the same occupation classification system from the CPS as the ECEC data uses. Note that some service occupations are not covered by the ECEC so I place zero weight on these occupations and scale up the other weights proportionally so that the total weights equal one.[22] Compute the ratios of the 2003 to 1987, and the 2015 to 2004, values of the average wage for each skill level, and do the same for average total compensation. Finally multiply each 2003 to 1987 ratio by the corresponding 2015 to 2004 ratio to get estimates of the 2015 to 1987 ratios.

C.7 Entrepreneur employment share

The share of employment in the entrepreneur sector is estimated using data from the BDS and CPS. For 1987 the idea is to create a mapping from self-employed people in the CPS

[21] For one occupation group (Construction, extraction, farming, fishing and forestry) the data is missing for 2004 to 2006. I impute values for average compensation and average wages for this occupation by assuming that their growth rates from 2004 to 2007 were equal to their average growth rates from 2007 to 2015. The occupational crosswalk used for the mapping between CPS and ECEC occupations is available on request.

[22] One mismatch between the CPS sample and the ECEC data arises because the ECEC data for 2004–15 groups construction and extraction occupations with farming, fishing and forestry, which I exclude from the CPS sample. To deal with this I assume that the relative growth rates of compensation and wages are the same for these two types of occupations.

	Low-skill	*High-skill*
Wage growth	2.017	2.405
Compensation growth	2.080	2.597

Table C3: **Gross wage and compensation growth by skill, 1987–2015.** This table presents the gross growth rate of average wage and salary income and average total compensation for low-skill employees and high-skill employees for 1987 to 2015. 2.00 means that the relevant variable grew by 100%. The data is from the Employer Cost of Employee Compensation dataset from the BLS.

to establishments in the BDS, since the BDS provides richer information on size. Since the BDS covers the universe of private sector employer firms in the US, I use the full CPS sample for these calculations so that the coverage of the two datasets matches up, rather than restricting the sample based on age. From the CPS the public and agricultural sectors are omitted, as is the case for all of the analysis, and the agriculture sector is omitted from the BDS as well. The BDS does not include the public sector. For the mapping between the CPS and BDS I assume that each self-employed person in the CPS accounts for one establishment in the BDS at a firm in the same size class as the self-employed person's firm. Some support for this assumption is that for 1992 the number of owners per firm at firms with 10–99 employees was similar to the number of establishments per firm, at 1.35 and 1.23 respectively.[23] From a theoretical perspective the idea is that there is one person responsible for each establishment, who is also an owner. This would be the case, for example, under a partnership or franchise structure where each member of the partnership or franchise operates a location for the business. To give a sense of the implication of this for large firms, it implies that in 1992 self-employed people operated 17.2% of establishments of firms with at least 1000 employees.

This mapping provides an estimate of the share of establishments in each firm size class in the BDS that are run by self-employed people. To translate the establishment share into an estimate of the employment share of the self-employed I assume that within firm size classes in the BDS, each establishment is equal to the average size.[24] Since the size classes of firms used in the CPS change over time, they do not line up exactly with the BDS size classes in every year. However, they do line up for 1991, which is close to the start of the period of analysis. For this year the estimated share of employment at firms of the self-employed is 49.5%. Based on this, in the calibration for 1987 I use a share of employment at entrepreneur firms of 50%.

To provide some context for this estimate, using the Longitudinal Business Database from the Census Bureau and Computstat, Davis et al. (2006) estimate that privately held firms accounted for 75% of private sector employment in 1990. This value should be higher than the estimate just outlined since not every privately held firm will have a self-employed person operating it. For example, there may be large privately owned

[23] See Section 2 in the main text for a discussion of the value for owners per firm.

[24] For example, if there were 100 establishments at firms with 10–24 employees in the BDS and the total employment of firms in this size class was 1500, then the average establishment size would be 15.

firms who are managed on a day-to-day basis by employed managers and executives, and therefore won't have a self-employed person under the CPS definition. Given this, an estimate of 50% seems reasonable.

For 2015 the estimate is based on the fact that the size distribution of firms of the self-employed was stable over the period of analysis (see Figure 1(b)). This implies that the percentage change in the share of employment at entrepreneur firms (firms of self-employed people with employees) equaled the percentage change in ratio of the share of people who are entrepreneurs to the share of people who are employees. After making adjustments for female labor force participation (discussed above), I estimate that this share declined by 21.1%. This implies a entrepreneur share of employment of 39.5% in 2015. As further validation of the methodology that I adopted for computing this employment share for 1987, I have repeated the calculations for 2015 and get a share of 39.0%. The fact that the two approaches to estimating the employment share of entrepreneurs in 2015 provide very similar answers supports the use of these estimates.

C.8 Entry rate

Since the March CPS provides annual cross-sectional samples that change each year, it is not suitable for measuring the entry rate of people into entrepreneurship. To estimate this moment I therefore make use of the BDS. Despite the BDS including non-entrepreneur firms, this doesn't create an issue for computing the entry rate. The reason for this is that we know from the data presented in Section 2 that the vast majority of firms with less than 100 employees are run by a single self-employed person, and that there is about one self-employed person for each of these firms. These firms also account for virtually all new firms in the BDS each year, and virtually all firms of all ages. For example, in 1987 firms with less than 100 employees account for 99.8% of new firms and 98.1% of all firms. Therefore the entry rate in the BDS is very similar to the rate of firm creation by entrepreneurs. The one issue that this doesn't address is that there could be entrepreneurs who close one firm and start another within a year. To the extent that this occurs, the BDS entry rate will overestimate the entry rate of people into entrepreneurship. While this could affect the level of the entry rate, the more important assumption for the purposes of the analysis in this paper is that the difference between these rates does not change over time, so that the trend in the firm entry rate is a good measure of the trend in the entrepreneurship entry rate.

The BDS data is collected for the pay period that includes March 12 each year. Therefore the best estimate of the entry rate for calendar year t is the entry rate between March in year t and March in year $t+1$ in the BDS. The formula for the entry rate is:

$$\text{entry}(t) = \frac{\text{entrants}(t+1)}{0.5(\text{firms}(t) + \text{firms}(t+1)},$$

where entry(t) is the entry rate in year t, entrants(t) is the number of entrants in the BDS

Parameter	*Value* 1987	2015	*Remark*
ω	0.779	0.651	Non-college share of agents from CPS
β	0.985		
ν	2.0		
δ	0.025		Expected working life of 40 years
ρ_l, ρ_h	0.95		Storesletten et al. (2004)
γ	-1.5		Guided by Krusell et al. (2000) and vom Lehn (2015)
$\alpha + \eta$	0.85		Atkeson and Kehoe (2005)
r_o	0.082	0.121	Eden and Gaggl (2018)
r_i	0.169	0.071	Eden and Gaggl (2018)
μ_l^N	-0.008		Normalized so that $E[z_l^N] = 1$
μ_h^N	-0.008		Normalized so that $E[z_h^N] = 1$
$\bar{\mu}_e^N$	0.0		Normalization

Table D4: **Parameter values: externally calibrated and normalized parameters.** 2015 values are the same as 1987 values unless stated otherwise. Where necessary, parameter values are rounded to three decimal places.

in year t, and firms(t) is the total number of firms in the BDS in that year.[25]

D Parameter values and model fit

D.1 Additional discussion of parameter values and calibration moments

Values for internally calibrated parameters and the calibration moments are presented in Tables 2 and 3 of the main text. Table D4 summarizes the values of externally calibrated parameters. The moments presented in Table 3 illustrate some of the differences by skill and education. College educated people do better along many dimensions. They are much more likely to be high-skill workers than non-college educated people (60% compared to 13%) and high-skill workers earn more (45% more on average compared to low-skill). The college educated also earn more conditional on skill: the average high-skill college educated worker earns 29% more than the average high-skill non-college worker, and for low-skill workers this education premium is 40%. The model captures this with different means of the productivity distributions for the two education levels.

The parameters controlling the correlation between worker and entrepreneur productivities are estimated to be small, positive for college-educated agents and negative for non-college agents. The implied correlations between z_s, $s \in \{l, h\}$, and z_e for non-college and college agents are -0.31 and 0.23, respectively.[26] Recall that these parameters are pri-

[25] I keep the agriculture sector in the data for this analysis since the total number of firms increases when the data is split by sector—presumably some firms are being counted in two sectors. Repeating the calculations excluding this sector produces virtually identical results.

[26] For a given education level, there are small differences between the correlation of z_e with z_l and z_h, but

Moment	Model	Data
Non-college		
High-skill:low-skill averages	1.60	1.47
Entrepreneur:high-skill averages	1.49	1.40
College		
High-skill:low-skill averages	1.50	1.43
Entrepreneur:high-skill averages	1.90	1.87
College to non-college ratios		
Low-skill average income	1.39	1.54
High-skill average income	1.31	1.51
Entrepreneur average income	1.67	2.01

Table D5: **2015 income moments.** This table presents values for relative average incomes for various groups of workers for 2015. The moments in the non-college and college sections provide relative average incomes for low-skill employees, high-skill employees and entrepreneurs within education groups. The college to non-college ratios provide the relative average incomes of low-skill employees, high-skill employees and entrepreneurs between the two education groups. The high to low-skill income ratios are scaled up from the raw data to account for greater growth in non-wage income for high-skill agents between 1987 and 2015—see the discussion of wage growth calculations in Section C.6 for the details.

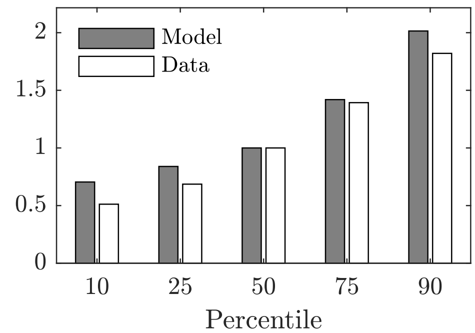
marily determined by the relative level of entrepreneur income and worker income. The negative correlation between worker and entrepreneur productivity for non-college educated agents is driven by the income premium for entrepreneurs in this education group being relatively low. From the perspective of the model, this implies that it is relatively low-productivity people in this education group who choose to be entrepreneurs. In terms of the quantitative importance of worker productivity in determining entrepreneur productivity, its role is modest. For the four education-skill groups, variation in worker productivity only accounts for 5.0–13.4% of the variance of entrepreneur productivity.[27]

The estimated elasticity of substitution between low-skill labor and IT capital ($\frac{1}{1-\tau}$) is 2.56. As a point of comparison, Krusell et al. (2000) estimate the elasticity of substitution between capital equipment and low education labor to be 1.67. Since the capital and labor inputs in this paper are defined more specifically to capture their substitutability, a higher elasticity of substitution makes sense. vom Lehn (2015) estimates the elasticity of substitution between routine labor and capital equipment at 1.39. While the labor input in this paper and vom Lehn (2015) are slightly different, the higher value that I estimate suggests that IT capital is more substitutable for lower skill labor inputs than capital equipment in general.

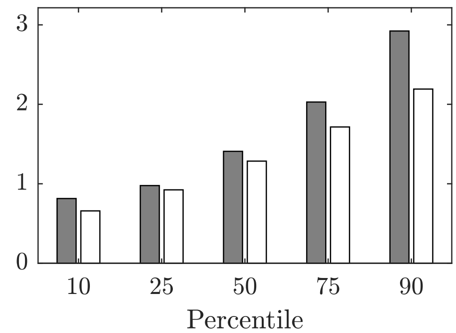
The main text notes that fixed costs are estimated to have increased by a factor of 1.9 from 1987 to 2015, and entry costs by a factor of 3.1. De Ridder (2019) provides some analysis to put these numbers in context. That paper analyses the trend in fixed costs using a range of methodologies, and documents consistent increases. Under that paper's

they're very small. For college educated agents, for example, the correlations are 0.231 and 0.237.

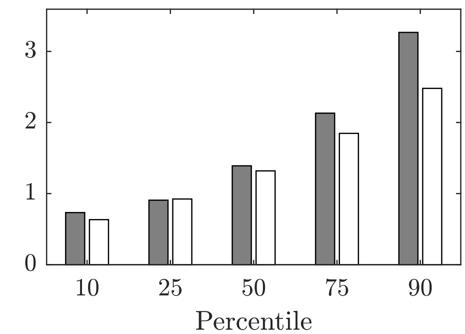
[27]These shares are computed by comparing the counterfactual variance in z_e if χ_n or $\chi_c = 0$ with the variance in the full model.



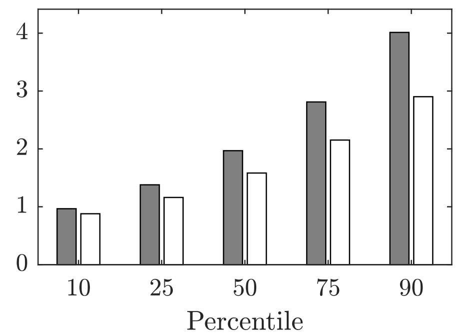
(a) Non-college, low-skill, 1987



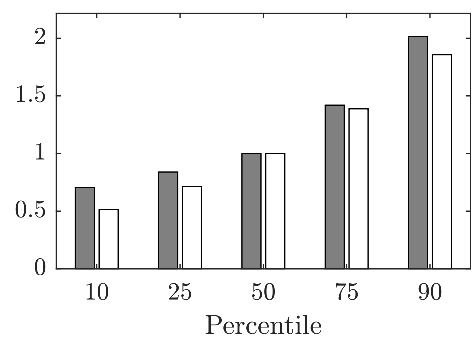
(b) Non-college, high-skill, 1987



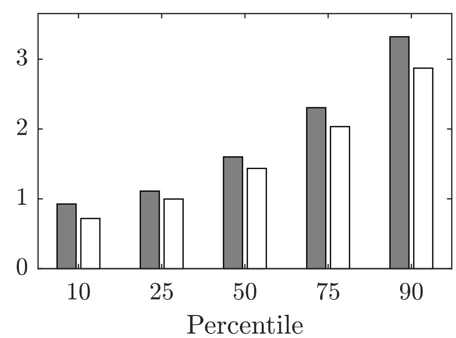
(c) College, low-skill, 1987



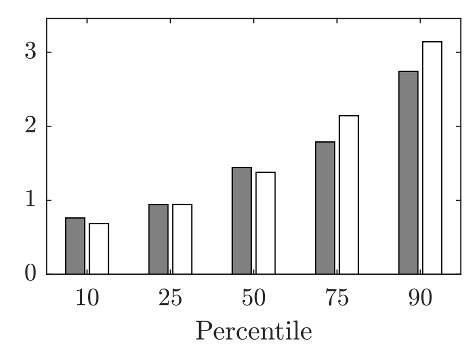
(d) College, high-skill, 1987



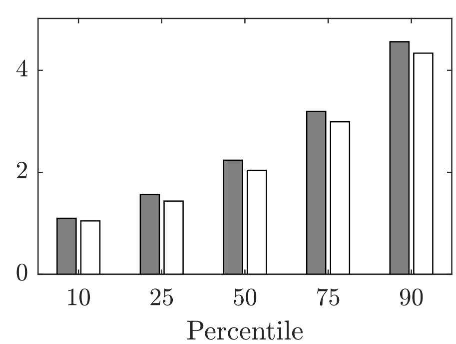
(e) Non-college, low-skill, 2015



(f) Non-college, high-skill, 2015



(g) College, low-skill, 2015



(h) College, high-skill, 2015

Figure D12: **Income distributions for model and data.** Each percentile is plotted relative to the 50$^{\text{th}}$ percentile of the non-college, low-skill distribution for the same year. For example, a value of 2.0 for the 75$^{\text{th}}$ percentile for non-college, high-skill people in 2015 means that this percentiles is twice as large as the 50$^{\text{th}}$ percentile for non-college, low-skill people in 2015. High-skill income in 2015 is scaled up to account for greater growth in non-wage compensation compared to that of low-skill workers—see the discussion of wage growth calculations in Section C.6 for the details.

baseline approach, the ratio of fixed costs to total costs increased by 64% for US public firms from 1979 to 2015, and by 47% for the universe of French firms from 1994 to 2016. In broad terms the model is consistent with these patterns, generating a slightly more modest growth rate of 25% for this ratio for entrepreneur firms from 1987 to 2015.[28]

D.2 Untargeted moments

Most moments of the occupation distributions for 1987 and 2015 are calibration targets, but there are a few free moments to check. The shares of college and non-college agents who are high-skill employees are targeted in 1987, but free for 2015. These moments don't change much over time in the data, and the model is consistent with this. In 2015 the values for the model are 13.2% and 60.9% for non-college and college agents, respectively, while the corresponding data values are 13.8% and 58.4% (values for 1987 are in Table 3). The out of labor force shares conditional on education are untargeted in both years. This moment closely matches the data in both years for non-college agents: 17.2% and 16.8% for 1987 for the model and data, respectively, and 31.8% and 29.8% for 2015. For college agents the out of labor force share is a few percentage points lower than in the data, but exhibits a similar proportional increase over time. It goes from 6.2% to 11.1% in the model, compared to 9.2% to 16.0% in the data.

Figure D12 and Table D5 present a range of moments of income distributions for 1987 and 2015, for the model and the data, to further assess the fit of the model.[29] Figure D12 provides income distributions for all education-skill pairs, for both 1987 and 2015. For each year all percentiles of the distributions are plotted relative to the median low-skill, non-college income for the relevant year. In this way the figure provides information on relative income between groups, as well as dispersion within groups. These relative incomes are obviously important for the occupation decisions of agents in the model. Panels (a)–(d) provide the results for 1987. In the calibration two moments of each distribution are targeted, so there are many more moments than targets presented. Overall the model fits the data reasonably. As indicated by the moments in Table 3, high-skill income is a little higher in the model than the data, and this is also true for the income of college agents. The model has more dispersion in the right tail of the distributions than the data.

Panels (e)–(h) of Figure D12 present the income distributions for 2015. The only moments related to these distributions that were targeted in the calibration were the growth of average low and high-skill income, so the moments in these panels are almost entirely

[28]The analog of De Ridder (2019)'s fixed costs to total costs ratio in the model is fixed costs to variable costs plus fixed costs. Scaling fixed costs in this way decreases the growth rate because firms, and their costs, have grown larger over time. I am comparing growth rates rather than levels of costs, since the latter are not comparable. Fixed costs are defined more narrowly in the model than in the empirical estimates.

[29]Note that in Figure D12 income distributions for entrepreneurs are not included since the relevant distributions from the data are not available. This is because the data does not distinguish between self-employed people without employees, and those with 1–9 employees, so the left tails of the entrepreneur income distributions in the model can't be easily mapped to the data.

	Prod. growth	Education	OLF value	r_o	2015
Entrepreneur share	1.05	1.10	1.02	0.93	0.71
Entry rate	0.99	0.93	0.93	0.92	0.72
Entrepreneur emp. share	1.01	1.08	1.07	1.06	0.80
College:non-college entrep. share	0.97	1.05	1.18	1.34	0.85
OLF share	0.87	0.69	1.28	1.56	1.66
w_l	1.14	1.33	1.36	1.21	–
w_h	1.24	0.94	0.93	0.79	–
Av. low-skill income	1.13	1.31	1.43	1.30	1.166
Av. high-skill income	1.22	1.00	1.04	0.93	1.443

Table E6: **Effects of changes in productivity, education and the out of labor force value, and SBTC.** All moments are presented relative to their 1987 values. For the *Productivity growth* column ζ is changed to its 2015 value and z_f, ψ, ψ_e and b are scaled by the the same percentage amount. For the next three columns, several parameters are changed to their 2015 values additively. For *Education* ω is changed to its 2015 value, for *OLF value* b is also changed to its 2015 value, and finally r_i and r_o are changed to their 2015 values as well in the *SBTC* column. The 2015 column provides moment values for 2015 relative to 1987 from the data.

untargeted. Given this, the model does a very good job of fitting the data. Relative income between groups are close to the data and the dispersion of income within groups is also similar. Table D5 provides additional moments for relative average incomes across groups for 2015, including for entrepreneurs. The model and data are reasonably close, with the main differences being that the premium for high-skill agents conditional on education is a little larger in the model, while the college premium is somewhat smaller. Overall the model replicates the relative incomes of the various types of agents in the model well, suggesting that it is doing a good job of capturing the tradeoffs that agents face when making their occupation choice decisions.

E Quantitative results

E.1 Effects of secondary parameter changes

The effects of the secondary parameter changes on key moments are presented in Table 4 of the main text. Table E6 decomposes these effects into the contribution of each type of parameter change and also adds wages, and average low and high skill income to the set of moments. There are four types of parameter changes in the decomposition, which are done in sequence, in a *cumulative* way. The first column shows just the effects of productivity growth, the second column shows the effects of productivity growth *and* the change in education, etc. For comparison, the final column of the table shows values for 2015 from the data. All values are presented relative to their 1987 values (i.e. 1.20 means a 20% increase), as in Table 4.

The parameter changes in the education, out of labor force value and r_o columns are straightforward. They involve changing the share of agents with a non-college education

(ω), the out of labor force value (b) and the non-IT capital rental rate from their 1987 to 2015 values (refer back to Table 2 for these). The parameter changes in the productivity growth column are slightly more involved. The objective in this column is to account for the effects of general productivity growth in the economy. To this end, the main parameter that changes is ζ, which changes the productivity level of all entrepreneurs by the same factor. Specifically, ζ is increased so that average entrepreneur productivity equals its 2015 value.[30] To simulate a general rise in productivity, rather than just for entrepreneurs, I increase z_f and the out of labor force value by the same factor. I also scale fixed costs and entry costs by the same factor so that their relevance is not diminished.

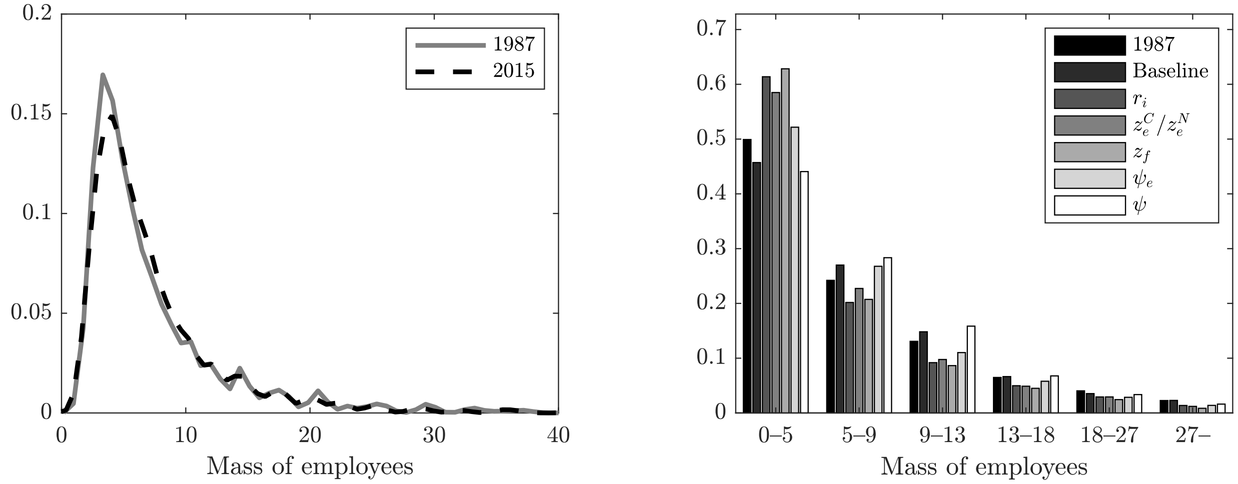
The main text explains that increasing education accounts for most of the decline in the entry rate resulting from the secondary parameter changes, and most of the increase in the entrepreneur employment share. The results in Table E6 confirm this, with the increase in education being the only parameter change that affects these moments by more than one percent.

The secondary parameter changes push the entrepreneur share down slightly, and the ratio of college to non-college entrepreneur shares up substantially. As mentioned in the main text, the main forces driving these changes are the increasing out of labor force value and the increasing cost of non-IT capital. To explain the mechanisms in more detail, the increasing out of labor force value has a direct effect on these moments in the following way. It attracts people out of entrepreneurship into not working, pushing the entrepreneur share down. This effect is stronger for less educated entrepreneurs because more or them have low enough profits for this this change to be relevant.[31] As for the increase in the rental rate of non-IT capital, it also pushes profits down, causing the entrepreneur share to fall. In equilibrium, fewer entrepreneurs means less demand for labor, so wages fall. This offsets the decline in the entrepreneur share, but only partially. This offsetting effect is larger for high-skill agents, because their wage declines by a larger percentage. This is why the increase in r_o increases the relative entrepreneur share of the college educated.

The out of labor force share increases significantly with the secondary parameter changes, almost fully accounting for the change in the data from 1987 to 2015. Productivity growth and increasing education work against this trend by pushing up the wages of low-skill people, and increasing the share of high-skill agents (who earn more on average). The increases in the out of labor force value and the non-IT capital rental rate have sufficiently

[30]There are two parameters changing from 1987 to 2015 affecting entrepreneur productivity: ζ and $\bar{\mu}_e^C$. For the secondary parameter changes being discussed here, it is ζ that increases so that average entrepreneur productivity changes from its 1987 to 2015 value. This requires ζ changing from its 1987 value of 1.0 to 1.122. The primary parameter changes, discussed in the main text, include a change in the relative entrepreneur productivity of college and non-college educated agents. This is achieved by changing $\bar{\mu}_e^C$ from its 1987 to 2015 value, and increasing ζ (to its 2015 value) to offset the effect of this on average entrepreneur productivity.

[31]The increase in the out of labor force value also pushes up the low-skill wage, strengthening these effects. However the direct effect, with wages held fixed, is quantitatively more relevant.



(a) Size distributions, 1987 and 2015

(b) Decomposition of 1987–2015 changes

Figure E13: **Entrepreneur firm size distributions.** Panel (a) presents the size distribution of entrepreneur firms in the model for 1987 and 2015. Panel (b) presents discretized size distributions for entrepreneur firms. The darkest bars are for the model in 1987. The second set of bars are for the baseline economy, after the secondary parameter changes. The remain bars capture the distribution after each of the primary parameter changes, in a cumulative way, such that the last set of bars represent the 2015 economy.

strong effects to offset these, and account for most of the increase in the out of labor force share in the data. The connection between this moment and the out of labor force value is straightforward, and this change accounts for 61% of the increase in the out of labor force share that is needed to match the 2015 data, once the countervailing effects of productivity growth and increasing education are accounted for. The increasing cost of non-IT capital is also quantitatively important, accounting for 29%. This effect primarily operates through the negative impact on wages.

As a final comment on the results for the secondary parameter changes, the last two rows show that these changes work against the increase in the relative income of high-skill employees. The gaps to the 2015 data are almost fully accounted for by SBTC (the declining cost of IT capital). This comes from the negative effect that this has on low-skill wages due to the substitutability between this type of capital and low-skill labor, and the positive effect on high-skill wages due to complementarity.

E.2 Firm size distribution

The main text discusses that, despite fixed and entry costs increasing over time, the model has a firm size distribution that is quite stable. Increases in fixed and entry costs make entrepreneur firms larger through two channels. They increase the productivity threshold for becoming an entrepreneur and, because wages decrease, they increase the size of entrepreneur firms conditional on productivity. At first glance this seems at odds with the stable entrepreneur size distribution documented in Figure 1(b). However, this ignores the fact that there are other changes to the economy occurring at the same time. These other changes may not matter much for the entry rate, for example, but can still influence the size distribution. Figure E13 presents information on how the size distribution changes in the model. Panel (a) shows the size distributions for 1987 and 2015, and they are very similar. Panel (b) discretizes the distribution and shows how various parameter changes from 1987 to 2015 change it. The darkest bars show the 1987 size distribution and the other bars show the effects of various parameter changes in a cumulative way. The increase in fixed and entry costs clearly shift the distribution to the right, as expected. However, SBTC (bars labeled r_i) offsets most of this effect.

SBTC affects entrepreneur firm size through several channels. The following expression for average firm size is useful for understanding these:

$$\bar{n} = \sum_{(\xi,s)\in\{N,C\}\times\{l,h\}} \omega_e(\xi,s) \int \omega_e(z_e|\xi,s) n(z_e)\, dz_e,$$

where $\bar{n}$ is average employment at entrepreneur firms, $\omega_e(\xi,s)$ is the share of entrepreneurs with education level ξ and skill level s, $\omega_e(z_e|\xi,s)$ is the p.d.f. for z_e for agents with education level ξ and skill level s who choose to be entrepreneurs, and $n(z_e)$ is the mass of employees at a firm with productivity z_e. SBTC changes all three variables in a way that decreases average firm size. It increases the share of people who are entrepreneurs within education-skill groups, so people with lower entrepreneur productivity choose to operate firms and this increases the share of small firms. $\omega_e(z_e|\xi,s)$ captures this effect. SBTC also causes the employment level of firms, conditional on productivity to decrease. The essence of this is that low-skill labor is substituted for IT capital and the increase in high-skill labor doesn't fully offset this. The third change is that SBTC shifts entrepreneurship towards people with low skills and education instead of high ones, and the former have smaller firms on average.

E.3 Effects of changes in labor force growth rate on results

Karahan et al. (2021) and Hopenhayn et al. (2021) have argued that a decline in the labor force growth rate has affected some of the moments of entrepreneurship studied in this paper. My approach to accounting for this theory is to take estimates of the share of changes in various moments that it accounts for, and then recalibrate the model to target the changes that remain. The rationale for this approach is explained in the main text.

The main quantitative exercise performed in the paper uses seven moments of the data from 1987 to 2015 to discipline parameters changes in the model. The labor force growth theory would definitely affect two of these moments—the entrepreneur share and the entry rate—and may affect a third—the entrepreneur share of employment. Karahan et al. (2021) and Hopenhayn et al. (2021) have results for the effect of their theory on the entry rate, so this is easy to quantify. They do not have direct results about the entrepreneur share, however their results for average firm size can be mapped to this moment (see the end of this section for an explanation of this mapping). Their models do not distinguish

Parameter	*Main calibration*	*Alternative calibrations* 1	2	3
b	0.423	0.430	0.432	0.438
z_f	1.338	1.312	1.346	1.318
ψ	0.290	0.329	0.187	0.144
ψ_e	0.981	0.755	0.594	0.411
$\bar{\mu}_e^C$	0.128	0.119	0.128	0.138
ζ	1.136	1.157	1.127	1.136

Table E7: **2015 parameter values for alternative calibrations.** Where necessary, parameter values are rounded to three decimal places. Parameters not listed maintain their values from Table 2.

between entrepreneur and non-entrepreneur firms, so several cases are considered for the effect of labor force growth on the share of employment in the entrepreneur sector.

In total I consider three scenarios for the share of the changes in these moments accounted for by the labor force growth theory. For scenario one, based on the results from Karahan et al. (2021), I assume that this theory accounts for 45% of the change in the entry rate and 75% of the change in average firm size.[32] For the share of employment at entrepreneur firms, since the shift in economic activity to non-entrepreneur firms may be related to the increasing size of firms, as a baseline I assume that the theory accounts for the same share of the change in this moment as the average size of firms. As an alternative I also consider the case in which this theory does not generate any change in this moment (scenario two). For the last scenario, I allow the theory to generate a greater increase in average firm size than has occurred in the data. Hopenhayn et al. (2021) find that the theory generates approximately twice the increase in average firm size as in the data from 1987 to 2014. I consider a scenario between the Karahan et al. (2021) and Hopenhayn et al. (2021) estimates, in which decreasing labor force growth generates 150% of the increase in average firm size that has occurred from 1987 to 2015 in the data.

The new parameter values for these scenarios and the updated calibration moments are presented in Tables E7 and E8. The value of the growth of average low-skill income from 1987 to 2015 is also included. This is the one other moment from the main calibration exercise whose value changes in these exercises. It changes a little, but it remains quite close to its value in the data. For each scenario the main quantitative results from Figure 7 are recomputed. They are presented in Figure E14. The results now inform us about the

[32]For the entry rate, Karahan et al. (2021) estimate that the labor force growth theory accounts for 1/3 to 60% of the decline in the data from 1979 to 2007, and these estimates decrease if you weaken their free entry assumption. I take approximately the mid-point of the estimated range, assuming that this theory accounts for 45% of the decline in the entry rate. An implicit assumption is that their results for 1979–2007 also hold for 1987–2015. For average firm size, Karahan et al. (2021) do not have results for this for the full dynamic exercise. However, from their comparative statics exercise, a decline in the labor force growth rate equal to the data generates a decline in the entry rate that's about 60% as large as in the data (similar to the result from the dynamic model) and a decline in average firm size that is close to the data. Since I discount the 60% estimate for the entry rate to 45%, I discount the average firm size result proportionally, to 75%.

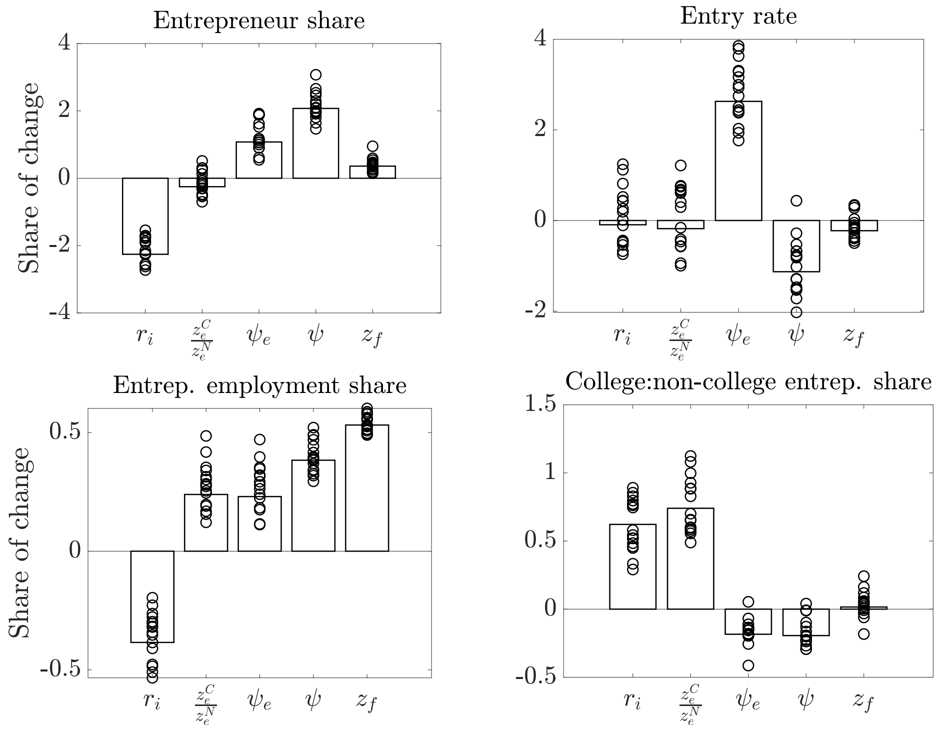
	1		2		3	
Targeted moments	*Model*	*Target*	*Model*	*Target*	*Model*	*Target*
1987–2015 growth of av. high-skill income	44.2%	44.3%	45.4%	44.3%	44.8%	44.3%
2015:1987 out of labor force share	1.66	1.66	1.65	1.66	1.66	1.66
2015:1987 entrepreneur share	0.83	0.83	0.84	0.83	1.00	1.01
2015:1987 entrep. share of employment	1.06	1.05	1.22	1.21	1.06	1.05
2015:1987 entry rate of entrepreneurs	0.84	0.84	0.85	0.84	0.84	0.84
2015:1987 college to non-college entrep. share	0.85	0.85	0.84	0.85	0.85	0.85
Untargeted moments	*Model*	*Data*	*Model*	*Data*	*Model*	*Data*
1987–2015 growth of av. low-skill income	19.1%	16.6%	19.7%	16.6%	20.7%	16.6%

Table E8: **Calibration moments for alternative calibrations.** Colons denote ratios. For example, '2015:1987 entrepreneur share' is the ratio of the 2015 to 1987 entrepreneur shares. Income growth rates are for real income. Full details of how the data moments are computed are in Section C.

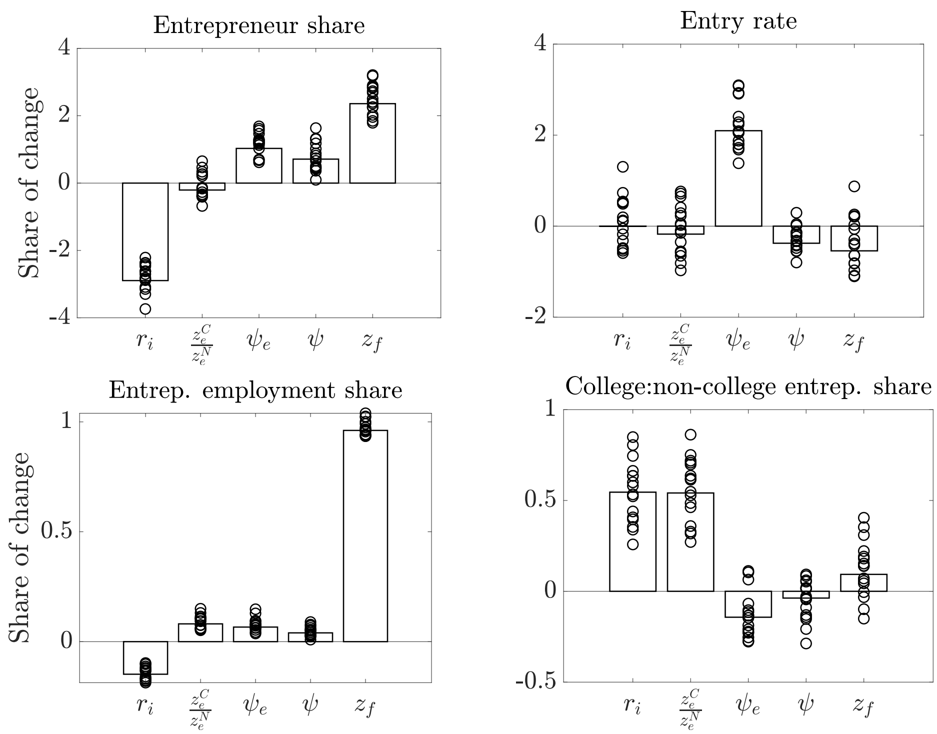
relative contribution of the forces in the model in accounting for the changes in the data that are not explained by the change in the labor force growth rate.

For scenarios one and two, the main results hold. A combination of increasing fixed costs, entry costs and increasing non-entrepreneur productivity generate the decline in the entrepreneur share; the decline in the entry rate is primarily due to increasing entry costs; the most important factor for explaining the decline in the entrepreneur share of employment is the increase in non-entrepreneur productivity; and the decline in the relative entrepreneur share of the college educated is due, in roughly equal measures, to SBTC and declining relative entrepreneur productivity of the college-educated. Taking scenario one as an example, the updated calibration targets require the model to generate smaller decreases in the entrepreneur share, the employment share of entrepreneurs, and the entry rate, and the relative sizes of some of these changes are different to what they were. This changes the size of the estimated increases in fixed costs, entry costs and non-entrepreneur productivity, but the general message that they have increased remains. Furthermore, the roles that they play in explaining the changes in the moments of entrepreneurship being studied remain broadly the same.

For scenario three, there is one qualitative difference. In this case the decrease in the labor force growth rate is assumed to increase average firm size more than has occurred in the data. This corresponds to a decrease in the entrepreneur share that is slightly larger than in the data. Consequently, the changes to the economy from 1987 to 2015 that are studied in the model need to increase this moment slightly, instead of decreasing it (Table E8). The model achieves this with more modest increases in fixed and entry costs (13% and 20% as large as under the main estimates, respectively). The slight increase in the entrepreneur share is achieved by SBTC pushing this share up, and increases in entry costs and non-entrepreneur productivity partially offsetting this effect (top left panel of Figure E14(c)). The remainder of the results from the main calibration are preserved.

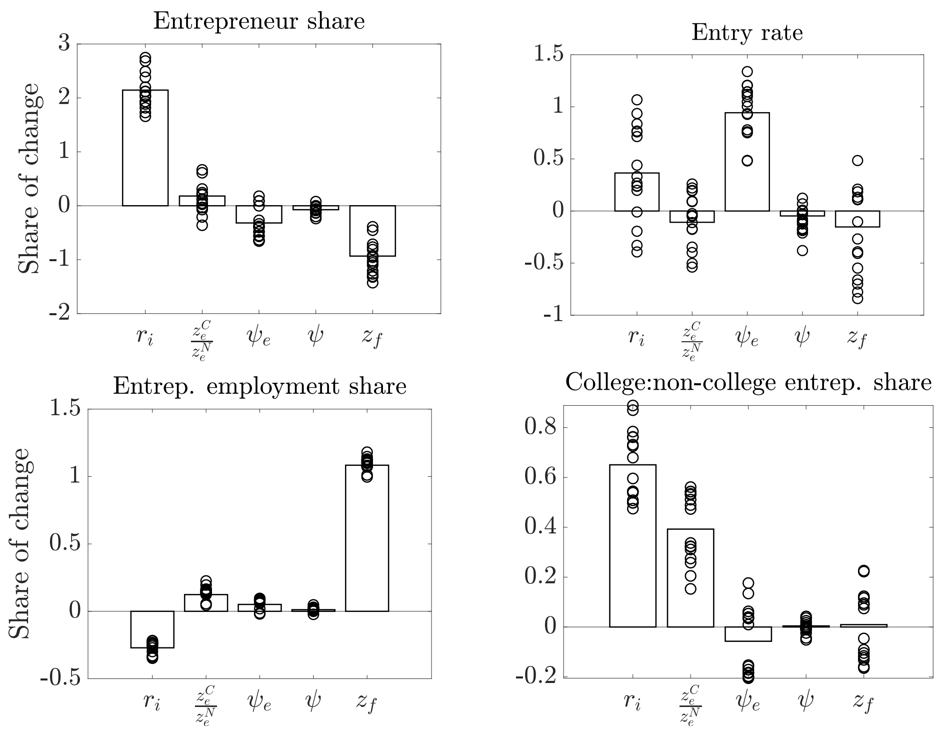


(a) Alternative calibration 1



(b) Alternative calibration 2

Figure E14: Continued on next page



(c) Alternative calibration 3

Figure E14: **Effects of changes to the economy on entrepreneurship for alternative calibrations.** This figure replicates Figure 7 for the alternative calibrations. The vertical scale is the share of the change in the relevant moment from its baseline value to its 2015 value in the model under the relevant calibration. The baseline scenario is defined in the same way as in the main text, with the parameter values differing according to the alternative calibrations.

Increasing entry costs is still the primary factor explaining the decline in the entry rate, increasing non-entrepreneur productivity accounts for most of the shift in employment to this sector, and SBTC and declining relative entrepreneur productivity of the college educated still drive the decline in their relative entrepreneur share.

Overall, allowing for changes in the labor force growth rate to account for some of the changes in entrepreneurship in the data reduces the quantitative importance of the factors studied in this paper. However, for the changes in the data not accounted for by that theory, the insights from the analysis about the relative importance of the factors studied in this paper generally hold. The main exception is that if the decrease in labor force growth generates a much larger increase in average firm size than has occurred in the data, then the factors studied in this paper are not needed to generate the decline in the entrepreneur share. However, they are still relevant for changes in the other moments of entrepreneurship.

Average firm size to entrepreneur share mapping As mentioned in the previous section, the papers on the labor force growth theory present results for changes in the average size of firms over time, which I map to changes in the the entrepreneur share. For this purpose, I approximate the average firm size in the model with

$$\frac{1 - \text{out of labor force share}}{\text{entrepreneur share}}.$$

This is slightly different to the measure of average firm size in the firm data and in the labor force growth papers for three reasons, but the impact should be minor. The first is that this measure omits non-entrepreneur firms from the firm count in the denominator. This only has a small effect, since, as discussed in Section C.8 of this appendix, close to 100% of firms in the economy have less than 100 employees and nearly all of these are associated with the self-employed. The second difference arises from the fact that some firms have multiple self-employed people associated with them, which will increase the firm count in the denominator and decrease average firm size. Since the vast majority of the self-employed are associated with firms with less than 100 employees, and in this size category there is close to one-self-employed person per firm (see the discussion of Table 1 in the main text), this also should not make a large difference. The third difference is to due to the sample being restricted to people aged 25 to 65. To quantify this difference, the average firm size measure outlined above implies a change in average firm size from 16.7 in 1987 to 20.8 in 2015 in the model. In the CPS the change over the same period is from 20.5 to 23.8.

The procedure for mapping average firm size to the entrepreneur share is as follows. Take an example in which it is assumed that changes in the labor force growth rate account for 50% of the increase in average firm size. For the calibration of 2015 parameters, I therefore target an average firm size of 18.75, instead of 20.8. Using the out of labor force value for 2015 of 25.0%, the equation above implies the target value for the entrepreneur share.

F Interpreting cost changes

F.1 RegData

The idea for this dataset is to take the Code of Federal Regulations, which contains all federal level regulations in the U.S., and separate it into its parts. For each part, textual analysis is performed to determine a relevance weight for the part for each industry, and the number of restrictions in the part. For each industry, a measure of regulation for each year is constructed by multiplying the relevance of each part by the number of restrictions in it, and then summing over parts. See McLaughlin and Sherouse (2018) for full details.

F.2 IT and regulation-related occupations

Table F9 lists the occupations that are treated as regulation-related and IT-related for the purposes of the analysis in Section 7. The occupation codes are from the 1990 Census Bureau Occupational Classification System.

Code	Occupation
	Regulation-related occupations
008	Human resources and labor relations managers
023	Accountants and auditors
027	Personnel, HR, training, and labor relations specialists
035	Construction inspectors
036	Inspectors and compliance officers, outside construction
178	Lawyers
234	Legal assistants, paralegals, legal support, etc
328	Human resources clerks, except payroll and timekeeping
337	Bookkeepers and accounting and auditing clerks
375	Insurance adjusters, examiners, and investigators
376	Customer service reps, investigators and adjusters, except insurance
796	Production checkers and inspectors
	IT-related occupations
044–059	Engineers
064–068	Mathematical and computer scientists
069–083	Natural scientists (Physicists and astronomers, chemists etc.)
213–223	Engineering and related technologists and technicians
224–225	Science technicians
229	Computer software developers
233	Programmers of numerically controlled machine tools
308	Computer and peripheral equipment operators
525	Repairers of data processing equipment

Table F9: **Regulation-related occupations** This table listed the occupations from the 1990 Census Bureau Occupational Classification System that are treated as regulation-related or IT-based in the analysis.

F.3 Industry definitions and sample size

The analysis requires consistent definitions of industries across datasets. The industry definitions from the BEA detailed fixed assets tables are used (a combination and two and three digit ISI codes) and industry codes from other datasets are harmonized with these. This results in a maximum of 144 observations. Some regressions have fewer observations because some industry years have small cell counts that don't allow all variables to be estimated. RegData provides information for fewer industries so any analysis including that data has fewer observations.

References

Acemoglu, Daron and David Autor (2011), "Skills, tasks and technologies: Implications for employment and earnings." *Handbook of Labor Economics*, 4, 1043–1171.

Atkeson, Andrew and Patrick Kehoe (2005), "Modeling and measuring organization capital." *Journal of Political Economy*, 113, 1026–1053.

Autor, David and David Dorn (2013), "The growth of low-skill service jobs and the polarization of the US labor market." *American Economic Review*, 103, 1553–1597.

Autor, David, Frank Levy, and Richard Murnane (2003), "The skill content of recent technological change: an empirical exploration." *Quarterly Journal of Economics*, 118, 1279–1333.

Bhandari, Anmol and Ellen McGrattan (2021), "Sweat equity in US private business." *Quarterly Journal of Economics*, 136, 727–781.

Cortes, Guido Matias, Nir Jaimovich, and Henry Siu (2017), "Disappearing routine jobs: Who, how, and why?" *Journal of Monetary Economics*, 91, 69–87.

Davis, Steven, John Haltiwanger, Ron Jarmin, and Javier Miranda (2006), "Volatility and dispersion in business growth rates: publicly traded versus privately held firms." *NBER Macroeconomics Annual 2006*, 21, 107–180.

De Ridder, Maarten (2019), "Market power and innovation in the intangible economy." Working paper.

Eden, Maya and Paul Gaggl (2018), "On the welfare implications of automation." *Review of Economic Dynamics*, 29, 15–43.

Goos, Maarten, Alan Manning, and Anna Salomons (2014), "Explaining job polarization: routine-biased technological change and offshoring." *American Economic Review*, 104, 2509–2526.

Hopenhayn, Hugo, Julian Neira, and Rish Singhania (2021), "From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share." Working paper.

Hurst, Erik, Geng Li, and Benjamin Pugsley (2014), "Are household surveys like tax forms? Evidence from income underreporting of the self-employed." *Review of Economics and Statistics*, 96, 19–33.

Jaimovich, Nir and Henry Siu (2020), "Job Polarization and Jobless Recoveries." *The Review of Economics and Statistics*, 102, 129–147.

Karahan, Fatih, Benjamin Pugsley, and Aysegul Sahin (2021), "Demographic origins of the startup deficit." Working paper.

Krusell, Per, Lee Ohanian, José-Víctor Ríos-Rull, and Giovanni Violante (2000), "Capital-skill complementarity and inequality: A macroeconomic analysis." *Econometrica*, 68, 1029–1053.

Lee, Sang Yoon Tim and Yongseok Shin (2016), "Horizontal and vertical polarization: task-specific technological change in a multi-sector economy." Working paper.

Levine, Ross and Yona Rubinstein (2017), "Smart and illicit: who becomes an entrepreneur and do they earn more?" *Quarterly Journal of Economics*, 132, 963–1018.

McLaughlin, Patrick A. and Oliver Sherouse (2018), *RegData US 3.1 Annual (dataset)*. https://quantgov.org/regdata-us/, QuantGov, Mercatus Center at George Mason University, Arlington, VA.

Polivka, Anne and Stephen Miller (1998), "The CPS after the redesign: Refocusing the economic lens." In *Labor Statistics Measurement Issues* (John Haltiwanger, Marilyn Manser, and Robert Topel, eds.), 249–289, University of Chicago Press.

Storesletten, Kjetil, Chris Telmer, and Amir Yaron (2004), "Cyclical dynamics in idiosyncratic labor market risk." *Journal of Political Economy*, 112, 695–717.

vom Lehn, Christian (2015), "Labor market polarization, the decline of routine work and technological change: a quantitative analysis." Working paper.