Under the Hood of Routine Share Decline*

Nir Jaimovich
University of California, San Diego and *NBER, CEPR*
nijaimovich@ucsd.edu

Miao Ben Zhang
University of Southern California
Miao.Zhang@marshall.usc.edu

Nicolas Vincent
HEC Montréal
nicolas.vincent@hec.ca

February 2023

Abstract

Using establishments' occupational data, we quantify the role of entrants, exiters, and incumbents in driving the decline in the share of routine occupations (R-share) in the U.S. First, entrants have a higher R-share than incumbents, casting doubt on a "creative destruction" mechanism whereby entrants drive this decline. Second, exiters have a higher R-share than their peers, supporting a "positive selection" mechanism. Finally, as incumbents age, they experience a fall in their R-share, which is not due to their size, consistent with the "technology-adoption" mechanism. Quantitatively, we show that incumbents are the primary drivers of the aggregate decline in R-share.

Keywords: Routine occupations, Establishments' dynamics.

*Vincent gratefully acknowledges financial support from the Fondation HEC Montréal and the Social Sciences and Humanities Research Council of Canada (grant number: 435 − 2022 − 0800). Jaimovich gratefully acknowledges financial support from the Swiss National Science Foundation (grant number: 100018 − 197799).

1 Introduction

The decline in the share of the population employed in routine jobs, those occupations focused on a limited set of tasks that can be performed by following a well-defined set of instructions and procedures, has been at the center of academic and policy discussions during the last two decades (see, for instance, Autor et al. (2006), Goos and Manning (2007), Goos et al. (2009), and Acemoglu and Autor (2011)).

Yet, evidence about the evolution of the share of routine occupations (*R-share* hereafter) at the establishment level is scant. Using administrative microdata of establishments' occupational employment, we address this gap. Our key contribution is to quantify the role of entrants, exiters, and incumbents in driving the R-share decline. We view our findings as instructive for future researchers who study the R-share decline phenomenon and those interested in learning how establishments adjust their employment composition over time. In what follows, we discuss our findings in detail.

First, over time, we document that entrants and exiters exhibit a fall in their R-share; that is, in *time series*, new cohorts of entrants (exiters) are characterized by a lower R-share than previous entrants' (exiters') cohorts. Similarly, incumbents tend to reduce their R-share.

Second, perhaps surprisingly, we find that in the *cross section* entrants have a higher R-share than incumbents. This holds beyond entry: even seven years after birth, entrants do not feature lower R-share than incumbents. Hence, entrants do not appear to contribute to the overall decline in R-Share via a "creative destruction" process.

Third, exiters exhibit selection patterns whereby establishments that exit tend to have a higher R-share than surviving incumbents, even many years before exiting. This evidence is supportive of a positive selection mechanism.

Taken together, the second and third findings imply that ceteris paribus, entrants tend to increase the R-share in the economy, while exiters tend to decrease it; we find that overall these two channels (i.e., the "net entry" margin) roughly cancel each other, and if anything, the net-entry margin contributes to an *increase* in the R-share.

These results suggest that incumbents are central to the adjustment of the aggregate R-share over time. Indeed, our fourth finding is that the R-share exhibits a significant fall *within* incumbent establishments. Importantly, this fact holds (i) across different establishments cohorts and (ii) after controlling for the size of the establishments.

The crucial role of incumbents is confirmed by our quantitative decomposition: in our fifth and final finding, we show that the fall in the within incumbents' R-share is the driving force for the overall R-share decline.

The paper is organized as follows. In Section 2, we describe our data set. Section 3 discusses our empirical findings regarding the evolution of the R-share within incumbents, entrants, and exiters. Section 4 concludes.

2 Data and measurement

2.1 Data

Our administrative microdata from the OEWS covers BLS surveys that track occupational-level employment and wage rates in approximately 1.2 million establishments stratified to represent the economy from 1988 to 2013.[1] Each establishment is surveyed every three years, resulting in about 400,000 establishments surveyed every year. For each establishment, we have information about the number of employees and average hourly wage rate per employee in each occupation, as well as the establishment's unique identifier, sampling weight, employer identification number (EIN), government ownership, county code, and industry code.

Our main sample covers an unbalanced panel of establishments from 2002 to 2013. This sample period allows us to have consistent occupation classifications and industry classifications across the years. We obtain an establishment's entry and exit years from the Quarterly Census of Employment and Wages (QCEW) microdata in the BLS, which tracks the *universe* of establishments from 1990 to 2013 in 40 of the 50 U.S. states.[2] We compute each establishment's first and last years in the QCEW universe and merge the information with our OEWS sample using the BLS's unique establishment identifier. Our final sample thus includes 24.2 million establishment-occupation-year observations for non-government-owned establishments for which QCEW information is available, covering about 257,000 establishments each year, from 2002 to 2013. In our econometric analysis, we weigh each establishment using the product of the establishment's total employment and the BLS sampling weight.

2.2 Measures

We measure each establishment's share of routine-task labor as follows: occupations are regarded as **routine (R)** if they are "sales and related occupations," "office and administrative support occupations," "production occupations," "transportation and material moving occupations," "construction and extraction occupations," and "installation, maintenance, and repair occupations."[3]

Our main variable is an establishment's share of total employment in routine-task labor **(R-share)**:

[1] OEWS represents the Occupational Employment and Wage Statistics program at the Bureau of Labor Statistics (BLS). Our accessed microdata ends in 2013. See more details of the data in Online Appendix A.1 and also in Zhang (2019) and Tuzel and Zhang (2021).

[2] The ten states for which we do not have the QCEW data include Florida, Kentucky, Massachusetts, Michigan, Mississippi, Missouri, New Hampshire, New York, Pennsylvania, and Virginia.

[3] We follow the definitions of Jaimovich and Siu (2020). In this definition Non-routine occupations are composed of **non-routine cognitive (C)** if they are "management, business, and financial operations occupations" and "professional and related occupations," and **non-routine manual (M)** if they are "service occupations." See more details in Online Appendix A.2.

$$\text{R-share}_{i,t} = \frac{\sum_o 1[o \in R] * emp_{o,i,t}}{\sum_o emp_{o,i,t}}. \tag{1}$$

where o, i and t respectively refer to an occupation, establishment, and year.

The establishment-level R-share has a simple average of 0.62 and a weighted average of 0.56 in our sample. These values are consistent with estimates from prior literature using the Census individual-level data.

3 The evolution of the routine share over time

In this section, we address the central question of this paper: how has the R-share evolved over time within establishments? We begin by regressing establishment-level R-share on year dummies, while controlling for NAICS3 fixed effects for establishments in each age group from 2002 to 2013, where 2002 serves as the benchmark year:

$$\text{R-share}_{i,t} = \sum_{t=2003}^{2013} \beta_t \times Year_t + FE_{NAICS3} + \epsilon_{i,t}, \tag{2}$$

All observations are weighted by the product of the establishment's total employment and the BLS sampling weight. We report robust standard errors.

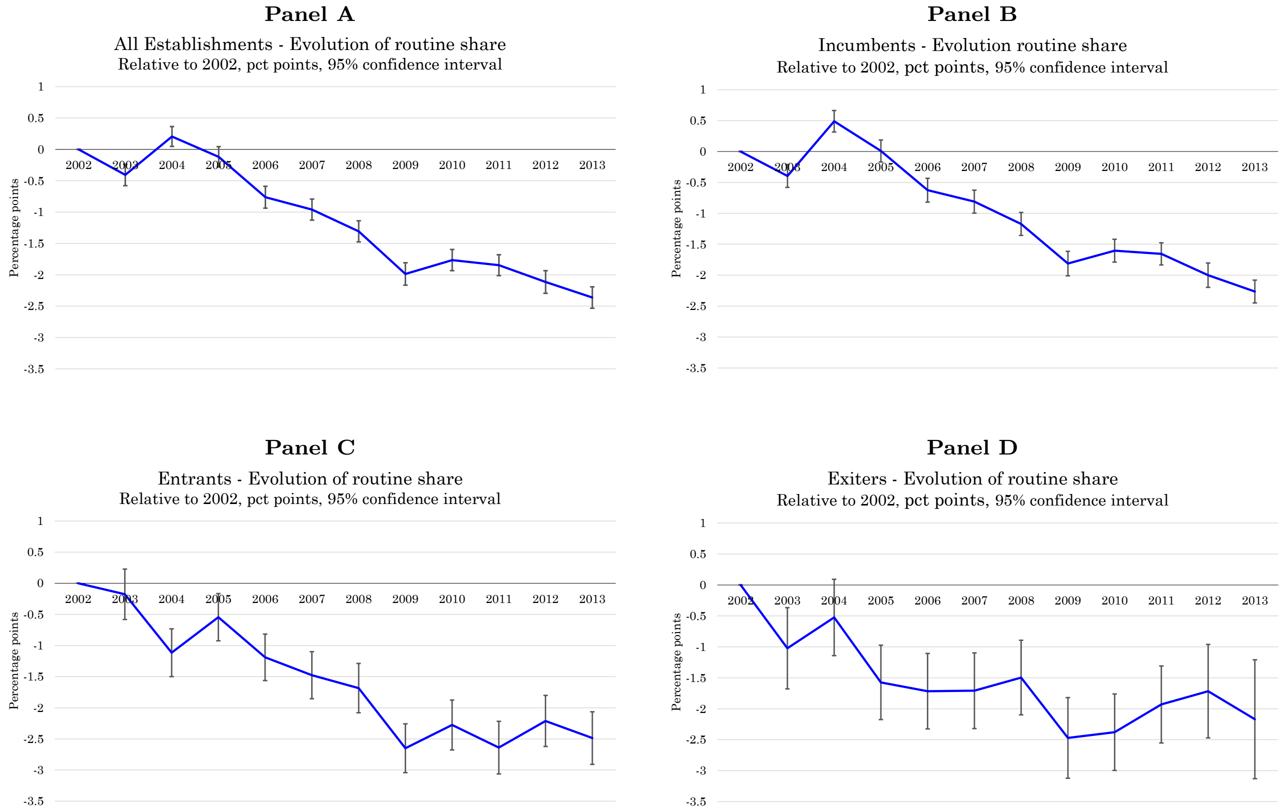
Panels A to D in Figure 1 depict the year-specific coefficients β_t for each establishment type.[4] Specifically, Panel A depicts the evolution of the coefficient β_t when estimated across all establishments, relative to the 2002 baseline. The figure identifies a clear downward trend in the coefficient values over time, becoming particularly pronounced around 2006. Overall, our estimates indicate a fall in R-share of slightly more than 2 percentage points over the decade under study.

Panels B to D each focuses on a type of establishment by depicting the evolution of the β_t coefficient from running the regression described in equation (2) for each subsample of incumbents, entrants, and exiters. These figures look strikingly similar: the decline in R-share is statistically significant for all three establishment types and very comparable in size.

In sum, Figure 1 shows that the average R-share of U.S. establishments fell over time. Yet, by itself, this decline can be the result of three non-exclusive processes. First, incumbents could exhibit a fall in their R-share as they age and adopt routine-biased technologies. Second, it could be due to a positive selection process that sees a disproportionately high survival rate among low-R-share establishments. Finally, it could happen if entrants, who later on to become incumbents, are "disruptors" in the sense that they enter with a lower R-share than currently existing incumbents. The rest of our analysis explores these channels in turn, starting with the specific role played by incumbents.

[4] The point estimator of β_t is presented relative to the benchmark year 2002 with a 95% confidence interval. See Table IA.1 in the Online Appendix for the detailed regression results.

Figure 1: **Evolution of Routine Share**



Notes: Panels A-D plot the evolution of establishment R-share for all establishments, all continuing incumbent establishments, all entrants, and all exiters, respectively, while controlling for NAIC3 industry fixed effects (see equation (2)). The vertical bars represent the robust standard errors of the point estimates. Online Appendix Table IA.1 presents details of these estimates.

3.1 The evolution of the routine share of incumbents

To investigate the specific role played by incumbents, we start by reporting in Table 1 the results of regressing an establishment's R-share on its age and establishment fixed effects. Specifically, we run the following regression:

$$\text{R-share}_{i,t} = \sum_j \gamma_j AgeGroup_j + FE_{Est} + \epsilon_{i,t}, \tag{3}$$

where *AgeGroup* specifies seven 3-year age bins: [3-5], ..., [18-20], [21-22]. Hence, the γ_j coefficients capture an establishment's R-share as it ages, with the [0-2] age group serving as the baseline. Adding these fixed effects (recall that our regression in (2) only included NAICS3 fixed effects) allows us to disentangle how much of the R-share evolution comes from an establishment's life cycle (age effect) relative to time (year effect).

The first column of Table 1 clearly shows that *within* surviving establishments, the R-share falls as the typical establishment ages: relative to the level right after birth (0 and 2 years old), a typical establishment has an R-share that is 2.4 ppt lower once it is in the [12-14] age group and 4 ppts lower once it reaches the age of 20. Those differences are all significant at the 1% level.

Table 1: **Evolution of Routine Share within Incumbents Over Time**

	All	1990 Cohort	1995 Cohort	2000 Cohort	2005 Cohort	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age[3-5]	-0.623***			-0.692***	-0.539***	-0.623***	-0.625***
	(0.105)			(0.166)	(0.147)	(0.105)	(0.105)
Age[6-8]	-1.235***		-0.544**	-1.343***	-1.135***	-1.234***	-1.237***
	(0.123)		(0.229)	(0.173)	(0.225)	(0.124)	(0.123)
Age[9-11]	-1.914***	-0.826	-1.333***	-1.861***		-1.912***	-1.917***
	(0.142)	(0.809)	(0.229)	(0.195)		(0.143)	(0.142)
Age[12-14]	-2.414***	-1.199	-1.806***	-2.797***		-2.413***	-2.419***
	(0.163)	(0.806)	(0.237)	(0.362)		(0.163)	(0.163)
Age[15-17]	-2.975***	-1.926**	-2.183***			-2.974***	-2.978***
	(0.184)	(0.828)	(0.266)			(0.184)	(0.184)
Age[18-20]	-3.658***	-2.604***	-2.029***			-3.657***	-3.660***
	(0.211)	(0.815)	(0.582)			(0.211)	(0.211)
Age[21-22]	-3.998***	-2.897***				-3.997***	-4.001***
	(0.310)	(0.850)				(0.310)	(0.310)
Log(Emp)						-0.015	0.064
						(0.110)	(0.058)
N	1,280,804	292,775	363,542	402,364	212,605	1,280,804	1,280,804
R^2	0.91	0.91	0.91	0.90	0.92	0.91	0.91

Notes: This table reports the results of regressing establishment routine share on its age with establishment fixed effects (see equation (3)). The t *Cohort* represents the sample of establishments born between t and $t + 4$. The benchmark age for all columns is Age[0-2], except for Columns (2) and (3) where the benchmark is Age[6-8] and Age[3-5], respectively. Columns (6) and (7) further control the natural logarithm of establishment employment and firm employment, respectively, where a firm is defined by the employer identification number (EIN). The sample period is from 2002 to 2013. All regressions are weighted by the product of the establishment's total employment and the BLS sampling weight. Robust standard errors are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

Arguably, one may be worried that these results are driven by a specific *cohort* of incumbents. To assuage this concern, we report the results for four separate cohorts of establishments in Columns (2)-(5). For example, the 1990 cohort in column (2) represents establishments born between 1990 to 1994, and which are therefore between 8 and 22 years old during our 2002-2013 sample period. For this cohort, we thus use age group [6-8] as the benchmark for estimating γ_js. We notice a strongly significant decline in R-share of almost 3 ppt over their fifteen years of life. A similar relationship with age is noticed for the 1995 (baseline is [3-5] age group), 2000, and 2005 (baseline is [0-2]) cohorts.

3.1.1 Time vs. size

It is well known that, on average, surviving establishments grow as they age. Hence, one potential confounding factor behind the result of Table 1 is that it is the growth of these establishments over time that accounts for the fall in their R-share, and not time by itself. We thus augment the regression in equation (3) by including an establishment's size as a control in an effort to discriminate between the age and size factors.

Comparing Column (6) to Column (1) in Table 1, we see that the age coefficients are unchanged after controlling for establishment size. In Column (7), we instead control for the *parent firm's* size (where a firm is measured at the EIN level following Song et al. (2019)), which addresses concerns that changes in the employment composition result from changes in firm-level size and not the particular establishment's size. Again, comparing the results from Column (7) to Column (1), we see no change in the age coefficients. Hence, the evidence suggests that it is indeed the life-cycle dimension of incumbents and not their growing size that is behind the fall in their R-share.

3.2 Entrants and exiters: routine share relative to incumbents

Panel C of Figure 1 indicated that *over time*, new establishments enter with a lower R-share. Yet, this does not necessarily imply that creative destruction plays a role in lowering the aggregate R-share; that is, by itself this finding does not imply that the R-share of entrants is lower than that of incumbents and hence driving the overall decline in the R-share.

To address this question, we investigate how the R-share of entrants at the time of entry compares to that of incumbent establishments. To do so, we use the following regression specification:

$$\text{R-share}_{i,t} = \theta_1 Entrant_{i,t} + FE_{NAICS3\times Year} + \epsilon_{i,t}, \tag{4}$$

where $Entrant_{i,t}$ is a dummy variables that is equal to 1 if establishment i is an entrant in year t. Industry interacted with time fixed effects are included.

Column (1) of Table 2 reports the coefficient θ_1 when we only include the entrant dummy. The result shows that, on average, entrants are characterized by a *higher* R-share than their incumbent industry peers, a difference of 0.34 ppt. These findings cast doubt on the role of entrants in the context of a "creative destruction" theory of routine-biased technological change, whereas entrants would be more likely to adopt newer technologies (and have lower R-share) than incumbents.

We also repeat the same analysis for exiters , i.e. we run the regression

$$\text{R-share}_{i,t} = \theta_2 Exiter_{i,t} + FE_{NAICS3\times Year} + \epsilon_{i,t}, \tag{5}$$

Maybe less surprisingly, we show in Column (2) that in their last year of existence, the R-share of the average exiter was 0.44 ppt higher than that of its peers. The results are very similar once both dummies are included in the same regression, with coefficients on the entrant and exiter dummies of 0.31 and 0.41 respectively, see Column (3). All coefficients are statistically significant at the 1% level.[5]

Table 2: **Routine Share Comparison: Entrants vs. Incumbents in Cross-Section**

	All Counties			Counties with Low Small Bank Share	Counties with High Small Bank Share
	(1)	(2)	(3)	(4)	(5)
Entrant	0.335***		0.310***	0.293***	0.161
	(0.083)		(0.083)	(0.089)	(0.189)
Exiter		0.439***	0.412***	0.412***	0.622**
		(0.127)	(0.128)	(0.136)	(0.267)
N	3,032,548	3,010,740	3,010,740	2,574,997	435,737
R^2	0.66	0.66	0.66	0.65	0.73

Notes: This table reports the results of regressing establishment routine share on an entrant dummy and an exiter dummy with NAICS3-Year fixed effects (see equation (4)). Columns (4) and (5) report results for establishments in U.S. counties with small bank shares below and above the median in the year, respectively. See Section 3.2.1 for definitions of counties' small bank shares. The sample period is from 2002 to 2013. All regressions are weighted by the product of the establishment's total employment and the BLS sampling weight. Robust standard errors are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

3.2.1 Entrants and financial frictions

Can external financing frictions contribute to entrants not having an R-share as low as existing incumbents? Our motivation to investigate this question stems from the fact that young entrant businesses have been shown to face greater challenges in obtaining external financing due to, for example, a lack of credit history.

Our measure of financial frictions is based on Chen et al. (2017) who show that in counties where the largest banks had a high market share, the aggregate flow of small business credit fell from 2008 to 2014, and Berger et al. (2017) who show that young small businesses rely heavily on relationship lending from small banks due to their superiority in processing soft, qualitative information (rather than hard information such as financial ratios from audited statements).[6] We thus adopt the measure of county-level *small bank share* from Berger et al. (2017) and Tuzel and Zhang (2021), and split our

[5]In Online Appendix Table IA.2, we conduct further robustness checks by restricting the sample to meaningful entrants, i.e., entrants with at least 20 employees, and by comparing entrants with incumbents with similar sizes by imposing NAICS3-Year-EmpBin fixed effects. We find very similar results, suggesting that our results are not driven by size differences between entrants and incumbents.

[6]Stein (2002) argues that small banks are superior at using soft information because such information is easier to communicate within a small organization with fewer layers of management.

sample into those establishments from counties with small bank shares below and above the median in the year.[7]

With this categorization at hand, we rerun the regression in equation (4) for establishments that reside in "constrained" locations (i.e., counties with a low small bank share) vs. those in "unconstrained" locations (i.e., counties with a high small bank share). The results, which are reported in Columns (4)-(5) of Table 2, reveal that entrants in unconstrained areas cannot be distinguished from incumbents in their R-share. In contrast, those entrants in constrained locations have a statistically significant higher R-share than incumbents in the same area. This finding is consistent with the view that some establishments face financial barriers in the adoption of the most advanced technology.

3.2.2 The routine share dynamics of entrants and exiters

So far, we have seen that both entrants and exiters have a higher R-share than their incumbent peers at the time of birth and death. Next, we turn our attention to their respective dynamics following entry and before exit.

We first focus on the entry dimension and ask a simple question: could the disrupting role of entrants come with a delay? More specifically, does the R-share of entrants fall rapidly in the years following birth, even if the relative R-share of entrants is initially high? Does their R-share eventually fall below that of incumbents?

To answer these questions, for each entering establishment in period t, we track its R-share in $t+1, ..., t+7$ relative to the existing incumbent peers in that specific year by estimating the following regression:

$$\text{R-share}_{i,t} = \lambda_\tau^E Entrant(-\tau)_{i,t} + FE_{NAICS3-Year} + \epsilon_{i,t}, \tag{6}$$

where $Entrant(-\tau)_{i,t}$ is a dummy variable that equals to one if establishment i was an entrant τ years ago, and zero if the establishment was an existing incumbent τ years ago. We exclude establishments younger than τ in this analysis. We run this regression for each τ from 0 to 7. As usual, observations are weighted by the product of the establishment's total employment and the BLS sampling weight.

Panel A of Figure 2 depicts λ_τ^E. The figure shows that the R-share of entrants becomes statistically indistinguishable from that of their older peers after about two years. More generally, we find no evidence that in the seven years following entry, new establishments on average ever display a lower

[7] Following Berger et al. (2017), we define a county's small bank share as the county's deposit share from banks with total assets below \$1 billion dollars based on information from bank Call Reports and Summary of Deposits. The median of the small bank share across counties in each year averages 56% from 2002 to 2013. Counties with a higher small bank share have a lower number of establishments, resulting in different numbers of observations in the two subsamples. Online Appendix Table IA.3 shows that the results are robust to alternative splits of the subsamples; specifically, we also experimented with cutoffs below and above the median, e.g. where the share of small banks is 50%, 55%, and 60%, and got similar results.

R-share than other establishments. This appears to confirm that the entry margin does not contribute to the overall decline in R-share in the economy.

Note that there are multiple potential explanations behind the convergence in R-share of the average entrant. First, it could be due to entrants quickly reducing their R-share over the life cycle, consistent with the within-firm R-share analyses in Section 3.1. Alternatively, it may also be the product of high-R-share entrants quickly exiting the economy, leading to a selection effect. To address the second hypothesis, we repeat the exercise focusing only on entrants that never exit later in the sample. The results can be found in Figure IA.1 of the Online Appendix: we also find for this sample of "surviving entrants" a convergence of the R-share towards that of their incumbent peers, though only after 5 years.

Next, we turn our attention to the dynamics of establishments prior to exit. Are these exiters negatively selected ahead of time? Or are they simply subject to a shock in the last year that forces them to exit? For each exiting establishment in period t, we go back in time and calculate its R-share in $t-7, ..., t-1$ relative to incumbents that survive at t in that specific year by estimating the following regression:

$$\text{R-share}_{i,t} = \lambda_\tau^X Exiter(\tau)_{i,t} + FE_{NAICS3-Year} + \epsilon_{i,t}, \tag{7}$$

where $Exiter(\tau)_{i,t}$ is a dummy variable that is equal to one if establishment i will be an exiter in τ years, and zero if the establishment will survive beyond τ years. We exclude establishments that exit in less than τ years in this analysis. We run this regression for each τ from 1 to 7 and rely on the usual weighting procedure.

Panel B of Figure 2 depicts the coefficients λ_τ^X, which capture the relative R-share of future exiters in the years before they exit. The figure demonstrates that exiters have a significantly higher R-share relative to their incumbent peers many years before their eventual death. This evidence suggests that exiters fall behind their surviving peers in terms of the R-share evolution years before their exit, and are not simply the victims of an exit-inducing shock.

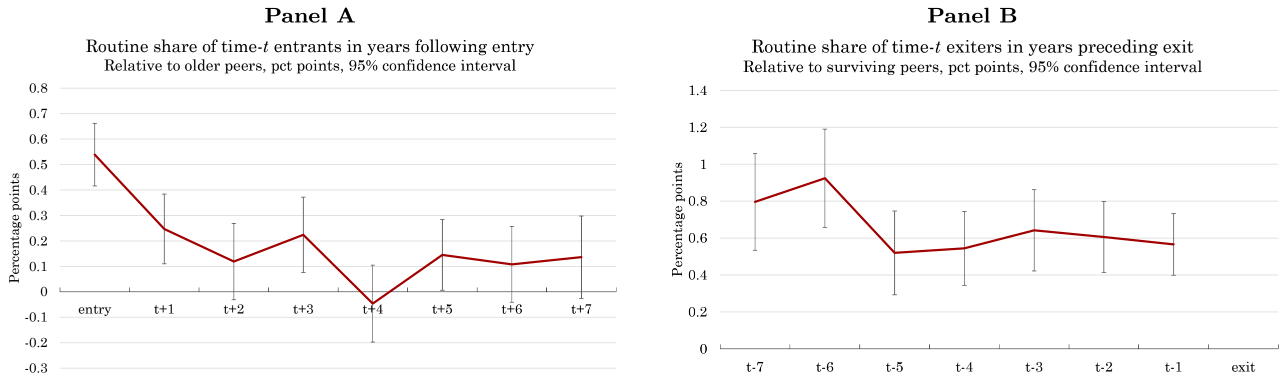
3.3 A decomposition of the evolution of the routine share

In the last section, we further highlight and quantify the role of the different types of establishments in shaping the evolution of the R-share. First, we express the R-share of a given industry at a given point in time, S_t, as

$$S_t = \sum_{i \in I} \omega_{i,t} S_{i,t}, \tag{8}$$

where i is an establishment in a given period t; $\omega_{i,t}$ denotes the establishment's weight in the industry; and $S_{i,t}$ denotes the establishment's R-share. It then follows that the n-year change in an industry's R-share can be decomposed as

Figure 2: **Routine Share Dynamics of Entrants and Exiters**



Notes: Panel A plots entrants' R-share relative to their existing incumbent peers in seven years following the entry (see equation (6)). Panel B plots exiters' R-share relative to their surviving incumbent peers in seven years preceding the exit (see equation (7)). The vertical bars represent the robust standard errors of the point estimates.

$$\Delta S_t = S_t - S_{t-n} = \underbrace{\sum_{i\in I} \widetilde{\omega}_{i,t-n} \Delta S_{i,t}}_{\text{Within}} + \underbrace{\sum_{i\in I} \Delta\widetilde{\omega}_{i,t} S_{i,t-1}}_{\text{Chg. Weight}} + \underbrace{\sum_{i\in I} \Delta\widetilde{\omega}_{i,t}\Delta S_{i,t}}_{\text{Cross Term}}$$
$$+ \underbrace{\sum_{i\in E} \omega_{i,t}\left(S_{i,t} - S_t^I\right)}_{\text{Entry}} - \underbrace{\sum_{i\in X} \omega_{i,t-n}\left(S_{i,t-n} - S_{t-n}^I\right)}_{\text{Exit}}, \qquad (9)$$

where E and X denote entrants and exiters respectively, and S_t^I is the industry's R-share defined based on equation (8) but using only establishments that exist in both year $t-n$ and t (i.e., non-entrants and non-exiters from $t-n$ to t). The first three terms in equation (9) quantify the shift-share decomposition of the R-share change within the continuing incumbent establishments in the industry, while the last two components quantify the impact of entry and exit on the industry's R-share change.

We decompose each NAICS3 industry's change in R-share from the beginning to the end of our sample period following equation (9).[8] We then aggregate the NAICS3-level ΔS_t to more aggregate levels using the industry's initial employment at the beginning period of our sample as weight.

Table 3 reports the results of this decomposition in the aggregate economy and within the manufacturing and non-manufacturing sectors separately. We observe that the bulk of the adjustment in the fall of the R-Share over time is due to the first "within" term. That is, holding constant the size

[8]Note that the OEWS program surveys each establishment every 3 years. To more precisely represent each industry, we follow the aggregation method from the OEWS and treat the observations from 2002-2004 as the beginning period sample, and the observations from 2011-2013 as the end period sample for this estimation. This aggregation approach is suggested by the OEWS and is also used by the OEWS to produce yearly industry-level statistics for public use at http://www.bls.gov/oes/tables.htm.

of incumbents, the main driving factor behind the fall in the R-share is the change in the R-share within incumbents, consistent with the evidence presented earlier. The second, "Changing Weight" term also plays an important role: it indicates that those incumbents with an initially lower R-share relative to the average incumbent grew more than their incumbent counterparts. This suggests that a selection margin is also at play among incumbents. Finally, echoing the fact that entrants are, on average, characterized by a higher R-share than incumbents, the results in the table confirm that the entry margin has *slowed down* the fall of the R-share in the economy. Table 3 indicates that these patterns are similar in the manufacturing and non-manufacturing sectors and are not solely driven by one industry.

Table 3: **Decomposition of Routine Share Change**

Industry	Total	Within	Chg.Weight	Cross-Term	Net Entry	Net Entry	
						Entry	Exit
All	-1.99	-1.74	-1.03	0.10	0.68	0.69	0.01
Manuf	-2.39	-2.48	-0.30	0.34	0.04	0.49	0.45
Non-Manuf	-1.93	-1.63	-1.15	0.07	0.78	0.72	-0.06

Notes: This table reports the results of decomposing changes in industry R-share from the beginning of our sample period to the end of our sample period following equation (9). We report results for all NAICS3 industries, industries in the manufacturing sector, and industries in the non-manufacturing sector. We use each industry's initial employment at the beginning period of our sample as weights to produce the broader level estimates.

4 Conclusions

Our findings regarding the evolution of the R-share can be summarized as follows. First, over time, all types of establishments (incumbents, entrants, exiters) exhibit a reduction in their routine employment share. Second, incumbents show a decline in the R-share as they age; that is, there is an adjustment in the R-share within surviving incumbents throughout their life cycle. Third, in the cross-section, entrants have a higher R-share than their incumbent peers, and we find evidence that is consistent with financial constraints preventing entrants from entering with an R-share as low as existing incumbents. Fourth, exiters have a significantly higher R-share than their surviving peers up to many years before they exit. Fifth, the driving factor for the R-share fall is a change within incumbents.

References

Acemoglu, Daron and David H. Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, 2011, *4*, 1043–1171.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "The Polarization of the U.S. Labor Market," *American Economic Review*, May 2006, *96* (2), 189–194.

Berger, Allen N., Christa H. S. Bouwman, and Dasol Kim, "Small Bank Comparative Advantages in Alleviating Financial Constraints and Providing Liquidity Insurance over Time," *Review of Financial Studies*, October 2017, *30* (10), 3416–3454.

Chen, Brian S., Samuel G. Hanson, and Jeremy C. Stein, "The Decline of Big-Bank Lending to Small Business: Dynamic Impacts on Local Credit and Labor Markets," Technical Report, National Bureau of Economic Research 2017.

Goos, Maarten, Alan Manning, and Anna Salomons, "Job Polarization in Europe," *American Economic Review*, May 2009, *99* (2), 58–63.

Goos, Maarten and Alan Manning, "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, February 2007, *89* (1), 118–133.

Jaimovich, Nir and Henry E. Siu, "Job Polarization and Jobless Recoveries," *Review of Economics and Statistics*, March 2020, *102* (1), 129–147.

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

Stein, Jeremy C., "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms," *Journal of Finance*, 2002, *57* (5), 1891–1921.

Tuzel, Selale and Miao Ben Zhang, "Economic Stimulus at the Expense of Routine-Task Jobs," *Journal of Finance*, 2021, *76* (6), 3347–3399.

Zhang, Miao Ben, "Labor-Technology Substitution: Implications for Asset Pricing," *Journal of Finance*, 2019, *74* (4), 1793–1839.

Online Appendix for

Under the Hood of Routine Share Decline

Nir Jaimovich, Miao Ben Zhang, and Nicolas Vincent

A Details on Data and Measures

A.1 BLS Microdata

We use two sets of confidential microdata from the Bureau of Labor Statistics (BLS) in this study. The first one is the BLS Occupational Employment and Wage Statistics (OEWS) microdata, which is a stratified sample of about 1.2 million establishments from a universe of approximately 6.8 million non-farm establishments from the Quarterly Census of Employment and Wages (QCEW) database.[9] The QCEW database is provided to the BLS by state workforce agencies that collect unemployment insurance (UI) reports from employers. Employers are required by law to file these reports to the state where each establishment is located. The establishments in the OEWS sample frame are stratified by establishments' industry, geography, and size, and each establishment is given a sampling weight by the OEWS.

Out of the 1.2 million establishments in the sample frame, the OEWS program surveys about 400,000 establishments each year, with one establishment surveyed in every 3 years to reduce the response burden. That is, an establishment in the sample framework is surveyed in years $t-3$ and t but not in between. From each OEWS survey, we obtain the establishment's number of workers in each occupation and wage bin, whereas occupation is defined based on over 800 categories, and wage bin is specified by the BLS based on about 12 bins of hourly wages.[10] From the survey results, we obtain an establishment's total number of employees in each occupation.

The OEWS survey maintained a consistent industry classification (based on the NAICS) and occupation classification (based on the SOC) after 2002. The BLS surveys about 200,000 establishments in May and November of each year and reports aggregate statistics in May of each year by weighting establishments in the previous six surveys. Because the establishment weights are assigned in May of each year, we thus regard survey results from November of year t and May of year $t+1$ as the

[9] BLS uses this microdata to produce aggregate occupation statistics at https://www.bls.gov/oes/tables.htm. The employees in the covered establishments represent 62% of the U.S. The survey covers all industries except for agricultural workers, private households, and unincorporated self-employed workers without employees.

[10] See a recent form of the OEWS survey at https://www.bls.gov/respondents/oes/pdf/forms/uuuuuu_fillable.pdf. See more details of the survey methods and statements for each year from the BLS documentation archives at https://www.bls.gov/oes/oes_doc_arch.htm

observations for the year ends at t. For instance, our last year 2013's sample includes surveys done in November 2013 and May 2014, where May 2014 is the last period of the microdata we accessed. Our OEWS sample thus includes establishments from 2002 to 2013. For each establishment, we have information about the number of employees and average hourly wage rate per employee in each occupation, as well as the establishment's unique identifier, sampling weight, employer identification number (EIN), government ownership, county code, and industry code.

The second microdata we accessed is the establishment identifiers of the QCEW universe for all but ten states from 1990 to 2014.[11] This microdata is helpful for us to identify the entry and exit years of each establishment in our OEWS sample. Specifically, we define the first year of the establishment in the QCEW universe as the entry year of the establishment and the last year in the QCEW universe as the exit year. For establishments that exist in the QCEW universe before 1990, we do not know their precious entry year, and we treat them as incumbents throughout our sample period.

Merging the QCEW and the OEWS microdata results in our final sample of 24.2 million establishment-occupation-year observations for non-government-owned establishments with the QCEW information (i.e., from states outside the ten states), covering about 257,000 establishments each year from 2002 to 2013. In all of our analyses, we weigh each establishment using the product of the establishment's total employment and the BLS sampling weight.

A.2 Measuring Routine Occupations

We follow the definition of Jaimovich and Siu (2020) and categorize occupations into three categories. In particular, occupations are regarded as **routine (R)** if they are "sales and related occupations (SOC2=41)," "office and administrative support occupations (SOC2=42)," "production occupations (SOC2=51)," "transportation and material moving occupations (SOC2=53)," "construction and extraction occupations (SOC2=47)," and "installation, maintenance, and repair occupations (SOC2=49)." Non-routine occupations are further divided into **non-routine cognitive (C)** if they are "management, business, and financial operations occupations (SOC2=11)" and "professional and related occupations (from SOC2=13 to SOC2=29)," and **non-routine manual (M)** if they are "service occupations (from SOC2=31 to SOC2=39)."

A.3 Measuring County-Level Small Bank Share

Berger et al. (2017) show that young small businesses rely heavily on relationship lending from small banks due to their superiority in processing soft, qualitative information (rather than hard information

[11]The ten states that we do not have the QCEW data include Florida, Kentucky, Massachusetts, Michigan, Mississippi, Missouri, New Hampshire, New York, Pennsylvania, and Virginia.

such as financial ratios from audited statements).[12] We adopt the measure of county-level small bank share from Berger et al. (2017) to proxy for the availability of external finance to new businesses.

We use quarterly bank Call Reports to obtain the asset level of banks and the Summary of Deposits data to obtain the deposits and county location of banks' branches. We define a bank as a *small bank* if its total assets are below \$1 billion. Then, we compute the small bank share of a county as the deposit share of small banks in the county.[13]

We split our full sample into establishments from counties with a small bank share above or below the cross-sectional median (which averages 56% from 2002 to 2013). Since more businesses are located in low small bank share counties, this procedure results in about four-fifths of the sample being categorized in low small bank counties. As a robustness check, we also experimented with alternative cutoffs of 50%, 55%, and 60% (reported in Online Appendix Table IA.3), and we observe similar results to Table 2.

B Additional Results

Table IA.1 presents the detailed regression results behind the panels of Figure 1. They show a distinct fall in the level of the R-share over time for incumbents, entrants, and exiters alike.

[12] Stein (2002) argues that small banks are superior at using soft information because such information is easier to communicate within a small organization with fewer layers of management.

[13] The 10th percentile, median, and 90 percentile for small bank shares across all counties from 2002 to 2013 are 5%, 56%, and 100%, respectively.

Table IA.1: **Evolution of Routine Share Over Time**

	All	Incumbent	Entrant	Exiter
Year=2003	-0.400**	-0.395**	-0.163	-1.044
	(0.170)	(0.186)	(0.405)	(0.652)
Year=2004	0.181	0.458***	-1.136***	-0.594
	(0.159)	(0.174)	(0.384)	(0.609)
Year=2005	-0.149	-0.025	-0.587	-1.678***
	(0.162)	(0.178)	(0.381)	(0.595)
Year=2006	-0.759***	-0.626***	-1.193***	-1.794***
	(0.175)	(0.194)	(0.374)	(0.604)
Year=2007	-0.958***	-0.812***	-1.494***	-1.793***
	(0.169)	(0.186)	(0.378)	(0.607)
Year=2008	-1.328***	-1.197***	-1.724***	-1.548***
	(0.170)	(0.186)	(0.398)	(0.596)
Year=2009	-1.999***	-1.828***	-2.689***	-2.532***
	(0.181)	(0.200)	(0.394)	(0.645)
Year=2010	-1.764***	-1.608***	-2.314***	-2.434***
	(0.169)	(0.185)	(0.403)	(0.611)
Year=2011	-1.835***	-1.653***	-2.665***	-1.950***
	(0.166)	(0.181)	(0.424)	(0.616)
Year=2012	-2.112***	-2.005***	-2.252***	-1.726**
	(0.182)	(0.199)	(0.413)	(0.745)
Year=2013	-2.368***	-2.275***	-2.523***	-2.111**
	(0.171)	(0.186)	(0.424)	(0.957)
N	3,032,551	2,532,651	499,899	245,918
R^2	0.66	0.67	0.62	0.59

Notes: This table reports the results of regressing an establishment's share of routine-task labor, in percentage, on year dummies while controlling for industry fixed effects at the NAICS3 level (see equation (2)). All regressions are weighted by a product of establishment employment and the sampling weight of the establishment assigned by the BLS. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2013, where 2002 serves as the benchmark year.

Table IA.2 provides additional results for the entry/exit regressions. In particular, we restrict the sample to larger establishments in Column (4) and control for establishment size in Column (5). See table notes for details.

Table IA.2: **Add'l Results: Routine Share Comparison Between Entrants and Incumbents**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Entrant	0.335***		0.310***	0.604***	0.511***	0.302***	0.137
	(0.083)		(0.083)	(0.112)	(0.106)	(0.091)	(0.142)
Exiter		0.439***	0.412***	0.764***	0.697***	0.433***	0.374**
		(0.127)	(0.128)	(0.163)	(0.153)	(0.140)	(0.159)
N	3,032,548	3,010,740	3,010,740	1,264,412	1,264,339	2,415,111	577,964
R^2	0.66	0.66	0.66	0.69	0.71	0.65	0.76

Notes: This table reports the results of regressing establishment routine share on an entrant dummy and an exiter dummy with NAICS3-Year fixed effects (see equation (4)). Column (4) restricts the sample to establishments with at least 20 employees, while Column (5) further includes NAICS3-Year-EmpBin fixed effects where the establishment employment bins are [20-49], [50-99], [100-249], [250-499], and [500+]. Columns (6) and (7) report baseline results for establishments in U.S. counties with high and low shares of small banks (see Section 3.2.1). The sample period is from 2002 to 2013. All regressions are weighted by the product of the establishment's total employment and the BLS sampling weight. Robust standard errors are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table IA.3 provides robustness checks of Table 2 by altering the cutoff for defining counties with a high small bank share.

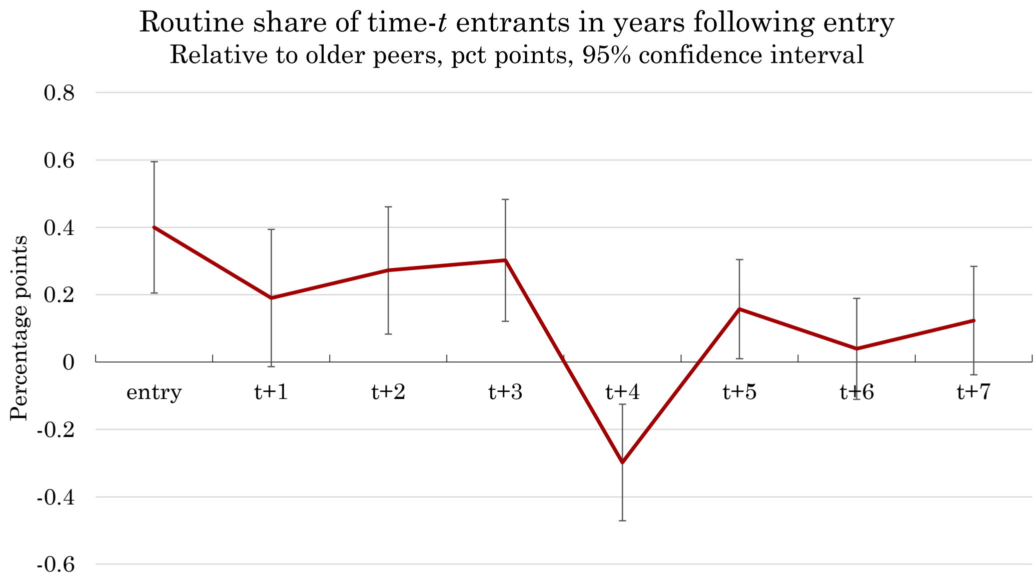
Table IA.3: **Robustness using Alternative Cutoffs for High Small Bank Share Counties**

	Counties with Low Small Bank Share			Counties with High Small Bank Share		
	Cutoff = 50%	Cutoff = 55%	Cutoff = 60%	Cutoff = 50%	Cutoff = 55%	Cutoff = 60%
	(1)	(2)	(3)	(4)	(5)	(6)
Entrant	0.302***	0.299***	0.293***	0.137	0.163	0.183
	(0.091)	(0.090)	(0.089)	(0.142)	(0.153)	(0.164)
exit_proxy	0.433***	0.417***	0.408***	0.374*	0.477**	0.435**
	(0.140)	(0.138)	(0.137)	(0.199)	(0.212)	(0.218)
N	2,415,111	2,493,478	2,577,937	577,964	499,598	415,139
R^2	0.65	0.65	0.65	0.76	0.76	0.76

Notes: This table reports the robustness of Table 2 by altering the cutoff for defining counties with a high small bank share. Instead of splitting the sample by the median of counties' small bank shares (which averaged 56% from 2002 to 2013), we adopt a fixed cutoff of 50%, 55%, and 60% as alternative approaches to split the sample. See Section 3.2.1 for more details of the definitions. The sample period is from 2002 to 2013. All regressions are weighted by the product of the establishment's total employment and the BLS sampling weight. Robust standard errors are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

Figure IA.1 depicts the dynamics of entrants following birth, focusing only on entering establishments that never exit afterward. This allows for ruling out a role for post-entry exit through selection.

Figure IA.1: **Robustness: The Routine Share Dynamics of Entrants Among Survivors**



Notes: This figure plots Panel A of Figure 2 using establishments that survive at $t + 7$. The vertical bars represent the robust standard errors of the point estimates.