Leading Firms and the Future of Work*

Max Brès, Fabrizio Colella, Philipp Kircher and David Koll

March 26, 2026

Preliminary Draft - Please Do Not Cite Without Permission

Abstract

This paper develops an approach to forecasting occupational change by identifying *leading firms* whose workforce composition anticipates the future occupational structure of their industry as other firms catch up over time. We develop a theoretical model with technology diffusion and frontier-laggard dynamics showing that leading firms are early adopters of technology and foreshadow broader labor demand adjustments. Using administrative data for French firms, we document substantial changes in occupational structure over time, mostly arising within industries and firms. To forecast occupational change, we first identify leaders using a machine-learning model based on firm characteristics and compare this approach with a productivity-based leader definition. We then estimate a forecasting model that combines current and lagged occupational structures of industries and leaders. In out-of-sample assessments, the machine-learning, leader-based approach improves forecasting accuracy by more than 21% relative to a baseline model without leader information. The predicted occupational changes are largely orthogonal to those implied by automation-risk indices, suggesting that the forces reshaping labor demand are not reducible to a single technological channel. Finally, we apply the model to predict the occupational structure of the French economy in 2034.

Keywords: Occupational change; leading firms; future of work; labor demand; occupational forecasting; machine learning; technology diffusion; firm heterogeneity

JEL: C53; C55; E24; J21; J23; J62; O33

*Brès: Sveriges Riksbank. Colella: USI Lugano and RFBerlin. Kircher: Cornell University and NBER. Koll: International Monetary Fund. We published a pre-analysis plan for this project, which can be accessed at https://osf.io/5gtsw/overview. Nicola Robatto and Leonardo Peñaloza-Pacheco provided outstanding research assistance. Vinzenz Ziesemer made invaluable contributions to an earlier version of this project. The authors are grateful for feedback of seminar participants at Cornell, Louvain-la-Neuve, IMF, Ski & Labor Workshop Laax, and AIEL Milan. The authors gratefully acknowledge funding through the European Research Council Grant No. 818859, and Kircher thanks the Universite Catholique de Louvain where he conducted much of this work. Access to confidential data, on which this work is based, has been made possible within a secure environment offered by Centre d'accès sécurisé aux données (Ref. 10.34724/CASD). The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of Sveriges Riksbank and of the IMF, its Executive Board, or its management.

1 Introduction

Accurately forecasting occupational change is critical for determining the appropriate skills to teach workers of the future, effectively reskilling the current workforce and unemployed, and enabling policymakers to navigate the evolving employment landscape and design targeted retraining programs.

Despite the clear benefits of predicting future labor market trends, the literature has yet to establish a comprehensive method for forecasting labor market changes. Previous research has attempted to forecast occupational trends by studying specific mechanisms, such as robotization, artificial intelligence, or trade shocks. However, a general approach independent of a specific mechanism has not yet been developed.

This paper fills this gap by introducing an algorithm to forecast the labor market's occupational structure based on the identification of leading firms. By *leading firms*, we mean firms whose occupational structure is already ahead of the current distribution in their industry. Identifying these firms allows us to extrapolate the aggregate labor market's occupational structure from those that adjust first.

The idea that innovations spread gradually from early adopters to the broader economy has a long intellectual history. Technology diffusion patterns have been documented at least since the influential work of Griliches (1957), Mansfield (1961), and Rogers (1962), and recent evidence confirms that adoption lags remain quantitatively large across technologies and countries (Comin and Hobijn, 2010). These lags imply that at any given time, a subset of firms has already adjusted its production and workforce to technologies that the majority of firms in the same industry have not yet fully adopted. The rise of information technology in the late 1990s illustrates the mechanism: firms such as Cisco and Dell reorganized their workforce around ICT-based supply chain coordination years before their competitors. A similar pattern is emerging with artificial intelligence, where some firms have already automated substantial parts of routine operations, while most have barely begun. This diffusion logic motivates our focus on occupational structure, because it captures employment dynamics directly relevant to retraining and educational policy, though the same approach could be applied to other outcomes shaped by adoption dynamics.

To build on this logic, our study examines whether identifying "leading firms" and postulating a catch-up process of other firms to leading firms improves the accuracy

of forecasting occupational structures within industries, and ultimately for the whole economy.

We implement this approach using matched employer-employee and balance sheet data covering the universe of private sector firms in France from 1994 to 2023. To identify leaders, we compare each firm's occupational structure in a base year with the occupational structure adopted by its industry several years later. We then use a random forest to learn which firm characteristics predict leadership, and classify as leaders those firms whose observable characteristics most resemble those of past leaders. The key assumption is that the characteristics associated with leadership are stable over time, so that the classification learned from a past episode generalizes to the present. We ground this identification strategy in a simple model of technology adoption with heterogeneous firms, which also delivers the catch-up forecasting equation that we take to the data. As an alternative, we define leaders as the most productive firms in their industry, motivated by the idea that a technological edge can be captured by standard productivity measures.

We estimate the model on 1997–2006 and evaluate it out-of-sample on 2010–2019, a period and set of firms not used in estimation. The out-of-sample evaluation was preregistered in a pre-analysis plan written before we accessed the testing data.[1] The machine learning specification improves forecasting accuracy by more than 21% relative to a baseline that uses only the contemporaneous and lagged occupational structure of the industry, while the productivity-based alternative yields modest gains. The improvement is robust to a range of alternative specifications, including changes in industry granularity, firm size thresholds, forecast horizons, and productivity measures. The forecasting gains are concentrated in a narrow subset of firms and are not a mechanical artifact of the leader definition: varying the leader share from 2.5% to 40% of industry employment yields a hump-shaped profile of out-of-sample accuracy, with the best performance at approximately 10%.

The characteristics that define leaders are multidimensional and stable over time. The machine learning algorithm selects leaders along multiple margins simultaneously: relative to followers within their industry, leaders are roughly three times larger, invest nearly twice as much per worker, hold twice the capital per worker, and pay 25% higher wages. No single observable drives the classification. The profile of leadership is stable across the estimation and testing samples, separated by more than a decade, suggest-

[1] The pre-analysis plan can be accessed through this link: https://osf.io/5gtsw/overview.

ing that the algorithm recovers a structural feature of the firm distribution rather than sample-specific patterns. By contrast, defining leaders solely by total factor productivity captures a narrower margin: productivity leaders dominate on TFP and profitability but are not systematically larger or more capital intensive, and their profile shifts between the two samples.

The predicted occupational changes from the leader model are largely orthogonal to those implied by automation risk indices. When we compare the employment changes predicted by our model to those derived from the risk scores of Frey and Osborne (2017) and Nedelkoska and Quintini (2018), the R^2 values are approximately 0.02 and 0.05 and the slope coefficients are statistically insignificant. The two sets of predictions thus capture different dimensions of occupational change, consistent with the view that the forces reshaping labor demand are not reducible to a single technological channel.

We apply the estimated model to predict the occupational structure of the French economy over the next decade. A natural limitation is that the model can only capture technologies and organizational practices that are already present in the economy: it has nothing to say about innovations that have not yet materialized. However, conditional on the current state of technology, the model does not require expert judgment to assess which occupations will be affected or how. Instead, it uses the cross-section of firms today to extract information from those that already operate at the frontier. In this sense the approach is complementary to methods that forecast employment losses on the basis of expert assessment, task content analysis, or patent-based measures of technology exposure. Those approaches can, in principle, anticipate the effects of technologies that do not yet exist, but must specify which technology matters and how it maps into occupational outcomes. The leader model takes the opposite route: it is silent on the source of change but captures whatever forces are already reshaping the occupational distribution, as revealed by the firms that have moved first.

While we apply the methodology to a particular dataset, future work could easily apply it to other datasets as long as they contain firm-level information on occupational structure. To the extent that leader characteristics are similar across countries, such work could also use our estimated leader characteristics to define leaders directly and then apply only the forecasting step to predict the future.

Related literature This paper relates to three strands of literature. The first studies the impact of technology on labor demand. A large body of work documents how automa-

tion, computerization, and offshoring reshape the occupational distribution, through skill biased technological change (Acemoglu and Autor, 2011), job polarization (Autor and Dorn, 2013; Goos et al., 2014), and robotization (Acemoglu and Restrepo, 2020). The task based framework of Autor et al. (2003) provides the canonical mapping from technological change to occupations, and underlies a large class of forecasts of employment losses, whether through expert assessment (Frey and Osborne, 2017), task level survey data (Nedelkoska and Quintini, 2018), individual level task measures that account for within occupation heterogeneity (Arntz et al., 2016), or patent based measures of exposure to automation and artificial intelligence (Webb, 2020). These approaches isolate a single technological channel. Our method is agnostic about the source of change: it captures the net effect of all forces reshaping occupational structures, as revealed by the behavior of leading firms.

The second strand studies technology diffusion and adoption lags across firms. The observation that innovations spread gradually from early adopters to the broader economy dates back to Griliches (1957); Mansfield (1961); Rogers (1962), and recent work confirms that adoption lags remain quantitatively large across technologies and countries (Comin and Hobijn, 2010). A growing literature documents persistent productivity gaps between frontier and laggard firms within narrowly defined industries (Andrews et al., 2017), consistent with slow diffusion of best practices. Complementary investments in organizational capital and workplace reorganization condition how quickly firms absorb new technologies (Bresnahan et al., 2002), and the returns to information technology itself depend on scale and intangible intensity (Lashkari et al., 2024). We build on this evidence by showing that leading firms' occupational structures contain exploitable information about the direction of diffusion, not only its speed.

The third strand examines the role of firm heterogeneity in shaping aggregate labor market outcomes. The rise of superstar firms and the associated decline in the aggregate labor share (Autor et al., 2020), and the role of between firm sorting in driving wage inequality (Song et al., 2019), are well documented. Hubmer and Restrepo (2021) show that the joint dynamics of firm size and factor shares rationalize aggregate trends once firm heterogeneity is taken into account, and Patel (2024) provides a firm-based account of job polarization in French data. Event study evidence further shows that technology adoption episodes trigger sizable within firm occupational reallocation (Humlum, 2022). Our contribution is to exploit firm heterogeneity for forecasting rather than for explaining contemporaneous outcomes: the same dispersion in firm characteristics that

generates cross sectional inequality also contains forward looking information about occupational reallocation.

2 Model

There are time periods $t \in \{..., -2, -1, 0, 1, ...\}$. Fix an industry $j \in J$. In this industry, there is a distribution of productivity z with a given technology and innovativeness I in adopting new technologies among firms, with stationary cumulative distribution function $G(z, I)$ with support in a subset of $\mathbb{R}^2_{++}$. This distribution arises, for example, if firm characteristics (z', I) evolve to (z', I') with markov probability $\pi(z', I'|p, I)$ and $G(z, I)$ denotes the ergodic distribution. We further assume that innovativeness might be linked to other observables $s_1, s_2, ..s_N$, so that $I(s_1, s_2, ..., s_N)$ which might themselves follow a markov chain.

Firms can operate with an existing technology θ_0 which is costless to operate, or can upgrade to some new technology θ_+ associated with additional weakly positive operating cost $c(I + t)$. Such a new technology can be anything that requires changes to the overall production structure of the firm, ranging for disruptive information technology adoption to using more outsourcing as part of the production process. As more is known about the new technology over time, firms find it easier to operate it, captured by the assumption that $c(.)$ is decreasing with $c(T) = 0$ for some finite T. That is, at some point also the new technology is so well understood that it is costless to operate it. Firms that are intrinsically more innovative as captured through higher values of I are "ahead" of their time and find it less costly to adopt the new technology.

Firms also invest in different amounts of capital K_m across several capital classes $m \in \{1, 2, ..., m\}$ and - more relevant for our analysis - employ number of workers L_o in different occupations $o \in \{1, 2, ..., O\}$. Assume that the indistry under consideration is small so that its factor demands do not affect factor prices r_m for different types of capital and w_o for different types of labor. For illustration, assume that these factor prices are constant over time, a point we will return to in the subsequent discussion.

A firm that operates technology θ produces output according to a production function $zF(L_1, .., L_O, K_1, .., K_r|\theta)$ that is homothetic and decreasing returns to scale in $(L_1, .., L_O, K_1, .., K_r)$. Even though we suppressed it, the production function and type distribution

should be indexed by the industry in consideration. So the maximization problem for firm with type (z, I) in this industry at time t reduces to

$$\begin{aligned}\pi(z,\theta) &= \max_{L_1,..,L_O,K_1,..,K_r} zF(L_1,..,L_O,K_1,..,K_M|\theta) - \sum_{o=1}^{O} w_o L_o - \sum_{m=1}^{M} r_m K_m \\ \Pi(z,I) &= \max_{\theta\in\{\theta_0,\theta_+\}} \pi(z,\theta) - C(I+t,\theta)\end{aligned}$$

where cost $C(I+t,\theta_0) = 0$ and $C(I+t,\theta_+) = c(I+t)$. Assume that technology θ_+ is strictly better in the sense that $F(L^1,..,L^O,K^1,..,K^M|\theta_0) < F(L^1,..,L^O,K^1,..,K^M|\theta_+)$ for all strictly positive input factors $(L^1,..,L^O,K^1,..,K^M)$.

The following insights immediate emerge from this simple model: First, all firms that operate the old technology θ_0 optimally choose the same share of its workers from occupation o, independent of their type z. This follows from homotheticity of the production function - see Appendix A for a short review on this. The same holds true for the firms that operate the new technology θ_+. So for any firm j at time t with $L_t^{o,j}$ in occupation o and $L_t^j = \sum_o L_t^{o,j}$ the occupational distribution

$$\mathcal{S}_t^j = (L_t^{o,j}/L_t^j, L_t^{2,j}/L_t^j, ..., L_t^{O,j}/L_t^j)$$

is only a function of technology choice θ, and we can simply write $S(\theta)$. For notational convenience with will analogously call $\mathcal{S}_t^J$ the occupational distribution across a set of firms J, and if all firms have the same technology θ we can again simply write $S(\theta)$. We will denote by $\mathcal{S}_t^i$ the overall occupational distribution in industry i at time t, and omit the industry subscript when it is obvious that we focus on one particular industry (which we do throughout this theory section).

Second, if a firm with characteristic (z, I) adopts the new technology in period t, then a firm with characteristics (z', I') such that $z' \geq z$ and $I' \geq I$ also finds it profitable to operate the new technology. This follows directly because a higher innovativeness decreases the fixed costs of adopting the innovative technology, and a higher z increases the benefits of moving to the new technology as it gets amplified by a higher total factor productivity.

Third, early adopters of the new technology will choose an internal division of labor across occupations that resembles the occupational mix in the overall industry later in time when the laggards have also caught up with the new technology: Under the assumption of constant factor prices and homothetic production, firms labor distribution

$L_o/L_{o'}$ is only a matter of technology choice. Initially early adopters choose the new technology, but over time more and more firms catch up, and latest by $t > T$ all firms adopt the new technology. So early adopters in this setting are firms whose occupational distribution foreshadows the distribution of occupations in the overall industry. It will be useful to denote the set of adopters a technology at time t as $\mathcal{A}_t(\theta)$, and to denote the complement of all other firms that have not adopted that technology at time t as $\mathcal{C}\{\mathcal{A}_t(\theta)\}$. The occupational distribution of early adopters ($\mathcal{S}_0^{\mathcal{A}_0}$) compared to initial laggards ($\mathcal{S}_0^{\mathcal{A}_{C,0}}$) is then useful in forecasting the occupational evolution in the overall industry $\mathcal{S}_t$:

$$\mathcal{S}_t = \beta_{L,t} \underbrace{\mathcal{S}_0^{\mathcal{A}_0(\theta_+)}}_{\text{occ. distr among ealy adoptors at time 0}} + \beta_{S,t} \underbrace{\mathcal{S}_0^{\mathcal{C}\{\mathcal{A}_0(\theta_+)\}}}_{\text{occ. distr of laggards at time 0}} \quad (1)$$

The occupational distribution at time t is a linear combination between the occupational distributions of early adopters and initial laggards which represent the occupations under the new and old technology, respectively. The weights $\beta_{L,t}$ and $\beta_{S,t}$ reflect the share of overall industry employment in adopting firms and lagging firms at time t, respectively. Over time, a_t grows, shifting more weight to the employment distribution present in early adopters. Our full specification will allow for some imperfect classification of leaders who might not all yet use the new technology, as well as misclassifications of laggards and the potential presence of even older technologies among laggards (see first extension), but for now we use this simple model to draw some implications from this for forecasting.

For forecasting, consider some future time $t_1 > T$ where technology θ_a is the costless base technology (where θ_a could equal θ_+). At time t_1 another even newer technology θ_{++} becomes available. Assume as before that operating the new technology costs $C(I + t, \theta_{++}) = c(I(t - t_1))$, so again, as time progresses it is easier to adopt the next new technology. If the benefits of adopting θ_{++} rather than θ_a resemble those of adopting θ_+ rather than θ_0 (i.e., $\pi(z, \theta_{++}) - \pi(z, \theta_a) = \pi(z, \theta_+) - \pi(z, \theta_0)$), then the fraction of new adopters in this new setting exactly equals that in the previous setting when θ_+ was introduced rather than θ_0. Even if the benefits are not exactly identical as last time, still again the high (z, I) firms are the first to adopt, and the overall industry will slowly follow in adoption over time until all firms adopt the improved technology. So for $t > t_1$

the industry evolves according to

$$\mathcal{S}_t = \hat{\beta}_{L,t} \underbrace{\mathcal{S}^{\mathcal{A}_{t_1}(\theta_{++})}}_{\text{occ. distr among early adopters at time } t_1} + \hat{\beta}_{S,t} \underbrace{\mathcal{S}^{\mathcal{C}\{\mathcal{A}_{t_1}(\theta_{++})\}}}_{\text{occ. distr of laggards at time } t_1} \tag{2}$$

where $\hat{a}_t$ is now the employment share in the newest technology θ_{++}. If one could observe $D(\theta_{++})$ and $D(\theta_{+})$ at time t_1, one could forecast the future employment distribution in the industry based on the previous experience in (1) by assuming that the adoption at time τ after the new technology introduction is similar:

$$\hat{\beta}_{L,\tau+t_1} = \beta_{L,\tau+t_0} \;\&\; \hat{\beta}_{S,\tau+t_1} = \beta_{S,\tau+t_0}, \tag{3}$$

and this assumption would be approximately correct if the adoption dynamics are similar over time. Under this assumption, our forecasting equation for the occupational distribution at time $t = \tau + t_1$ would be

$$\mathcal{S}_{\tau+t_1} = \beta_{L,\tau+t_0} \underbrace{\mathcal{S}_{t_1}^{\mathcal{A}_{t_1}(\theta_{++})}}_{\text{occ. distr among early adopters at time } t_1} + \beta_{S,\tau+t_0} \underbrace{\mathcal{S}_{t_1}^{\mathcal{C}\{\mathcal{A}_{t_1}(\theta_{++})\}}}_{\text{occ. distr of laggards at time } t_1} \tag{4}$$

where all variables on the right side are known at time t_1 if one could identify early adopters at that time.

The challenge that arises to use the choices of early adopters and their occupational structure to identify the identity of the early adopters: we first have to find them at time t_0, and then we have to find them again at time t_1 because firm types can change and so the firms that adopt early might change. We usually do not observe the technology choice of firms in administrative data. And even if this is recorded, it is usually vague: when a firm reports use of ICT there is a difference whether they do this a bit or whether they are really early adopters who have re-thought the structure of the firm to make the best use of ICT. The same holds for outsourcing: firms might report to do some outsourcing but only a few firms have really thought through how to use this optimally and reshaped their organizational structure accordingly. Again similarly, firms might report to use AI in its workforce, but only few of them might yet have thought through how to optimally reshape their workflows and workforce to make best use of this technology.

If data on actual technology adoption is missing or unreliable, what else could we use to identify early adopters? Our theory identifies two ways which we will explore in

empirical work.

1) Early adopters as productive firms. Early adopters are high (z, I) firms in our theory. These tend to be the more productive firms in our theory. So one can try to back out a measure of productivity, and use that to identify such firms. Then one can look at the distribution of labor within the set of productive firms at time t_0 and call it $\mathcal{S}_0^{\mathcal{A}_0(\theta_+)}$, while the the distribution of labor in all less productive firms is taken as a proxy for $\mathcal{S}_0^{\mathcal{C}\mathcal{A}_0(\theta_+)}$ (and we explore misclassification in more detail in the first extension). And then one can see if less productive firms converge in occupational distribution to the more productive firms over time after $t = 0$ to back out the adjustment dynamics $\beta_{L,t}$ and $\beta_{S,t}$. At time t_1, one again tries to estimate which firms in the industry are the most productive, call the occupational distribution within such firms $\mathcal{S}_0^{\mathcal{A}_0(\theta_{++})}$ and the occupational distribution in the less productive firms $\mathcal{S}_0^{\mathcal{C}\mathcal{A}_0(\theta_{++})}$, and use relationship (2) and assumption (3) to forecast future occupational change according to (4). The downside of this approach is that this method might classify some low-z but high-I firms as "low-productive" even though they do innovate, and might classify some high-z but low-I firms as "high-productive" even though they do not innovate. In the model, the second mis-classification implies that the occupational those firms identified as "high-productive" would not fully foreshadow the future occupational distribution.

2) Alternatively, we can classify leaders as those whose occupational distribution precedes the future distribution in the whole industry, and learn the characteristics that are associated with such leaders. The idea to learn about the characteristics of industry leaders follows from this simple model: Observe the occupational distribution $\mathcal{S}_0^j$ at time $t = 0$ for each firm j in the industry. Observe the occupational distribution $\mathcal{S}_t$ of the whole industry at some future point in time $t > T$. In our model, those firms whose internal occupational distribution at time $t = 0$ resembles the overall occupational distribution at the future point t (i.e., firms with $\mathcal{S}_0^j = \mathcal{S}_t$ must have been the early adopters). Given that we now know the early adopters, we can use machine learning to identify which observables characterize early adopters. This could include size (in our theory larger firms tend to be earlier adopters, but also other observables $s_1, s_2, ..s_N$ that are systematically linked to higher innovativeness. These could include firm age or their spending on research and development, among many other things. Once we identified the observable characteristics of leaders, we assume according to our theory that this also identifies leaders in general (and in particular in the future at time t_1, which is

correct in our model if the benefits of adoption remain similar across periods). Again, the occupational distribution of leaders at $t = 0$ identifies $\mathcal{S}_0^{\mathcal{A}_0(\theta_+)}$ and the occupational distribution of the other firms identifies $\mathcal{S}_0^{\mathcal{CA}_0(\theta_+)}$, and we can back out $\beta_{L,t}$ and $\beta_{S,t}$ for the evolution of occupational employment in the industry. At time t_1 we want to use that to project the occupational distribution into the future.

At time t_1 we can identify those firms with "leader" observables. Call their occupational distribution $\mathcal{S}_0^{\mathcal{A}_0(\theta_{++})}$. Those firms that do not have observables that would correspond to leaders identify the laggards, and their average occupational distribution identifies $\mathcal{S}_0^{\mathcal{CA}_0(\theta_+)}$. Then equations (2) and (3) allow us again to forecast the overall occupational distribution in the industry as outlined in (4).

The advantage of the second method is that one can use machine learning techniques in the early from t_0 to T on the identified leaders to uncover which characteristics are associated with those leaders, where the set of characteristics under consideration can be broad and far exceed simple ideas such as "productive firms must be early adopters". Then one can use these leader characteristics to identify leaders in t_1 based on a broad set of characteristics. The downside is that the initial machine learning assumes that those firms that foreshadow the future occupational distribution are indeed leaders. There could be reasons outside the scope of this model that would have some firms' occupational distribution look early on like the distribution of the whole industry that have nothing to do with visionary adaptation of new ideas and technologies. For example, it could be shear coincidence that some firms by chance only managed to hire in occupations that happened to look similar to occupations that are popular overall in the future. Therefore, it is important to test this. One can do that if one tries to forecast from t_1 into the future, but actually can cross-check against the success of the projection that actually happens after t_1. This is the step we take in the empirical approach.

Before going there, we discuss some extensions to our approach that clarify the extended empirical model we use in our application. We also discuss robustness to other variants of technology adoption including those where the amount of early adopters makes it easier for later ones to adopt.

Extension and Robustness Discussion: As a first extension, and justification for our full empirical strategy, we take into account the presence of some even older technologies and some mis-classification of leaders in the two ways of leader-identification discussed above. Assume that our leader identification at time $t = 0$ is not perfect: our

set of "leaders" $\mathcal{A}_0$ at time $t = 0$ has a distribution of (z, I) that stochastically dominates the distribution in the set of "laggards", but assignment is not perfect and some of the leaders might still use the old technology while some laggards use the new technolgy. Further, assume that apart from θ_+ and θ_0 there are older technologies θ_-, θ_{--} ect. that are associated with a successively lower technology frontier with costs $c(I+t+\tau_\theta)$ where τ_θ is higher for older and less productive technologies. In particular, θ_0 is not free for all firms, and some firms still use older technologies. If there are up to four technologies present in the economy, under mild regularity conditions one can show that the occupational distribution at the industry level can now be written as:

$$\mathcal{S}_t = \beta_{L_1,t}\mathcal{S}_0^{\mathcal{A}_0} + \beta_{L_2,t}\mathcal{S}_{-h}^{\mathcal{A}_0} + \beta_{S_1,t}\mathcal{S}_0^{\mathcal{C}\{\mathcal{A}_0\}} + \beta_{S_2,t}\mathcal{S}_{-h}^{\mathcal{C}\{\mathcal{A}_0\}} \tag{5}$$

where S_{-h} denotes the occupational distribution of the firms $-h$ periods earlier. So here we use both the current as well as the lagged occupational distribution of leaders and followers. This is helpful because the shift over time among leaders allows to identify the increased shift into new technology, and the change over time for the laggards allows to identify the shift out of the oldest technology. The reason that this decomposition is possible is that $\mathcal{S}_t$ is a weighted average of the occupational distributions $S(\theta_+)$, $S(\theta_0)$, $S(\theta_-)$ and $S(\theta_{--})$, and the four variables $\mathcal{S}_\tau^\cdot$ on the right side of (5) are also mixtures of these distributions. As long as the $\mathcal{S}_\tau^\cdot$ are not collinear, representation (5) holds.[2] Note that equation (5) is quite flexible and can under particular parameter constellations capture a difference equation instead of a specification in levels.[3] Assuming a similar change in the future results in adjusting our forecasting equation (4) into:

$$\mathcal{S}_{\tau+t_1} = \beta_{L_1,\tau+t_1}\mathcal{S}_{t_1}^{\mathcal{A}_{t_1}} + \beta_{L_2,\tau+t_1}\mathcal{S}_{t_1-h}^{\mathcal{A}_{t_1}} + \beta_{S_1,\tau+t_1}\mathcal{S}_{t_1}^{\mathcal{C}\{\mathcal{A}_{t_1}\}}\beta_{S_2,\tau+t_1}\mathcal{S}_{t_1-h}^{\mathcal{C}\{\mathcal{A}_{t_1}\}} \tag{6}$$

Additional robustness questions concern the cost function, and we will briefly discuss how our logic carries over to alternative specifications. Adopting the new technology might be easier as more firms adopt it. This can be captured with extended cost function $c(I + \alpha A_{t,\theta} + \beta t + \tau_\theta)$, where $A_{t,\theta}$ is the fraction of firms that have already adopted the

[2]Formally, for any t the vector $\mathcal{S}_t$ is a linear combination of the vectors $S(\theta_+)$, $S(\theta_0)$, $S(\theta_-)$ and $S(\theta_{--})$ which form a basis. Each of the vectors $\mathcal{S}_0^{\mathcal{A}_0}, \mathcal{S}_{-h}^{\mathcal{A}_0}, \mathcal{S}_0^{\mathcal{C}\{\mathcal{A}_0\}}$ and $\mathcal{S}_{-h}^{\mathcal{C}\{\mathcal{A}_0\}}$ can also be represented as a linear combination of the basis vectors. If they are linearly independent, by The Basis Theorem, $\mathcal{S}_0^{\mathcal{A}_0}, \mathcal{S}_{-h}^{\mathcal{A}_0}, \mathcal{S}_0^{\mathcal{C}\{\mathcal{A}_0\}}$ and $\mathcal{S}_{-h}^{\mathcal{C}\{\mathcal{A}_0\}}$ form an alternative basis. So they also linearly represent $\mathcal{S}_t$.

[3]For example, in special cases it can reduce to $\mathcal{S}_t - \mathcal{S}_{t-h} = \beta_{L,t}[\mathcal{S}_0^{\mathcal{A}_0} - \mathcal{S}_{-h}^{\mathcal{A}_0}] + \beta_{S,t}[\mathcal{S}_0^{\mathcal{C}\{\mathcal{A}_0\}} - \mathcal{S}_{-h}^{\mathcal{C}\{\mathcal{A}_0\}}]$.

technology. It is easy to see that the steps leading to equations (1) - (4) follow in the same way, only that the timing is now endogenous for each of the technology adoptions. Since decisions in our model are static for each period, it does not affect who adopts in the instant when the new technology becomes available, and so adoption dynamics follow an (endogenous) differential equation in time. For the individual firm this simply forms a time path with which costs fall, and higher (z,I) firms still adopt earlier and their choice foreshadows the distribution in the overall industry.

Alternatively, one could imagine that technolology is embedded in a particular capital type - say capital type M - and the cost $c(I + t + \tau_\theta)$ is not a fixed cost but the marginal cost of operating this capital with the new technology. This might be applicable if the new capital is, e.g., robots where there is a cost to install robots but the cost depends on the familiarity with such new technology. Then our basic maximization problem reads reads in analogy to (1): $\max_{\theta\in\{\theta_0,\theta_+\}} \max_{L_1,..,L_O,K_1,..,K_r} zF(L_1,..,L_O,K_1,..,K_M|\theta) - \sum_{o=1}^{O} w_o L_o - \sum_{m=1}^{M-1} r_m K_m - (r_M + C(I + t))K_M$. In this setting firms with very high innovativeness $I > T$ face simply the cost r_M and adopt factor shares according to the relative input prices. Firms with $I < T$ initially distort their production decision, but over time they converge to the production decision of the early adopters and when $t > T$ all firms adopt the decision of early adopters. So early full adopters are still those that best foreshadow the future of the occupational structure in the industry, and we can still learn their characteristics according to our second approach. The industry will eventually catch up with their occupational distribution, though the occupational distribution is no long just a weighted average of the early and late adoptors as in 2 since some firm now adopt different capital M over time. In such a case an attempt to estimate equation (2) with an error on the right hand side would amount to linearly projecting the actual evolution into the space spanned by the occupational distribution of early full adopters and others.

Finally, one might imagine that capital M which embeds the technology is formed in a putty-clay fashion where previous investment is a stock and stuck in the old technology, while new investment can choose either technology at a fixed cost but is then again locked in for the next periods. Assume that depreciation is sufficiently high that firms invest in every period. All other inputs are fully flexible. Assume that firms combine each type of capital M with other inputs in a constant returns to scale fashion to produce intermediate inputs, and these are aggregated according to some decreasing returns to

scale function into final output.[4] While we do not cover this setting formally due to the dynamic nature of the choice problem, it is evident that those firms that switch part of their capital to θ_+ choose a workforce that is a weighted average of $S(\theta_+)$ and $S(\theta_0)$, and foreshadow the future evolution of the industry better than other firms that only operate the old technology. We can again learn their type. Since the leaders no longer purely operate the new technology, the weights in (2) no longer correspond to the employment share of early adopters, but now reflect the employment share in the new technology. Still we can estimate (2) or its extended version (5) to uncover the industry dynamics.

To conclude, we acknowledge that we fixed the factor prices throughout. In a competitive economy changing labor demand will lead to changing factor prices. In particular the price w_o for labor from particular occupations might change because the aggregated demand across all industries for labor from different occupations changes. While it is possible to write models in which labor demand is fully elastic and wages are pinned down by the differential disutility of work across occupations, most models of occupational choice will only have partially elastic labor supply and a full model would need to account for the changes in input costs for various labor types that the firms will face. Still these partial equilibrium derivations give an idea of how demand across occupations might evolve, and wage effects might diminish but hopefully not fully reverse these effects.

3 Data

Data Sources. Our analysis combines matched employer–employee information from the *Déclarations Annuelles de Données Sociales* (DADS, (Déclarations Annuelles de Données Sociales), "Annual Declarations of Social Data") with firm balance-sheet and income-statement information from *FICUS/FARE* (*Statistique structurelle annuelle dentreprises*) provided by the *Institut national de la statistique et des études économiques* (INSEE) and the *Direction générale des Finances publiques* (DGFiP). The DADS provides annual matched

[4]Recall that a firm that produces with one technology achieves output $zF(L_1,..,L_O,K_1,...,K_M|\theta) = f(h(L_1,..,L_O,K_1,...,K_M))$ where $h(.)$ is constant returns to scale and $f(.)$ is concave, which is the definition of homotheticity. In principle both $h(.)$ and $f(.)$ can depend on technology θ. Here we restrict attention to functions where only $h(.)$ depends on θ, and the output of a firm that operates both technologies is given by $f(h(L'_1,..,L'_O,K'_1,...,K'_M|\theta') + h(L''_1,..,L''_O,K''_1,...,K''_M|\theta''))$ where variables indexed by $'$ are combined with the capital K'_M that uses technology θ', and those indexed by $''$ refers to the inputs combined with capital K''_M that embeds the new technology.

employer–employee records covering the universe of French workers over 1994–2023. It contains detailed worker outcomes—including annual earnings and occupation—and allows us to construct aggregate outcomes at the firm–year–occupation level, such as the number of workers and the wage bill, while retaining rich worker-level information (notably hours worked) for complementary analyses. The FICUS/FARE data are built from mandatory administrative fiscal filings and provide harmonized firm tax balance sheets and income statements for the universe of French firms. These data include standard measures of firm performance and production—such as total assets, sales, value added, profits/profitability, capital stock, investment in physical capital, cash flows, and interest payments—which we use to characterize firms and track their outcomes over time.

Sample Construction. We first harmonize variable definitions over time within each source. In the DADS, we restrict attention to firms with at least 20 employees in every year of the sample and aggregate worker information to the firm–year–occupation cell by summing (i) the number of workers and (ii) total annual earnings (wage bill). In the FICUS/FARE, we harmonize accounting variables across years, drop firms missing industry information in any year, and exclude non-relevant sectors (e.g., public sector, extraterritorial activities, and home production). We then merge the two sources at the firm–year level, yielding a panel that links detailed labor-market outcomes by occupation to rich firm-level balance-sheet information. The merged panel contains 6.97 million firm–year observations spanning 29 years (1994–2022) and 542,063 distinct firms (SIREN). We split the panel into two parts: an *estimation sample* spanning 1997–2006 and a *testing sample* spanning 2010–2023. We exclude the years 2007–2009 from the analysis to avoid distortions from the global financial crisis and related disruptions in the data that could confound our forecasting exercise.

Table 1: Main information on the sample (1994–2022)

Statistic	**Value**
Number of observations (firm–year–occupation)	6.97 mn
Number of years	29
Number of distinct firms (SIREN)	542 k
Number of occupations (2-digit)	29
Number of industries	187

Descriptive Statistics. Table 2 reports summary statistics for the key variables in our merged panel over 1994–2022. The median firm employs 22 workers and pays an average annual wage of roughly €17,300, with substantial right-skewness in employment. Median sales and value added are €1,573k and €605k respectively, and the interquartile range of profitability (0.04 to 0.30) indicates that most firms generate positive margins despite a negative unconditional mean driven by the left tail. Firms hold considerably more tangible than intangible capital at the median (€248k versus €25k), while median investment is modest at €18k. Median exports are zero, consistent with the well-documented finding that only a minority of firms engage in international trade. ACF productivity, available for the subset of firms with sufficient data to estimate production functions ($N = 467{,}932$), has a median of 3.9 log points with relatively low dispersion (interquartile range of 1.6).

Table 2: Descriptive statistics for key variables (1994–2022)

Variable	**Mean**	**SD**	**P25**	**P50**	**P75**
Number of workers	83	2,061	13	22	44
Mean annual wage	19,844	20,071	9,765	17,294	25,480
Firm age (years)	18	17	6	13	25
Sales	12,565	698,834	633	1,573	4,262
Value added	4,219	672,229	259	605	1,392
Gross profits	1,905	666,589	7	69	239
Profitability	−1.8	1,032	0.04	0.15	0.30
Exports	2,021	81,437	0	0	3
Tangible assets	5,767	285,604	70	248	783
Intangible assets	1,017	61,138	1	25	154
Investment	551	28,888	3	18	74
ACF productivity	3.8	1.7	3.2	3.9	4.8

Notes: Variables are defined at the firm–year level. Monetary values are in euros for wages, and in thousands of euros for other financial variables. Profitability is gross profits over sales. ACF productivity is estimated following Ackerberg, Caves and Frazer (2015).

4 Occupational Change Over Time and its Drivers

This section documents the evolution of the occupational structure over time and decomposes it into its sources. The goal is to quantify how much of the observed reallocation across occupations can be attributed to shifts across industries, firms, and within firms. The section also motivates the forecasting exercise in the next part of the paper by showing that occupational change is driven to a large extent by firm-level adjustments rather than by sectoral reallocation alone.

4.1 Evolution of the Occupational Structure

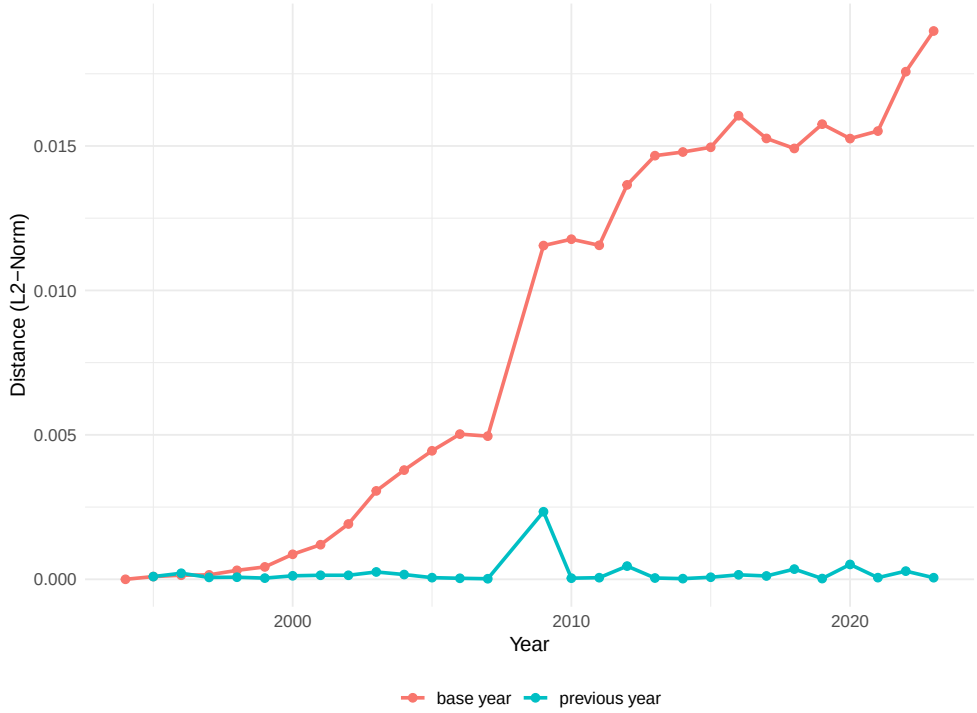
We define the occupational structure of the economy at time t as the vector $S_t = (S_t^o)_{o \in O}$ where S_t^o denotes the share of total employment in occupation o. Changes in the occupational structure between two dates t and s are measured by the difference $S_t^o - S_s^o$. To summarize changes over time, we measure the distance between occupational distributions using the L_2 norm,

$$\|S_t - S_s\|_2 = \sqrt{\sum_{o \in O} \left(S_t^o - S_s^o\right)^2} \tag{7}$$

This metric captures the extent to which the allocation of work across occupations changes over time. Figure 1 highlights a persistent transformation of the occupational structure between 1994–2023. The distance to the base year 1994 rises steadily over time, indicating that occupational change reflects a sustained structural trend rather than transitory fluctuations with mean-reversion.

The underlying occupation-level changes are also economically meaningful. Figure 2 shows that between 1994 and 2023, high-skill occupations such as engineers, executives, and other professional occupations expand, whereas middle-skill white-collar occupations and (un)skilled industrial employment contract. At the same time, the employment share of some lower-skill service occupations increases. These patterns appear consistent with occupational polarization.

Figure 1: Distance between Occupational Structures over time with 1994 as base year



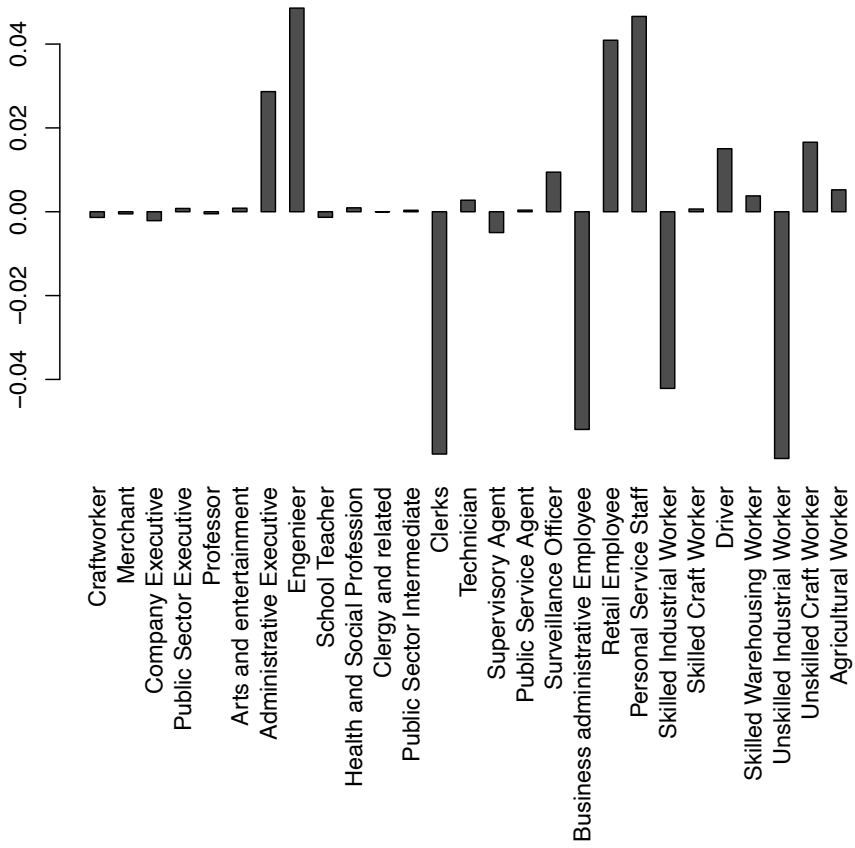
Notes: The Figure shows the distance between the occupational structures in two years measured by the L2-Norm, first, relative to the base year 1994 and, second, relative to the previous year.

4.2 Decomposing Occupational Change

To understand the forces behind these changes, we decompose the variation in occupation shares into several components. The change in the aggregate occupation share S_t^o between periods t and $t+1$ can be written as the sum of six terms: a between-industry component, a between-firm component within industries, a within-firm component, an entering-firm component, an exiting-firm component, and a net-entry component.

This decomposition separates reallocation occurring across sectors from changes that arise inside industries and firms. The between-industry term captures the role of shifts in industry employment shares. The between-firm term captures the reallocation of employment toward firms with different occupational mixes within the same industry and is driven by differential firm growth. The within-firm term captures changes in the firm's organization of production reflecting adjustments in its occupational composition. The entry and exit terms capture that entering and exiting firms might differ in their occupational structure from the average structure in the industry. The net-entry

Figure 2: Changes in the 2-digit Occupational Structures between 1994–2023



Notes: The Figure shows the difference in the employment shares of 2-digit PCS occupations between 2023 and 1994. A positive number reflects a higher share in 2023.

term summarizes a re-weighting due to different entry and exit shares, thereby capturing the effect of firm dynamics. Hence, we can decompose the change in the share of hours in occupation o, S_t^o, between period t and $t+1$ as:

$$S_{t+1}^o - S_t^o = \underbrace{\sum_i (S_{t+1}^i - S_t^i) S_{t+1}^{o/i}}_{\text{Between industry}}$$

$$+ \underbrace{\sum_i \left(\sum_{f \in C_i} \left(\frac{S_{t+1}^f}{S_{t+1}^i} - \frac{S_t^f}{S_t^i} \right) \frac{S_{t+1}^{o,f}}{S_{t+1}^f} \right) S_t^i}_{\text{Between firm}} + \underbrace{\sum_i \left(\sum_{f \in C_i} \left(\frac{S_{t+1}^{o,f}}{S_{t+1}^f} - \frac{S_t^{o,f}}{S_t^f} \right) \frac{S_t^f}{S_t^i} \right) S_t^i}_{\text{Within firm}}$$

$$+\underbrace{\sum_i \left(\sum_{f \in N_i} \left(\frac{S^{o,f}_{t+1}}{S^f_{t+1}} - \frac{S^{o,i}_t}{S^i_t} \right) \frac{S^f_{t+1}}{S^i_{t+1}} \right) S^i_t}_{\text{Entering firms}} - \underbrace{\sum_i \left(\sum_{f \in X_i} \left(\frac{S^{o,f}_t}{S^f_t} - \frac{S^{o,i}_t}{S^i_t} \right) \frac{S^f_t}{S^i_t} \right) S^i_t}_{\text{Exiting firms}}$$

$$+\underbrace{\sum_i \frac{S^{o,i}_t}{S^i_t} \left(\sum_{f \in N_i} \frac{S^f_{t+1}}{S^i_{t+1}} - \sum_{f \in X_i} \frac{S^f_t}{S^i_t} \right) S^i_t}_{\text{Net entry at distribution in } t}$$

where we denote by S^i_t the share of workers in industry i at time t, by $S^{o,i}_t$ the share of workers in occupation $o \times$ industry i at time t, and by $S^{o/i}_t$ the share of workers in occupation o within industry i at t. In addition, S^f_t is defined as the share of workers in firm f at time t, $S^{o,f}_t$ as the share of workers in occupation $o \times$ firm f at time t, and $S^{o/i}_t$ as the share of workers in occupation o within industry i at t. Finally, C_i is the set of firms in industry i that exist in period t and $t+1$, X_i is the set of firms in the industry i existing in t but not in $t+1$, and N_i is the set of firms in i not existing in t but existing in $t+1$.

Results of the Decomposition. The decomposition result in Table 3 highlights four central findings. First, a substantial share of occupational change is accounted for by the between-firm reallocation within industries. This indicates that employment shifts toward firms with systematically different occupational structures. Second, within-firm adjustments also contribute to aggregate occupational change by reorganize their workforce composition. Third, the difference in the occupational structure of entering firms appears to play a substantial role in aggregate occupational change as well. Fourth and finally, the aggregated within-industry reallocation amounts to 51–85% depending on the measure. This has an important implication for the following forecasting exercise. If occupational change were driven mainly by shifts in employment across industries, then forecasting could rely primarily on sectoral growth differentials. Instead, our results suggests that much of the action takes place within industries and within firms. This motivates our focus on identifying firms whose occupational structures are already tilted toward the future of their industry. Such firms are natural candidates for "leaders" in the process of occupational change. Their workforce composition may therefore contain information about the direction of future aggregate adjustments.

Table 3: Decomposition of Occupational Change

	Single Components						
	Between industry	Between firms	Within firms	Entering firms	Exiting firms	Net entry	Within industry
1995 - 2001							
$\frac{1}{N_o}\sum_o \frac{x_{o,c}}{\sum_c x_{o,c}}$	21.6%	44.2%	17.6%	23.5%	-2.5%	-4.5%	78.4%
$\frac{1}{N_o}\sum_o S_t^o \frac{x_{o,c}}{\sum_c x_{o,c}}$	23.7%	-7.4%	52.0%	16.0%	-4.4%	20.1%	76.3%
2002 - 2006							
$\frac{1}{N_o}\sum_o \frac{x_{o,c}}{\sum_c x_{o,c}}$	14.8%	129.5%	-91.6%	12.4%	33.6%	1.3%	85.2%
$\frac{1}{N_o}\sum_o S_t^o \frac{x_{o,c}}{\sum_c x_{o,c}}$	49.3%	7.5%	42.2%	7.2%	7.9%	-14.1%	50.7%

Notes: $x_{o,c}$ with $c \in [1, ..., 6]$ reflects the six components of the decomposition: between industry, between firms, within firms, entering firms, exiting firms, and net entry. The within-industry component is the sum of the five components between firms, within firms, entering firms, exiting firms, and net entry.

5 Forecasting Occupational Change – the Role of Leading Firms

Forecasting the future of work and changes in the occupational structure of the economy is a hard task. This section lays out the structure and reports results for forecasting an industry's future occupational structure relying on "leading" firms: identifying leading firms allows to map out a catch-up process for other firms to forecast occupational change of an entire industry. We define leading firms as those whose occupational structures are ahead of broader economic trends in their industry. Our main methods for identifying leading firms use a machine learning approach and a productivity-based approach.

To validate our methodology, we first use an earlier "training" window (1997–2007) to (i) establish a method to identify leading firms and (ii) estimate the mapping from leaders occupational composition to subsequent changes in the industrys occupational structure. We then apply the estimated models to the subsequent assessment window (2010–2024), i.e., the estimated model to identify leaders based on their characteristics

and the model which predicts the future occupational structure.

5.1 Variables

Our key outcome variable is the occupational structure at the industry level, defined by the vector $S_t^i = [S_t^{1,i}, \ldots, S_t^{O,i}]$, where $S_t^{o,i}$ is the share of employees in each of 27 occupational categories o at time t in industry i, defined at the three-digit level (224 sectors). To identify leading firms, we also compute the occupational structure at the firm level, S_t^f. We measure the distance between two occupational structures S_t and S_s using the L_2 norm: $\|S_t - S_s\|_2 = \sqrt{\sum_{o \in O}(S_t^o - S_s^o)^2}$.

We construct a set of firm-level predictor variables to identify leading firms. All predictors are computed over a three-year sub-period at the start of each window: 1997–1999 for the training sample, and 2010–2012 for the estimation sample. To avoid the Covid period we only use 2022–2023 as the sub-period for an out-of-sample prediction into the future.

We include three groups of predictors. The first consists of averages and growth rates of core firm-level variables: number of employees, total sales, value added, gross operating profits, export sales, labor cost, tangible and intangible assets, and investment. The second group captures structural characteristics of the firm through time-invariant ratios computed over the same sub-period: productivity—both labor productivity (value added per worker) and total factor productivity estimated following Ackerberg et al. (2015)—capital intensity (tangible assets relative to sales, tangible and intangible assets per worker), investment intensity (investment relative to tangible assets and per worker), profitability (gross operating surplus over sales, gross profits per worker, sales per worker), export orientation (export share of sales, exports per worker), and the average annual wage. The third group consists of firm age and two-digit industry classification, both measured at the start of each window. In total, the predictor set comprises 32 variables.

5.2 Identifying the Leading Firms

A leading firm is one whose current occupational structure resembles the future occupational structure of its industry. Intuitively, if a firm's workforce composition today looks

like where its industry is heading, that firm is "ahead" of the curve. We operationalize this idea using two complementary approaches: a machine learning method that identifies leaders based on a broad set of firm characteristics, and a productivity-based method that uses total factor productivity as the sole criterion.

Machine learning approach. For each firm f in industry i, we compute the L_2-Norm between the firm's occupational structure in year t and the leave-one-out industry occupational structure h years later:

$$\left\| S_t^f - S_{t+h}^{(i_f,-f)} \right\|_2 = \sqrt{\sum_{o \in O} \left(S_t^{o,f} - S_{t+h}^{o,(i_f,-f)} \right)^2}, \tag{8}$$

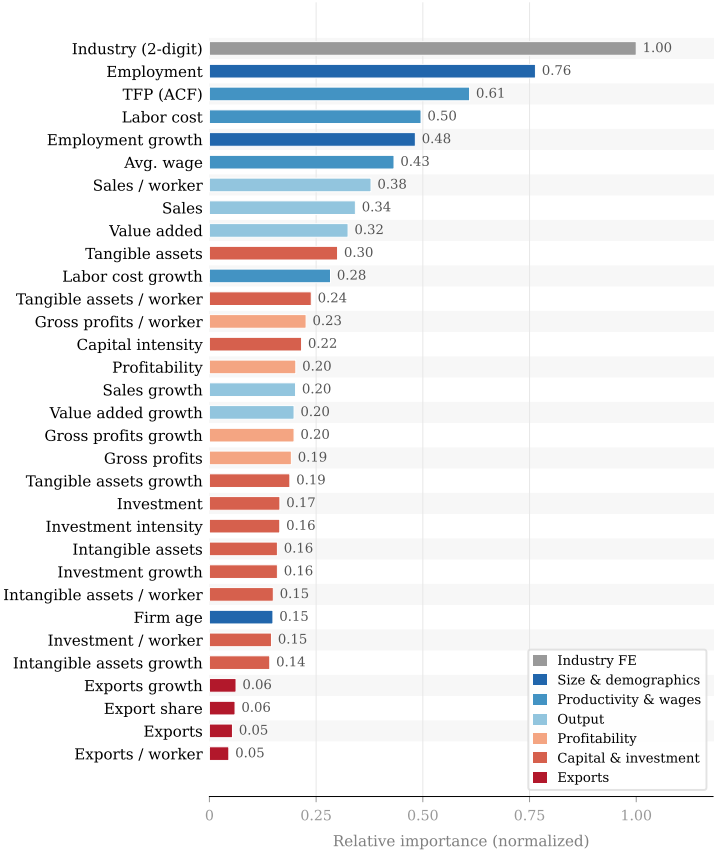
where $S_{t+h}^{(i_f,-f)}$ denotes the occupational structure of all firms in industry i other than f at time $t+h$. A firm with a small distance is one whose workforce composition already anticipates where its industry is heading.

We train a random forest algorithm (Breiman, 2001) to predict this distance using the set of firm-level predictors described in Subsection 5.1. The model is trained on firms in the estimation sample, using characteristics averaged over 1997–1999 to predict occupational distances to 2006 in our baseline seven-year horizon. We grow 3,000 trees, each using seven randomly selected features at each split. Within each industry, we rank firms by their predicted distance and classify as "ML-leaders" those with the lowest predicted distances whose cumulative employment reaches 20% of industry employment.

To account for potential differences in the determinants of leadership across the firm size distribution, we split each industry into two groups of approximately equal total employment—subject to each group containing at least 20 firms in every year and the largest firm in the smaller group employing at least 50 workers. We identify leaders separately within each size group and then pool them, so that the merged set of leaders comprises roughly 20% of total industry employment.

Figure 3 reports the variable importance measures from the random forest estimation. Employment is by far the most important predictor, followed by ACF productivity, the average wage, and labor costs—consistent with the view that leading firms are distinguished primarily by their scale and the quality of their workforce.

Figure 3: Variable Importance of Random Forest Model



Notes: Impurity-based variable importance from the random forest model trained on the estimation sample (1997–1999 firm characteristics predicting occupational distances in 2006). Higher values indicate that a variable contributes more to reducing prediction error across the 3,000 trees.

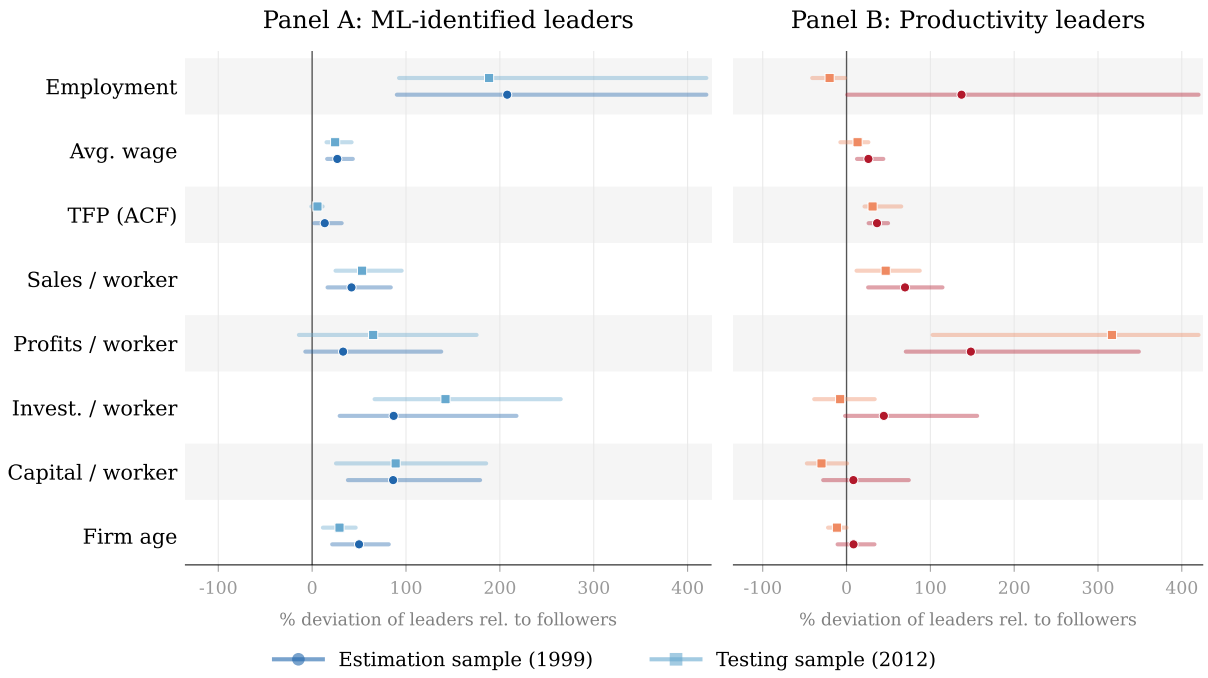
Productivity approach. Our secondary approach uses productivity as the sole criterion for identifying leaders. We estimate total factor productivity using the Ackerberg et al. (2015) procedure, proxying output with value added, labor with the number of employees, capital with tangible assets, and intermediate inputs with raw materials and customs. We estimate separate production function parameters for the manufacturing and services sectors. Within each industry-size cell, firms whose cumulative employment-weighted productivity ranks in the top 20% are classified as "productivity leaders." We restrict the sample to firms observed in at least two out of three consecutive years.

Characterizing leaders and followers. Figure 4 compares the median characteristics of leaders and followers across all 184 industries, for both the estimation (1999) and

testing (2012) samples. ML leaders are roughly three times larger than followers, invest nearly twice as much per worker within their industry, hold twice the capital per worker, pay 25% higher wages, and are older. These patterns are stable across the two periods, separated by more than a decade, suggesting that the ML algorithm recovers a consistent profile of leadership. The within industry investment premium implies a younger capital stock, which embodies more recent production technologies (see, e.g., Greenwood et al., 1997) and is associated with new forms of labor organization (Bartel et al., 2007), providing a natural channel through which these firms may lead occupational change.

Productivity leaders dominate on TFP and profitability, earning two to four times the profits per worker, but are not systematically larger or more capital intensive. In the testing sample they are smaller and less capital intensive than followers, and their profile shifts between 1999 and 2012. The ML method thus captures a broader and more stable dimension of leadership, while the productivity approach isolates an efficiency margin that does not generalize to other firm characteristics.

Figure 4: Characteristics of Leaders Relative to Followers



Notes: Each dot reports the median percent deviation of leaders' characteristics relative to followers, computed across 184 three-digit industries. Horizontal bars span the interquartile range. Panel A shows ML-identified leaders; Panel B shows productivity leaders. Circles denote the estimation sample (1999); squares denote the testing sample (2012).

5.3 Estimation of the Forecasting Model

Once we have defined our algorithms to identify leading firms, we estimate the forecasting model of industry-level occupational structure using linear regression. These models include as covariates contemporaneous and lagged occupational structures of each industry as well as the contemporaneous and lagged occupational structures of leading firms defined according to the algorithms outlined in the previous subsection. Hence, we estimate the following model:

$$S_{t+k}^{o/i} = \alpha + \beta_{L_1} L_{t,t}^{o/i} + \beta_{L_2} L_{t,t-h}^{o/i} + \beta_{S_1} S_t^{o/i} + \beta_{S_2} S_{t-h}^{o/i} + \epsilon, \tag{9}$$

where $L_t^{o/i}$ is the share of workers in occupation o within leading firms in industry i at time t, while $S_t^{o/i}$ refers to the entire industry. In our baseline setting we set $t = 1999$, $t - h = 1997$, and $t + k = 2006$.

The outcomes of interest are the estimates $\widehat{\alpha}$, $\widehat{\beta_{L_1}}$, $\widehat{\beta_{L_2}}$, $\widehat{\beta_{S_1}}$, and $\widehat{\beta_{S_2}}$ which we will use or prediction in the next step. We estimate different models for the productivity-leaders and the ML-leaders. We also estimate an alternative model without including leaders which serves as our baseline model:

$$S_{t+k}^{o/i} = \tilde{\alpha} + \tilde{\beta}_{S_1} S_t^{o/i} + \tilde{\beta}_{S_2} S_{t-h}^{o/i} + \epsilon, \tag{10}$$

with the outcomes of interest being $\widehat{\tilde{\alpha}}$, $\widehat{\tilde{\beta}_{S_1}}$, and $\widehat{\tilde{\beta}_{S_2}}$.

Note that the setup of equation (9) nests the setting with only contemporaneous levels because if $\beta_{L2} = \beta_{S2} = 0$ then $S_{t+h}^{o/i} = \alpha + \beta_{L_1} L_{t,t}^{o/i} + \beta_{S_1} S_t^{o/i} + \epsilon$. Second, it can also nests a model using differences since with $\beta_{L1} = -\beta_{L2}$ and $\beta_{S1} = 1 - \beta_{S2}$), the equation can be rewritten as $S_{t+h}^{o/i} - S_t^{o/i} = \alpha + \beta_{l_1} (L_{t,t}^{o/i} - L_{t,t-h}^{o/i}) + \beta_{s_1} (S_t^{o/i} - S_{t-h}^{o/i}) + \epsilon$.

Table 4 displays the regression output of the linear forecasting model. Both models which use Machine Learning Leaders put lower weight on the industry-wide occupational structures and emphasize the structures of the leading firms. Even though this is less prevalent for productivity leader models, they still emphasize the contemporaneous industry structure less than the baseline model. Contemporaneous leaders add in-sample predictive power in both leader setups as highlighted by the adjusted R^2. Note that the sum of coefficients is close to 1 for all five different models.

Table 4: Forecasting Models of the Future Occupational Structure by Industry, $S_{t+h}^{o/i}$

	Baseline	Machine Learning Leaders		Productivity Leaders (ACF)	
$S_t^{o/i}$	1.3018^{***}	0.4960^{***}	0.3411^{***}	1.0628^{***}	1.0388^{***}
	(0.0294)	(0.0255)	(0.0244)	(0.0296)	(0.0301)
$S_{t-h}^{o/i}$	-0.3333^{***}	-0.2289^{***}	0.0519^{**}	-0.3492^{***}	-0.3065^{***}
	(0.0293)	(0.0220)	(0.0228)	(0.0283)	(0.0299)
$L_{t,t}^{o/i}$		0.6683^{***}	1.2408^{***}	0.2485^{***}	0.3344^{***}
		(0.0107)	(0.0228)	(0.0084)	(0.0203)
$L_{t,t-h}^{o/i}$			-0.6932^{***}		-0.1068^{***}
			(0.0249)		(0.0230)
(Intercept)	0.0012^{***}	0.0024^{***}	0.0022^{***}	0.0014^{***}	0.0015^{***}
	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Adj. R^2	0.9469	0.9703	0.9743	0.9556	0.9558
Observations	4968	4941	4941	4509	4455

Notes: OLS estimates of equations (9) and (10). The dependent variable is the share of workers in occupation o within industry i in 2006. $S_t^{o/i}$ and $S_{t-h}^{o/i}$ are contemporaneous (1999) and lagged (1997) industry occupational structures; $L_{t,t}^{o/i}$ and $L_{t,t-h}^{o/i}$ are the corresponding structures among leading firms. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.4 Assessment of the Forecasting Models

To assess the quality of the forecasting models, we use the estimated models–precisely the estimates α and β–and generate out-of-sample forecasts for the period 2010–2019. We use the algorithms constructed in Section 5.2 to identify leading firms in 2012. We then use the models in Equation 9 and 10 and the coefficients estimated in Section 5.3 to forecast the occupational structure of the industry in 2019. As covariates, we use the occupational structures of industries and leaders in 2010 and 2012. We assess the accuracy of these forecasts by comparing the actual occupational structure of the industry with the forecasted one using the Root Mean Square Error (RMSE). We follow the standard in the literature and compute the Theil index to compare models that incorporate leading firm information with those that do not, measuring the additional predictive value of our leader-approach:

$$\textbf{Theil ratio} = \frac{\text{RMSE (Leader Model)}}{\text{RMSE (Baseline Model)}}$$

Table 5: Assessment of the Forecasting Models

	Baseline	Machine Learning Leaders		Productivity Leaders (ACF)	
Panel A. In-Sample Assessment					
RMSE (times 10^{-2})	2.110	1.577	1.466	1.933	1.928
Theil Ratio (vs Baseline)	—	0.748	0.695	0.916	0.914
Panel B. Out-of-Sample Assessment					
RMSE (times 10^{-2})	3.333	2.744	2.620	3.052	3.032
Theil Ratio (vs Baseline)	—	0.823	0.786	0.916	0.910
Included Regressors					
$S_t^{o/i}$	✓	✓	✓	✓	✓
$S_{t-h}^{o/i}$	✓	✓	✓	✓	✓
$L_{t,t}^{o/i}$		✓	✓	✓	✓
$L_{t,t-h}^{o/i}$			✓		✓

Notes: Panel A reports in-sample results using the estimation period 1997–2006. Panel B reports out of sample results, applying coefficients estimated on 1997–2006 to predict 2019 occupational structures using 2010–2012 data. RMSE is the employment weighted root mean square error across industry-occupation cells (multiplied by 10^2). The Theil ratio is the RMSE of each leader model divided by the RMSE of the baseline model; values below one indicate improved forecasting accuracy.

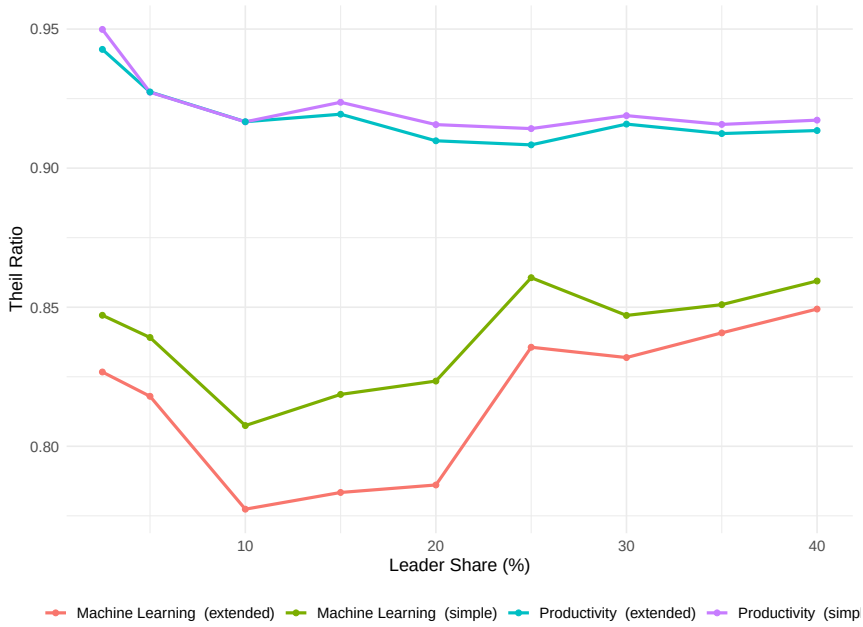
Table 5 reports the results for both the in-sample estimation period (1997–2006) and the out-of-sample testing period (2010–2019). We discuss each in turn. Panel (a) shows the in-sample fit. The extended ML model reduces the RMSE by roughly 30% relative to the baseline (Theil ratio of 0.695), while the productivity leader models reduce it by less than 10%.

The out-of-sample results in Panel (b) are the central test. Applying the coefficients estimated on 1997–2006 to predict 2019 occupational structures using 2010–2012 data, the extended ML model achieves a Theil ratio of 0.786, a 21% gain in forecasting accuracy over the baseline. The simple ML model yields a ratio of 0.823. Productivity leaders improve on the baseline only marginally, with ratios around 0.91. Two features of these results stand out. First, the ML models retain a large share of their in-sample advantage when applied to a different period and a different set of firms, indicating that the estimated relationship between leader and industry occupational structures generalizes

beyond the training sample. Second, the gap between ML and productivity leaders widens out of sample, consistent with the view that the broader set of firm characteristics captured by the random forest carries predictive content for occupational change that productivity alone does not.

Different Leader Shares. The forecasting gains are concentrated in a narrow subset of firms and are not a mechanical artifact of the leader definition. Our baseline sets the leader share at 20% of industry employment; Figure 5 varies this threshold from 2.5% to 40%. The ML Theil ratios exhibit a U-shaped profile, with the lowest out-of-sample values at a leader share of 10%. Narrower definitions yield noisier estimates of the leader occupational structure, while broader definitions dilute the leader set with less informative firms and reduce forecasting accuracy. If the gains were mechanical, for instance from selecting a sufficiently large group to approximate the industry average, expanding the set should not deteriorate performance. The optimum around a 10% share indicates that the algorithm identifies a specific subset of firms with distinct information about the future of work. The productivity leader models exhibit a flatter, almost monotonic profile: varying the share has little effect on the forecasting performance.

Figure 5: Out-of-sample Assessment with changing Leader Shares



Notes: Out of sample Theil ratios as a function of the cumulative employment share used to define leaders within each 3-digit industry. The baseline specification uses a 20% leader share. Values below one indicate improved forecasting accuracy relative to the baseline model without leaders.

5.5 Connection to the Theoretical Framework

The theoretical model in Section 2 rests on two assumptions about the behavior of leading firms that can be assessed empirically. First, firm types (z, I) follow a Markov process, implying partial persistence in leader identity with stochastic turnover. Second, the distribution $G(z, I)$ that governs technology adoption is stationary, so that the characteristics defining leaders should be stable across periods even if specific firms enter and exit the leader set. We examine whether the data are consistent with each assumption.

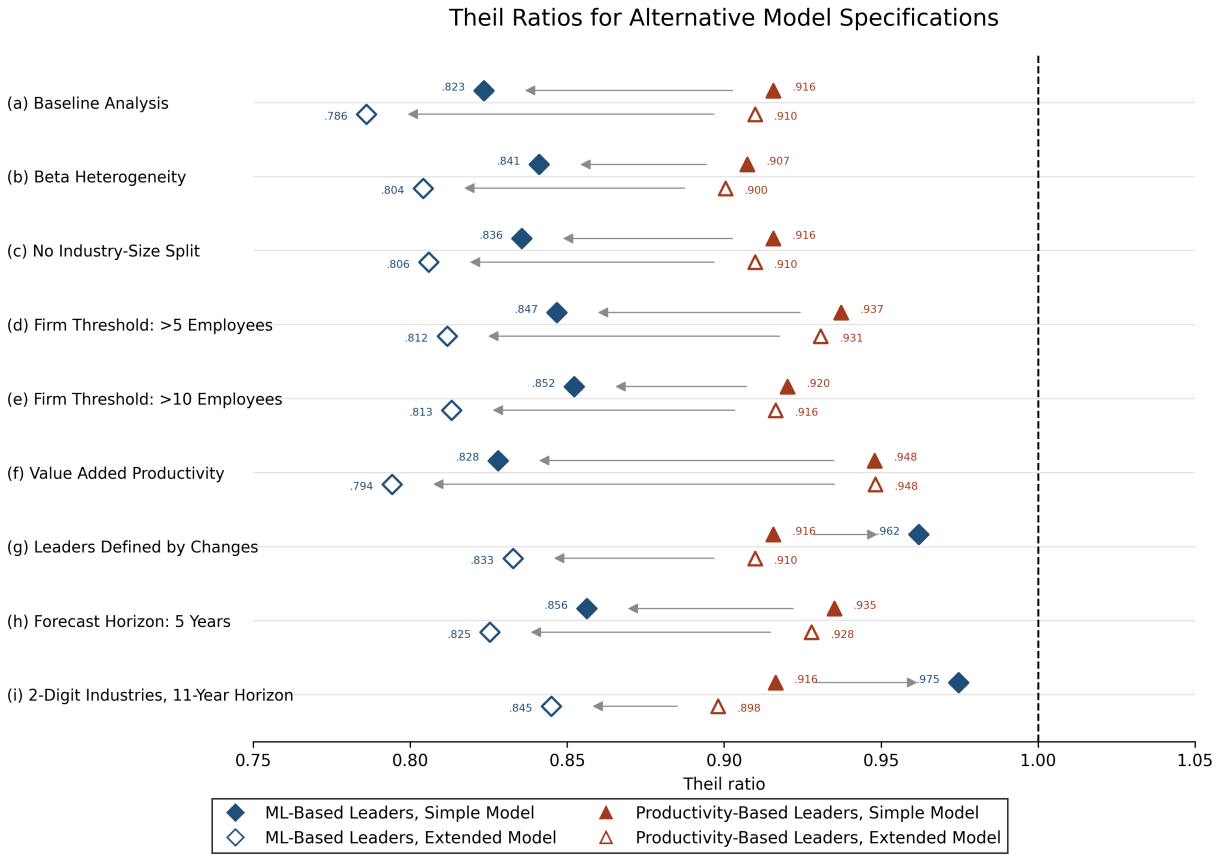
Leader turnover. The Markov assumption on (z, I) implies partial persistence in leader identity: high type firms tend to remain high type, but stochastic transitions generate turnover. We find substantial turnover. Of the 29,071 firms classified as ML leaders in 1999, about 60% survive to 2012, a rate comparable to the overall firm survival rate of 58% (both in terms of firm counts). Among these surviving leaders, only 33% retain their leader status in 2012; roughly two thirds of former leaders are reclassified as followers. This degree of turnover is consistent with a process in which firms move in and out of the adoption frontier over a 13 year window.

Stability of the selection criterion. The stationarity of $G(z, I)$ implies that the characteristics defining leaders should be stable across periods, even if the specific firms change. Three pieces of evidence are consistent with this assumption. First, the ML criterion is robust to the choice of forecast horizon: leaders identified using a five year target overlap with those identified using a seven year target by 89–93% of firms, in both the estimation (1999) and testing (2012) samples. Second, when we retrain the random forest on 2012 characteristics—using a model estimated entirely within the testing period—and compare the resulting leader set with the one obtained by applying the 1999 trained model to the same 2012 cross section, the overlap is 42% of firms (42% by employment). When comparing leaders identified for the common target year 2019 using the estimation sample model versus the testing sample model, the overlap rises to 61% of firms (63% by employment). The two models, trained more than a decade apart, identify a substantial common core of firms despite using independent training data. Third, leader characteristics are stable across the estimation and testing samples: ML leaders in both 1999 and 2012 are roughly three times larger than followers, invest nearly twice as much per worker, and pay 25% higher wages (Figure 4). The profile of leadership is preserved even as the composition of the leader set changes.

6 Alternative specifications

This section reports a set of alternative specifications and robustness checks for the forecasting exercise. Figure 6, panels (b)–(i), summarizes the corresponding Theil ratios, while panel (a) plot the Theil ratios of our baseline analysis reported in Table 5.

Figure 6: Theil Ratios for Alternative Model Specifications



Notes: The figure reports out of sample Theil ratios across the baseline specification and a range of alternative model specifications. For each specification, diamonds denote ML based leaders and triangles denote productivity based leaders; filled markers correspond to the simple model, hollow markers to the extended model. Thin arrows connect the productivity based measure to the corresponding ML based measure within each specification. The vertical dashed line at 1 indicates equality with the baseline; values below 1 indicate improved forecasting accuracy.

Beta heterogeneity. We allow the coefficients of the forecasting regression to vary by one-digit industry in order to capture sectoral differences in the catch-up process. This modification leaves the main pattern unchanged: the ML-based models continue to outperform the baseline, with Theil ratios of 0.841 in the simple specification and 0.804 in

the extended specification, while the productivity-based models remain closer to the benchmark at 0.907 and 0.900. This result, shown in panel (b) of Figure 6, suggests that allowing for heterogeneous betas modestly weakens performance relative to the baseline specification, but does not overturn the advantage of the ML-based approach.

Industry granularity. Our baseline definition uses 3-digit industries interacted with firm size. As an alternative, we remove the industry-size split and define leaders at the industry level only. As shown in panel (c) of Figure 6, the resulting Theil ratios are 0.836 and 0.806 for the ML-based simple and extended models, and 0.916 and 0.910 for the corresponding productivity-based models. The results are very close to the baseline case, indicating that the predictive gains are not driven solely by the industry-size split.

Firm sample criteria. In the main analysis, we exclude firms with fewer than 20 employees. We then relax this threshold and re-estimate the models using firms with more than 5 employees and, separately, firms with more than 10 employees. Panels (d) and (e) of Figure 6 show that with the > 5 cutoff, the ML-based models yield Theil ratios of 0.847 and 0.812, while the productivity-based models yield 0.937 and 0.931; with the > 10 cutoff, the corresponding values are 0.852 and 0.813 for ML-based leaders and 0.920 and 0.916 for productivity-based leaders. Loosening the firm-size restriction reduces predictive performance, but the ML-based models remain clearly below 1 and continue to dominate the productivity-based alternatives.

Productivity measures and ML inputs. We also replace the productivity measure used to define productivity-based leaders with a simpler measure of labor productivity, namely value added per worker. We continue to include productivity-related information among the ML inputs, but assess whether this simpler definition changes forecasting performance. As showed in panel (f) of Figure 6, the ML-based models remain strong under this alternative, with Theil ratios of 0.828 and 0.794, whereas the productivity-based models both deliver a ratio of 0.948. This result indicates that the ML-based approach is robust to this alternative productivity measure, while the productivity-only approach moves noticeably closer to the benchmark.

Changes versus levels. Our baseline specification predicts levels of occupational shares. As a sensitivity check, we instead define leaders using changes in occupational structure between $t - h$ and t, aligned with the changes in the industry occupational structure between t and $t+k$. Results are in panel (g) of Figure 6. Under this definition, the ML-based

simple model yields a Theil ratio of 0.962, the ML-based extended model 0.833, and the productivity-based simple and extended models 0.916 and 0.910, respectively. Defining leaders with changes weakens the simple ML specification substantially, while the extended ML model still performs clearly better than the benchmark.

Forecast horizon. We also consider a shorter forecasting horizon of five years. Panel (h) of Figure 6 shows that the ML-based simple and extended models produce Theil ratios of 0.856 and 0.825, while the productivity-based simple and extended models produce 0.935 and 0.928. All four models remain below 1, but the gains are somewhat smaller than in the baseline specification, especially for the productivity-based models.

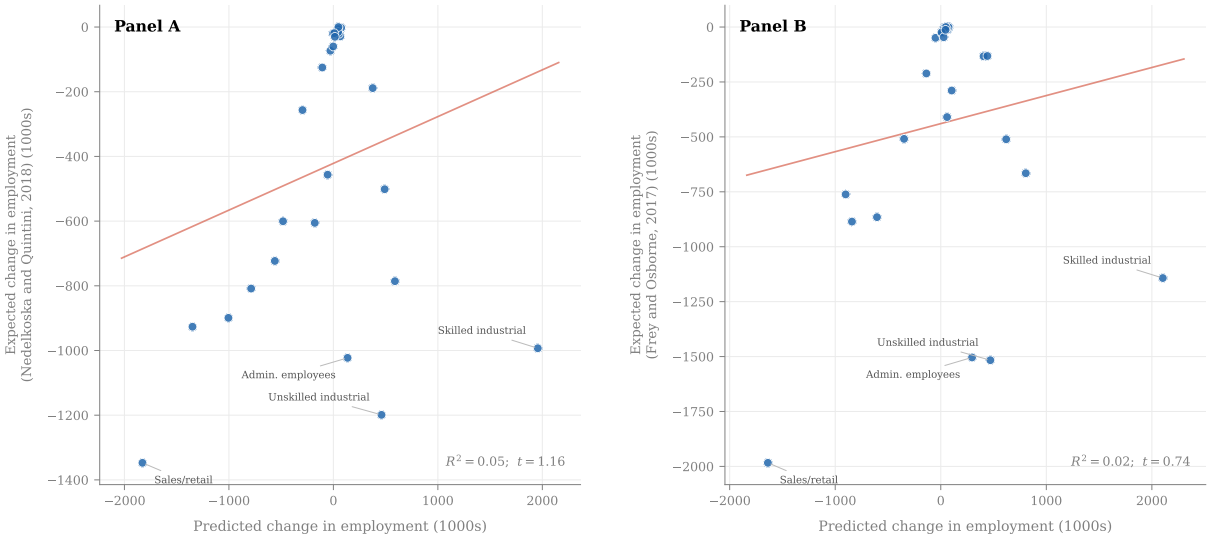
Two-digit industries and a longer horizon. Finally, we re-estimate the forecasting exercise at the two-digit industry level, where the NAF classification remains largely unchanged over 1997–2023. This allows us to use 1997–2010 as the estimation sample and to evaluate an out-of-sample forecast over an 11-year horizon without extrapolating coefficients across a classification break. In this case, the ML-based simple and extended models yield Theil ratios of 0.975 and 0.845, while the productivity-based simple and extended models yield 0.916 and 0.898. As panel (i) of Figure 6 makes clear, the extended ML model remains comfortably below the benchmark, whereas the simple ML model moves much closer to 1 over the longer horizon.

7 Comparison to Other Drivers or Predictors of Occupational Change

The occupational changes predicted by the leader model are largely orthogonal to those implied by automation risk indices. A prominent literature forecasts employment losses by assessing the technological susceptibility of occupations to automation, either through expert evaluation (Frey and Osborne, 2017) or task-level data from worker surveys (Nedelkoska and Quintini, 2018). These approaches answer a different question than the leader model: they identify occupations where displacement is technologically feasible, whereas the leader model captures where reallocation is actually occurring within industries. We compare the two by plotting the employment changes implied by each automation risk index against those predicted by our model.

To construct comparable economy-level predictions, we aggregate the industry-level predicted occupational shares to the economy-level using industry-employment weights, multiply by total employment in 2021 to obtain predicted occupation-level employment, and compute the change relative to actual 2015 levels. We then regress the predicted job losses from each automation index on the employment changes predicted by the leader model. Figure 7 presents the results. Both comparisons yield an R^2 of approximately 0.04, with slope coefficients that are positive but statistically insignificant ($t \approx 1.0$). The occupations that the leader model predicts will grow or shrink bear essentially no relationship to those flagged as at risk by the automation indices.

Figure 7: Comparison to Automation Risk Approaches



Each point represents a 2 digit occupation (PCS classification). The horizontal axis shows the predicted change in employment from the leader model (in thousands). The vertical axis shows the expected change in employment based on the automation risk score from Nedelkoska and Quintini (2018) (Panel A) and Frey and Osborne (2017) (Panel B). Automation risk scores are matched to 2 digit PCS occupations and multiplied by actual employment in the reference year (2016 for Nedelkoska and Quintini, 2010 for Frey and Osborne). The predicted change from the leader model is computed as the difference between predicted employment in 2021 and actual employment in the reference year. Labels identify the four largest outliers from the OLS regression line.

Three features of the automation risk framework help explain the disconnect. First, automation risk scores measure the feasibility of substituting specific tasks, not the net effect on occupations. Within occupation task recomposition, where firms automate some tasks but retain the occupation with a reweighted task bundle, weakens the link between risk scores and realized employment changes. Second, occupational reallocation reflects forces beyond automation: offshoring, organizational restructuring, demand shifts, and

regulatory changes all reshape the occupational structure of industries, and the leader model captures the net effect of all these margins. Third, the time horizons differ. Frey and Osborne (2017) describe a long run technological frontier; the leader model forecasts realized change over a seven-year window. Much of the predicted displacement may not have materialized within the sample period, or may follow adjustment paths that differ substantially from the feasibility assessments.

The low correlation is consistent with the theoretical framework in Section 2. The type space (z, I) that defines leaders encompasses all dimensions of production organization, not only automation adoption but also offshoring decisions, managerial practices, and human capital strategies. The catch up mechanism requires only that followers converge toward the occupational structure of leaders, regardless of which forces shaped that structure. An index built on a single dimension of I should therefore have weak explanatory power for the overall direction of change whenever multiple dimensions matter. The low R^2 suggests that the forces driving occupational reallocation are indeed multidimensional, which is precisely the setting where the leader model adds value over approaches that isolate one channel. For policy purposes, targeting retraining programs, or anticipating local labor market adjustment, a forecast grounded in the revealed behavior of firms is more directly actionable than an index of technological susceptibility.

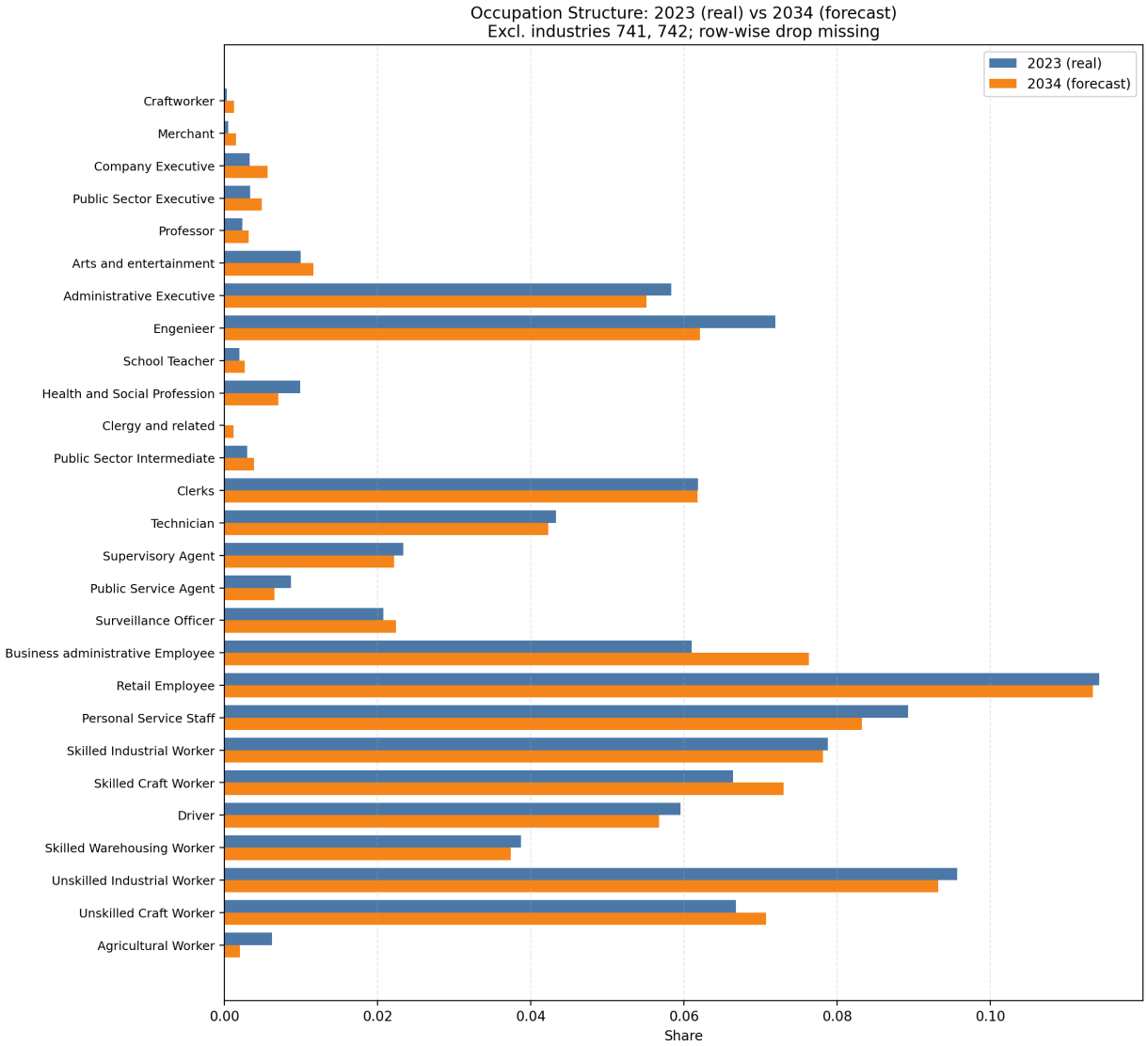
8 Predicting the Future Occupational Structure

Figure 8 compares the industry-weighted distribution of employment across detailed occupations in 2023 (realized structure) with the 2034 forecast from the extended machine-learning model. The bars read as shares of total employment.

Retail employees are expected to remain the largest single group in both years (about 11.4% in 2023 vs. 11.3% in 2034—a small absolute dip). Unskilled industrial workers stay the next-largest bucket; personal service staff, skilled industrial workers, and engineers sit in the next tier by size in 2023, with a mix of small upward and downward moves by 2034.

The clearest gains are predicted for business administrative employees (roughly $+1.5$ percentage points, from about 6.1% to 7.6%) and, to a lesser extent, skilled and unskilled craft workers (on the order of $+0.7$ and $+0.4$ points). Several smaller white-collar and service categories also tick up slightly in level (e.g. company executives, surveillance

Figure 8: Predicted occupational structure in 2034 versus observed structure in 2023



Notes: Actual occupation structure in 2023 and forecast occupation structure for 2034. Shares aggregate industryoccupation cells to national occupation shares using 2023 industry employment shares as weights.

officers, arts and entertainment). Clerks are essentially flat (a negligible change of about -0.05 points), so the figure does not point to a large clerical vs. administrative swap within this breakdown.

The largest absolute predicted reduction in share is for engineers (about -1.0 percentage point, from roughly 7.2% to 6.2%, for a reduction of approximately 14%). Personal service staff is also forecast to decline markedly (about -0.6 points). Among the mid-sized

technical and managerial categories, administrative executives and drivers fall on the order of -0.3 points, with smaller reductions for some health and social roles and public-sector agents. Unskilled industrial workers decline by roughly 0.25 points—material in absolute terms given their mass, but not a dramatic collapse in levels. Agriculture is predicted to fall noticeably in share from an already small base (0.4 percentage points, which equals a drop of almost 70%). Taken together, the picture is a joint reallocation across services, craft, and office work rather than a single story in one occupation.

Overall, the comparison shows that several categories are predicted to move by meaningful percentage-point margins, which suggests that the occupation mix is not very stable over the forecast window.

9 Conclusion

This paper shows that forecasts of the future occupational change can be improved by exploiting heterogeneity across firms and, in particular, by identifying "leading firms" whose workforce composition anticipates the future occupational structure of their industry. We develop a novel theoretical model building on the literature on technology diffusion and leader-laggard dynamics to show that early technology adopters are already employing the workforce of the future, thereby allowing a glimpse into the future of work. Using administrative data for France, we apply a machine-learning procedure that identifies leading firms based on a broad set of firm characteristics and substantially outperforms both a baseline model without leader information and a model using a productivity-based leader definition, reducing out-of-sample forecast error by about 21 percent in the preferred specification. The evidence further suggests that leading firms are larger, older, better paid, more productive, and have higher sales, profits, investment, and capital per worker than other firms in narrowly defined industries. Hence, the notion of leading firms is economically meaningful, and leadership is relatively stable in its observable characteristics over time.

The occupational changes predicted by the leader model are not strongly correlated with those implied by well-known automation-risk indices, indicating that the evolution of labor demand reflects a broader set of forces than just automation or routinization. More broadly, our results suggest that looking at frontier firms provides a practical and scalable way to discipline forecasts of the future of work, while remaining complementary

to task-based and expert-based approaches.

A natural limitation is that the approach can only capture technologies and organizational practices that are already present in the economy, not innovations that have yet to emerge. But since it relies on revealed adjustments in the cross section of firms rather than ex-ante judgments about specific technologies, the framework offers an alternative tool for anticipating future labor-market change and for informing education, retraining, and workforce policy.

We also apply the model to predict the occupational structure of the French economy in 2034, starting from the latest year available in our database, 2023. This exercise illustrates how the framework can be used in practice to generate forward-looking predictions of the future of work based on currently observed differences across firms.

References

Acemoglu, Daron and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4, Amsterdam: Elsevier, 2011, pp. 1043–1171.

___ **and Pascual Restrepo**, "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M Shapiro, "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 2017, *132* (4), 1553–1592.

Arntz, Melanie, Terry Gregory, and Ulrich Zierahn, "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis," OECD Social, Employment and Migration Working Papers 189, OECD Publishing, Paris 2016.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Autor, David H. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 2013, *103* (5), 1553–1597.

___, Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 2003, *118* (4), 1279–1333.

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw, "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills," *The Quarterly Journal of Economics*, 2007, *122* (4), 1721–1758.

Breiman, Leo, "Random Forests," *Machine Learning*, 2001, *45* (1), 5–32.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376.

Comin, Diego and Bart Hobijn, "An Exploration of Technology Diffusion," *American Economic Review*, 2010, *100* (5), 2031–2059.

Frey, Carl Benedikt and Michael A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?," *Technological Forecasting and Social Change*, 2017, *114*, 254–280.

Goos, Maarten, Alan Manning, and Anna Salomons, "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, 2014, *104* (8), 2509–2526.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell, "Long-Run Implications of Investment-Specific Technological Change," *The American Economic Review*, 1997, *87* (3), 342–362.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 1957, *25* (4), 501–522.

Hubmer, Joachim and Pascual Restrepo, "Not a Typical Firm: The Joint Dynamics of Firms, Labor Shares, and Capital–Labor Substitution," Working Paper 28579, National Bureau of Economic Research March 2021. Accepted for publication in

American Economic Journal: Macroeconomics (forthcoming); accepted-article DOI: 10.1257/mac.20230325.

Humlum, Anders, "Robot Adoption and Labor Market Dynamics," Study Paper 175, The ROCKWOOL Foundation Research Unit May 2022.

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information Technology and Returns to Scale," *American Economic Review*, 2024, *114* (6), 1769–1815.

Mansfield, Edwin, "Technical Change and the Rate of Imitation," *Econometrica*, 1961, *29* (4), 741–766.

Nedelkoska, Ljubica and Glenda Quintini, "Automation, Skills Use and Training," Social, Employment and Migration Working Papers 202, OECD 2018.

Patel, Aseem, "The Role of Firms in Shaping Job Polarization," Working Paper hal-04637258, HAL Open Science 2024.

Rogers, Everett M., *Diffusion of Innovations*, New York: Free Press of Glencoe, 1962.

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

Webb, Michael, "The Impact of Artificial Intelligence on the Labor Market," 2020. Working paper.