What's Driving the Decline in Entrepreneurship?

Online Appendix

Contents

A Empirics **1**
- A.1 Sample, variable definitions, and owners per firm estimate 1
- A.2 Composition effects on the entrepreneur share 1
- A.3 Decline in entrepreneurship by sector 2
- A.4 Broader definitions of an entrepreneur 3
- A.5 Robustness exercises for the change in entrepreneurship by education . . . 5

B Model and proofs **6**
- B.1 Optimal input choices and profit function for entrepreneurs 6
- B.2 Parameter identification 7
- B.3 Proofs of propositions 9
- B.4 Alternative formulation of non-entrepreneur sector 12

C Calibration details and model fit **13**
- C.1 Additional details for mapping model to data 13
- C.2 Full description of calibration approach 13
- C.3 Additional discussion of parameter values and calibration moments 16
- C.4 Untargeted moments 18

D Calibration moments **19**
- D.1 Entrepreneur share 19
- D.2 Out of labor force share and female participation 21
- D.3 Occupation distribution 22
- D.4 1987 income moments 23
- D.5 1987–2015 income growth 25
- D.6 Entrepreneur employment share 26
- D.7 Entry rate 28

E Quantitative results **28**
- E.1 Effects of secondary parameter changes 28
- E.2 Firm size distribution 30
- E.3 Effect of changes in the labor force growth rate on results 32

F Interpreting changes in fixed and entry costs **35**
- F.1 Data and methodology 36
- F.2 Results 37
- F.3 Additional information about the data 37

G Additional results for policy analysis **39**

A Empirics

A.1 Sample, variable definitions, and owners per firm estimate

The sample period of 1987–2015 has been chosen to ensure that self-employment can be measured consistently over time. Prior to 1987 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. 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.[1]

The estimates for the number of owners per business in various size categories are from the 1992 Characteristics of Business Owners Survey from the Census Bureau. 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 in the firms count, so I am assuming that they account for a negligible number of the businesses with less than 100 employees owned by self-employed people.

A.2 Composition effects on the entrepreneur share

In this section I show that the decline in entrepreneurship is not driven by changes in the composition of the population over time. To do this, I compute the entrepreneur share holding the composition of the economy fixed along several dimensions. Specifically, the entrepreneur share in year t can be written as

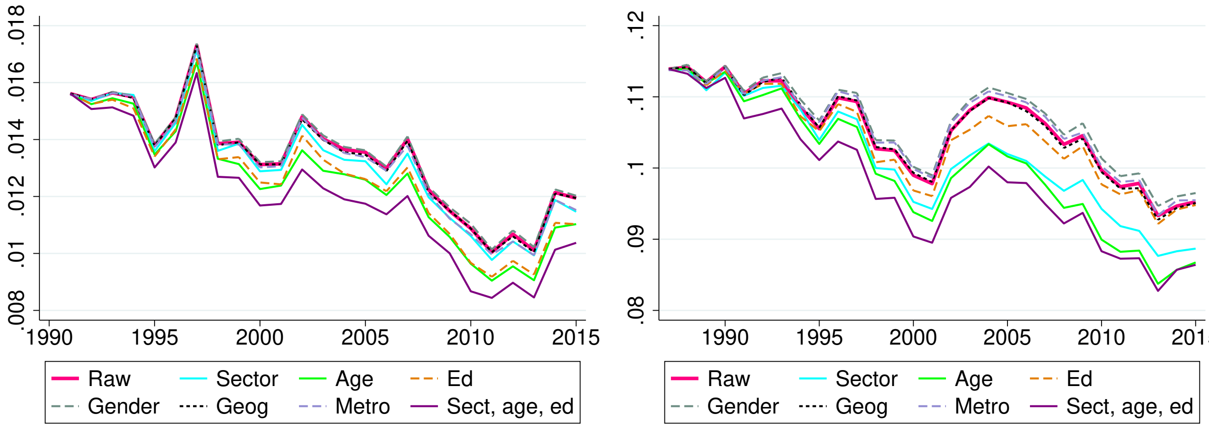
$$e_t = \sum_{g \in \mathcal{G}} \omega_{g,t} e_{g,t}$$

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

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

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

[1] 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) Entrepreneur share (b) Self-employment share

Figure A1: **Entrepreneur and self-employment shares with composition controls.** In panel (a) 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). Panel (b) is the same as panel (a), except that the results are for the self-employment share rather than the entrepreneur share.

distributions. The results for $e_{\mathcal{G},t}$ for each of these composition controls are presented in Figure A1(a): 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.[2] The results are in Figure A1(a) and the decrease in the entrepreneur share is, again, larger under these joint controls.

This exercise has been replicated for the self-employed share in Figure A1(b). The main message is the same.

A.3 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

[2]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. 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.

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.

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. 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 economy is held fixed. For sector g this is

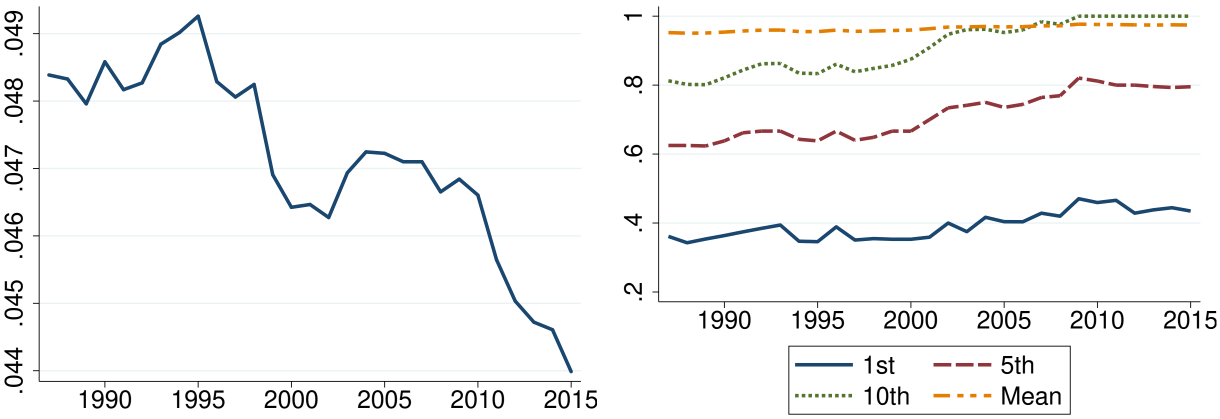
$$\frac{\omega_{g,1991}(\bar{e}_{g,2015} - \bar{e}_{g,1994})}{\bar{e}_{\mathcal{G},2015} - \bar{e}_{\mathcal{G},1994}}$$

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.4 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



(a) Ratio of employer firms to the labor force

(b) Share of income from main job

Figure A2: **Firms to labor force ratio and income share of main job for employees.** Panel (a) presents the 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. Panel (b) 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.

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.[3] This ratio is presented in Figure A2(a) and shows a decline over time.

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 A2(a) 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.[4] In Figure A2(b) 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. Looking at the mean only could hide a decrease in the share of income from the main job for people

[3]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.

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

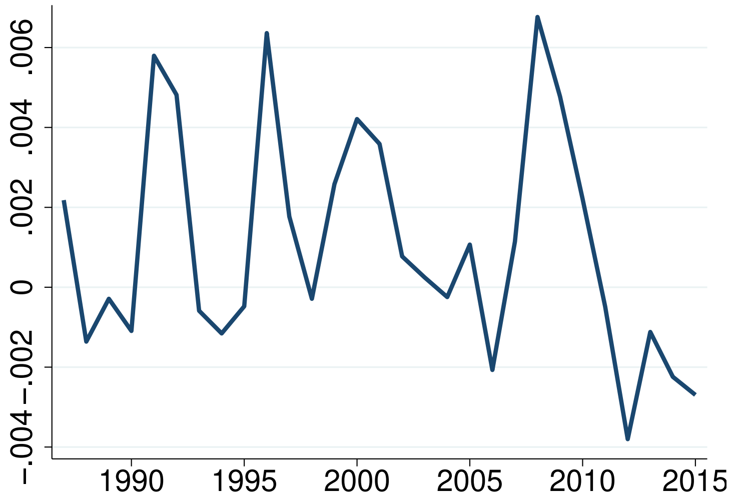


Figure A3: **Net merger and acquisition rate.** This is the implied net merger and acquisition rate from the BDS, computed using equation (A2). It is a net rate because it measures M&A less firm splits. A negative value implies that there were more firms splitting than merging or being acquired.

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 on the number of firms each year, the number of new firms, and the number of firm exits. A firm exit occurs when all 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:

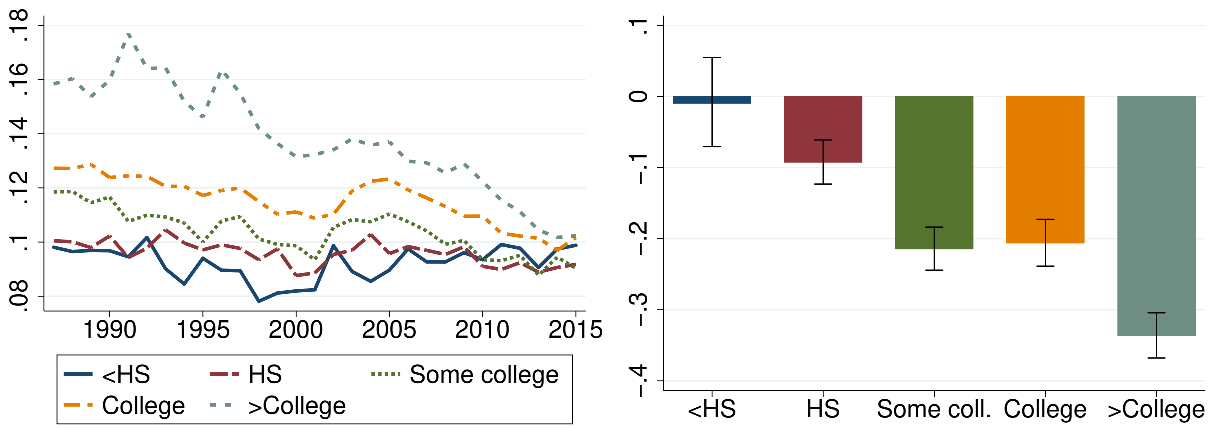
$$\text{M\&A rate}(t) = \frac{\text{firms}(t) - \text{deaths}(t) + \text{entrants}(t+1) - \text{firms}(t+1)}{\text{firms}(t)}, \tag{A2}$$

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 A3 and shows that there is not an upward trend over time.

A.5 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



(a) Self-employmed share, 1987–2015

(b) Self-employment, 1987/90–2012/15

Figure A4: **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).

each education group. This analysis is reproduced for the self-employed share in Figure A4 to show that the results hold for this broader measure of entrepreneurship as well.[5]

A second potential concern is that the larger decline in entrepreneurship for more educated people could be driven by a number of professional services industries—such as legal services, accounting, 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. the point estimates in Figure A5 are very similar to the main results. The confidence intervals are, of course, wider, because of the smaller sample.

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:

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

$$\Gamma_{\ell_h} = D_3^{\frac{1}{1-\alpha}} \Gamma_{k_o}^{\frac{\eta}{1-\alpha}}, \quad \Gamma_{\ell_l} = \left(D_1 D_2^{\frac{\gamma-\tau}{\tau}}\right)^{\frac{1}{1-\gamma}} \Gamma_{\ell_h}, \quad \Gamma_{k_i} = \left[\left(\frac{\lambda}{1-\lambda}\right)\left(\frac{w_l}{r_i}\right)\right]^{\frac{1}{1-\tau}} \Gamma_{\ell_l},$$

[5]This analysis goes back to 1987, rather than starting in 1991, since it does not require firm size information.

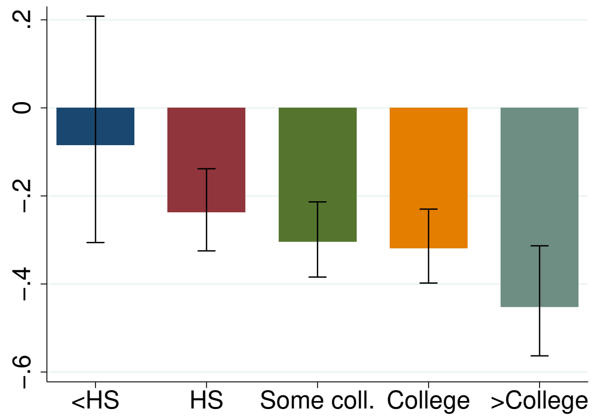


Figure A5: **1991–2015 percentage change in entrepreneur share by education, omitting professional services and FIRE.** This figure presents the relative change in the entrepreneur share from 1991–94 (pooled date) to 2012–15 for each education group with the professional services and FIRE sectors omitted (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 A4.

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

where

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

B.2 Parameter identification

This section formalizes the identification strategy discussed in Section 4.4 of the main text. Let $\mathcal{P}$ denote a set of values for the parameters of the model and let $x(\mathcal{P})$ denote the value of parameter x in $\mathcal{P}$. Allow functions to be conditional on parameters so that, for example, the occupation choice function for parameter set $\mathcal{P}$ is $o(\mathbf{z}, \epsilon|\mathcal{P})$. Now consider the employment by entrepreneurs of workers of skill type $s \in \{l, h\}$ and restrict attention to entrepreneurs who have skill type $s' \in \{l, h\}$ themselves (recall that an agent with skill type l, for example, has $z_l > 0$ and $z_h = 0$). Let the employment of skill type s by such entrepreneurs under parameters $\mathcal{P}$ be defined as:

$$L_s^{s'}(\mathcal{P}) \equiv \int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_{o}(e|\mathcal{P})\ell_s(z_e|\mathcal{P})\, dQ(\mathbf{z}, \epsilon|\mathcal{P}),$$

where $\mathbb{R}_+$ denotes the set of strictly positive real numbers. The following proposition provides a result regarding the relative effects of changes in non-entrepreneur productivity and fixed costs on $L_s^{s'}$, for a given change in the entrepreneur share for agents of type s'.

Proposition 3. *Let $\mathcal{P}$, $\mathcal{P}_{z_f}$ and $\mathcal{P}_\psi$ be sets of parameter values and take an $s' \in \{l, h\}$. Assume that $\partial w_s/\partial z_f > 0$ and $\partial w_s/\partial \psi < 0$ for all $s \in \{l, h\}$. For $\mathcal{P}_{z_f}$, $x(\mathcal{P}_{z_f}) = x(\mathcal{P})$ for all parameters x except z_f,* *and $z_f(\mathcal{P}_{z_f}) > z_f(\mathcal{P})$. For $\mathcal{P}_\psi$, $x(\mathcal{P}_\psi) = x(\mathcal{P})$ for all parameters x except ψ, and define $\psi(\mathcal{P}_\psi)$ to satisfy*

$$\int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_{o}(e|\mathcal{P}_\psi)\, dQ(\mathbf{z},\epsilon|\mathcal{P}_\psi) = \int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_{o}(e|\mathcal{P}_{z_f})\, dQ(\mathbf{z},\epsilon|\mathcal{P}_{z_f}). \tag{B3}$$

If ψ_e is sufficiently small, then, for all $s \in \{l, h\}$,

$$\frac{L_s^{s'}(\mathcal{P}_\psi)}{\ell_s(z_f|\mathcal{P}_\psi)} > \frac{L_s^{s'}(\mathcal{P}_{z_f})}{\ell_s(z_f|\mathcal{P}_{z_f})}. \tag{B4}$$

This proposition starts by taking increases in non-entrepreneur productivity and fixed costs, relative to a benchmark set of parameters $\mathcal{P}$, such that they generate the same entrepreneur share for agents of skill type s'. The result is that this increase in fixed costs generates a higher level of employment of both low and high-skill labor in the entrepreneurial sector, relative to the non-entrepreneurial sector, than the increase in non-entrepreneur productivity. The reason for this is that the two changes to the economy cause very different types of agents to switch from choosing entrepreneurship to being workers or out of the labor force.

This is illustrated in Figure B6, which plots the entrepreneur thresholds for type s' agents for the special case of $\psi_e = 0$. There are thresholds for an initial set of parameter values $\mathcal{P}$, and for increases in non-entrepreneur productivity and fixed costs, $\mathcal{P}_{z_f}$ and $\mathcal{P}_\psi$. For each set of parameters there is only one threshold, since with $\psi_e = 0$ the problems of agents with and without an endowment of a business are the same. Under both $\mathcal{P}_{z_f}$ and $\mathcal{P}_\psi$ the entrepreneur share is the same,[6] but $\mathcal{P}_\psi$ is associated with more higher-productivity entrepreneurs (area B) and fewer lower-productivity ones (area A). This higher productivity is associated with firms that employ more workers, as Proposition 3 provides.

There are three reasons for the different effects of increasing non-entrepreneur productivity and fixed costs. These can be seen by looking at the equation for the slope of $\underline{z}_e^{s'}(z_{s'},\epsilon)$ to the right of the kink point:

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

The first reason is that an increase in ψ decreases the slope of the threshold because it cuts into the profits of lower-productivity entrepreneurs more, in relative terms, than for higher-productivity entrepreneurs. The second difference between increases in non-entrepreneur productivity and fixed costs is how they affect operating profits, which is captured by Γ_π. An increase in fixed costs causes these to increase, due to lower wages, while higher non-entrepreneur productivity increases wages and decreases profits. These effects scale with entrepreneur productivity, so that they shift the entrepreneur thresh-

[6]The figure is drawn with a uniform distribution over $(z_{s'}, z_e)$ in mind, so that the masses of agents in areas A and B are equal.

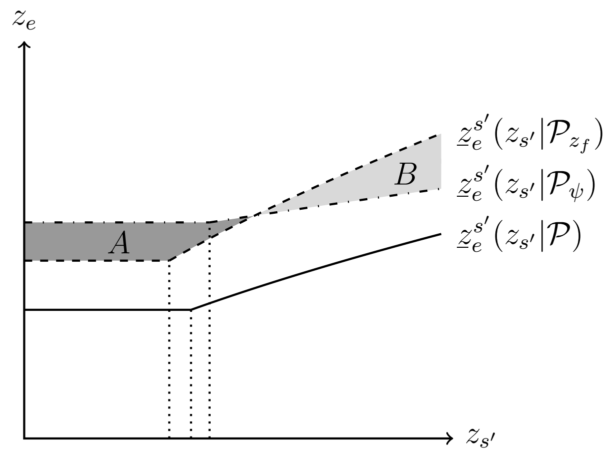


Figure B6: **Occupational choice when non-entrepreneur productivity or fixed costs increase.** This is a stylized representation of the entrepreneur threshold for agents of type s' for the sets of parameter values $\mathcal{P}$, $\mathcal{P}_{z_f}$ and $\mathcal{P}_{\psi}$, introduced in Proposition B4, when $\psi_e = 0$.

old more for high-productivity agents than low-productivity ones. For increasing fixed costs, this effect flattens the entrepreneur threshold, while higher entrepreneur productivity makes it steeper. The third effect operates through changes in employee income, which is one of the outside options for entrepreneurs. This effect is captured by $w_{s'}$ in the above equation. A change in $w_{s'}$ has a larger effect on the entrepreneur threshold for high-productivity agents because its effect is scaled by employee productivity, $z_{s'}$. When increasing fixed costs lower wages, this flattens the entrepreneur threshold, while increasing wages as a result of increasing non-entrepreneur productivity makes the threshold steeper.

Turning back to parameter identification, if Proposition 3 holds for both skill types simultaneously then, in the aggregate, an increase in fixed costs will result in a larger entrepreneur share of employment than an increase in non-entrepreneur productivity, for a given change in the entrepreneur share.[7] This distinction between the effects of these parameters is what will allow them to be separately identified. The quantitative analysis verifies this strategy for the full model, and also shows that the same distinction exists between increasing entry costs and increasing non-entrepreneur productivity. For entry costs, the same intuition applies as for fixed costs.[8]

B.3 Proofs of propositions

Proposition 1

[7]For the aggregate, the additional consideration is that increasing fixed costs and non-entrepreneur productivity are likely to cause different changes in the entrepreneur shares, conditional on skill type. This reallocation of entrepreneurship between skill groups can work against the result. The quantitative analysis will confirm that, to the extent that this happens, the effect is not strong enough to undo this feature of the economy.

[8]Potential changes in the shares of entrepreneurs with and without a business endowment make the result less general for this case, so it is verified quantitatively.

(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 (10) and the proof of part (a) of this proposition. From equation (10), $\underline{z}_e^s(z_s, 0) - \underline{z}_e^s(z_s, 1) > 0$ 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 (9) with respect to ψ.

[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).

For the second inequality, it follows from equation (10) 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 (9), 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 (9), it follows that $\partial\underline{z}_e^s(z_s,1)/\partial\psi_e < 0$.

Proposition 3 Using equation (5) for $\ell_s(z)$, inequality (B4) can be expressed as

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

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

$$\int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_{\mathscr{e}}(e|\mathcal{P}_\psi) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z},\epsilon|\mathcal{P}_\psi) > \int_{\mathbb{Z}} \mathbb{1}_{z_{s'}}(\mathbb{R}_+)\mathbb{1}_{\mathscr{e}}(e|\mathcal{P}_{z_f}) z_e^{\frac{1}{1-\alpha-\eta}}\ dQ(\mathbf{z},\epsilon|\mathcal{P}_{z_f}). \quad \text{(B5)}$$

To condense notation, let

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

so that (B5) 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 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 B3), 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 val-

[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.

ues, $\tilde{Z}_e^{s'}(\mathcal{P}_\psi) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f})$ can be expressed as

$$\tilde{Z}_e^{s'}(\mathcal{P}_\psi) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) = \sum_{\epsilon \in \{0,1\}} \int_{z_{s'}^*(\epsilon)}^{\infty} \int_{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_\psi)}^{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{z_f})} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} Q(\mathbf{z},\epsilon) dz_e dz_{s'}$$
$$- \int_0^{z_{s'}^*(\epsilon)} \int_{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_{z_f})}^{\underline{z}_e^{s'}(z_{s'},\epsilon|\mathcal{P}_\psi)} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} Q(\mathbf{z},\epsilon) dz_e dz_{s'}, \quad \text{(B6)}$$

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

$$\lim_{\psi_e \to 0} \tilde{Z}_e^{s'}(\mathcal{P}_\psi) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) = \int_{z_{s'}^*(0)}^{\infty} \int_{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_\psi)}^{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{z_f})} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} g(\mathbf{z}) dz_e dz_{s'}$$
$$- \int_0^{z_{s'}^*(0)} \int_{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_{z_f})}^{\underline{z}_e^{s'}(z_{s'},0|\mathcal{P}_\psi)} \mathbb{1}_{z_{s'}}(\mathbb{R}_+) z_e^{\frac{1}{1-\alpha-\eta}} g(\mathbf{z}) dz_e dz_{s'}. \quad \text{(B7)}$$

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 (B3) ensures that the weights placed on these two sets of values of z_e are equal, therefore $\lim_{\psi_e \to 0} \tilde{Z}_e^{s'}(\mathcal{P}_\psi) - \tilde{Z}_e^{s'}(\mathcal{P}_{z_f}) > 0$.

B.4 Alternative formulation of non-entrepreneur sector

Section 5.2 discusses a microfoundation for the non-entrepreneur sector using a continuum of firms. This section provides this formulation, showing how the firms aggregate to a representative firm with the same production technology as used in the main text. I also show that a change in the number of firms, which is a constant returns to scale (CRS) expansion of the sector, is isomorphic to a change in the productivity of the representative firm.

Assume that the non-entrepreneur sector is composed of a mass m of firms with productivities drawn from a Pareto distribution with scale parameter $\underline{z}_f > 0$ and shape parameter $\kappa > 0$. Denote the c.d.f. of this distribution by $G_f(z)$. Input demands for a firm in this sector are given by equation (5). Output has the form

$$y(z) = \Gamma_y z^{\frac{1}{1-\alpha-\eta}}, \quad \Gamma_y \equiv \Gamma_o^\eta [\phi \Gamma_h^\gamma + (1-\phi)(\lambda \Gamma_i^\tau + (1-\lambda)\Gamma_l^\tau)^{\frac{\gamma}{\tau}}]^{\frac{\alpha}{\gamma}}.$$

Note that I am using the same notation for this function as for the output of entrepreneur firms because the functions are indeed the same. We can aggregate the demand for input x from the non-entrepreneur sector with

$$\Gamma_x \int_{\underline{z}_f}^{\infty} z^{\frac{1}{1-\alpha-\eta}} dG_f(z),$$

and aggregate output similarly. This results in expressions for output and input demands with the forms $y(z_f)$ and $x(z_f)$, where

$$z_f \equiv m^{1-\alpha-\eta} \left(\frac{\kappa}{\kappa - \frac{1}{1-\alpha-\eta}} \right)^{1-\alpha-\eta} \underline{z}_f.$$

This shows that a non-entrepreneur sector composed of a continuum of firms is equivalent to a representative firm with the form used in the main text.

Finally, consider an increase in the mass of firms m. This is a constant returns to scale (CRS) expansion of this sector. In all of the functions for input demands and output, m only enters through the productivity term z_f. Therefore a CRS expansion of this sector through an increase in m is isomorphic to an increase in productivity of the representative firm.

C Calibration details and model fit

C.1 Additional details for mapping model to data

Data The main dataset 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. 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] I use non-routine cognitive occupations as high-skill occupations and the rest as low-skill occupations.

C.2 Full description of calibration approach

This section expands on Section 5.2 of the main text, providing a full description of how the parameter values are determined.

The share of the population without a college education can be computed with the CPS and is 77.90% in 1987 and 65.1% in 2015.[12] The death rate is set to a value of 0.025 to achieve an expected working life of 40 years. Given this value, β is chosen so that the effective annual discount rate is 4%. The CRRA parameter is set to 2.0. The value for the parameter controlling the persistence of employee productivity is assumed to be equal for low and high skill agents, and is given a value of 0.95 in accordance with the estimate of Storesletten et al. (2004).

The returns to scale of the production function are given by $\alpha + \eta$. Atkeson and Kehoe (2005) provide an extensive discussion of returns to scale for single plants and settle on a value of 0.85, which is used here as well. The rental rates for IT capital are 16.9% in

[11] For a detailed discussion of these categories see Autor et al. (2003) and Acemoglu and Autor (2011)

[12] A college education is defined as having completed at least a bachelor's degree.

1987 and 7.1% in 2015, and for non-IT capital they are 8.2% and 12.1%, respectively (Eden and Gaggl, 2018).[13] For productivities, the average productivity of low-skill workers, high-skill workers and entrepreneurs can be normalized for one of the education levels. I make this normalization for non-college agents, setting μ_l^N and μ_h^N so that average low and high-skill productivities for this group are equal to 1. $\bar{\mu}_e^N$ is normalized to zero. ζ can also be normalized for 1987 and is set to one.

The following parameters are calibrated internally using 1987 data. While the parameters are determined jointly by simulated method of moments, the approximate mapping between the moments and parameters is as follows. The consumption level for agents who are out of the labor force is set to target the out of labor force share.[14] The production function parameters η, ϕ and λ affect the demand for the various production inputs. To determine their values I use moments related to the division of income among inputs: the share of income going to employees, the ratio of the average high-skill income to average low-skill income, and the IT share of capital.[15] Given η, the value of returns to scale implies the value for α. The productivity level of the non-entrepreneur sector z_f, the fixed cost ψ, and the entry cost ψ_e are pinned down using the identification strategy outlined in Section 4. Regarding the moments used for this, the share of employment at entrepreneur firms is estimated using data from the CPS and Business Dynamics Statistics (BDS), and the share of agents who are entrepreneurs comes from the CPS.[16] To estimate the entry rate into entrepreneurship, the entry rate of firms in the BDS is used since, as discussed earlier, self-employed people account for a large share of firms.[17]

Now turn to parameters relating to skill shares and productivities. The share of agents who are high-skill conditional on education, θ_h^ξ for $\xi \in \{N, C\}$, is chosen to target the share of people in the relevant education group who work in high-skilled occupations.[18] The

[13]Note that these rental rates depend on the prevailing rate of return in the economy, as well as physical depreciation rates and changes in capital prices. So they are not directly comparable to the return on equities, for example. Since the Eden and Gaggl (2018) series end in 2013, I use the 2013 values as estimates for 2015. I also deflate their nominal rental rates using the GDP Implicit Price Deflator from the BEA to get real rental rates.

[14]In computing the out of labor force share in the data, I correct for the trend decline in this share for women up until the late 1990s. See Section D.2 for details. Since the model is solved on a discrete grid for z_e, z_l and z_h, a small amount of noise is added to the out of labor force value, b, to smooth out occupational choice functions. Specifically, for each agent in each period, b is drawn from normal distribution with mean equal to the calibrated value of b and standard deviation of 0.01. This helps with solving and calibrating the model and has virtually no effect on the aggregate moments of interest.

[15]The first moment is from the BEA data on value-added by industry. The second moment is from the CPS. Since there is no variation in hours worked in the model, moments of the empirical income distributions are computed using average hourly income for each person. Full details of income calculations are in Section D.4. The third moment is from the BEA detailed fixed assets tables.

[16]In the model an entrepreneur is a person who spends their time managing a firm with employees, so in the data I define an entrepreneur as a self-employed person (which means that they spend the majority of their working hours in self-employment) with at least one employee. See Section D.1 for details on how this entrepreneur share is estimated.

[17]Additional details for these moments are provided in Section D.7.

[18]See Sections D.2 and D.3 for details of the occupation distribution calculations in the data.

parameters that determine the level of low and high-skill productivity for college educated agents, μ_l^C and μ_h^C, are chosen to target the ratio of average income for college and non-college people in each of these skill groups. The level of entrepreneur productivity for college agents, $\bar{\mu}_e^C$, determines the share of college agents who are entrepreneurs. χ^ξ affects the correlation between worker and entrepreneur productivity for agents with education level ξ. A higher correlation increases the productivity of entrepreneurs, so this parameter is chosen to target the ratio of average entrepreneur to average high-skill employee income for this education level. There are six standard deviation parameters: for each education level there is one for each skill level and one for entrepreneurship. These determine the coefficient of variation of income for people in the corresponding occupation-education group. The persistence of entrepreneur productivity shocks affects the persistence of entrepreneur income. From the data I use the fraction of continuing entrepreneurs who remain in the same decile of the entrepreneur income distribution from one year to the next (37.5%), from DeBacker et al. (2018).

For parameters that take different values in 2015, the share of agents without a college education, ω, and the capital rental rates, r_o and r_i, are taken directly from the data. The consumption level of agents who are out of the labor force, the level of non-entrepreneur productivity, and the fixed and entry costs are all calibrated internally using the 2015 values of the same moments as are used for 1987. The level of entrepreneur productivity for college-educated agents $\bar{\mu}_e^C$ is chosen to target the relative entrepreneur shares of college and non-college agents in 2015.

The remaining parameters are the two elasticity of substitution parameters (τ and γ), which take the same value for both years, and the level of entrepreneur productivity ζ for 2015. These parameters are key for determining how the wages of low and high-skill workers change from 1987 to 2015. These moments are crucial since wages are fundamental for the tradeoff between being a worker and an entrepreneur. I fix one of the elasticity of substitution parameters, γ, with guidance from the literature and use the other two parameters to target the changes in average real income of low-skill workers and high-skill workers from 1987 to 2015. Since the CPS omits non-wage income, I adjust the growth rates using data on non-wage compensation from the Bureau of Labor Statistics' Employer Costs of Employee Compensation dataset. Using similar production functions to in the present model, Krusell et al. (2000) and Vom Lehn (2020) have estimated the elasticity of substitution between high-skill workers, defined on the basis of education or occupation, and capital equipment, generating estimates of 0.67 and 0.13 respectively.[19] γ is set to achieve an elasticity in the middle of this range (0.4).

[19]In Krusell et al. (2000) the group of workers that most closely corresponds to the high-skilled is those with a college education, which that paper labels "skilled." In Vom Lehn (2020) the corresponding category of people perform "abstract" occupations, which are defined in a very similar way to high-skilled occupations in this paper. While the production functions in those papers are not identical to one presently in use, they provide elasticity of substitution estimates to guide the choice of γ.

C.3 Additional discussion of parameter values and calibration moments

Parameter values and calibration moments are presented in Tables 2 and 3 of the main text. 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.

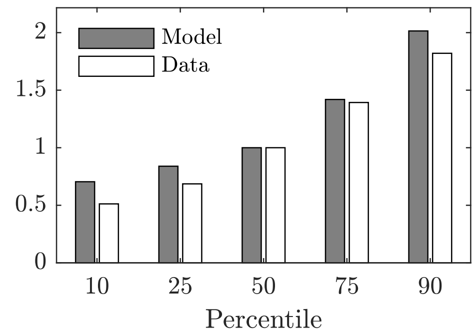
The parameters controlling the correlation between worker and entrepreneur productivities are estimated to be small, positive for college-educated agents and negative for non-college agents. The implied correlations between z_s, $s \in \{l, h\}$, and z_e for non-college and college agents are -0.31 and 0.23, respectively.[20] Recall that these parameters are primarily 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.[21]

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 (2020) estimates the elasticity of substitution between routine labor and capital equipment at 1.39. While the labor input in this paper and Vom Lehn (2020) 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.

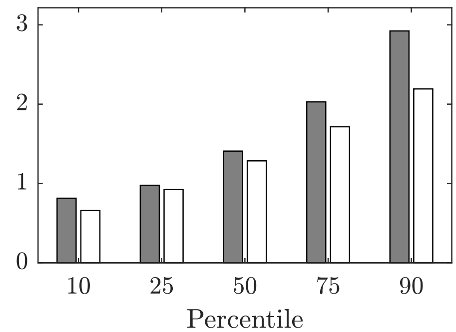
The main text notes that fixed costs are estimated to have increased by a factor of 1.9 from 1987 to 2015, and entry costs by a factor of 3.1. De Ridder (2024) provides some analysis to put these numbers in context. That paper analyses the trend in fixed costs using a range of methodologies, and documents consistent increases. Under that paper's baseline approach, the ratio of fixed costs to total costs increased by 64% for US public

[20] For a given education level, there are small differences between the correlation of z_e with z_l and z_h, but they're very small. For college educated agents, for example, the correlations are 0.231 and 0.237.

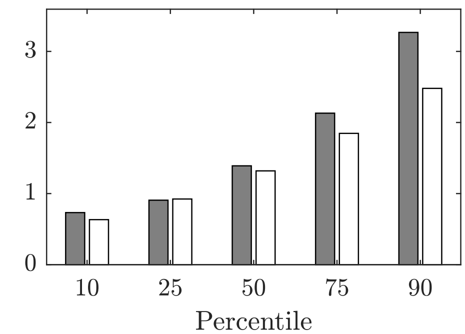
[21] These shares are computed by comparing the counterfactual variance in z_e if χ_n or $\chi_c = 0$ with the variance in the full model.



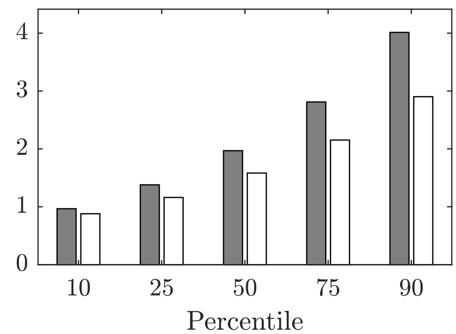
(a) Non-college, low-skill, 1987



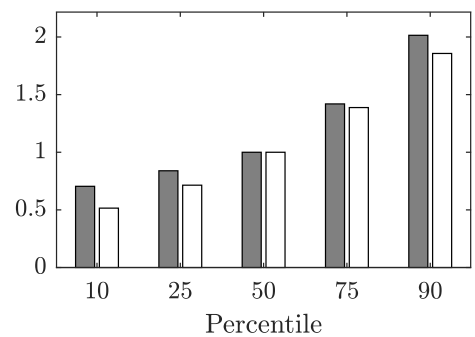
(b) Non-college, high-skill, 1987



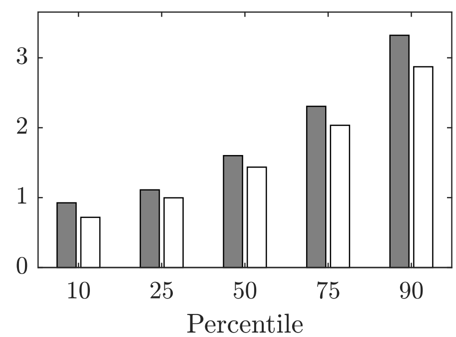
(c) College, low-skill, 1987



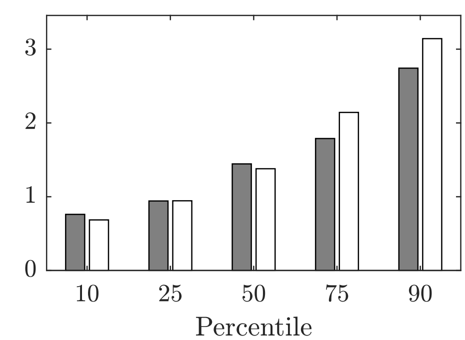
(d) College, high-skill, 1987



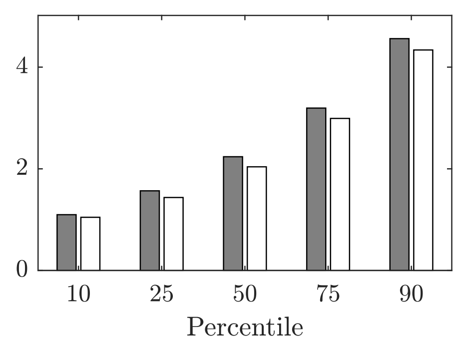
(e) Non-college, low-skill, 2015



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



(g) College, low-skill, 2015



(h) College, high-skill, 2015

Figure C7: **Income distributions for model and data.** Each percentile is plotted relative to the 50^{th} percentile of the non-college, low-skill distribution for the same year. For example, a value of 2.0 for the 75^{th} percentile for non-college, high-skill people in 2015 means that this percentiles is twice as large as the 50^{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 D.5 for the details.

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 C2: **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 D.5 for the details.

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.[22]

C.4 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 C7 and Table C2 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.[23] Figure C7

[22]The analog of De Ridder (2024)'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.

[23]Note that in Figure C7 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

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 C7 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 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 C2 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.

D Calibration moments

This section provides details of how the moments used for calibrating the model have been computed.

D.1 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.

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

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.

Figure D8: **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

basis up to 2014, including establishments with 1–9 employees.[24] 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.[25] This gives me an estimate of the number of self-employed people with businesses with 1–9 employees each year.

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.[26] 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 with less than 10 employees and the number of firms with 1–9 employees are presented in Figure D8. 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

[24]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.

[25]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.

[26]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.

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.[27] 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.

D.2 Out of labor force share and female participation

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.

This is a complicated issue to address. In principle, one would like to construct a counterfactual economy for 1987 in which the barriers to female labor force participation are the same as in 2015. Considering this from the perspective of the model, start by holding wages fixed. With women facing different labor market opportunities, their education choices and the resulting education distribution may change. The share of women participating in the labor force and the occupational choices of women in the labor force are also likely to change. If we allow wages to change, then this will affect the labor force participation and occupation choices of all agents, and possibly the education distribution as well. To fully account for these effects, one would need to construct a general equilibrium counterfactual model for 1987 with a richer decision space than the present model. This is beyond the scope of this paper, so I make a partial adjustment. I correct female labor force participation and hold other choices and equilibrium wage distributions fixed. The results should therefore be taken with the caveat that this is an imperfect correction.

The approach to adjusting the out of labor force share 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

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

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. 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 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.[28]

This procedure implies that there are additional people in the in the labor force that need to be allocated to occupations. I assume that, conditional on being in the labor force, these people would have had the same occupational distribution as the people in the labor force.[29] I also assume that their income distributions conditional on education and occupation would have been the same.

D.3 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.[30] This gives me estimates of the occupation distribution for each education level, consisting of the shares of people who are out of the labor force, low-skill employees, high-skill employees and entrepreneurs.[31] 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 D3.

[28] Figures illustrating this exercise are available on request.

[29] For example, if non-college educated people in the labor force in 1987 were allocated to low skill, high skill and entrepreneur occupations with shares of 0.6, 0.3 and 0.1, then I would apply to same shares for the additional labor force participants.

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

[31] 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.

	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 D3: **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.

D.4 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.[32] 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 they 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 two additional issues with the income data that are addressed. First is top coding. First, 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 second 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

[32]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 of respondents. The replacement values for the 1988 March CPS have been taken from the CPS IPUMS website: https://cps.ipums.org/cps/income_cell_means.html, accessed 4 May 2020.

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 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.[33] 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).

D.5 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.[34] 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 12th 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.

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 D4.
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

[33]This assumes that the return on savings invested in private businesses and elsewhere are similar.
[34]The data used in this paper come from ECEC Table 9 for 1987–2003 and Table 15 for 2004–15.

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

Table D4: **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.

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.[35] 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.[36] 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.

D.6 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 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

[35]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.

[36]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.

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.[37] 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.[38] 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. 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.

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

[38] 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.

D.7 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.[39]

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 in year t, and firms(t) is the total number of firms in the BDS in that year.[40]

E Quantitative results

E.1 Effects of secondary parameter changes

Table E5 provides details of the effects of the secondary parameter changes, parameter by parameter. 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).

The parameter changes in the education, out of labor force value and r_o columns are

[39] 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.

[40] 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.

	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 E5: **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.

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 increases so that average entrepreneur productivity equals its 2015 value.[41] 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 changes in the main moments of entrepreneurship are modest. Starting with the entry rate, it falls by a total of eight percent, mostly due to the increase in education. This change increases the supply of high skill workers and drives down their wage. This increases the gap between the values of entrepreneurship and high skill work, resulting in less churn between these occupations. The share of employment at entrepreneur firms increases by six percent, going against the trend in the data. This is also primarily due to education increasing because more educated entrepreneurs have larger firms on average.

The secondary parameter changes push the entrepreneur share down by seven percent, and the ratio of college to non-college entrepreneur shares up by 34 percent. The

[41] 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.

increase in this ratio goes against the data, implying that the decline in the relative entrepreneurship rate of the college-educated has been even larger than it first appears. The main forces driving these changes are the increasing out of labor force value and the increasing cost of non-IT capital. 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 change to be relevant.[42] 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.

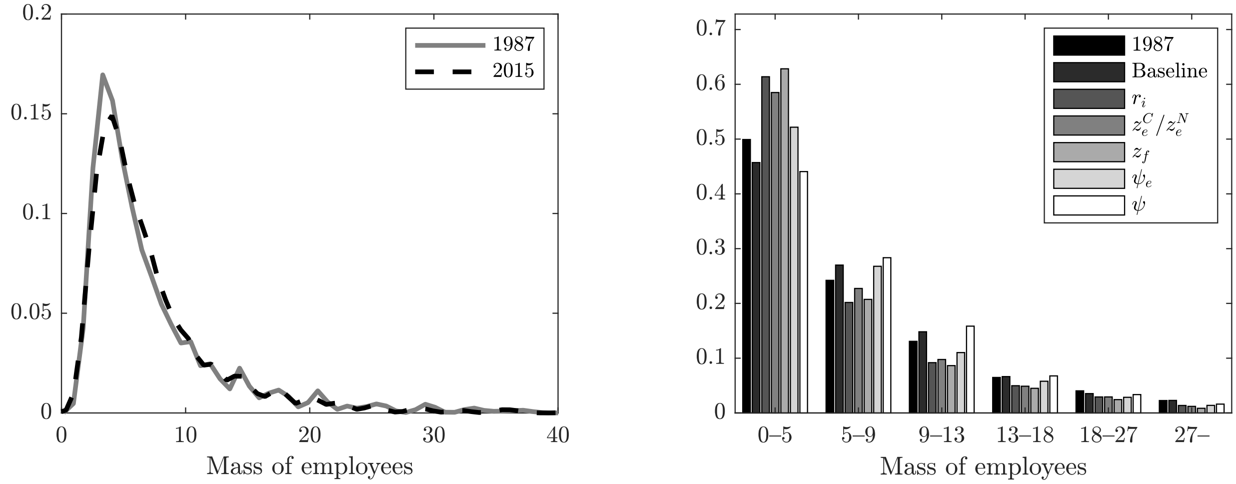
The out of labor force share increases significantly with the secondary parameter changes, almost fully accounting for the change in the data from 1987 to 2015. Productivity growth and increasing education work against this trend by pushing up the wages of low-skill people, and increasing the share of 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 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. These results imply that the changes to the economy generating the increase in the out of labor force share are not closely related to those driving changes in entrepreneurship.

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 thresh-

[42]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 E9: **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.

old 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 E9 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 Effect of changes in the labor force growth rate on results

While the analysis in the main text has considered a range of factors that could explain the changes in entrepreneurship from 1987 to 2015, there are possibilities outside the framework. One that has been considered in the literature is changes in the growth rate of the labor force (see Karahan et al., 2024; Hopenhayn et al., 2022).[43] The approach to assessing the effects of this theory on the results is to start by estimating how much of the changes in the moments used in the quantitative exercise can be accounted for by this theory. I then reestimate the model for 2015 so that it generates the residual changes in the relevant moments. The results inform us about the contribution of the factors considered in this paper to explaining changes in entrepreneurship not accounted for by the assumed effects of changes in the labor force growth rate.

Ordinarily, an issue with this approach would be that changes in the labor force growth rate could interact with the changes in parameters being studied in the model in this paper, such that they cannot be studied independently in this way. However, under the theory, changes in the labor force growth rate generate changes in the entry rate of firms, while having little or no impact on prices.[44] This absence of price effects means that this change in the economy should not interact with the changes studied in this paper, because they all act through changes in prices.

The quantitative exercise presented in Sections 5 and 6 uses changes in seven moments of the data from 1987 to 2015 to discipline parameter 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. (2024) and Hopenhayn et al. (2022) 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—the details are discussed below. Their models do not distinguish 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

[43] In a more recent contribution, Peters and Walsh (2021) also study this theory. For the purpose of the exercises undertaken here, I focus on Karahan et al. (2024) and Hopenhayn et al. (2022) since they use models that are closer to this paper's.

[44] In Hopenhayn et al. (2022) the impact is precisely zero, while in Karahan et al. (2024) it is small.

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 E6: **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.

consider three cases based on the estimates from these papers.

For scenario one, based on the results from Karahan et al. (2024), I assume that this theory accounts for 45% of the change in the entry rate and 75% of the change in average firm size (which maps to 43% of the change in the entrepreneur share, as I explain below).[45] 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 take scenario one and allow the theory to generate a greater increase in average firm size than has occurred in the data. Hopenhayn et al. (2022) find that the theory generates approximately twice the increase in average firm size as in the data from 1987 to 2014. I consider a case between the Karahan et al. (2024) and Hopenhayn et al. (2022) 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 (mapping to 104% of the change in the entrepreneur share over this period).

To map average firm size to the entrepreneur share, I approximate the average firm size in the model with

$$\text{Average firm size} = \frac{1 - \text{out of labor force share}}{\text{entrepreneur share}}.$$

This is slightly different to the measure of average firm size in the firm data and in the labor force growth papers, but the impact should be minor.[46] With this mapping one can

[45]For the entry rate, Karahan et al. (2024) 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. (2024) 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%.

[46]There are three reasons for the difference. The first is that this measure omits non-entrepreneur firms

	1		2		3	
Targeted moments	*Model*	*Target*	*Model*	*Target*	*Model*	*Target*
1987–2015 growth of av. high-skill income	44.23	44.30	45.40	44.30	44.84	44.30
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	0.94	0.95	0.78	0.79	0.94	0.95
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%

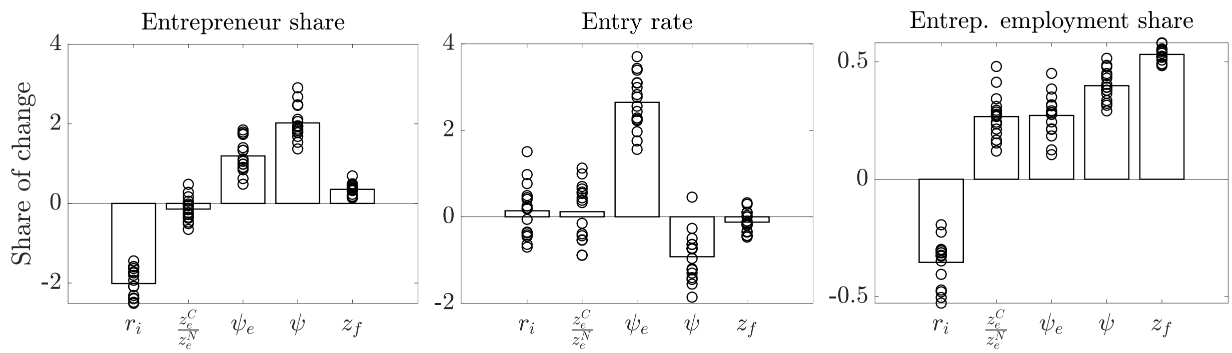
Table E7: **Calibration moments for alternative calibrations.** Colons denote ratios. For example, '2015:1987 entrepreneur share' is the ratio of the 2015 to 1987 entrepreneur shares. Income growth rates are for real income. Full details of how the data moments are computed are in Section D.

compute average firm size for the calibrated model in 1987 and 2015, and determine what share of the change in the entrepreneur share is implied by a given share of the change in average firm size.

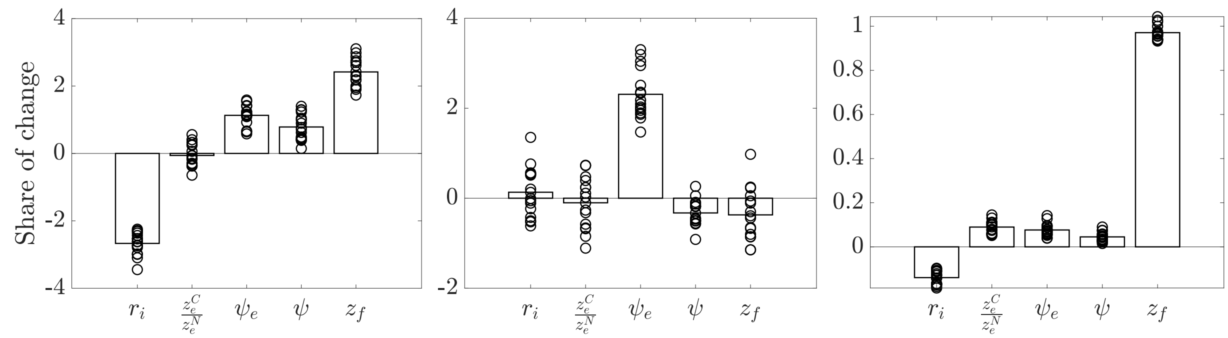
The new parameter values for these scenarios and the updated calibration moments are presented in Tables E6 and E7. The value of the growth of average low-skill income from 1987 to 2015 is also included. This is the one other moment from the main calibration exercise whose value changes in these exercises—but the change is small and its value remains quite close to the data.

The results of the analysis are presented in Figure E10. I focus on the entrepreneur share, the entry rate and the entrepreneur share of employment since these are the moments that are primarily affected by labor force growth. To summarize, while the absolute magnitude of the changes in these moments that the forces in the model account for are different under these calibrations—which is by assumption—their relative importance mostly stays the same. In all cases, rising entry costs are the most important factor for the decline in the entry rate, and rising non-entrepreneur productivity is the main factor driving the decline in the entrepreneur employment share. For the entrepreneur share it remains the case that rising fixed costs, entry costs and non-entrepreneur productivity account for the decline under alternative calibrations one and two. Under alternative calibration three, the change in the labor force growth rate is assumed to generate a larger

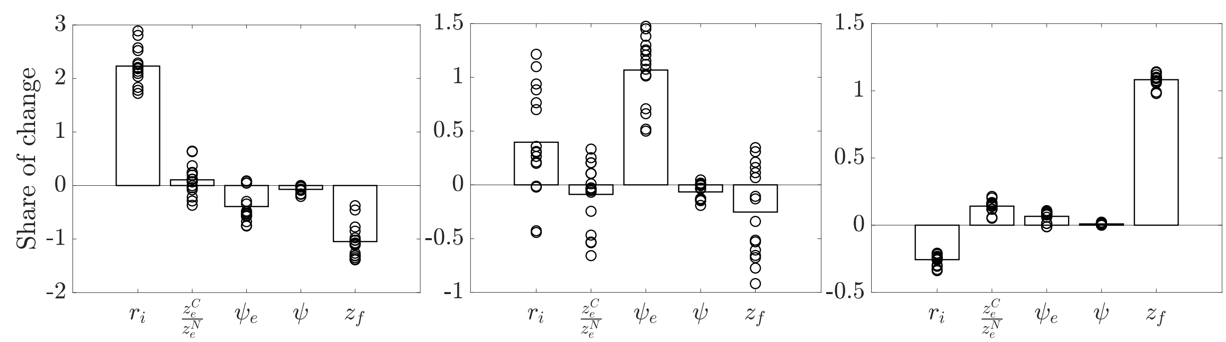
from the firm count in the denominator. This only has a small effect, since, as discussed in Section D.7 of this appendix, close to 100% of firms in the economy have less than 100 employees and nearly all of these are associated with the self-employed. The second difference arises from the fact that 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.



(a) Alternative calibration 1



(b) Alternative calibration 2



(c) Alternative calibration 3

Figure E10: **Effects of changes in primary parameters for alternative calibrations.** This figure replicates the panels for the entrepreneur share, the entry rate and the entrepreneur employment share from Figure 6 for the alternative calibrations.

decline in this moment than has occurred in the data. So the forces in the model need to increase this moment and, as discussed in the main text, SBTC is able to do this.

F Interpreting changes in fixed and entry costs

The quantitative results in the main text show that increases in both fixed and entry costs have contributed to the declines in the entrepreneur share and the entry rate, with increasing entry costs being particularly important. This raises the question of what factors are behind the rise in these costs. In Section 4 I discuss how increases in these costs could be driven by increasing regulation, or changes in technology that increase upfront costs and reduce variable costs. Other possibilities include that it has become more expensive to generate new ideas (Bloom et al., 2020), which could raise the cost of creating and expanding businesses, and that it is becoming more costly to build a customer base (Bornstein, 2021), which is essential for a firm's sales. While it is beyond the scope of this paper to fully assess these explanations, here I evaluate whether there is suggestive empirical support for regulations and changing production technologies driving up these costs.

F.1 Data and methodology

The strategy is to assess the relationship across industries between changes in entrepreneurship and measures of changes in regulations and technologies that could have driven up fixed and entry costs. The period of analysis is 1987–2015. To measure entrepreneurship I use the share of the labor force in an industry who are self-employed from the CPS. Unlike in Section 2, I do not restrict attention to self-employed people with at least 10 employees because at the industry level this would leave too few observations to construct reliable entrepreneur shares.

To quantify changes in regulations at the industry level I use two measures. The first is the measure of the number of Federal regulations at the industry level from the RegData dataset, constructed from the Code of Federal Regulations by McLaughlin and Sherouse (2018).[47] For the second measure I construct a proxy for the level of industry regulations by computing the share of employees in regulation-related occupations using the CPS. These are occupations in which people are likely to be performing tasks related to regulatory compliance, such as legal, human resources, accounting and auditing occupations. For the full list of occupations that I classify as regulation-related see Section F.3 below.

For changes in technology that could drive the increase in fixed and entry costs, I focus on a particular theory. The idea is that improvements in IT technology have allowed firms to adopt technologies with higher upfront costs and lower marginal costs (see Aghion et al., 2023; Hsieh and Rossi-Hansberg, 2023; De Ridder, 2024). Under this theory measures of IT technology adoption should be positively related to the rise in fixed and entry costs. I use three such measure at the industry level. There are two measure of IT capital intensity: the ratio of the IT capital stock to value added, and the real capital stock per employee.[48] The third measure is based on the occupation composition of each industry. I identify occupations in the CPS data that are IT-related and compute the share of employees in each industry in these occupations.[49] The idea is that if an industry is adopting more IT technology over time then it should also have more employees in these occupations.

To assess the relationship between changes in entrepreneurship, and changes in regu-

[47]See the Section F.3 for a discussion of how this measure is constructed.

[48]The IT capital stock is taken from BEA detailed fixed assets tables. Value added is also from the BEA and employment is from the CPS.

[49]See the Section F.3 for a list of these occupations.

lations and technology, I use the following regression:

$$\Delta \log e_{jt} = \alpha + \boldsymbol{\beta}_1' \boldsymbol{\Delta x}_{jt} + \boldsymbol{\beta}_2' \boldsymbol{\Delta y}_{jt} + \varepsilon_{jt} \tag{F8}$$

where $\Delta \log e_{jt}$ is the change in the log of the entrepreneur share from an earlier period (specified shortly) to period t for industry j,[50] $\boldsymbol{\Delta x}_{jt}$ is a vector of changes in IT and regulation measures (in most regressions it will just have one element), and $\boldsymbol{\Delta y}_{jt}$ is a vector of changes in control variables: changes in the average age of people working in each industry, the share who are males, the share who have a college degree, and the share who live in a metropolitan area. I divide the sample into three sub-periods to increase the number of observations, and average each variable over two years at the start and end points to smooth them. The sub-periods are 1988-89 to 1999/2000, 1999/2000 to 2005/06, and 2005/06 to 2014/15. With the exception of the final endpoint, each sub-period starts and ends just before a business cycle peak to reduce the risk of higher frequency fluctuations contaminating the results. Of course the data does not contain another peak after 2007, so the last years of the dataset are used for the final endpoint.

F.2 Results

The results are presented in Table F8. In columns (1)–(5) I take one measure of technological change, or changes in regulation, and regress it on the change in the entrepreneur share. The main result is that the coefficients on all variables are negative, consistent with both the increasing use of IT technology and increasing regulation driving up fixed and entry costs, and pushing entrepreneurship down. As expected with a small number of observations, the statistical power of the results is generally low, so the evidence should only be taken as suggestive. To give a sense of magnitudes, one percentage point increases in the IT employment share and the regulatory employment share in an industry are associated with 7.0% and 2.6% declines in the self-employed share, respectively. When I include measures of both changes in IT technology adoption and changes in regulations (focusing on the measures that had the highest statistical significance in the individual regressions), both variables have negative coefficients (column 6). Their p-values are 5% and 11%, which I do not interpret to be sufficiently different to favor one theory over the other. I therefore take the data as providing suggestive support for both theories.

F.3 Additional information about the data

RegData This dataset is described fully by McLaughlin and Sherouse (2018). To summarise, the idea 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

[50] After harmonization across datasets, there are 48 industries. Section F.3 discusses this and the sample size further.

	(1)	(2)	(3)	(4)	(5)	(6)
IT employment share	-7.022^{***}					-4.919^{*}
	(2.353)					(2.506)
log(IT capital per employee)		-0.109				
		(0.069)				
IT capital/Value-added			-0.072			
			(0.388)			
log(Regulations)				-0.254^{*}		-0.230
				(0.144)		(0.144)
Regulatory employment share					-2.587	
					(1.616)	
Average age	0.018	0.054	0.041	0.004	0.037	-0.027
	(0.044)	(0.042)	(0.041)	(0.048)	(0.042)	(0.052)
College share	0.958	0.023	-0.125	-0.704	0.082	0.309
	(1.126)	(1.051)	(1.066)	(1.230)	(1.091)	(1.410)
Male share	0.426	-0.054	0.231	0.345	-0.037	0.653
	(1.004)	(0.997)	(0.996)	(1.086)	(1.022)	(1.120)
Metropolitan share	2.223	1.534	1.282	1.700	1.560	2.944^{**}
	(1.149)	(1.018)	(1.014)	(1.084)	(1.080)	(1.295)
Constant	-0.101	-0.063	-0.098	0.032	-0.097	0.028
	(0.082)	(0.081)	(0.079)	(0.093)	(0.080)	(0.100)
Observations	139	144	144	102	140	98
R^2	0.083	0.035	0.018	0.063	0.037	0.119
Adjusted R^2	0.048	0.001	-0.018	0.014	0.002	0.061

Table F8: **Relationship between changes in the self-employment share, and changes in IT technology and regulations.** The regression is specified in equation (F8). The unit of observation is industry-time. Observations are for three time periods: 1988–89 to 1999/2000, 1999/2000 to 2005/06, and 2005/06 to 2014/15. Variables are averaged over the two years at the start and end of each period.*IT employment share* is the share of employees in an industry in IT related occupations. *IT capital per employee* is the real value of the IT capital stock (2012 dollars) in an industry divided by the number of workers. *IT capital/Value-added* is the nominal value of the IT capital stock divided by nominal value-added. *Regulations* is a measure of the number of regulations from RegData. *Regulatory employment share* is the share of workers in an industry who are regulation-related occupations. *College share*, *male share* and *metropolitan share* and the shares of workers who have a college degree, are male, and live in a metropolitan area, respectively. *Average age* is the average age of workers (employees and the self-employed). Standard errors are in parentheses. *** and ** denote statistically significant differences from 0 at the 1% and 5% levels, respectively.

constructed by multiplying the relevance of each part by the number of restrictions in it, and then summing over parts.

IT and regulation-related occupations Table F9 lists the occupations that are treated as regulation-related and IT-related. The occupation codes are from the 1990 Census Bureau Occupational Classification System.

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.

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.

G Additional results for policy analysis

Figure G11 shows the flows in and out of entrepreneurship that result from the entry costs subsidy, as referred to in Section 7 of the main text. This figure shows that approximately half of the flow into entrepreneurship is coming from people who would be out of the labor force in the absence of the subsidy, while the remaining half comes in equal shares from the low and high-skill occupations. There is also a small outflow of marginal entrepreneurs who switch to being employees as a result of modestly higher wages.

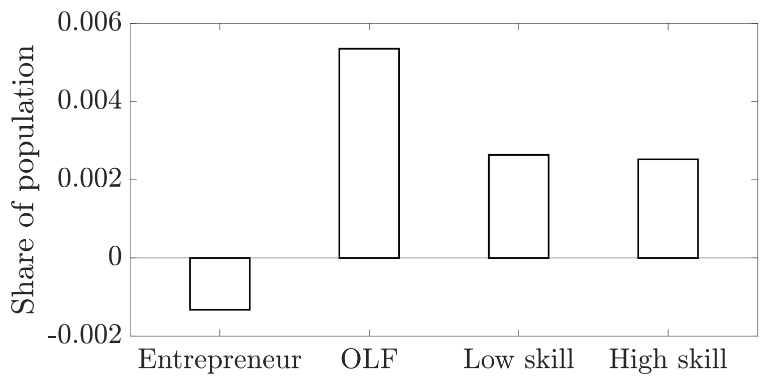


Figure G11: **Flows in and out of entrepreneurship due to entry subsidy, by initial occupation.** The horizontal axis is the 2015 occupations of agents. The vertical axis is the mass of agents (as a share of the population) flowing into or out of entrepreneurship from each occupation. A positive value is an inflow and a negative value an outflow.

References

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

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter Klenow, and Huiyu Li (2023), "A theory of falling growth and rising rents." *Review of Economic Studies*, 90, 2675–2702.

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

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.

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 (2024), "Market power and innovation in the intangible economy." *American Economic Review*, 114, 199–251.

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

Hopenhayn, Hugo, Julian Neira, and Rish Singhania (2022), "From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share." *Econometrica*, 90, 1879–1914.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg (2023), "The industrial revolution in services." *Journal of Political Economy Macroeconomics*, 1, 3–42.

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.

Karahan, Fatih, Benjamin Pugsley, and Aysegul Sahin (2024), "Demographic origins of the startup deficit." *American Economic Review*, Forthcoming.

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

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 (2020), "Labor market polarization, the decline of routine work, and technological change: A quantitative analysis." *Journal of Monetary Economics*, 110, 62–80.